Spectral investigations of surface ordering in ultrathin molecular films

V. K. Dolganov,*) V. M. Zhilin, and K. P. Meletov

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

(Submitted 30 November 1998)
Zh. Éksp. Teor. Fiz. **115**, 1833–1842 (May 1999)

Surface molecular ordering in ultrathin molecular films is investigated. The optical transmission spectra of molecular films ranging in thickness from 2 to 13 smectic layers (6.7–43 nm) in the region of the electronic absorption bands in the smectic *A* phase of cyanobiphenyl CB9 are measured. The thickness and temperature dependences of the permittivity are determined. It is found that the orientational ordering of the molecules depends on the film thickness. The penetration depth of the surface molecular orientational order does not exceed two smectic layers (<7 nm). © *1999 American Institute of Physics.* [S1063-7761(99)02205-2]

1. INTRODUCTION

A great deal of attention has been devoted in the last few years in both pure and applied investigations to the study of surface molecular layers. The difference of the interaction of molecules with the environment on the surface and in the interior can change the structure of the layer, change the collective and molecular dynamics, shift the phase transition temperatures, and induce new phases on the surface. Depending on the relative magnitude of various types of interatomic and intermolecular interactions, surface layers can melt at higher or lower temperatures compared with the interior of the sample. For example, investigations of the structures of SmA and SmO[1,2] have shown that cooling the isotropic phase produces layered smectic order on the surface at a temperature several degrees above the phase transition temperature in the interior. As the sample is cooled further, layer-by-layer phase transitions eventually cause the smectic phase to fill the entire sample.

Free-standing molecular films are convenient objects for investigating near-surface and dimensional effects in organic materials.[3,4] The two flat surfaces of these films are bounded by air, and the films themselves can be prepared with various (rigorously determined) numbers of molecular layers. Immediately after preparation a film can contain defects (dislocations, thickness nonuniformity, and so on). However, because of the comparatively high mobility of the molecules, the quality of the surface and of the film itself can be substantially improved by holding the film for several hours near the temperature of the transition to the isotropic or nematic phase. This makes it possible to obtain films ($\sim 1\ \mathrm{cm}^2$) that are uniform over their thickness and contain a definite number of smectic layers. The surface of such a film is a single, continuous smectic plane. This advantageously distinguishes the surface of a free-standing organic film from inorganic structures, where as a rule it is difficult to prepare perfect surfaces of adequate size. For this reason, molecular films, aside from their intrinsic interest, can also serve as model objects for investigating surfaces, two-dimensional structures, and phase transitions in finite-size samples. It has been shown that in molecular films, phase transitions accompanied by a change in the structure of near-surface smectic layers[5–9] and transitions to a crystalline phase[4,10,11] occur at temperatures 10–30 °C above transitions occurring in the interior of the sample. Thin molecular films also melt at a higher temperature than bulk samples.[12]

Theoretical calculations have shown[13–20] that the influence of a surface on translational and orientational molecular ordering must be taken into account in a description of the structure and dynamics of thin films. At the same time, up to now there have been no direct experimental observations of the differences in orientational order of molecules on a surface and in the interior, and in consequence no direct observations of a dependence of the orientational ordering on film thickness. In the present work such investigations were performed using optical methods. The optical transmission spectra were measured in the region of the electronic absorption bands for films with various thickness. A smectic-*A* structure (Fig. 1), in which the ''director'' **n** (the direction of the predominant orientation of the long axes of the molecules) is perpendicular to the plane of the layers and the film surface, was investigated. The measurements were performed in the UV region of the spectrum, where intense absorption bands of the molecules forming liquid-crystal structures are found. This made it possible to observe the electronic absorption in ultrathin samples (thickness down to ~ 7 nm) and to determine from the experimental spectra the imaginary part of the permittivity $\epsilon_2(\omega)$ for films with various thicknesses. The quantity

$$E = \int \epsilon_2(\omega)\, d\omega,$$

where the integral extends over an electronic transition, is proportional to the squared projection of the dipole moment of the electronic transition on the direction of polarization of the light. The permittivity along the principal directions (parallel and perpendicular to the ''director'') depends on the orientational order. This makes it possible to characterize the orientational ordering of a structure on the basis of the relative value of the permittivity.

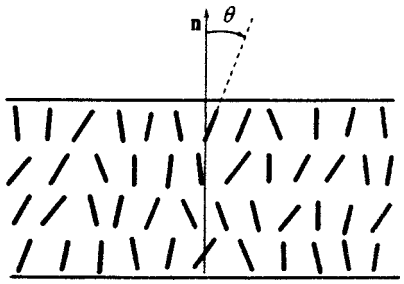

FIG. 1. Smectic-*A* film. The ''director'' **n** is perpendicular to the surface; θ is the disorientation angle of the long axes of the molecules relative to the ''director.''

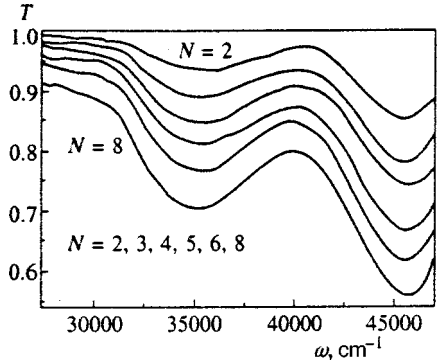

FIG. 2. Transmission spectra of films with thickness $N=2$, 3, 4, 5, 6, and 8 smectic layers at temperature $T=47.7\,^{\circ}\mathrm{C}$.

2. EXPERIMENTAL RESULTS AND ANALYSIS

The measurements of the transmission spectra were performed on films of 4-cyano-4-n-alkylbiphenyl CB9 (nine is the number of carbon atoms in a chain). This substance forms a smectic-*A* phase with the smectic-*A* – nematic–isotropic-liquid phase transition temperatures 48 °C and 49.5 °C, respectively, in the interior of the sample. The films were placed in a 5-mm opening of a thin metal plate. Layer-by-layer thinning by heating a thick film above the temperature of the phase transition to the isotropic liquid in the interior was used to obtain thin films with the required number of molecular layers.[12,21,22] The transmission spectra were measured with the light polarized perpendicular to the ''director.'' The spectra $T(\omega)=I(\omega)/I_0(\omega)$ presented in this paper are the spectra $I(\omega)$, normalized to the lamp spectrum $I_0(\omega)$, of the light transmitted through a film. To take account of the reflection of light from the quartz windows of the heat-bath vessel, the spectrum $I_0(\omega)$ was measured in the same geometry as $I(\omega)$ without the film. To determine the film thickness (the number of smectic layers), the optical reflection was measured in the transmission region. The spectral dependence of reflection under normal incidence and ''backward'' reflection is given by[23]

$$\frac{I_r(\omega)}{I_0(\omega)}=\frac{(n^2-1)^2\sin^2(2\pi nNd\omega)}{4n^2+(n^2-1)^2\sin^2(2\pi nNd\omega)}, \tag{1}$$

where N is the number of smectic layers in the film, d is the interplanar distance ($\approx$3.3 nm in the smectic-*A* phase of CB9), n is the index of refraction, and $\omega=1/\lambda$. In thin films ($N\leqslant 6$), the reflected intensity is proportional to the squared film thickness:

$$I_r(\omega)/I_0(\omega)\approx N^2d^2\pi^2(n^2-1)^2\omega^2. \tag{2}$$

In this case the number of smectic layers was determined according to the relative reflection intensities for films with various thicknesses.

Figure 2 shows the results of the measurements of the transmission spectra of films with thickness $N=2$, 3, 4, 5, 6, and 8 smectic layers. The spectra consist of two electronic bands F_a ($\omega_a\approx 3.5\times 10^4$ cm^{-1}) and F_b ($\omega_b\approx 4.6\times 10^4$ cm^{-1}). In cyanobiphenyls, the electronic transition dipole moment for the low-frequency band F_a is parallel to the long axis of the molecule. The transmission spectrum of a film with $N=13$ layers for two temperatures 37.6 °C and 47 °C is shown in the top half of Fig. 3. The intensity of the spectrum in the frequency range displayed in Figs. 2 and 3 is determined by electronic absorption and interference effects in the reflection of light by the film surface. Even in the transmission range the transmission spectrum exhibits a strong spectral dependence $T(\omega)=[1-I_r(\omega)/I_0(\omega)]$ [see the expression (1)]. This dependence is even more complicated in the vicinity of the absorption bands, since the absorption coefficient, the refractive index, and the change in phase of the light wave for transmission through a film depend on ω. Interference effects make a large contribution to the intensity and frequency-dependence of the spectrum. Thus, the deviation of $T(\omega)$ from 1 in the low-frequency part of the spectra is completely due to interference. On account of interference effects, the intensity of the light transmitted through a film in the absorption region does not follow the Lambert–Beer law, and the spectra must be analyzed taking account of energy dissipation in the film and interference.

The transmission spectrum and the response of the film[21] to a light wave are described by the permittivity $\epsilon(\omega)$, which depends on the orientational ordering of the molecules in the film. Even though the expressions relating the transmission spectrum of the absorbing film with $\epsilon(\omega)$ were derived comparatively long ago and are presented in the classic monographs (see, for example, Refs. 24 and 25), thus far the optical spectra of the ultrathin films have not been analyzed using these expressions because of the lack of experimental data.

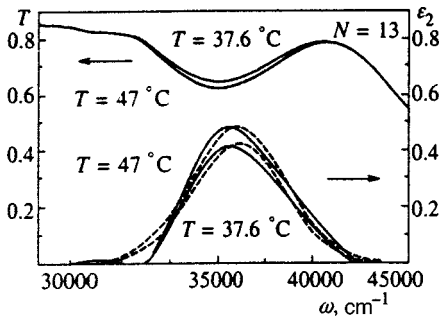

FIG. 3. Transmission spectra of a film with thickness $N=13$ smectic layers at temperatures $T=37.6\,^{\circ}\mathrm{C}$ and 47 °C (top half of the figure). The bottom half of the figure shows the imaginary part of the permittivity, obtained using Gaussian (dashed curves) and arbitrary (solid curves) forms for the spectral curves $\epsilon_2(\omega)$.

For an absorbing medium the light transmission coefficient $T(\omega)$ of a film is given by the expression[24,25]

$$T(\omega)=\frac{16(n^2+\kappa^2)\exp(-4\pi d\kappa\omega)}{[(n+1)^2+\kappa^2]^2[1-2\rho\exp(-4\pi d\kappa\omega)\cos[2(\beta+2\pi dn\omega)]+\rho^2\exp(-8\pi d\kappa\omega)]}, \tag{3}$$

where

$$\rho=\frac{(n-1)^2+\kappa^2}{(n+1)^2+\kappa^2} \tag{4}$$

is the light reflection coefficient of the surface of a semi-infinite medium, and

$$\beta=\tan^{-1}\frac{2\kappa}{n^2-1+\kappa^2} \tag{5}$$

is the phase delay due to energy dissipation in the absorbing film. It is well known that the absorption coefficient $\kappa(\omega)$ and the refractive index $n(\omega)$ can be expressed in terms of the real and imaginary parts of the permittivity $\epsilon(\omega)=\epsilon_1(\omega)+i\epsilon_2(\omega)$:

$$n=\frac{1}{\sqrt{2}}[(\epsilon_1^2+\epsilon_2^2)^{1/2}+\epsilon_1]^{1/2}, \tag{6}$$

$$\kappa=\frac{1}{\sqrt{2}}[(\epsilon_1^2+\epsilon_2^2)^{1/2}-\epsilon_1]^{1/2}. \tag{7}$$

Dipole electronic excitations bands are often described using for the permittivity the simple ''classical dispersion'' form

$$\epsilon(\omega)=\epsilon_0-\frac{f/2\omega}{\omega-\omega_0-i\gamma}, \tag{8}$$

where f is a parameter characterizing the oscillator strength of the electronic transition, ω_0 is the resonance frequency, γ is a decay constant, and ϵ_0 is the permittivity due to all other electronic states. This representation of $\epsilon(\omega)$ greatly simplifies the calculations, since the imaginary and real parts of $\epsilon(\omega)$ are given by analytic expressions that depend on the same parameters. We have attempted to describe the experimental transmission spectra (Figs. 2 and 3) using the permittivity for both transitions in the form (8). The quantities ϵ_0, f, ω_0, and γ served as adjustable parameters in the calculation of the transmission spectra. However, we were not able to obtain a satisfactory description of the experimental spectra. In the region of the absorption bands the transmission spectrum decreases much more rapidly than Lorentzian curves. This is due to the fact that, strictly speaking, the expression (8) is applicable for an isolated electronic-excitation band. Intrinsic absorption bands in the condensed state are, as a rule, a superposition of a large number of vibronic transitions with participation of intramolecular vibrations. Individual vibronic transitions are broadened because of structural disordering and temperature, forming a wide structureless band. The absorption bands F_a and F_b are a superposition of such vibronic transitions. In this case the permittivity $\epsilon(\omega)$ in the form (8) can be used to describe the spectrum when the structureless contour formed by a superposition of vibronic transitions accidently happens to be close to a Lorentzian curve. For this reason, the calculations of the transmission spectrum were performed with a more general form of $\epsilon(\omega)$ for each band. In this case, first, the form of only the imaginary part $\epsilon_2(\omega)$ of the permittivity was given and $\epsilon_1(\omega)$ was calculated on the basis of the analytic properties of the function $\epsilon(\omega)$. The imaginary and real parts of the permittivity are related by the Kramers–Kronig relation. This makes it possible to determine $\epsilon_1(\omega)$ from the given function $\epsilon_2(\omega)$ as[24]

$$\epsilon_1(\omega)=\epsilon_0+\frac{2}{\pi}\int\frac{x\epsilon_2(x)}{x^2-\omega^2}dx, \tag{9}$$

where we take the Cauchy principal value of the integral. The spectral dependence $\epsilon_1(\omega)$ was calculated by numerically integrating of Eq. (9) for two transitions F_a and F_b. Much better agreement with experiment was obtained by using Gaussians for $\epsilon_2(\omega)$. On the basis of this analysis, the calculation of the optical transmission spectra of films with various thicknesses and the determination of $\epsilon_2(\omega)$ from the experimental spectrum were performed in two steps according to the following scheme.

First, $\epsilon_2(\omega)$ was approximated by a sum of two Gaussians:

$$\epsilon_2(\omega)=\sum H_i\exp\left[-\ln 2\left(\frac{\omega-\omega_i}{c_i}\right)^2\right]. \tag{10}$$

The initial values of the adjustable parameters $H_{a,b}$, $c_{a,b}$, and $\omega_{a,b}$ were fixed, and the spectral dependence $\epsilon_1(\omega)$ was determined from the Kramers–Kronig relation (9) by numerical integration. The permittivity ϵ_0 served as another adjustable parameter. The functions $\epsilon_2(\omega)$ and $\epsilon_1(\omega)$ obtained in this manner were used to calculate $n(\omega)$and $\kappa(\omega)$ [Eqs. (6) and (7)] as well as the transmission spectrum (9). The optimal values of the parameters of the Gaussians and ϵ_0 were obtained by a least-squares fit of the computed curves (3) to the experimental spectra (Figs. 2 and 3). In Fig. 4 the experimental spectra (dots) are compared with the computed spectra (dashed curves) for films with $N=3$ and $N=8$. The total intensity of the transmission spectrum and the decrease in intensity with increasing film thickness can be described satisfactorily in the absorption and transmission regions of the film. However, the differences in the position of the peaks and in the shape of the low-frequency band F_a cannot be eliminated by using Gaussians to describe $\epsilon_2(\omega)$.

In the second step of the calculations, the spectral dependence $\epsilon_2(\omega)$ for the band F_a was not approximated by an analytic expression. To work with arbitrary curves the imaginary part of the permittivity was given by cubic spline interpolation over a set of discrete values $\epsilon_2(\omega_i)$ $(1\leqslant i\leqslant 15)$. This number of points was sufficient to describe the form of $\epsilon_2(\omega)$. The real part of the permittivity $\epsilon_1(\omega)$ was calculated

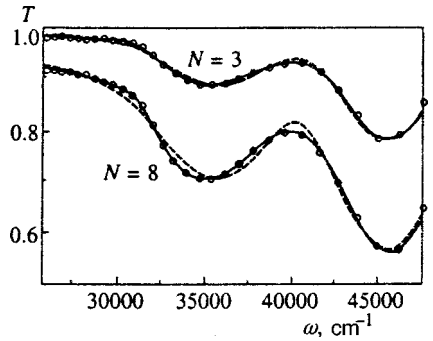

FIG. 4. Comparison of the experimental spectra (dots) with the computed spectra for films with $N=3$ and $N=8$ smectic layers. Dashed curves — calculation with $\epsilon_2(\omega)$ given by a Gaussian. Solid lines — calculaton with an arbitrary form of $\epsilon_2(\omega)$.

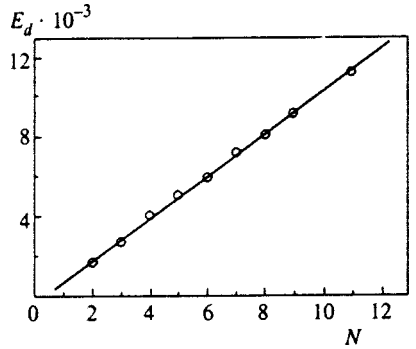

FIG. 6. $E_d = d\int \epsilon_2(\omega) d\omega$ versus film thickness N.

from the Kramers–Kronig relation (9). In the second step of the calculations the ordinates $\epsilon_2(\omega_i)$ served as adjustable parameters. Next, $T(\omega)$ [the expression (3)] was fit to the experimental spectrum using the same scheme as in the first step of the calculations. The spectra obtained are presented in Fig. 4 (solid lines). It follows from this figure that this method makes it is possible to obtain in the region of the F_a band a computed spectrum that is essentially identical to the experimental spectrum. Figures 5 and 6 display $E = \int \epsilon_2(\omega) d\omega$ and $E_d = d\int \epsilon_2(\omega) d\omega$ as a function of N. Figure 7 shows the temperature variation of $E = \int \epsilon_2(\omega) d\omega$ for a film with 13 smectic layers. The figure also shows for comparison the results obtained in the first computational step, where $\epsilon_2(\omega_i)$was described by a Gaussian. One can see from Fig. 3 that the spectral dependence $\epsilon_2(\omega_i)$ obtained in the second computational step is appreciably different from a Gaussian. However, the integrated intensities do not differ much (Fig. 7) and they show the same temperature variation.

3. DISCUSSION

A characteristic feature of the behavior of $E = \int \epsilon_2(\omega) d\omega$ as a function of film thickness is that in ultrathin films E decreases (the fact that the straight line drawn through E_d in Fig. 6 does not pass through zero is due to this effect). As already mentioned, $E = \int \epsilon_2(\omega) d\omega$ characterizes the orientational ordering of the molecules. Since the dipole moment of an optical transition for the F_a band is parallel to the long axis of a molecule, lower values of E correspond to lower values of the angle $\overline{\theta}$ of disorientation of the molecules relative to the ''director'' **n** (Fig. 1) and therefore a higher degree of orientational ordering. As temperature decreases, E decreases (Fig. 7, $N=13$), i.e., the orientational order of the molecules in the film increases. The temperature behavior of the orientational order in a film is similar to the change in the ordering in bulk samples in the smectic and nematic phases.[26,27] The fact that the molecules in a film become ordered on cooling can also be seen qualitatively by comparing the transmission spectra directly (Fig. 3). The decrease in the intensity of the F_a band on cooling is due to the increase in orientational order.

The quantity $P_2 = (1/2)(3\langle\cos^2\theta\rangle - 1)$, called the degree of orientational ordering,[26] can be used to characterize the orientational order of molecules in bulk samples. In our case (the electronic transition dipole moment is parallel to the long axis of the molecules) the dependence of P_2 on E has the simple form

$$P_2 = (1 - 3E/E^{\parallel}), \tag{11}$$

where $E^{\parallel} = \int \epsilon_2^{\parallel}(\omega) d\omega$ and $\epsilon_2^{\parallel}(\omega)$ is the permittivity in a direction parallel to the ''director'' **n** in a completely ordered structure, i.e., with $P_2 = 1$. In thick films P_2 should correspond to its value in the interior. For P_2 of the order of 0.65 in films with $N \geqslant 8$, the relative variation of the orientational ordering P_2 in ultrathin films ($N=2,3$) can be estimated using Eq. (11) and the values of E for $N=2$, 3 and $N \geqslant 8$. The degree of orientational ordering obtained in this manner is $P_2 \approx 0.72$ ($N=2$) and $P_2 \approx 0.69$ ($N=3$). The sharp depen-

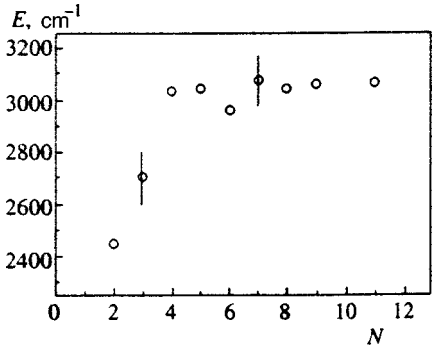

FIG. 5. Total intensity $E = \int \epsilon_2(\omega) d\omega$ versus film thickness N.

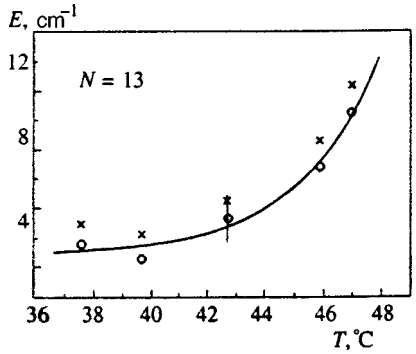

FIG. 7. $E = \int \epsilon_2(\omega) d\omega$ versus temperature for a film with $N=13$ smectic layers (○) and the values of $E = \int \epsilon_2(\omega) d\omega$ obtained at the first step of the calculations where $\epsilon_2(\omega)$ was given by a Gaussian (×).

dence $E(N)$ for $N \leqslant 4$ (Fig. 5) shows that the effect of the surface is limited to a thin near-surface layer of the order of one or two smectic planes.

A number of models have been proposed to describe the layer and orientational order of molecules near a surface. Rosenblatt and Ronis[13,14] were the first to use a lattice model to study the effect of a surface on the structure of and phase transitions in films. The material parameters of the bulk samples were used in the numerical calculations. A fundamentally important result of these works is that the order parameter near a surface is different from its value in the interior, and the thickness-averaged orientational order parameter in thin films is larger than in the interior. Surface ordering is often interpreted in terms of the wetting of the interface between media by various phases.[16–18] It has been shown that near the phase transition temperature in the interior this process can occur by continuous and layer-by-layer increase in the number of smectic layers near the surface.[17,18] In the mean-field model (the analog of the McMillan theory[28] for bulk samples) the influence of the surface was also taken into account by introducing an effective orienting field acting on a molecule at the surface

$$W_s(\theta) = -\frac{W_s}{2}(3\cos^2(\theta) - 1).$$

The absolute value of P_2 on the surface depends on the ratio W_s/W_0, where W_0 is the intermolecular interaction constant. The large arbitrariness in the choice of these quantities precludes a quantitative comparison of the experimental and computed values of P_2 at a surface. More important is the theoretical dependence of P_2 on the distance from the surface. The change in P_2 from its value at the surface to its value in the interior occurs mainly at the transition from the first (surface) to the second molecular layer,[20] in agreement with our data.

The change of the collective and molecular dynamics at the surface makes a large contribution to the ordering of the molecules. The amplitude of the fluctuations of the smectic layers in the interior of a film is ~ 0.45 nm. Surface tension suppresses collective (long-wavelength) fluctuations, decreasing their amplitude by 0.1–0.2 nm relative to the amplitude of the fluctuations in the interior of the film.[19] Precise measurements of the surface tension have been performed on thin films.[29] According to these data the surface tension does not depend on the film thickness, right down to two molecular layers, i.e., the mechanism responsible for surface tension is localized in a layer near the surface.[29] According to the calculations in Ref. 19 the amplitude of the surface fluctuations depends weakly on the film thickness. The bulk of the change in the amplitude of the fluctuations occurs at the surface over one or two molecular layers. Our results on the effect of a surface on orientational ordering agree with this model.

In summary, a spectral method for investigating surface orientational ordering was proposed in this paper. The first measurements of the transmission spectra of thin, free-standing, molecular films in the region of electronic absorption bands were performed.[1)] The degree of orientational order in ultrathin films ($N = 2,3$) is greater than in thick films. Our results on the thickness dependence of the transmission spectra and permittivity show that the effect of a surface on the orientational structure is localized in one or two smectic layers near the surface.

This work was supported by the Russian Fund for Fundamental Research (Project 98-02-16639) and the State Scientific and Technical Program ''Statistical Physics.''

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

1)The electronic spectra of molecules forming liquid crystals are investigated, as a rule, in solutions. The intrinsic absorption spectrum in the liquid-crystal state is difficult to measure by the conventional method, i.e., when the substance is in a cell, because of the virtually complete absorption even in thin samples $\sim 1\mu$m. Free-standing films make it possible to perform such measurements and can be used to investigate intrinsic electronic and vibronic absorption.

[1] B. M. Ocko, A. Braslau, and R. S. Pershan, Phys. Rev. Lett. **57**, 94 (1986).
[2] Y. Galerne and L. Liebert, Phys. Rev. Lett. **64**, 906 (1990).
[3] P. Pieranski *et al.*, Physica A **194**, 364 (1993).
[4] T. Stoebe and C. C. Huang, Int. J. Mod. Phys. B **9**, 2285 (1995).
[5] S. Henekamp, R. A. Pelcovits, E. Fontes, E. Yi. Chen, and R. Pindak, Phys. Rev. Lett. **52**, 1017 (1984).
[6] S. M. Amador and P. S. Pershan, Phys. Rev. A **41**, 4326 (1990).
[7] Ch. Bahr and D. Fleigner, Phys. Rev. A **46**, 7663 (1992).
[8] P. O. Andreeva, V. K. Dolganov, and K. P. Meletov, JETP Lett. **66**, 442 (1997).
[9] D. Sclauf and Ch. Bahr, Phys. Rev. E **57**, R1235 (1998).
[10] A. J. Jin, T. Stoebe, and C. C. Huang, Phys. Rev. E **49**, 4791 (1994).
[11] V. K. Dolganov, R. Fouret, and C. Gors, JETP Lett. **63**, 285 (1996).
[12] T. Stoebe, P. Mach, and C. C. Huang, Phys. Rev. Lett. **73**, 1384 (1994).
[13] D. Ronis and C. Rosenblatt, Phys. Rev. A **21**, 1687 (1980).
[14] C. Rosenblatt and D. Ronis, Phys. Rev. A **23**, 305 (1981).
[15] K. Binder, *Phase Transitions and Critical Phenomena*, C. Domb and J. L. Leibowitz (eds.), Academic Press, London (1986), Vol.8.
[16] D. E. Sullivan and M. M. Talo de Gama, *Fluid Interfacial Phenomena*, C. A. Croxton (eds.), Wiley, New York (1986), p. 45.
[17] Z. Pawlowska, G. F. Kventsel, and T. J. Sluckin, Phys. Rev. A **36**, 992 (1987).
[18] Z. Pawlowska, G. F. Kventsel, and T. J. Sluckin, Phys. Rev. A **38**, 5342 (1988).
[19] R. Holist, Phys. Rev. A **44**, 3692 (1991).
[20] L. M. Mirantsev, Phys. Lett. A **205**, 412 (1995).
[21] E. I. Demikhov, V. K. Dolganov, and K. P. Meletov, Phys. Rev. E **52**, R1285 (1995)
[22] V. K. Dolganov, E. I. Demikhov, R. Fouret, and C. Gors, Phys. Lett. A **220**, 242 (1996).
[23] M. Born and E. Wolf, *Principles of Optics*, Pergamon Press, New York (1964).
[24] L. D. Landau and E. M. Lifzhitz, *Electrodynamics of Continuous Media*, 2nd ed., Pergamon Press, New York (1984).
[25] A. S. Davydov, *The Theory of Solids*, Nauka, Moscow (1976).
[26] P. G. de Gennes, *The Physics of Liquid Crystals*, Clarendon Press, Oxford (1975).
[27] L. M. Blinov, *Electro-Optical and Magneto-Optical Properties of Liquid Crystals*, Wiley, New York (1983).
[28] W. L. McMillan, Phys. Rev. A **4**, 1238 (1971).
[29] P. Mach, S. Grantz, D. A. Debe, T. Stoebe, and C. C. Huang, J. Phys. II France **5**, 217 (1995).

Translated by M. E. Alferieff

Coulomb screening of impurity charge and anomalous tunneling transparency

L. A. Manakova*)

Kurchatov Institute, 123182 Moscow, Russia
(Submitted 17 September 1998)
Zh. Éksp. Teor. Fiz. **115**, 1843–1859 (May 1999)

This paper analyzes the effect of the screened Coulomb interaction between metallic electrons in the sidewalls, on the one hand, and a localized electron in an impurity level, on the other, on the tunneling in doped quantum structures with an intrinsic two-dimensional continuum. We show that Mahan's non-Fermi-liquid singularity at the Fermi level is unstable against additional scattering due to tunneling. As a result, the current–voltage characteristic changes radically when the Fermi level in the sidewalls is approached by the edge of the two-dimensional band. Specifically, the peak due to the non-Fermi-liquid singularity with a section of negative differential resistance is replaced with a step-like or a two-step feature, which corresponds to a single or split Fermi-liquid resonance near the edge of the 2D band involved in the tunneling process. © *1999 American Institute of Physics.* [S1063-7761(99)02305-7]

1. INTRODUCTION

Among the physical realizations of the non-Fermi-liquid behavior, the models that have been studied most extensively by both theoretical and experimental methods are generalized versions of Anderson's impurity model, along with the multichannel Kondo model, both spin and orbital, and in particular, the two-channel Kondo model, a prototype of all non-Fermi-liquid impurity models (see the review, for example, in Ref. 1).

Tunneling experiments open up an opportunity for direct detection of non-Fermi-liquid effects in such systems using features of current–voltage characteristics.[2] For this reason, studies of tunneling spectra of impurity systems with non-Fermi-liquid ground states are of fundamental importance. Detecting a crossover from a non-Fermi-liquid to Fermi-liquid state is probably feasible, since the non-Fermi-liquid state is unstable against all perturbations that lift degeneracy in orbital or spin degrees of freedom.

In particular, two mechanisms of instability in the non-Fermi-liquid state have been described by the two-channel orbital (quadrupole) Kondo model. The first is instability against distortions of the impurity center which lower its symmetry (Jahn–Teller effect or pseudo-effect)[3] and lift the orbital degeneracy of an impurity level, and accordingly return the system to the Fermi-liquid behavior at low temperatures.

The second mechanism[4] is responsible for instability of the non-Fermi-liquid state due to the anisotropy of scattering channels (recall that the scattering channel indices in the two-channel orbital Kondo model correspond to two projections of the electron spin). The channel anisotropy arises under an external magnetic field, and the corresponding crossover from the non-Fermi-liquid to Fermi-liquid state has been detected experimentally.[2]

It was shown previously[5,6] that a new instability mechanism of the non-Fermi-liquid state occurs in tunneling structures. Tunneling mechanisms in doped quantum structures with an intrinsic two-dimensional continuum were analyzed. A new physical realization of the two-channel Kondo model was described, and a crossover from a non-Fermi-liquid to Fermi-liquid state due to a variation in either the separation between the Fermi level and two-dimensional band edge or the impurity level depth was detected.

The physical reason for the crossover is the existence of Fermi-liquid resonances near the two-dimensional band edge. The resonances are generated in the process of tunneling due to the scattering of electrons of the defect layer of the quantum structure by non-Fermi-liquid excitations from the Fermi level in the sidewalls. In this case, the impurity level is the lowest and does not resonate with the Fermi level in the sidewalls.

Thus, it has been proven[5,6] that the non-Fermi-liquid state is unstable against impurity scattering (both resonant and potential) engendered by tunneling in the situation under discussion.

The contribution of the edge resonances to the tunneling current is considerably larger than the current in the non-Fermi-liquid state. For this reason, the described instability of the non-Fermi-liquid state sidewalls to an anomalous increase in the tunneling transparency and current.

The key role of the two-dimensional continuum in the tunneling in doped quantum structures was first demonstrated in Refs. 7 and 8.

One example of a system with non-Fermi-liquid excitations at the Fermi level[9,10] that has been well known for a long time is a system with Coulomb interaction between a localized electron trapped at a nondegenerate impurity level and conduction electrons in the sidewalls. Matveev and Larkin[11] considered the Coulomb interaction between a localized electron in an impurity level and metallic electrons in the sidewalls in the context of tunneling via a resonant level under a barrier. In this case, when the dominant role is played by Mahan's resonance at the Fermi level,[12] the current–voltage characteristic is a power law, and it contains a section of negative differential resistivity.

The present paper considers the effect of a screened Coulomb interaction between metallic electrons in sidewalls and a localized electron in an impurity level on the tunneling in doped quantum structures with an intrinsic two-dimensional continuum.

It turns out that the tunneling mechanisms in quantum structures with an intrinsic two-dimensional continuum are radically different from those taking place in systems with both tunneling via a resonant level under a barrier[11] and a non-Fermi-liquid state due to the two-channel orbital Kondo scattering.[5,6] An experimentally important result is that the current–voltage characteristic radically changes when the Fermi level in the sidewalls approaches the two-dimensional band edge: the peak with the section of negative differential resistivity due to Mahan's non-Fermi-liquid feature at the Fermi level is replaced by a step-like or a two-step feature, which are due to the contribution of Fermi-liquid edge resonances to the tunneling current.

2. STATEMENT OF THE PROBLEM

1. Consider a quantum structure with a band-edge profile similar to that of a double-barrier quantum well (DBQW), such as GaAlAs/GaAs/GaAlAs layered structures, where the interior GaAs layer is a quantum well with an intrinsic two-dimensional continuum of spatially quantized band states.

A transition-metal impurity generates a deep level with a binding energy E_d within the band gap of the DBQW inner layer. This layer also has a continuum of two-dimensional states with dispersion $\varepsilon_{\mathbf{k}_\perp}$. We are considering the situation when the Fermi level in the sidewalls is close to the conduction band edge in the inner layer. The system Hamiltonian has the form

$$H=H_{00}+H_t+H_{\rm int}, \tag{1}$$

where $H_{00}=H_{00}^{\nu}+H_{00}^{d}+H_{00}^{c}$ is the Hamiltonian of the decoupled sidewalls and quantum well.

The tunneling Hamiltonian for the system under consideration can be written in the form

$$H_t=H_{td}+H_{tc}=\sum_{\mathbf{k}\nu\sigma}(T_{\mathbf{k}d}^{\nu}a_{\mathbf{k}\nu\sigma}^{+}d_\sigma+\text{H.c.})$$

$$+\sum_{\mathbf{k}\nu\sigma}\sum_{\mathbf{k}_\perp'}(T_{\mathbf{k}\mathbf{k}_\perp'}^{\nu}a_{\mathbf{k}\nu\sigma}^{+}c_{\mathbf{k}_\perp'\sigma}+\text{H.c.}). \tag{2}$$

Operators $a_{\mathbf{k}\nu\sigma}$ describe electron states in the left-hand (L) and right-hand (R) sidewalls of the tunneling junction. Operators d_σ and $c_{\mathbf{k}_\perp}$ correspond to wave functions of hybridized localized, $\psi_d(\mathbf{r})$, and band, $\Psi(\mathbf{k}_\perp,\mathbf{r})$, states.[13]

The tunneling matrix elements in Eq. (2) are given by

$$T_{\mathbf{k}d}^{\nu}=B(\mathbf{k}_\perp)T_d^{\nu}(k_l),\quad T_{\mathbf{k},\mathbf{k}_\perp'}^{\nu}=(T_0^{\nu}(k_l)\delta_{\mathbf{k}_\perp\mathbf{k}_\perp'}+T_{\mathbf{k}c}^{\nu}B(\mathbf{k}_\perp')),$$

$$T_{\mathbf{k}c}^{\nu}=T_c^{\nu}(k_l)B(\mathbf{k}_\perp). \tag{3}$$

Here $B(\mathbf{k}_\perp)=V_{\mathbf{k}_\perp d}/(E_d-\epsilon_{\mathbf{k}_\perp})$, $V_{\mathbf{k}_\perp d}$ is the matrix element of hybridization in the quantum well. In Eq. (3) $\mathbf{k}=\mathbf{k}_\perp,k_l$, and the longitudinal and transverse motions of electrons in the sidewalls are assumed to be decoupled: $\varepsilon_{\mathbf{k}}=\varepsilon_{\mathbf{k}_\perp}+\varepsilon_{k_l}$. As a result, the impurity contribution to $T_{\mathbf{k},\mathbf{k}_\perp'}^{\nu}$ is separable, which enables us to obtain an exact solution of the tunneling problem.

The ''bare'' tunneling between the sidewalls and quantum well, described by the term containing $T_0^{\nu}(k_l)$ in $T_{\mathbf{k},\mathbf{k}_\perp'}^{\nu}$ controls the restructuring of the electronic spectrum near the band edge in the quantum well. This is the region where evanescent edge states are formed,[7] whose density of states for $\varepsilon-\varepsilon_c<\gamma_0\ll W_c$ is

$$\rho_c(\varepsilon)=\frac{\rho_{0c}}{\pi}\left[\tan^{-1}\frac{\varepsilon-\varepsilon_c}{\gamma_0}-\tan^{-1}\frac{\varepsilon-W_c}{\gamma_0}\right]. \tag{4}$$

Here ρ_{0c} is the constant density of states in the unperturbed 2D band, $\gamma_0\sim\Sigma_\nu|T_0^{\nu}(\varepsilon_c)^2\rho_{0\nu}$, where $\rho_{0\nu}\sim W_a$ is the corresponding tunneling width, and W_a is the width of the conduction band in the sidewalls.

Thus, in the energy range of interest, we have in the tunneling Hamiltonian H_{tc} only the impurity term proportional to $T_{\rm c}^{\nu}(k_l)$ in Eq. (3), but the 2D-continuum density of states is expressed by Eq. (4).

The tunneling Hamiltonian H_t can be transformed to a ''single-band'' form, which is more convenient for analysis. This is done using a linear transform

$$a_{\mathbf{k}\sigma}=u_{\mathbf{k}}a_{\mathbf{k}\mathrm{L}\sigma}+v_{\mathbf{k}}a_{\mathbf{k}\mathrm{R}\sigma},\quad b_{\mathbf{k}\sigma}=u_{\mathbf{k}}a_{\mathbf{k}\mathrm{R}\sigma}-v_{\mathbf{k}}a_{\mathbf{k}\mathrm{L}\sigma},$$

$$u_{\mathbf{k}}=\frac{T_{\mathbf{k}d}^{\mathrm{L}}}{[(T_{\mathbf{k}d}^{\mathrm{L}})^2+(T_{\mathbf{k}d}^{\mathrm{R}})^2]^{1/2}},\quad u_{\mathbf{k}}^2+v_{\mathbf{k}}^2=1. \tag{5}$$

It can be verified immediately that in the new representation only quasiparticles of one sort represented by operators $a_{\mathbf{k}\sigma}$ are hybridized in both localized and continuum states.

The transformed tunneling Hamiltonian $H_t^{(a)}$ can be derived from H_t in Eq. (2) through the transformations

$$a_{\mathbf{k}\nu\sigma}^{+}\to a_{\mathbf{k}\sigma},\quad T_{\mathbf{k}d}^{\nu}\to T_{\mathbf{k}d}^{a}=[(T_{\mathbf{k}d}^{\mathrm{L}})^2+(T_{\mathbf{k}d}^{\mathrm{R}})^2]^{1/2},$$

$$T_{\mathbf{k}\mathbf{k}_\perp'}^{\nu}\to T_{\mathbf{k}\mathbf{k}_\perp'}^{a}=T_{\mathbf{k}\mathbf{k}_\perp'}^{\mathrm{L}}u_{\mathbf{k}}+T_{\mathbf{k}\mathbf{k}_\perp'}^{\mathrm{R}}v_{\mathbf{k}}.$$

Here the dispersion relations in the sidewalls are assumed to be identical: $\varepsilon_{\mathbf{k}}^{\mathrm{L}}=\varepsilon_{\mathbf{k}}^{\mathrm{R}}=\varepsilon_{\mathbf{k}\mathbf{a}}$. Since the tunneling Hamiltonian acts only on states $a_{\mathbf{k}\sigma}$, the interaction Hamiltonian is defined in their basis.

2. The singularities in the energy range of interest near the edge of the 2D band are caused by the term H_{tc} from the tunneling Hamiltonian and the Hamiltonian $H_{\rm int}$ of interaction between the electrons in the sidewalls and impurity states within the quantum well. Using the formalism developed earlier,[5–7] it is convenient to diagonalize Hamiltonian (1) in two stages. In the first stage, we diagonalize the Hamiltonian $H_0=H_{00}+H_{\rm int}$ and obtain multiparticle excitations at the Fermi level in the sidewalls. Then we take into account the additional scattering of quasi-two-dimensional electrons in the quantum well by these excitations due to the tunneling Hamiltonian H_t.

In this study, $H_{\rm int}$ is the Hamiltonian of screened Coulomb interaction due to polarization of conduction electrons in the sidewalls caused by recharging of impurity states in

the process of tunneling. By expressing $H_{\rm int}$ in terms of partial states $a_{k\mu\sigma}$, which are introduced for *d*-wave impurity states by the relationship

$$a_{\mathbf{k}\sigma}=\sum_{\mu} a_{k\mu\sigma}K_{d\mu}(\Omega_{\mathbf{k}}),$$

where $K_{d\mu}(\Omega_{\mathbf{k}})$ are cubic harmonics as functions of the solid angle $\Omega_{\mathbf{k}}$ and μ is the row number of the irreducible representation of the point group, we obtain

$$H_{ee}=\sum_{kk'\sigma}\sum_{\mu} V_{\mu}(kk')a^{+}_{k\mu\sigma}a_{k'\mu\sigma}d^{+}d. \tag{6}$$

As was shown previously,[8] the interaction H_{ee} is generated by two tunneling mechanisms, namely, $T^{a}_{\mathbf{k}d}$ and $T^{a}_{0}(k_l)$ between the sidewalls and the 2D continuum. The second mechanism is a manifestation of the ''Bloch tail'' of the impurity state wave function. Specifically, the recharging of Bloch states near the 2D band edge generates charge fluctuations at the deep level, hence the interaction described by Eq. (6).

This is the basic interaction (6) in the so-called non-Kondo regime, in the absence of the exchange scattering. At the *d*-level, such a situation is plausible when both the orbital and spin degeneracy of the level are lifted by the combined action of crystal fields, Jahn–Teller distortions, and Hund's rule.

It is known[9,11] that solving a problem with the Coulomb interaction requires a description of two processes of different physical character, which lead to an infrared divergence in the Green's function and tunneling characteristics. The first process is electron scattering by the potential

$$\sum_{kk'\sigma}\sum_{\mu} V_{\mu}(kk')a^{+}_{k\mu\sigma}a_{k'\mu\sigma},$$

acting over a finite time interval. This scattering always leads to a divergence in the Green's function $G_{kk'}(\varepsilon)$ of conduction electrons, which corresponds to Mahan's resonance at the Fermi level.[12] The second process is related to the ''orthogonality catastrophe:'' even a weak potential due to charge fluctuations at the impurity level modifies the multiparticle electron wave function so that it becomes almost orthogonal to the initial wave function. This process, which also effectively broadens the impurity level, spreads Mahan's resonance in the conduction electron spectrum.[9]

Below we will use the electron density of states in the sidewalls,

$$\rho_{a\mu}(\varepsilon)=-\frac{1}{\pi}\,\mathrm{Im}\sum_{k} G_{kk}(\varepsilon)\equiv-\frac{1}{\pi}\,\mathrm{Im}\,G(\varepsilon),\quad \varepsilon>0,$$

where $G(\varepsilon)$ is a sum over the momenta of delocalized electrons scattered by a potential that is turned on abruptly (at $t=0$). This function was calculated by Nozières and de Dominicis:[9]

$$G(\varepsilon)=i\rho_{0a}\exp(i\pi\alpha_{\mu})\Gamma(\alpha_{\mu})\left(\frac{\varepsilon_{\rm F}}{\varepsilon}\right)^{\alpha_{\mu}},\quad \alpha_{\mu}=\frac{2\delta_{\mu}}{\pi}, \tag{7}$$

hence the density of states is

$$\rho_{a\mu}(\varepsilon)=A_{\mu}\rho_{0a}\left(\frac{\varepsilon_{\rm F}}{\varepsilon}\right)^{\alpha_{\mu}},\quad A_{\mu}=(1/\pi)\sin(\pi\alpha_{\mu})\Gamma(\alpha_{\mu}). \tag{8}$$

Here $\rho_{0a}\sim\varepsilon_{\rm F}^{-1}$ is the density of states at the Fermi level for noninteracting electrons, $\Gamma(\alpha)$ is the gamma function, and δ_{μ} is the phase shift in the scattering channel μ.

Here we also recall that for $\alpha_{\mu}\to 0$, the last factor in Eqs. (7) and (8) should be written in a more accurate form, namely

$$[(\varepsilon_{\rm F}/\varepsilon)^{\alpha_{\mu}}-1],$$

so that the density of states is given by

$$\rho_{a\mu}(\varepsilon)\sim\rho_{0a}\alpha_{\mu}\ln\frac{\varepsilon_{\rm F}}{\varepsilon},$$

and the singularities at the Fermi level vanish in self-energy functions(see below), and accordingly in $T_0(z)$.

The Green's function $G_d(z)$ of the localized state also has a well-known form[9] determined by the modification of the Fermi sea caused by the impurity potential:

$$G_d(z)\sim\frac{1}{(\varepsilon_{\rm F})^{\alpha_d}(z-\overline{E}_d)^{1-\alpha_d}},$$

$$\alpha_d=2\sum_{\mu}\left(\frac{\delta_{\mu}}{\pi}\right)^{2},\quad \alpha_d\ll\frac{\gamma_d}{|\varepsilon_d|}\ll 1, \tag{9}$$

$$\overline{E}_d=E_d+\Sigma_d(\overline{E}_d)\equiv\varepsilon_d+i\gamma_d,$$

$$\Sigma_d(\overline{E}_d)\approx\sum_{\mu}|T^{a}_{k_{\rm F}d\mu}|^2\rho_{0a}\left(\frac{\varepsilon_{\rm F}}{\overline{E}_d-\varepsilon_c}\right)^{\alpha_{\mu}}. \tag{10}$$

In the case of a deep level discussed in this paper, the following conditions are satisfied: $|\varepsilon_d|\equiv\varepsilon_c-\varepsilon_d\sim\varepsilon_{\rm F}$, $\gamma_d\ll|\varepsilon_d|$.

The elementary excitations at the Fermi level described by the Green's function (7) are boson-like electron–hole pairs.[10]

3. GENERAL EXPRESSION FOR THE SCATTERING MATRIX AND ITS FEATURES

1. We now discuss additional scattering of quasi-two-dimensional electrons in the defect layer by electron–hole pairs at the Fermi level in the sidewalls due to the tunneling Hamiltonian H_{tc}.

The scattering matrix $\mathscr{T}^{cc}_{\sigma}(\mathbf{k}_{\perp},\mathbf{k}'_{\perp};z)$ for an electron inside the quantum well can be derived[5–7] from the Green's function

$$G^{cc}_{\sigma}(\mathbf{k}_{\perp},\mathbf{k}'_{\perp};z)=\delta_{\mathbf{k}_{\perp},\mathbf{k}'_{\perp}}G_{0\mathbf{k}_{\perp}}(z)+G_{0\mathbf{k}_{\perp}}(z)\times\mathscr{T}^{cc}_{\sigma}(\mathbf{k}_{\perp},\mathbf{k}'_{\perp};z)G_{0\mathbf{k}'_{\perp}}(z), \tag{11}$$

which yields

$$\mathscr{T}^{cc}_{\sigma}(\mathbf{k}_{\perp},\mathbf{k}'_{\perp};z)=\frac{T_0(z)}{1-T_0(z)J_c(z)}B(\mathbf{k}_{\perp})B^{*}(\mathbf{k}'_{\perp}), \tag{12}$$

$$T_0(z)=|\Sigma_{dc}(z)|^2G_{d\sigma}(z)+\Sigma_{cc}(z). \tag{13}$$

Equation (11) uses the notation $G_{0\mathbf{k}_\perp}(z)=[z-\tilde{\varepsilon}_{\mathbf{k}\perp}]^{-1}$, and $\tilde{\varepsilon}_{\mathbf{k}\perp}$ is the spectrum corresponding to the density of states described by Eq. (4).

It follows from the definition (12) of the scattering matrix that $T_0(z)$ acts as an effective scattering potential for 2D electrons. The first term of $T_0(z)$ is responsible for the resonant scattering involving virtual transitions between the 2D continuum and impurity states in the quantum well via electron states in the sidewalls. The second term in $T_0(z)$ describes potential scattering of quasi-2D electrons with amplitude $\Sigma_{cc}(z)$.

The functions $\Sigma_{cc}(z)$ and $\Sigma_{dc}(z)$ can be conveniently expressed in the form of a spectral representation of the conduction electron Green's function:[5,6]

$$\Sigma_{cc}(z)=\sum_{k\mu\sigma}\frac{|T^a_{kc\mu}|^2 f(\varepsilon_{ka})}{(z-\varepsilon_c)-(\varepsilon_{ka}-\varepsilon_c)}$$

$$=\sum_{\mu}|T^a_{k_{\rm F}c\mu}|^2\int_{-\infty}^{0}d\varepsilon\,\frac{\rho_{a\mu}(\varepsilon)}{(z-\varepsilon_c)-\varepsilon}. \qquad (14)$$

Here the energy is measured with respect to $\varepsilon_{\rm F}\to\varepsilon_c$, and $f(\varepsilon)$ is the Fermi distribution function.

It also follows from Eq. (14) that in an interacting system the functions $\Sigma_{cc}(z)$ and $\Sigma_{dc}(z)$ are Hilbert transforms of the multiparticle density of states; for this reason, they have features at the Fermi level corresponding to Mahan's peaks in the density of states. By substituting Eq. (8) in (14), we obtain

$$\Sigma_{cc}(z)=\sum_{\mu}|T^a_{k_{\rm F}c\mu}|^2 A_\mu\rho_{0a}\left(\frac{\varepsilon_{\rm F}}{z-\varepsilon_c}\right)^{\alpha_\mu},\quad A_\mu\sim 1. \qquad (15)$$

The form of the expression for $\Sigma_{dc}(z)$ is similar to Eq. (15) in the energy range of interest, but $|T^a_{k_{\rm F}c\mu}|^2$ is replaced by $T^a_{k_{\rm F}c\mu}T^{a*}_{k_{\rm F}d\mu}$.

The integral $J_c(z)$ is the Hilbert transform of the quasi-two-dimensional density of states $\rho_c(\varepsilon)$ defined by Eq. (4). In the region $|z-\varepsilon_c|/\gamma_0\ll 1$, this integral is logarithmically divergent:

$$J_c(z)=\frac{1}{2}\tilde{\rho}_{0c}\,\mathrm{Ln}\frac{\gamma_0}{z-\varepsilon_c},\quad \tilde{\rho}_{0c}=\rho_{0c}(\varepsilon_c)|B(\varepsilon_c)|^2. \qquad (16)$$

This logarithmic divergence in the self-energy part of $J_c(z)$ generates one-particle resonances in the same energy range where the multiparticle resonance responsible for the peak in the effective scattering potential $T_0(z)$ occurs. For this reason, poles of the scattering matrix are determined by the self-consistent equation

$$1-T_0(z)J_c(z)=0. \qquad (17)$$

In the absence of the interaction, $T_0(z)$ is a function slower than $J_c(z)$, so one-particle resonances near the edge of the 2D band obtained previously[7] are fully dependent on the logarithmic divergence in $J_c(z)$, and they are therefore exponentially narrow.

2. Consider solutions of Eq. (17) in the presence of Coulomb interaction. By substituting expressions (9) and (15) in that for $T_0(z)$ and considering for simplicity only one scattering channel μ_0 where the phase shift is maximal, we transform Eq. (17) for poles of the scattering matrix $\mathscr{T}^{cc}$ to

$$1-\left[\gamma_{d\mu_0}\left(\frac{\varepsilon_{\rm F}}{z-\varepsilon_c}\right)^{\alpha_{\mu_0}}+\gamma^2_{d\mu_0}\rho_{0a}\left(\frac{\varepsilon_{\rm F}}{|z-\varepsilon_c|}\right)^{2\alpha_{\mu_0}}\times\left(\frac{\varepsilon_{\rm F}}{z-\overline{E}_d}\right)^{1-\alpha_d}\right]J_c(z)=0, \qquad (18)$$

where $\gamma_{d\mu_0}\equiv|T^a_{k_{\rm F}d\mu_0}|^2\rho_{0a}A_{\mu_0}\sim\gamma_d\sim\gamma_0|B|^2$.

It is clear that Mahan's resonance and the orthogonality catastrophe contribute to different quantities that control $T_0(z)$. The resonance in the Green's function of conduction electrons determines the density of states in spectral distributions of self-energy functions $\Sigma_{cc}(z)$ and $\Sigma_{dc}(z)$. The channel related to the orthogonality catastrophe determines the Green's function $G_d(z)$ of the impurity state. We are interested in solutions $z_r=\varepsilon_r+i\gamma_r$ of Eq. (18) that correspond to resonances near the 2D band edge for the deep impurity level, so that

$$|z_r-\varepsilon_c|\ll|\overline{E}_d-\varepsilon_c|\sim\varepsilon_{\rm F}.$$

Under this condition, the major contribution to the effective scattering potential $T_0(z)$ is due to the resonant component, i.e., the second term in the brackets in Eq. (18), and the contribution of the power-law factor with $G_d(z)$ is of the order of unity. This means that the resonant scattering of electrons in the quantum well is fully determined by Mahan's peak at the Fermi level in the sidewalls.

The shape of the edge resonance is determined by the competition between features in the Hilbert transforms $\Sigma_{dc}(z)$ and $J_c(z)$ of the multiparticle and single-particle densities of states. The shape of these functions is determined by the interaction amplitude. If the interaction is weak and α_{μ_0} is sufficiently small, the changes in the function $\Sigma_{dc}(z)$ are much smaller than in $J_c(z)$ in the energy range of interest. In this case, an exponentially narrow resonance determined largely by the logarithmic divergence of $J_c(z)$ is feasible.[7] Using Eq. (18), one can easily prove that the condition for this resonance is

$$\Lambda^0_{dc}<\alpha_{\mu_0}\ll\Lambda^0_{dc}\ln\frac{1}{\Lambda^0_{dc}},\quad \Lambda^0_{dc}\equiv\gamma^2_d\rho_{0a}\rho_{0c}|B|^2\ll 1, \qquad (19)$$

and hereafter it is assumed that $\rho_{0a},\rho_{0c}\sim 1/\varepsilon_{\rm F}$. The first inequality in (19) is the condition that the density of states be a power-law function of energy.

For $\alpha>\alpha_c\equiv\Lambda^0_{dc}\ln(1/\Lambda^0_{dc})$ (hereafter we use the notation $\alpha_{\mu_0}\equiv\alpha$) the solution of Eq. (18) is determined by the multiparticle singularity in $\Sigma_{dc}(z)$ [or in other words, by the singularity in the effective potential $T_0(z)$], and the positions and widths of resonances are accordingly power-law functions of the tunneling structure parameters. The solutions of Eq. (18) in the energy range of interest can be written in the form

$$\varepsilon_{r^\pm}=\varepsilon_c\pm\gamma_r\cos\varphi,\quad \gamma_{r^\pm}=\gamma_r\sin\varphi, \qquad (20)$$

$$\gamma_r \approx \varepsilon_0 (\Lambda_{dc})^{1/2\alpha}, \quad \varepsilon_0 \sim \varepsilon_F, \quad \Lambda_{dc} \approx \frac{1}{2\alpha}\Lambda_{dc}^0 \ln\frac{1}{\Lambda_{dc}^0} \approx \frac{\alpha_c}{2\alpha}, \tag{21}$$

$$\varphi \approx \frac{1-\alpha_d}{2\alpha}\frac{\gamma_d}{|\varepsilon_d|}\ln\frac{1}{\Lambda_{dc}^0}. \tag{22}$$

The edge resonance is split into two components that are symmetric about the edge of the 2D band at $\varphi < \pi/4$, when $\gamma_{r^\pm} < |\varepsilon_{r^\pm} - \varepsilon_c|$. This condition defines the case of sufficiently strong scattering:

$$\alpha > 2\alpha_{c1}, \quad \alpha_{c1} \equiv \frac{1}{\pi}(1-\alpha_d)\frac{\gamma_d}{|\varepsilon_d|}\ln\frac{1}{\Lambda_{dc}^0} \gg \alpha_c. \tag{23}$$

Over a limited range of tunneling parameters, when $\pi/4 < \varphi \leqslant \pi/2$, or in other words for

$$\alpha_{c1} < \alpha \leqslant 2\alpha_{c1}, \tag{24}$$

Eq. (18) has a solution in the form of a single resonance of width γ_r at the edge of the 2D band.

Thus, over a significant range of tunneling parameters, the resonant scattering of quasi-two-dimensional electrons in the quantum well by the excitations from the Fermi level in the sidewalls, which are responsible for Mahan's resonance, gives rise to a split resonance in electronic Green's functions. For all admissible values of tunneling parameters,

$$|\varepsilon_r - \varepsilon_c|, \quad \gamma_r \ll \gamma_0. \tag{25}$$

Depending on the Fermi level position with respect to the 2D band edge in the quantum well, solutions of the self-consistent equation for poles of the scattering operator have the following forms. At $|\varepsilon_F - \varepsilon_c| < \gamma_d$ (in our notation, γ_d has the sense of Mahan's resonance width) the peak at the Fermi level is due to a power-law or exponential edge resonance. At $|\varepsilon_F - \varepsilon_c| > \gamma_d$ there is only a Mahan-like feature at the Fermi level, whose contribution to the tunneling current in the absence of a two-dimensional continuum was investigated by Matveev and Larkin.[11]

3. Along with the scattering operator $\mathscr{T}_\sigma^{cc}(\mathbf{k}_\perp \mathbf{k}_\perp^H z)$, the parameters whose behavior is determined by non-Fermi-liquid singularities in the density of states in the sidewalls are tunneling widths. They determine the tunneling transparency in the band channel at $|\varepsilon_F - \varepsilon_c| > \gamma_d$.

In the case of tunneling via the 2D continuum with a one-dimensional spectrum near the Fermi level, the tunneling widths have the form

$$\Gamma_\nu^c(\mathbf{p}_\perp; \varepsilon_k^\nu) = \sum_q \langle a_{q\nu}|H_{tc}|c_{\mathbf{p}_\perp}\rangle\langle c_{\mathbf{p}_\perp}|H_{tc}|a_{q\nu}\rangle$$

$$= \sum_k |T_{k\mathbf{p}_\perp}^\nu|^2 \delta(\varepsilon - \varepsilon_k^\nu). \tag{26}$$

By expressing $a_{k\mathrm{L,R}}$ in Eq. (26) in terms of a_k, b_k with the help of Eq. (5), we obtain a relation that will prove useful:

$$\Gamma_{\mathrm{L,R}}^c = \Gamma_a^c \frac{\Gamma_{\mathrm{L,R}}^{c0}}{\Gamma_{\mathrm{L}}^{c0} + \Gamma_{\mathrm{R}}^{c0}}, \tag{27}$$

where $\Gamma_{\mathrm{L,R}}^{c0}$ are tunneling widths in the absence of interaction,

$$\sum_\nu \Gamma_{0\nu}(\varepsilon_F) = \sum_\nu |T_0^\nu(\varepsilon_F)|^2 \rho_\nu(\varepsilon_F) \sim \gamma_0,$$

and the arguments of all functions are the same as in the previous equation.

It is clear that the tunneling widths $\Gamma_a^c(\mathbf{p}_\perp; \varepsilon)$ in Eqs. (26) and (27) can be written in the form

$$\Gamma_a^c(\mathbf{p}_\perp; \varepsilon_F) = \frac{1}{\pi}|B(\mathbf{p}_\perp)|^2 \operatorname{Im}\Sigma_{cc}(z - \varepsilon_c), \quad \operatorname{Re} z = \varepsilon_F. \tag{28}$$

In the absence of edge resonances, we have for the tunneling widths

$$\Gamma_c^a(\varepsilon_F) = \Gamma_{0c}^a(\varepsilon_F/\Gamma_{0c}^a)^{\alpha/(1+\alpha)}, \quad \Gamma_{0c}^a \sim \gamma_d. \tag{29}$$

This expression (29) is different from a more familiar one,[11,14] in that the exponent α does not contain component α_d associated with the orthogonality catastrophe. The reason is that the tunneling widths are determined in the band channel, but not for a localized impurity level.

In the presence of edge resonances, tunneling widths are cut off at γ_r, and in accordance with Eq. (28), take the form

$$\Gamma_c^a(\varepsilon_F) \approx A_{\mu_0}\gamma_0\left(\frac{\varepsilon_F}{\gamma_r}\right)^\alpha |B(\varepsilon_F)|^4 \sim \gamma_0\left(\frac{2\alpha}{\alpha_c}\right)^{1/2}|B(\varepsilon_F)|^4. \tag{30}$$

Note that ''exponential'' quasi-single-particle resonances exist at $\alpha < \alpha_c$, whereas ''power-law'' multiparticle resonances described by Eqs. (20)–(24) occur in the region

$$\frac{\alpha}{\alpha_c} \gg \left(\frac{\varepsilon_F}{\gamma_0}\right)|B(\varepsilon_F)|^{-4}. \tag{31}$$

Thus, Eq. (30) clearly shows how the character of the tunneling, along with that of edge resonances, changes with the interaction amplitude. Let us fix for definiteness the impurity level depth at $|B(\varepsilon_F)| \sim 1$.

Recall that γ_0 is an energy scale characterizing the change in $\rho_c(\varepsilon)$ within which $J_c(z)$ has a logarithmic divergence (16). For this reason, in the absence of interaction, exponentially narrow edge resonances are formed in the process of scattering of evanescent states, whose density of states is $\rho_c(\varepsilon)$ defined by Eq. (4), by electrons in the sidewalls.

A similar situation occurs at $\alpha < \alpha_c$, when the condition $\Gamma_c^a \ll \gamma_0$ holds. This means that, in the case of weak scattering, resonances at the 2D band edge are still due to the scattering of quasi-two-dimensional states from the region near the band edge with a width much smaller than γ_0. As a result, the domain of exponentially narrow quasi-one-particle resonances persists.

In the case of ''strong'' scattering at $\alpha \gg \alpha_c$, the condition $\Gamma_c^a \gg \gamma_0$ is satisfied. This means that, in this case, not only the states from the energy band much narrower than γ_0 participate in the scattering, but also the states from the ''tails'' of function $\rho_c(\varepsilon)$. As a result, the character of edge resonances is radically different from that in the case of one-particle resonances.

The character of tunneling also changes with the impurity level depth (at a fixed interaction amplitude). From Eq. (30) the following relations can be directly derived:

$$\Gamma_c^a \gg \gamma_0 \quad \text{for } 1 \gg |B| \gg \frac{\gamma_0}{\varepsilon_F}\left[\frac{\ln(\varepsilon_F/\gamma_0|B|^3)}{\alpha}\right]^{1/2}, \tag{32}$$

$$\Gamma_c^a \ll \gamma_0 \quad \text{for } |B| \ll \frac{\gamma_0}{\varepsilon_F}\left[\frac{\ln(\varepsilon_F/\gamma_0|B|^3)}{\alpha}\right]^{1/2}. \tag{33}$$

It follows from the above expressions that $\Gamma_c^a(z_r) \ll |T_0(z_r)|$. This condition means that the characteristic tunneling times in the interacting system, $\tau_t^{\text{int}} \sim (\Gamma_c^a)^{-1}(z_r)$, are much longer than the characteristic scattering times $\tau_{\text{sc}}^{\text{int}} \sim |T_0(z_r)|^{-1}$, so the electron lifetime within the quantum well is sufficient to form resonances due to the scattering defined by the Hamiltonian H_t.

In this case, the formation of edge resonances is largely controlled by the resonant scattering with amplitude $|\Sigma_{dc}|^2 G_d(z)$. Recall that, as was shown for the single-channel Kondo scattering,[8] i.e., in the case of Fermi-liquid excitations at the Fermi level, edge resonances are determined by the potential scattering.

One can see that, as in the case of the two-channel Kondo scattering,[5,6] in the problem under discussion, the additional scattering of electrons from the region near the 2D band edge by electron–hole pairs at the Fermi level due to the tunneling H_t^a generates a Fermi-liquid resonance at the 2D band edge in the quantum well. This resonance has the Fermi-liquid nature, since it corresponds to a simple pole in electron Green's functions.

Without the tunneling H_{tc}, there is only a power-law feature in the density of states (8) in the sidewalls and in the tunneling width owing to Mahan's non-Fermi-liquid resonance at the Fermi level.

Thus, a crossover from the Fermi-liquid regime of tunneling to the non-Fermi-liquid tunneling is possible when the separation between the Fermi level and 2D band edge in the quantum well varies. The conditions of this transition are identical to those for the existence of a solution to Eq. (17). In the case of Coulomb interaction, these conditions are less stringent than in the problem with the two-channel Kondo scattering. Indeed, solutions of Eq. (18) exist throughout the region of the exponent α where the density of states is a power-law function of energy [see Eqs. (8) and (19)], and at all admissible positions of the deep level we have $|B| \leqslant 1$.

In the case of the two-channel Kondo scattering,[5,6] edge resonances exist only at interaction amplitudes in the collective pseudospin channel above a certain critical value. Moreover, the condition determining the impurity level position is more restrictive, namely, $|B| \ll \gamma_0/\varepsilon_F$.

4. TUNNELING TRANSPARENCY AND CURRENT–VOLTAGE CHARACTERISTICS

1. The tunneling transparency and current are determined by the expressions

$$\sigma(\varepsilon_F) = 2e^2 \int dE\, \delta(E-\varepsilon_F) \sum_{\mathbf{k}_\perp, \mathbf{k}'_\perp} W(\mathbf{k}_\perp, \mathbf{k}'_\perp; E),$$

$$J = 2\pi e \int dE [f_L(E) - f_R(E)] \sum_{\mathbf{k}_\perp, \mathbf{k}'_\perp} W(\mathbf{k}_\perp, \mathbf{k}'_\perp; E), \tag{34}$$

$\varepsilon_{FL} - \varepsilon_{FR} = eV$, where V is the potential difference between the sidewalls.

The probability of elastic tunneling can be expressed in terms of the scattering matrix:

$$W(\mathbf{k}, \varepsilon_{\mathbf{k}}^L; \mathbf{k}', \varepsilon_{\mathbf{k}}'^R) = 2\pi |\mathscr{T}(\mathbf{k}, \varepsilon_{\mathbf{k}}^L; \mathbf{k}', \varepsilon_{\mathbf{k}}'^R)|^2 \delta(\varepsilon_{\mathbf{k}}^L - \varepsilon_{\mathbf{k}}'^R), \tag{35}$$

where $\mathscr{T} = H_t G H_t$ and $G = (z - \hat{H})^{-1}$ is the Green's function, and the $\mathscr{T}$-matrix describes tunneling of multiparticle excitations from the Fermi level via both band and localized impurity states of the quantum well with due account of all elastic scattering processes in the well, which determine the form of the Green's function G. Assuming that the major contribution to the tunneling amplitude is due to the matrix element containing the Green's function $\langle c_{\mathbf{p}}|G|c_{\mathbf{p}'}\rangle \equiv G_{cc}(\mathbf{k}_\perp, \mathbf{k}'_\perp; \varepsilon_{\mathbf{k}}^L)$, we obtain an expression for the tunneling amplitude in the band channel:

$$\begin{aligned}\mathscr{T}(\mathbf{k}, \varepsilon_{\mathbf{k}}^L; \mathbf{k}', \varepsilon_{\mathbf{k}}'^R) &= \sum_{\mathbf{p}\mathbf{p}'\sigma} \langle a_{\mathbf{k}L\sigma}|H_t|c_{\mathbf{p}\sigma}\rangle \langle c_{\mathbf{p}\sigma}|G|c_{\mathbf{p}'\sigma}\rangle \\ &\quad \times \langle c_{\mathbf{p}'\sigma}|H_t|a_{\mathbf{k}'R\sigma}\rangle \\ &= u_{\mathbf{k}} v'_{\mathbf{k}} \sum_{\mathbf{p}\mathbf{p}'\sigma} \langle a_{\mathbf{k}\sigma}|H_t|c_{\mathbf{p}\sigma}\rangle \langle c_{\mathbf{p}\sigma}|G|c_{\mathbf{p}'\sigma}\rangle \\ &\quad \times \langle c_{\mathbf{p}'\sigma}|H_t|a_{\mathbf{k}'\sigma}\rangle. \end{aligned} \tag{36}$$

By substituting Eqs. (11) and (36) in the formula for the tunneling probability, we find that the tunneling transparency contains the non-Fermi-liquid and resonant contributions: $\sigma(\varepsilon_F) = \sigma_c(\varepsilon_F) + \sigma_r(\varepsilon_F)$. The non-Fermi-liquid contribution $\sigma_0(\mu)$ is fully determined by the tunneling width, i.e., non-Fermi-liquid singularities in the density of states in the sidewalls:

$$\sigma_c(\varepsilon_F) = \frac{e^2}{4\pi} \Gamma_c^a(\varepsilon_F) \rho_c(\varepsilon_F) \tag{37}$$

(the tunneling widths $\Gamma_{0L,0R}^c$ in the absence of interaction are assumed to be equal). The tunneling widths $\Gamma_c^a(\varepsilon_F)$ are determined by Eq. (29). The non-Fermi-liquid contribution to the transparency at $|\varepsilon_F - \varepsilon_c| > \gamma_d$ dominates.

At $|\varepsilon_F - \varepsilon_c| < \gamma_d$, the prime contribution to the transparency is due to the resonances, so the character of tunneling is Fermi-liquid.

The contribution of the edge resonances to the transparency is determined by the second term in the Green's function $G_\sigma^{cc}(\mathbf{k}_\perp, \mathbf{k}'_\perp; z)$ in Eq. (11) and is given (at equal tunneling widths in the absence of interaction, $\Gamma_{0L} = \Gamma_{0R}$) by

$$\sigma_r^m(\varepsilon_F) = \frac{e^2}{4\pi} F_m \frac{\Gamma_a^{c2}(z_r)}{(\varepsilon_F - \varepsilon_m)^2 + \gamma_m^2} [(\varepsilon_m - \varepsilon_c)^2 + \gamma_m^2] I^2(z_r), \tag{38}$$

where F_m is a function of parameters, which is of the order of unity $m = r^{\pm}$ for a split resonance and $m = r$ for an unsplit resonance,

$$I(z_r)=\sum_{\mathbf{k}_\perp}|B(\mathbf{k}_\perp)|^4|G_{0\mathbf{k}_\perp}(\varepsilon_F)|^2$$
$$=\frac{1}{\pi}\tilde{\rho}_{0c}|B(\varepsilon_c)|^2\begin{cases}|\varepsilon_F-\varepsilon_c|^{-1} & \text{for } \gamma_0\gg|\varepsilon_F-\varepsilon_c|\gg\gamma_{r^\pm},\\ \pi/2\gamma_r & \text{for } \gamma_0\gg\gamma_r\gg|\varepsilon_F-\varepsilon_c|\end{cases} \tag{39}$$

for the split and unsplit resonance, respectively, and $\mathrm{Re}\, z_r=\varepsilon_F$.

2. By substituting the expressions for the tunneling (Γ_c^a) and resonant (γ_r) widths, we obtain the maximal contribution of a single edge resonance to the transparency at $\varepsilon_r=\varepsilon_F$:

$$\sigma_r^{\max}(\varepsilon_F)=F_{1r}\frac{e^2}{4\pi}\left(\frac{\Gamma_a^c}{\gamma_r}\right)^2\equiv\frac{e^2}{4\pi}S(\varepsilon_F), \tag{40}$$

where $S(\varepsilon_F)$ is the enhancement factor given by the expression

$$S(\varepsilon_F)=F_{1r}\left(\frac{\varepsilon_F}{\gamma_r}\right)^{2\alpha}\left(\frac{\gamma_0}{\gamma_r}\right)^2|B|^8\sim\frac{2\alpha}{\alpha_c}\left(\frac{\gamma_0}{\gamma_r}\right)^2\gg1, \tag{41}$$

where $F_{1r}\sim1$ and $|\mathrm{B}|\sim1$. Provided that condition (19) is satisfied, the widths γ_r are exponential functions of the tunneling parameters,[7] otherwise they are determined by Eqs. (20)–(22). In either case,

$$\sigma_r^{\max}(\varepsilon_F)\gg\sigma_c^{\max}(\varepsilon_F).$$

In the case of a split resonance, the enhancement factor is larger than the value given by Eq. (41) by a factor $(\sin\varphi)^{-1}$ [φ is defined by Eq. (20)].

In the absence of interaction, the contribution of exponentially narrow one-particle resonances at the 2D band edge yields the enhancement factor[7]

$$S_0(\varepsilon_F)\sim(\gamma_0/\gamma_r^{(0)})^2, \tag{42}$$

where $\gamma_r^{(0)}$ is the width of the one-particle resonance, which is an exponential function of the tunneling parameters.

Interestingly enough, this enhancement factor in the transparency is the greatest of those reported previously.[5,11,15]

Equations (29), (40), and (41) determine the transparency enhancement due to tunneling of electron–hole pairs from the Fermi level via the two-dimensional continuum. It is clear that in all cases discussed above, the enhancement factor

$$S_0(\varepsilon_F)\gg S(\varepsilon_F)\gg1. \tag{43}$$

The transparency enhancement factor due to the multiparticle tunneling is much smaller than that due to one-particle tunneling because the multiparticle resonances described above are wider than one-particle resonances, and this effect is more significant than the effect of the increase in the tunneling widths.

An important point is that, in a quantum structure with a two-dimensional continuum in its quantum well, the anomalous transparency enhancement factor $S(\varepsilon_F)\gg1$ is caused, in addition to the small width of edge resonances, by their proximity to the 2D band edge, which yields the additional factor $I^2(z_r)$.

Equations (29), (40), and (41) determine the transparency enhancement in an elementary (microscopic) tunneling event. The total contribution of impurities to the quantum well transparency is known to be $\sigma_{\mathrm{im}}=c_{\mathrm{im}}\sigma_r$, where c_{im} is the impurity concentration. As follows from the above expression, $\sigma_{\mathrm{im}}\gg\sigma_0$ for the reasonable values of the impurity concentration.

3. The foregoing tunneling mechanisms in a quantum structure with an intrinsic two-dimensional continuum dictate the changes in the current–voltage characteristic as the structure parameters are varied. Consider as examples two situations.

First, let us determine changes in tunneling spectra with increasing interaction amplitude, i.e., exponent α. Let us fix the Fermi level in the sidewalls in the position when its separation from the 2D band edge is smaller than Mahan's resonance width γ_d. In this case, as was shown above, there are edge resonances determined by solving Eq. (18).

As follows from Eq. (19), at small $\alpha<\alpha_c$, the main contribution to the transparency is due to the exponentially narrow quasi-one-particle edge resonance.

When $\alpha\gg\alpha_c$ [Eq. (23)], the transparency enhancement factor is determined by the wider but split edge resonance described by Eq. (20), whose characteristics are power-law functions of the tunneling parameters. Thus, a crossover from the regime of tunneling with the ''exponentially high'' quasiparticle transparency with an enhancement factor of the order of $S_0(\varepsilon_F)$ in Eq. (42) to the regime with a lower transparency with the enhancement factor approximately determined by Eq. (41) should occur as the interaction amplitude increases. But in this case the transparency will most probably have two peaks [compare Eqs. (21) and (25)] that are symmetric about the 2D band edge.

According to Eq. (34), which yields the current, the resonant contribution to the transparency generates a broadened ''step'' in the current–voltage characteristic as $eV\to eV_{\mathrm{th}}=\varepsilon_F-\varepsilon_m\leqslant\gamma_d$:

$$J=2\pi\gamma_m\sigma_m^{\max}\left(\tan^{-1}\frac{e(V-V_{\mathrm{th}}^m)}{\gamma_m}+\tan^{-1}\frac{eV_{\mathrm{th}}^m}{\gamma_m}\right). \tag{44}$$

This change in the transparency corresponds to a modification of the $J(V)$ curve: the step of height proportional to

$$J_r^{\max}(\varepsilon_F)\sim e^2\gamma_0\begin{cases}\gamma_0/\gamma_r^{(0)} & \text{for } \alpha<\alpha_c,\\ (2\alpha/\alpha_c)(\gamma_0/\gamma_r) & \text{for } \alpha_c<\alpha\leqslant2\alpha_{c1}\end{cases}$$

(Fig. 1) is replaced by a two-step feature symmetrical about the 2D band edge (Fig. 2b). The heights of the steps above and below the 2D band edge equal $\gamma_0(2\alpha/\alpha_c)(\gamma_0/\gamma_{r^\pm})$ and $2\gamma_0(2\alpha/\alpha_c)(\gamma_0/\gamma_{r^\pm})$, respectively, the separation between them is $\varepsilon_{r^+}-\varepsilon_{r^-}$ [Eq. (20)], and their widths are proportional to $\gamma_{r^\pm}$.

The crossover is most sensitive to changes in the heights (or widths) of the barriers in DBQW heterostructures. In

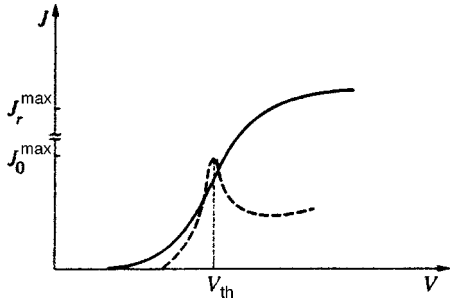

FIG. 1. $J(V)$ in the case of weak scattering: $\alpha<\alpha_c$ (at fixed positions of the Fermi level, $|\varepsilon_F-\varepsilon_c|<\gamma_d$, and of the impurity level). The dashed line represents the non-Fermi-liquid current.

point of fact, the matrix elements of interaction in Eq. (6) are proportional to γ_0^2 and are exponential functions of the barrier parameters.

Thus, by varying the barrier parameters at a fixed position of the Fermi level, as described above, we obtain a crossover between the two Fermi-liquid tunneling regimes with different $J(V)$ curves, whose overall shapes are shown in Figs. 1 and 2b.

The second situation to be discussed is the change in the $J(V)$ curve as the separation between the Fermi level in the sidewalls and 2D band edge varies. In this case, we fix the impurity level at the position determined by the condition $B\approx 1$.

In this situation, the tunneling transparency and $J(V)$ curve are determined by Mahan's resonance at the Fermi level as long as $|\varepsilon_F-\varepsilon_c|>\gamma_d$. The transparency in this case is given by Eqs. (37) and (29). When the Fermi level and band edge are so close that $|\varepsilon_F-\varepsilon_c|<\gamma_d$, the transparency and $J(V)$ curve are determined by Fermi-liquid edge resonances, and the transparency is given by Eqs. (40) and (41), which contain the specific ''Fermi-liquid'' factor $(\gamma_0/\gamma_r)^2$ (compare with the expression for S_0).

Mahan's feature at the Fermi level corresponds to a peak on a $J(V)$ curve at $V\to V_{th}$ with a section of negative differential resistivity (Fig. 2a). The current at the peak J_c^{max}, to order of magnitude, is

$$J_c^{max}\sim e^2\Gamma_c^a,$$

where Γ_c^a is given by Eq. (29). At all admissible values of the parameters, the height of Mahan's peak on a $J(V)$ curve is much smaller than the heights of steps in the resonant tunneling current: $J_r^{max}\gg J_c^{max}$. Thus, the crossover from the non-Fermi-liquid and Fermi-liquid tunneling regimes corresponds to a transformation of a peak with a section of negative resistivity to either a step at $V\to V_{th}$ or a two-step feature symmetric about $V\to V_{th}$. The transition leads to anomalous growth in the tunneling current, and it occurs at all interaction amplitudes.

Note also that at $V>V_{th}$ there is a slow logarithmic growth in $J(V)$ owing to the intrinsic conductivity of the structure with a two-dimensional continuum, which is proportional to $\gamma_0\rho_c(\varepsilon)$.

5. CONCLUDING REMARKS

1. The results of the reported study indicate that the mechanisms of tunneling in a quantum structure with an intrinsic two-dimensional continuum of electron states are radically different from those of tunneling via a resonant level under the barrier.[11] Let us discuss two essential differences.

First, there is a strong resonant tunneling in the case when the Fermi level in the sidewalls is not resonant with the localized impurity level lying deep within the band gap of the doped layer.

Second, the edge resonances associated with new tunneling channels are Fermi-liquid resonances, because they are described by simple poles in electron Green's functions [Eqs. (11) and (12)], instead of branch points, as in the case of Mahan's singularity at the Fermi level [Eq. (7)]. For this reason, the existence conditions for edge resonances are those of instability of the non-Fermi-liquid state against the interband impurity scattering, which takes place in the process of tunneling and is described by the term H_{tc} in Hamiltonian (2).

We can also summarize the physical differences between instability conditions for the non-Fermi-liquid state in the problem with the Coulomb interaction studied in this paper and in the problem with the two-channel Kondo scattering.[5,6]

As noted above, the stability condition common to these two cases is the existence of the potential, H_{tc}, and resonant, H_{td}, scattering due to tunneling. Nonetheless, the sufficient conditions for the existence of edge Fermi-liquid resonances, hence the instability of the non-Fermi-liquid state, are determined by ''intrinsic'' properties of the latter and differ considerably in the cases of the Coulomb interaction and two-channel Kondo scattering.

In the two-channel problem[5,6] the existence of edge resonances and instability are possible only when the non-Fermi-liquid peak at the Fermi level is widened sufficiently by screening in the pseudospin channel. In other words, instability can occur only as a result of the orthogonality catastrophe for collective fermion variables describing excitations in the two-channel problem.

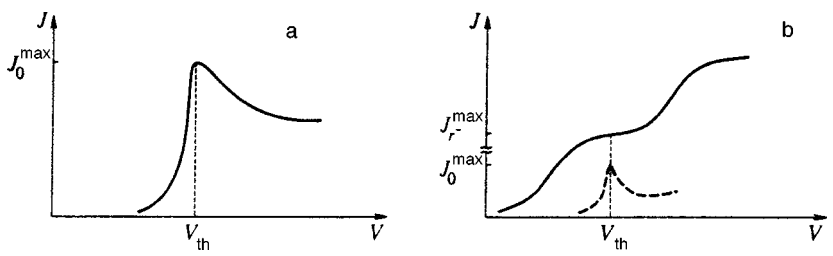

FIG. 2. $J(V)$ at different separations between the Fermi level in the sidewalls and the edge of the 2D band: a) $|\varepsilon_F-\varepsilon_c|>\gamma_d$; b) $|\varepsilon_F-\varepsilon_c|<\gamma_d$, $\alpha\gg\alpha_c$. The position of the impurity level is fixed. For comparison, the dashed line shows the Fermi-liquid contribution plotted in Fig. 1a.

In the Coulomb problem with tunneling in the band channel and a deep impurity level, the orthogonality catastrophe has little effect on the existence of edge resonances [Eq. (18)]. They exist owing to Mahan's peak at the Fermi level, i.e., sufficiently strong electron scattering by the potential due to recharging of impurities in the process of tunneling.

In conclusion, we emphasize that the existence conditions for the non-Fermi-liquid state in the case of the Coulomb interaction are significantly weaker than in the case of two-channel Condo scattering.[6] For this reason, it should be easier to detect this transition experimentally in systems with the Coulomb interaction.

It is a pleasure to express my gratitude to L. A. Maksimov, who read this paper in manuscript form and offered constructive criticism. This work was supported by the Russian Fund for Fundamental Research (Projects 98-02-16729 and 98-02-16730) and the INTAS–RFBR fund (Grant No. 97–1066).

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

[1] P. Schlottmann and P. D. Sacramento, Physica B **206-207**, 95 (1995).
[2] D. C. Ralph and B. A. Buhrman, Phys. Rev. Lett. **72**, 3401 (1994).
[3] D. J. Cox, Phys. Rev. Lett. **59**, 1240 (1987).
[4] M. Fabrizio, A. O. Gogolin, and Ph. Nozières, Phys. Rev. Lett. **74**, 4503 (1995).
[5] L. A. Manakova, JETP Lett. **67**, 1069 (1998).
[6] L. A. Manakova, Zh. Éksp. Teor. Fiz. **114**, 1466 (1998) [JETP **87**, 796 (1998)].
[7] K. A. Kikoin and L. A. Manakova, Phys. Rev. B **57**, 4863 (1997).
[8] K. A. Kikoin and L. A. Manakova, JETP Lett. **65**, 484 (1997).
[9] P. Nozières and C. T. de Dominicis, Phys. Rev. **178**, 1097 (1969).
[10] K. D. Schotte and U. Schotte, Phys. Rev. **182**, 479 (1969).
[11] K. A. Matveev and A. I. Larkin, Phys. Rev. B **46**, 15337 (1992).
[12] G. D. Mahan, Phys. Rev. **163**, 612 (1967).
[13] K. A. Kikoin and V. N. Fleurov, *Transition Metal Impurities in Semiconductors*, World Scientific, Singapore (1994).
[14] Yu. Kagan and N. V. Prokof'ev, Zh. Éksp. Teor. Fiz. **93**, 366 (1987) [Sov. Phys. JETP **66**, 211 (1987)].
[15] L. I. Glazman and M. E. Raikh, JETP Lett. **47**, 452 (1988); T. K. Ng and P. A. Lee, Phys. Rev. Lett. **61**, 1768 (1988).

Translation provided by the Russian Editorial office

Jahn–Teller effect on Sm^{3+} ions in SmB_6

T. S. Al'tshuler*)

E. K. Zavoĭskiĭ Physicotechnical Institute, 420061 Kazan', Russia

M. S. Bresler

A. F. Ioffe Physicotechnical Institute, Russian Academy of Sciences, 194021 St. Petersburg, Russia
(Submitted 5 November 1998)
Zh. Éksp. Teor. Fiz. **115**, 1860–1871 (May 1999)

Electron spin resonance on samarium ions with stabilized valence Sm^{3+} is investigated in the fluctuating-valence semiconductor SmB_6, both pure and doped with the rare-earth ions Eu^{2+}, Er^{3+}, and Gd^{3+}. The dynamic and static Jahn–Teller effects have been observed for the first time on rare-earth ions. The relation between the Jahn–Teller effect in a fluctuating-valence semiconductor and the excitonic nature of the ground state of such a semiconductor is discussed. © *1999 American Institute of Physics.* [S1063-7761(99)02405-1]

1. INTRODUCTION

Samarium hexaboride SmB_6 is a classical object in the physics of fluctuating-valence materials.[1] The average valence of the samarium ions in SmB_6 is $+2.6$, but the valence actually fluctuates with frequency $10^{13}-10^{14}$ Hz between the states Sm^{2+} and Sm^{3+}. The crystal structure of SmB_6 is similar to that of CsCl. It consists of two interpenetrating simple cubic lattices consisting of samarium atoms and boron octahedra (Fig. 1).

Materials with fluctuating valence are of great interest because this state is spatially homogeneous and at low temperatures it exhibits features of a coherent state, i.e., a macroscopically quantum state. Electron spin resonance (ESR) investigation of SmB_6 has been very effective. In Ref. 2 it was shown that the ground state of SmB_6 exhibits characteristics of an excitonic insulator. In Ref. 3 the unusual configuration $4f^75d^1$ of the gadolinium ion was observed. This configuration arises because an additional electron is localized on gadolinium.

It is well known that the Jahn–Teller effect is ordinarily not observed on rare-earth ions, if it is investigated by the ESR method. According to the conventional point of view, the strong spin-orbit coupling typical of rare-earth ions stabilizes the high-symmetry state and prevents the appearance of the Jahn–Teller effect.[4] At the same time, a dynamic Jahn–Teller effect has recently been observed on Er^{3+} and Gd^{3+} ions in SmB_6.[3,5,6] In the present paper we report the observation of the static and dynamic Jahn–Teller effects on Sm^{3+} ions in SmB_6. The preliminary results of this work have been reported at the LT-21 conference[7] and partially published in Ref. 8.

2. EXPERIMENTAL RESULTS

The ESR investigations were performed on samarium hexaboride single crystals, both pure and doped with rare-earth ions, namely, europium Eu^{2+} (with concentration $c \simeq 0.01$–0.04 at. %), gadolinium Gd^{3+} ($c \simeq 0.05$ at. %), and erbium Er^{3+} ($c \simeq 0.05$ at. %).

Samarium purified by fractional sublimination was used to prepare the samples. Spectral analysis showed that the total impurity content of rare-earth elements in samarium is less than 10^{-4} at. %. The crystals were grown by the flux solution method and consisted of $2\times1\times0.1$ mm^3 plates and $3\times0.2\times0.2$ mm^3 needles.

The measurements were performed on a Varian radiospectrometer at 9.3 MHz in the temperature range 1.6–4.2 K. Electron spin resonance spectra from the rare-earth dopant were observed on all samarium hexaboride samples. The ESR spectrum on the Eu^{2+} ions is described by a spin Hamiltonian with cubic symmetry and is similar to that given in Refs. 9 and 10. The ESR spectra on the trivalent ions Er^{3+} and Gd^{3+} had the same overall features as those observed in Refs. 3 and 5, where they were described by a dynamic Jahn–Teller effect. A weak, narrow (line width $\Delta H \approx 5$ Oe) signal with $g \approx 2$ is also seen in all samples, possibly from disordered boron with a dangling bond. Signals from defects, as described in Refs. 11 and 12, were not observed in any of our single crystals.

In all samples, rare-earth-doped and pure, ESR signals with g factor characteristic for Sm^{3+} ions were observed in high fields 8–16 kOe. The $\nu \sim 10^{13}-10^{14}$ Hz (much higher than the frequency of the ESR spectrometer, 10^{10} Hz) fluctuations of the samarium ions between the Sm^{3+} and Sm^{2+} states make it impossible to observe directly the ESR signal on these ions. However, it was established a long time ago that impurities and defects in SmB_6 stabilize the valence of some samarium ions in the paramagnetic state Sm^{3+}.[13] Therefore doping of SmB_6 with rare-earth impurities stabilized 0.04–0.1 at. % of the samarium ions in the Sm^{3+} state. In pure SmB_6 vacancies and defects probably play a stabilizing role.

The typical ESR spectrum for the Sm^{3+} ion in pure and Eu^{2+}-doped SmB_6 samples at $\theta = 45°$ (θ is the angle between the magnetic field and the [100] axis in the (100) plane) is displayed in Fig. 2. Only the ESR lines that are not marked by arrows in Fig. 2 are observed for samples doped

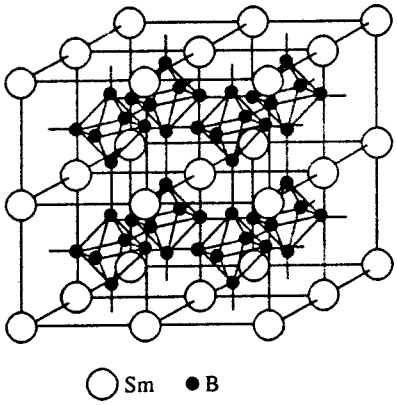

FIG. 1. Crystal structure of samarium hexaboride SmB_6.

with the trivalent ions Er^{3+} and Gd^{3+}. Depending on the angle θ, three to five ESR lines were observed simultaneously. Figure 3 shows the positions of the resonance lines versus angle at 1.6 K for all experimental samples.

The experimental data can be described theoretically in terms of two types of spectra: one corresponding to a cubic center (quartet) and the other to an anisotropic doublet (cubic symmetry is preserved by the presence of three types of such doublets with symmetry axes along the x, y, and z axes; the doublets Γ_{6x}, Γ_{6y}, and Γ_{6z} are shown in Fig. 3 by the lines E, F, and G). The parameters of these spectra were found to be $P=0.465$ and $Q=-0.1$ for the quartet and $g_{\parallel}=0.42\pm0.02$ and $g_{\perp}=0.79\pm0.02$ for the doublet. Note that the observed values of P and Q are close to the theoretical values for the spin-5/2 Sm^{3+} ion, $P_{\rm th}=0.525$ and $Q_{\rm th}=0.144$, i.e., the renormalization of these parameters that is ordinarily attributed to the closeness of the 7/2 level to the 5/2 ground state is not too large. For the quartet with effective spin 3/2 the theory predicts four possible transitions. At $\theta=0°$ these are $-3/2\rightarrow3/2$ (the line B in Fig. 3), $1/2\rightarrow3/2$ and $-3/2\rightarrow-1/2$ (the line C), $-1/2\rightarrow3/2$ and $-3/2\rightarrow1/2$ (the line D), and $-1/2\rightarrow1/2$ (the line A). However, only two transitions are observed experimentally: $-3/2\rightarrow3/2$ (the line B) and $-1/2\rightarrow1/2$ (the line A) in the angle range $\theta=30-60°$. The maximum field of the spectrometer was 16 kOe, so that the transition $-1/2\rightarrow1/2$ is not seen at other angles and the transition $-1/2\rightarrow3/2$ (the line D) is also experimentally inaccessible. The transition $1/2\rightarrow3/2$ (the line C) is not observed at $\theta=0°$ and $\theta=90°$ because the intensity is too low and noise is present at $H\sim12$ kOe; at all other angles it is completely or partially forbidden. It is evident from Fig. 3 that the proposed theoretical values of the parameters provide a good description of the angular dependence of the positions of the five observed ESR lines. The values of the g factors were identical for all samples, irrespective of whether or not the sample was doped, and they did not depend on the valence of the dopant.

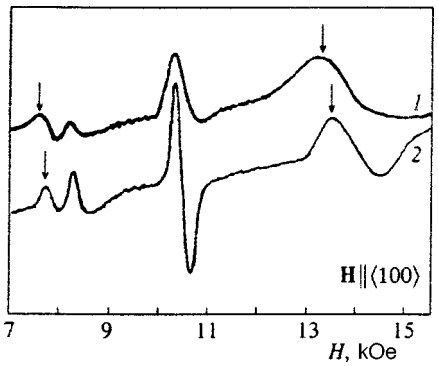

FIG. 2. Examples of traces of ESR signals on Sm^{3+} ions in SmB_6 at two temperatures, $T=3.8$ K (curve *1*) and $T=1.6$ K (curve *2*). The arrows mark lines belonging to the quartet Γ_8; all other lines correspond to the doublets Γ_6.

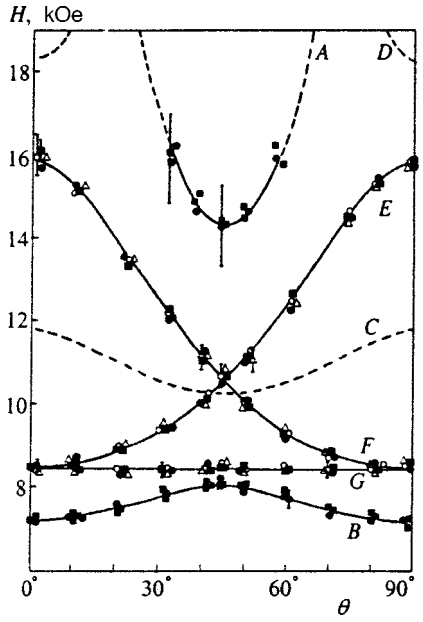

FIG. 3. Angular dependence of the positions of resonance lines of the Sm^{3+} ion with the magnetic field rotating in the (100) plane at $T=1.6$ K. The experimental positions of the lines are shown by filled squares for pure SmB_6, filled circles for Eu^{2+}-doped SmB_6, triangles and open circles for Gd^{3+} and Er^{3+}-doped SmB_6. The lines *A*, *B*, *C*, and *D* show the theoretically computed positions of the lines of the quartet Γ_8; the lines *E*, *F*, and *G* show the theoretically computed positions of the lines for the doublets Γ_{6x}, Γ_{6y}, and Γ_{6z}.

The intensities of the strongest lines of the doublets and quartets were estimated by double integration. Since the method for estimating the line intensities is relatively crude and the intensity itself is low in many cases because of the low values of the g factors, the error in the signal amplitude is much higher than the error in the line positions, i.e., the g factors. Nonetheless it can be asserted that the ratios of the line intensities also agree with experiment (Fig. 4).

Subsequent analysis of the results showed that the doublet state is not an independent center, but rather originates from splitting of the quartet Γ_8 by the axial (tetragonal) field. In this case, as is well known,[14] the quartet Γ_8 splits into two doublets, the g factors of these doublets being related to the parameters P and Q of the quartet. For one of these states we obtain $g_{\parallel}=+2Q=-0.2$ and $g_{\perp}=(1/2)|3P-Q|=0.747$, which is very close to the experimentally observed g factors (an ESR experiment gives the absolute value of the g factor). (We note that in Ref. 15 it was indicated that partial renor-

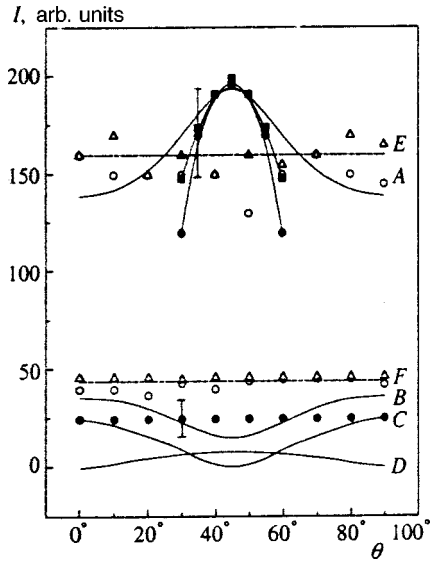

FIG. 4. Angular dependence of the intensity of ESR resonance lines of Sm^{3+} ions with the magnetic field rotating in the (100) plane at $T=1.6$ K. The intensities of the experimentally observed ESR signals are denoted as follows: filled squares for pure SmB_6, filled and open circles for Eu^{2+}-doped SmB_6 for, respectively, the quartet Γ_8 (the lines *A* and *B*) and the doublet Γ_6 (the lines *E* and *F*), and triangles for Gd^{3+}-doped SmB_6. Solid lines *A*, *B*, *C*, and *D* — theoretically computed intensities of ESR lines for the quartet Γ_8, dashed lines *E* and *F* — theoretically computed intensities of the ESR lines for the doublet Γ_6.

malization of the g factor occurs for such splitting, $g_\parallel$ being renormalized more strongly than $g_\perp$.)

Investigation of the temperature dependence of the intensities of the resonance lines for pure and Eu^{2+}-doped SmB_6 samples (only the strongest lines are used) reveals a relative decrease in the number of doublets and a relative increase in the number of quartets (the corresponding data are shown in Fig. 5). Thus, in these samples some ions are in a field with tetragonal symmetry, described by the doublets Γ_{6x}, Γ_{6y}, and Γ_{6z}, and the rest of the ions are in a field with cubic symmetry (the quartet Γ_8). As temperature increases, the number of centers described by the quartet Γ_8 increases and the number of tetragonal centers decreases. In SmB_6 containing the trivalent ions Er^{3+} and Gd^{3+} the Sm^{3+} ions are located only in a tetragonal field.

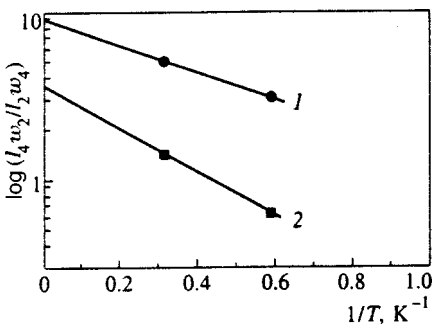

FIG. 5. Temperature dependence of the ESR signal intensity for Eu^{2+}-doped SmB_6 (curve *1*) and for pure SmB_6 (curve *2*). The intensity ratios are normalized to the theoretical value of the ratio of the transition probabilities, so that the segments on the ordinate directly give the ratio of the static rates N_4/N_2.

3. DISCUSSION

The formal description of the experimental results is based on the coexistence of a center with cubic symmetry Γ_8 and centers with tetragonal symmetry (doublet Γ_6), obtained from the splitting of a cubic center by a tetragonal field. The ground state is tetragonally distorted. This description, however, does not answer the question of the mechanism that in some cases gives rise to the splitting of the states while in others it makes it possible to observe an unsplit quartet also.

The splitting of the quartet state by a tetragonal field could be caused by breaking of local symmetry (for example, by the presence of an impurity or a defect near an ion with stabilized valence) as well as by the Jahn–Teller effect. We shall consider both possibilities.

1. The average valence of the fluctuating samarium ions is +2.6. In accordance with the charge-compensation principle,[13] there are $2.6/0.4=6.5$ (i.e., 6–7) Sm^{3+} ions per vacancy in the samarium sublattice. Europium (Eu^{2+}) doping of the SmB_6 crystal gives $0.6/0.4=1.5$ (1–2) Sm^{3+} ions per Eu ion. Similarly, introducing the trivalent ions Er^{3+} and Gd^{3+} should shift some fluctuating-valence ions into the Sm^{2+} state. In reality, introducing trivalent ions also results in the production of lattice defects, so that erbium or gadolinium introduced into the sample increases rather than decreases the intensity of the ESR signal on the Sm^{3+} ions. The Sm^{3+} ions (6–7 centers) compensating the vacancy charge can occupy positions both close to and far away from the vacancy. However, if the Sm^{3+} ions are located next to a vacancy, only some of the possible configurations will possess tetragonal symmetry; all other configurations of the Sm^{3+} ions in neighboring sites will have a lower symmetry. The absence of centers with low symmetry in the experiment shows that the Sm^{3+} ions are not clustered near vacancies whose charge they compensate.

On the other hand, the ESR of Eu^{2+},[9,10] Er^{3+},[5] and Gd^{3+} ions[10] exhibits exceptionally symmetric characteristics, indicating cubic symmetry of the environment around an ion, i.e., the Sm^{3+} ions compensating these charges are also located far from them.

Hence it follows that Sm^{3+} ions, as a rule, are located far from defects and impurities, i.e., in the symmetric environment of Sm^{3+} ions with fluctuating valence (with average valence +2.6), and the observation of the ground state of Sm^{3+} in a tetragonal field can be explained by the Jahn–Teller effect.

2. As temperature increases from 1.6 to 4.2 K, the relative number of quartets Γ_8 increases and correspondingly the number of doublet states of Sm^{3+} ions decreases. This result is completely natural for the Jahn–Teller effect and does not have a simple explanation in the case of local breaking of symmetry. In the latter case a temperature increase should result only in the observation of one more doublet state. Note that the experimental data can also be explained by assuming that the tetragonal field which splits the state of a quartet is due to the presence of a defect near an Sm^{3+} ion, and thermal excitation can transfer an Sm^{3+} ion into a symmetric environment. However, this explanation contradicts the arguments showned above, according to which it is unlikely that

an Sm^{3+} ion will be located next to a defect and, moreover, it makes it impossible to interpret the ESR of trivalent ions in SmB_6 from a unified standpoint.

3. If the tetragonal splitting of the state of a trivalent rare-earth ion in the cubic lattice of SmB_6 had been due to defects, then such splitting should have been observed in the isomorphic compounds with integral valence LaB_6, BaB_6, YbB_6, and CaB_6, which also contain defects, vacancies, and impurities. However, in contrast to SmB_6, the standard ESR signal for a rare-earth ion in a cubic field is observed in these crystals.[5]

4. In a recently published work on a Raman scattering investigation of SmB_6,[16] splitting of the state of an Sm^{3+} ion was observed and attributed to the dynamic Jahn–Teller effect.

The theory of the Jahn–Teller effect on Er^{3+} ions in samarium hexaboride can be used to describe the experimental results.[5,6] This theory is not significantly different from the theory of the interaction of an E state with lattice vibrations transforming according to the representation e_g, described in the book by Abragam and Bleaney.[14]

In the SmB_6 lattice the Sm^{3+} ion is in an octahedral environment of fluctuating Sm ions [the boron atoms form a rigid framework, tied together by homopolar bonds, and do not participate in the Jahn–Teller effect (Fig. 1)] (see, however, Refs. 11 and 12 concerning the influence of defects in the boron sublattice on the ESR of samarium ions in various charge states). The vibrations of an octahedron which transform according to the Γ_3 representation are well known and are described in, for example, Ref. 17. The contribution of the interaction with these vibrations to the Hamiltonian of the quartet Γ_8 can be constructed by the method of invariants and has the form

$$\hat{H}=A(Q_2\hat{S}_2+Q_3\hat{S}_3)+B[(Q_3^2-Q_2^2)\hat{S}_3-2Q_2Q_3\hat{S}_2], \tag{1}$$

where

$$\hat{S}_2=\begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix}, \quad \hat{S}_3=\begin{vmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{vmatrix}. \tag{2}$$

The part of the Hamiltonian that is linear in the vibrational coordinates Q_2 and Q_3 can be diagonalized by a unitary transformation

$$\hat{U}=\begin{vmatrix} 0 & \cos\Phi/2 & 0 & -\sin\Phi/2 \\ \sin\Phi/2 & 0 & -\cos\Phi/2 & 0 \\ 0 & \sin\Phi/2 & 0 & \cos\Phi/2 \\ -\cos\Phi/2 & 0 & -\sin\Phi/2 & 0 \end{vmatrix}. \tag{3}$$

Then, introducing as usual the polar coordinates ρ and Φ

$$Q_2=\rho\sin\Phi, \quad Q_3=\rho\cos\Phi, \tag{4}$$

we obtain the energies of an ion in the elastic field of the lattice vibrations:

$$E_L=\frac{1}{2}\omega_0^2\rho^2\pm A\rho. \tag{5}$$

The energy surfaces in the space E_L, ρ, and Φ will have the familiar sombrero form. The terms $\sim\rho^2$, ρ^3 containing an anisotropic contribution to the interaction energy of an ion with lattice vibrations (the factors $\sin\Phi$ and $\cos\Phi$) will give rise to energy minima in the (ρ,Φ) configuration space that in ordinary space correspond to deformations of octahedra of Sm ions along one of the principal cubic axes. Thus, a tetragonal field $\pm A\rho$ develops which splits the quartet Γ_8 into two doublets; the total cubic symmetry of the problem is preserved because the doublets are equally likely to be occupied in a tetragonal field directed along the x, y, and z axes. This is the static Jahn–Teller effect.

The degeneracy of ion states corresponding to different energy minima in configuration space is lifted by tunneling transitions between these states: if tunneling transitions are sufficiently effective, then the six-fold degenerate vibronic state splits into a vibronic quartet Γ_8 (generally speaking, with renormalized parameters different from the initial parameters) and a vibronic doublet Γ_6 (dynamic Jahn–Teller effect).

If the tunneling splitting Δ is large (compared with the temperature) and the ground state is a quartet, then the Hamiltonian of the vibronic quartet can be written in the form

$$H=\begin{vmatrix} -\Delta-XH_z+\frac{YZ}{4}H_z & 0 & -\frac{X}{2}H_- -\frac{YZ}{8}H_- & \frac{\sqrt{3}\,YZ}{8}H_+ \\ 0 & -\Delta-XH_z-\frac{YZ}{4}H_z & \frac{\sqrt{3}\,YZ}{8}H_+ & -\frac{X}{2}H_-+\frac{YZ}{8}H_- \\ -\frac{X}{2}H_+-\frac{YZ}{8}H_+ & \frac{\sqrt{3}\,YZ}{8}H_- & -\Delta+\frac{X}{2}H_z-\frac{YZ}{4}H_z & 0 \\ \frac{\sqrt{3}\,YZ}{8}H_- & -\frac{X}{2}H_++\frac{YZ}{8}H_+ & 0 & -\Delta+\frac{X}{2}H_z+\frac{YZ}{4}H_z \end{vmatrix}, \tag{6}$$

where $X=P-Q$, $Y=P+Q$, $Z=c_1+4c_2$, and the constants c_1 and c_2 describe the matrix elements

$$c_1=\langle a_0|\cos\Phi|a_0\rangle, \quad c_2=\langle a_0|\cos\Phi|a_1\rangle \tag{7}$$

between the states $|a_0\rangle$ and $|a_1\rangle$ corresponding to different minima in the (ρ,Φ) space. (Thus, the set $c_1=1$ and $c_2=0$ corresponds to the static Jahn–Teller effect; the values $c_1=0.7$ and $c_2=0.08$ were obtained for the dynamic Jahn–Teller effect on Er^{3+} ions.[5]) If the splitting Δ is large and the quartet–doublet interaction can be neglected, then only a combination of the numbers c_1 and c_2, specifically, $c=(c_1+4c_2)$, appears in the theory, and the Hamiltonian itself can be represented in the standard form for a quartet by introducing the parameters

$$\overline{P}=\frac{1}{2}(P-Q)+\frac{1}{4}(P+Q)c,$$

$$\overline{Q}=-\frac{1}{2}(P-Q)+\frac{1}{4}(P+Q)c. \tag{8}$$

(We note that a vibronic doublet is isotropic: its Zeeman splitting does not depend on the direction of the magnetic field H with respect to the crystal axes.)

Comparing with experiment shows that in order for the parameters of a vibronic quartet to describe the experiment they must be the same as the parameters of the initial quartet (in the absence of a Jahn–Teller effect), i.e., $c=2$.

Thus, the idea of the appearance of a Jahn–Teller effect in SmB_6 makes it possible to describe the experimental results satisfactorily. Assuming that the transition of doublets into quartets is associated with a transition from the static to the dynamic Jahn–Teller effect, we shall estimate the splitting E_{JT} of the quartet and doublet states using the formula for the intensity ratio of the corresponding ESR lines:

$$\frac{I_4}{I_2}=\frac{N_4 w_4 \tau_4 \exp(-E_{JT}/kT)}{N_2 w_2 \tau_2}, \tag{9}$$

where N_2 and N_4 are the static weights of the doublet and quartet states, w_2 and w_4 are transition probabilities, and τ_2 and τ_4 are the spin-relaxation times determining the line width. For a rough estimate we can set $\tau_2\approx\tau_4$ and the ratio w_4/w_2 can be taken from the theory. Then we obtain $E_{JT}\approx 2.8$ K and $N_4/N_2\approx 3.5$ for the pure sample and $E_{JT}\approx 1.8$ K and $N_4/N_2\approx 8.5$ for a europium-doped sample (see data in Fig. 5). To within the accuracy with which we can estimate in practice the desired quantities for large resonance linewidths and low signal amplitudes, they are close for the two samples, which once again confirms that the Jahn–Teller effect is responsible for the observed phenomena: for the case where local symmetry is broken, there should not be any special correlation between the results obtained for these quantities. The results obtained are somewhat more accurate than the data presented in our brief report in Ref. 8. The energy E_{JT} characterizes in order of magnitude the barrier separating the vibronic states responsible for the deformations of the octahedral environment of an Sm^{3+} ion along different cubic axes.

The ratio of the static weights of the quartet and doublet states should be three according to the theory, since three distorted states correspond to one symmetric state. Taking account of the large width of the resonance lines, the small amplitude of the signals, and the rough approximation made in finding the characteristics of the Jahn–Teller effect, it can be concluded that our estimate is close to the theory.

Thus, the observation of only one type of center, which at low temperatures is in a doublet state and at high temperatures passes into a quartet state, in the ESR spectrum proves that, for the first time for rare-earth ions, both the static and dynamic Jahn–Teller effects were observed on Sm^{3+} centers in the compound SmB_6.

Although now there is no doubt that a dynamic Jahn–Teller effect can be observed on trivalent rare-earth ions in an SmB_6 lattice, the interpretation of this effect is different in different models: Sturm, Elschner, and Hoeck[5] consider the standard mechanism involving lattice vibrations, whereas Weber, Sigmund, and Wagner[18] have proposed a new ''electronic'' mechanism. In Ref. 18 the interaction of an Er^{3+} ion with the electronic excitations of neighboring Sm ions, caused by fluctuations of the valence on these ions, is considered. Specifically (for the ions of an octahedron surrounding an Er^{3+} ion), collective wavefunctions with $\Gamma_1(\chi_0)$ and $\Gamma_3(\chi_1$ and $\chi_2)$ symmetries, describing these excitations, are introduced.

However, the static Jahn–Teller effect observed for the Sm^{3+} ion cannot be obtained directly from the Hamiltonian of Ref. 18. The reasons for this are, apparently, that Weber, Sigmund, and Wagner studied electronic excitations of the system, i.e., fermions, which unlike excitons cannot be reduced to the influence of a classical field (they are created and vanish in pairs, as a result of which the Hamiltonian of Ref. 18 is bilinear in the operators creating and annihilating electronic excitations). Nevertheless, for certain approximations the Weber–Sigmund–Wagner Hamiltonian can be put into a form similar to the Hamiltonian of the standard Jahn–Teller effect problem. If the electronic degrees of freedom are considered to be fast and averaging over them is performed in the spirit of the adiabatic approximation, replacing the creation and annihilation operators by occupation numbers t_{ik}, then dipole moments $P_2=t_{22}-t_{11}$ and $P_1=t_{12}+t_{21}$ of the electronic clouds can be introduced, as is done in the theory of two-level systems. Such dipole moments can be treated as classical fields acting on an ion in an octahedral environment. Correspondingly, the term $\hat{S}_2P_2+\hat{S}_3P_3$, similar to the interaction of an ion with lattice vibrations with Γ_3 symmetry, can be distinguished in the Hamiltonian of Ref. 18. Evidently, just as for lattice vibrations, such an interaction will give rise to a static Jahn–Teller effect, and terms of a higher order in P_2 and P_3 (which are not written out in the Weber–Sigmund–Wagner Hamiltonian) will have to be introduced in order to localize ions at the energy minima in (ρ,Φ) space. In such an approach the appearance of a dynamic Jahn–Teller effect in the model of Ref. 18 is due to breakdown of the conditions of adiabaticity.

Averaging over the fast (electronic) degrees of freedom therefore actually makes the influence of lattice vibrations on an ion indistinguishable from the influence of vibrations of electronic clouds associated with fluctuations of the valence of Sm^{3+} ions. As shown in Ref. 19, the electron–phonon

interaction plays an important role in the fluctuating-valence theory, and fast transitions of an ion from a $+3$ state into a $+2$ state strongly influence lattice stiffness and thus the frequency of lattice vibrations. For this reason the Jahn–Teller effect in SmB_6 probably arises as a result of the total effect of both factors. We note (as even Weber, Sigmund, and Wagner have emphasized) that in LaB_6, CeB_6, and YbB_6, where the valence of the rare-earth ions is stable, the Jahn–Teller effect is not observed, but unlike Weber, Sigmund, and Wagner we assume that this does not preclude the ''ordinary'' mechanism of the Jahn–Teller effect in SmB_6 but rather indicates that the two mechanisms are closely related. Actually, the electronic states of an Sm^{3+} ion interact not with pure phonon modes but rather with lattice vibrations renormalized by the coupling with valence fluctuations (see, for example, Ref. 19; experimentally mixed modes were observed in Refs. 20 and 21). For such mixed modes the method of invariants can be used to construct the interaction Hamiltonian, as is done in Refs. 5 and 6, but the dispersion law for these modes and the interaction constants will be different from the case of an interaction with purely vibrational modes.

In Ref. 16, where Raman scattering in SmB_6 was investigated, bound electronic states with energies in the band gap of this semiconductor were observed. Analysis of these results showed that on account of the magnetoelastic interaction of the ground-state quartet Γ_8 with a quasilocal vibrational mode t_{2g} the dynamic Jahn–Teller effect on Sm^{3+} ions is a possible explanation of the origin of this series of bound states. This explanation agrees with our results, which are interpreted above in the representation of static and dynamic Jahn–Teller effects on Sm^{3+} ions. In our case, however, we took account of the interaction of a Γ_8 quartet with the vibrational mode e_g. It should be noted that the symmetry properties of the quartet state allow for the state to interact with both modes, and our results could also be explained by taking account of the mode t_{2g} instead of e_g. However, we chose a simpler model, which satisfactorily describes our results and was used successfully to describe ESR on Er^{3+} ions.

In summary, the ESR of trivalent ions (Er^{3+}, Gd^{3+}, Sm^{3+}) in SmB_6 can be described from a unified standpoint by considering the Jahn–Teller effect. There is an interesting qualitative difference between the behavior of the Jahn–Teller effect of these trivalent ions. While only a dynamic Jahn–Teller effect is observed for the Er^{3+} and Gd^{3+} ions, both the dynamic and the static Jahn–Teller effects can be observed for Sm^{3+} ions in SmB_6. This can be described phenomenologically by a change in the constant A in Eq. (5). On a microscopic level this could be due to stabilization of the static Jahn–Teller effect by chaotic lattice deformations (see Ref. 14) in the case of the Sm^{3+} ion.

Although the average valence of the fluctuating Sm ion in SmB_6 is $+2.6$, i.e., closer to the valence of Er^{3+} and Gd^{3+} ions than to the divalent Eu^{2+} ion, it is the embedding of erbium ions (and the appearance of Sm^{3+} ions) that disrupts the state of the lattice more than the larger Eu^{2+} ions. However, the average radius of an ion with a fluctuating valence is apparently not a simple arithmetic mean of the radii of Sm^{2+} and Sm^{3+} ions. As shown in the excitonic model of a semiconductor with fluctuating valence,[22] a normal Sm ion maintains an electron in the nearest coordination sphere, even when the electron escapes from the f shell, i.e., when the samarium ion formally has a $+3$ valence. Actually, we are dealing with a small-radius exciton. The radius of the exciton should be appreciably greater than that of a free ion with valence $+3$ and can be close to the radius of an ion with valence $+2$. This explanation helps to explain the tendency of ions with valence $+3$ to break the local lattice symmetry (the Jahn–Teller effect), while the Eu^{2+} ion makes it possible for the cubic symmetry of the environment (i.e., the unperturbed state) to remain. Apparently, it is the stronger deformation of the crystal lattice by trivalent ions in SmB_6 that stabilizes the static Jahn–Teller effect and is responsible for the observation of only tetragonally split states of the Sm^{3+} ion in SmB_6 doped with Er^{3+} and Gd^{3+}.

In summary, our experimental results taken together indicate that the magnetic moments of the Sm^{3+} ions in SmB_6 interact with mixed electron–phonon vibrational modes and attests to an exciton–polaron nature of the ground state of a semiconductor with fluctuating valence.

4. CONCLUSIONS

Electron spin resonance on samarium ions with stabilized valence Sm^{3+} was investigated in the semiconductor SmB_6 with fluctuating valence. The measurements were performed on single crystals of both pure SmB_6 and SmB_6 doped with rare-earth ions with different valence: Eu^{2+}, Er^{3+}, and Gd^{3+}. The parameters of the spin Hamiltonian were obtained and the splitting E_{JT} between the quartet and doublet states in SmB_6 was estimated. It was established that all observed facts taken together can be explained by the existence of the dynamic and static Jahn–Teller effects in this compound. In addition, as far as we know, this is the first observation of a static Jahn–Teller effect on rare-earth ions. It was shown that the observation of the Jahn–Teller effect in a fluctuating-valence compound could be due to the excitonic–polaron nature of the ground state of such a semiconductor.

This work was supported by the Russian Fund for Fundamental Research (Project 97-02-16235).

*)E-mail: tatiana@dionis.kfti.kcn.ru

[1] D. I. Khomskiĭ, Usp. Fiz. Nauk **129**, 443 (1979) [Sov. Phys. Usp. **22**, 879 (1979)].
[2] T. S. Al'tshuler, G. G. Khaliullin, and D. I. Khomski Zh. Éksp. Teor. Fiz. **90**, 2104 (1986) [Sov. Phys. JETP **63**, 1234 (1986)].
[3] G. Wiese, H. Schaeffer, and B. Elschner, Europhys. Lett. **11**, 791 (1990).
[4] S. A. Al'tshuler and B. M. Kozyrev, *Electronic Paramagnetic Resonance* (Academic Press, New York, 1964) [Russian original, Nauka, Moscow, 1972].
[5] H. Sturm, B. Elschner, and K. H. Hoeck, Phys. Rev. Lett. **54**, 1291 (1985).
[6] H. Sturm, Dissertation, Darmstadt (1985).
[7] T. S. Altshuler and M. S. Bresler, Czech. J. Phys. **46**, Suppl. S4, 1985 (1996).
[8] T. S. Al'tshuler and M. S. Bresler, JETP Lett. **66**, 681 (1997).
[9] .S. Altshuler, V. N. Mironov, and M. M. Zaripov, J. Phys. **15**, 3785 (1982).
[10] S. Kunii, T. Uemura, Y. Chiba, T. Kasuya, and M. Date, J. Magn. Magn. Mater. **52**, 271 (1985).

[11] T. Uemura, Y. Chiba, M. Hagiwara, and M. Date, J. Phys. Soc. Jpn. **55**, 3737 (1986).
[12] S. V. Demishev, A. V. Semeno, N. E. Sluchanko, N. A. Samarin, J. Singleton, A. Ardavan, S. J. Blundell, W. Hayes, and S. Kunii, JETP Lett. **64**, 760 (1996).
[13] M. Kasaya, H. Kimura, Y. Isikawa, T. Fujita, and T. Kasuya, in *Valence Fluctuations in Solids*, edited by L. M. Falicov, W. Hanke, and M. B. Maple (North Holland, Amsterdam, 1981, p. 251).
[14] A. Abragam and B. Bleaney, *Electron Paramagnetic Resonance of Transition Ions* (Clarendon Press, Oxford, 1970) [Russian translation, Vols. 1, 2, Mir, Moscow, 1972, 1973].
[15] A. A. Antipin, I. N. Kurkin, L. Z. Potvorova, and L. Ya. Shekun, Fiz. Tverd. Tela (Leningrad) **7**, 3209 (1965) [Sov. Phys. Solid State **7**, 2596 (1965)].
[16] P. Nyhus, S. L. Cooper, Z. Fisk, and J. Sarrao, Phys. Rev. **55**, 12488 (1997).
[17] I. B. Bersuker, *Electronic Structure and Properties of Coordination Compounds* (Khimiya, Leningrad, 1976, p. 195).
[18] C. Weber, E. Sigmund, and M. Wagner, Phys. Status Solidi B **138**, 661 (1986).
[19] K. A. Kikoin and A. S. Mishchenko, Zh. Éksp. Teor. Fiz. **104**, 3810 (1993) [JETP **77**, 828 (1993)].
[20] P. A. Alekseev, A. S. Ivanov, B. Dorner *et al.*, Europhys. Lett. **10**, 457 (1989).
[21] P. A. Alekseev, A. S. Ivanov, V. N. Lazukov *et al.*, Physica **180–181**, 281 (1992).
[22] K. A. Kikoin, Zh. Éksp. Teor. Fiz. **85**, 1000 (1983) [Sov. Phys. JETP **58**, 582 (1983)].

Translated by M. E. Alferieff

Transmission times of wave packets tunneling through barriers

Yu. E. Lozovik*) and A. V. Filinov†)

Institute of Spectroscopy, Russian Academy of Sciences, 142092 Troitsk, Moscow Region, Russia
(Submitted 6 November 1998)
Zh. Éksp. Teor. Fiz. **115**, 1872–1889 (May 1999)

The transmission of wave packets through barriers by tunneling is studied in detail by the method of quantum molecular dynamics. The distribution of the arrival times of a tunneling packet in front of and behind a barrier and the momentum distribution function of the packet are calculated. The average position and average momentum of the packet and their spread are investigated. It is found that below the barrier a part of the packet is reflected, and a Gaussian barrier increases the average momentum of the transmitted packet and its spread in momentum space. © *1999 American Institute of Physics.* [S1063-7761(99)02505-6]

1. INTRODUCTION

The study of tunneling in nanostructures has acquired an important role in the last few years in connection with advances in nanoelectronics. The problem of tunneling of wave packets through a potential barrier arises in many cases, for example, in the study of the action of femtosecond light pulses on coupled wells. This problem is also important because of possible applications of scanning tunneling microscopes irradiated with femtosecond pulses for studies of nanostructures with high spatial and temporal resolution simultaneously.[1] Other interesting questions are the tunneling time in the ionization of a hydrogen atom by ultrashort laser pulses and the time for tunneling induced by the action of a laser pulse on low-lying nuclear energy levels. In the present paper we investigate the no less interesting question of the tunneling time in nanostructures. The tunneling time is of practical interest in this case because it permits the response time of semiconductor components to be estimated. In this connection we shall study the following problem: Let a laser pulse produce a wave packet of an excited electron near a tunneling barrier. The question is: When will the tunneling portion of the packet appear behind the barrier? The arrival of the wave packet can be detected by studying local variations of the optical properties using ultrashort probe pulses.

It is interesting that a number of effects which are absent in the time-independent case are observed when a packet passes through a tunneling barrier. The tunneling time of a packet is determined in general not by the reciprocal of the probability of time-independent tunneling, but rather is related to complicated processes—the change in the shape and behavior of the packet inside the barrier. Moreover, the transmission time through a barrier is not a unique function of the measured quantities, i.e., the type of experiment.

The investigation of the question of the residence time of a tunneling particle below a barrier started long ago,[2–4] and many theoretical and experimental methods for measuring the tunneling time have been proposed. For example, there exist approaches where the peak of the packet or the average position (the centroid) is chosen as the observed quantity while the reflection or transmission time is determined by their evolution. However, it has been shown in Refs. 5 and 6 and subsequent works that a wave-packet peak incident on a potential barrier is not transformed into the peak of the transmitted wave. In Ref. 7 a case is examined where the high-energy components reached the barrier before the other components because of dispersion of the wave packet in momentum space. Since the tunneling probability increases with energy, these components made the main contribution to the transmitted part of the packet. The initial parameters could be chosen so that the transmitted part of the packet left the barrier long before the arrival of the main peak, chosen as the observed quantity. This example demonstrates the breakdown of the causality principle, which is the basis of this method, and therefore limits the applicability of the method. Moreover, it is difficult to conceive of an experimental method for measuring the arrival time of a packet according to its peak or centroid.

There also exists a class of approaches that employ an ensemble of dynamical trajectories to find the tunneling time. These dynamical trajectories arise as a necessary apparatus of the description in the Feynman and Bohm interpretations of quantum mechanics. When Feynman trajectories were used,[8] the transmission time through a barrier was found as a path integral over all possible trajectories that start from a prescribed point to the left of the barrier and arrive at a certain time at a point located to the right of the barrier. The integrated function in the path integral contained the product of a classical residence time of a trajectory inside the barrier and a weight factor $\exp\{iS(x(t'))/\hbar\}$, where $S(x(t'))$ is the action related to the trajectory $x(t')$ under consideration. The computed times possess real and imaginary parts because of the multiplication by a complex weighting factor, and the question of how these times should be associated with the physically observable quantities, which are always real, arose. To explain the complex times, which arise in other methods also (for example, in the method of physical clocks (see below)), Sokolovski and Connor[9] examined so-called direct and indirect measurements. In indirect measurements, such as the method considered here, the quantities can be complex.

Approaches employing physical clocks have found wide application. Physical clocks are various additional degrees of freedom in the system that allow the residence time of a particle to be determined in a given region. Three types of clocks have been investigated in theoretical works. Baz' and Rybachenko[10,11] used spin precession in a weak uniform magnetic field inside a barrier. At first spin precession in a single plane was considered. Then Buttiker and Landauer[12] extended the analysis to three dimensions. Over the tunneling time a spin acquires a component in the direction of motion and parallel to the magnetic field. It is obvious that the intensities of the detected components with spin polarization in these two directions will be proportional to the residence time of the particle in the region with the magnetic field, i.e., in the region of the barrier. It was found that for a square barrier the tunneling times found in this manner are identical to the real and imaginary parts of the complex tunneling time introduced via Feynman path integrals.[8] The extension of this method to the case of an arbitrary potential barrier was made in Ref. 13. Buttiker and Landauer[14] considered as physical clocks an oscillating barrier in which the amplitude of the oscillations of the temporal component was much smaller than the barrier height. At low frequencies particles see an effective static barrier, since the transmission time through the barrier is much shorter than the period of the oscillations of the temporal component of the barrier. As the frequency increases, the delayed particles or wave-packet components see a slightly modified potential barrier. Finally, for some frequencies one or several periods of the oscillations influence the tunneling particles. The frequency at which a substantial difference from the adiabatic case corresponding to a time-independent barrier appears will be determined by the reciprocal of the interaction time with the barrier or the transmission time through the barrier. Martin and Landauer[15] chose as physical clocks the oscillating amplitude of the incident wave. For this, a wave function consisting of a superposition of two plane waves with different energies was chosen to the left of the barrier. It is obvious that in this case the wave function to the right of the barrier will also be a superposition of the tunneled parts of the plane waves, which, however, possess a different transmission amplitude, since the amplitude depends on the energy. The transmitted wave function will reproduce the incident wave function if the amplitudes of the tunneled plane waves differ very little; this corresponds to the adiabatic case. The energy difference between the initial plane waves for which the wave function behind the barrier does not produce the incident wave function makes it possible to find the transmission time through a potential barrier. The main advantage of this method is that it applies to all types of potentials, but it employs two values of the energy, so that it is not clear to which energy the tunneling time obtained should be ascribed.

Do all clocks give the same measurement result? Of course not. However, in many cases these results are identical or close to one another.[13,16–18] The main advantage of the approaches using physical clocks is that they try to find the tunneling time in terms of possible measurements in physical expeirments.

The search for time operators and the study of their properties is no less popular.[19–24] As first noted by Pauli,[25] the main difficulty is that a measurable hermitian time operator for a system Hamiltonian with a bounded spectrum does not exist. Various attempts have been made to construct operators that would describe the necessary properties of physical times. In order that the constructed operator satisfy the correspondence principle, relations from classical mechanics were taken as the basis for the operator construction. However, it is well known that the construction of an operator expression corresponding to a classical quantity is not unique, and its relation to the measurement process requires additional analysis.

In the present work the tunneling time was calculated as the difference of the average arrival and residence times of a wave packet (see Sec. 3) before and after the barrier. The method of quantum molecular dynamics was used to calculate these times and to investigate the dynamics of a tunneling wave packet.[26–28]

As is well known, molecular dynamics investigates the properties of classical systems in phase space. Therefore it is likewise natural to extend this method to quantum systems in phase space. The evolution of a system in phase space can be described, for example, in the Wigner formalism of quantum mechanics by the Wigner–Liouville equation. To solve the Wigner–Liouville equation written in integral form it is convenient to rewrite the equation in the form of an iterative series. Each term of this series can be treated as the weighted contribution of a trajectory consisting of segments of classical trajectories separated by finite disturbances of the momentum. In what follows we shall call such a trajectory a quantum trajectory. The statistical ensemble of quantum trajectories makes it possible to calculate the sum of all terms in the series. The Monte Carlo method is used to take account of only the trajectories making the main contribution. In the classical limit the quantum trajectories pass into classical trajectories, and the method of generalized molecular dynamics becomes identical to ordinary molecular dynamics. The principles of the method are presented in Sec. 2. The expressions for calculating the distributions of the arrival and residence times of a wave packet are presented in Sec. 3 on the basis of the Wigner formalism of quantum mechanics. The simulation results are discussed in Sec. 4. The one-dimensional case is considered in this paper, but the method employed makes it possible to perform similar calculations for multidimensional and multiparticle systems, where it has serious advantages from the standpoint of computer time over, for example, the solution of the time-dependent Schrödinger equation.

2. COMPUTATIONAL METHOD

To calculate the quantum-mechanical average of a quantity A for a time-dependent state $|\psi\rangle$ in the Wigner formulation of quantum mechanics it is necessary to calculate an integral in phase space[29]

$$A(t)=\langle\psi|\hat{A}(t)|\psi\rangle=\int\int dqdpA(q,p)W(q,p,t), \quad (1)$$

where, by definition, the Weyl symbol $A(q,p)$ is introduced for the operator $\hat{A}$ and $W(q,p,t)$ is the Wigner function, which is the Fourier transform of the off-diagonal density-matrix element:

$$A(q,p)=\int d\xi \exp\left(\frac{ip\xi}{\hbar}\right)\left\langle q+\frac{\xi}{2}|\hat{A}|q-\frac{\xi}{2}\right\rangle, \tag{2}$$

$$W(q,p,t)=\frac{1}{2\pi\hbar}\int d\xi \exp\left(-\frac{ip\xi}{\hbar}\right) \times\psi^*\left(q-\frac{\xi}{2},t\right)\psi\left(q+\frac{\xi}{2},t\right). \tag{3}$$

Differentiating the distribution function with respect to time, substituting it for the time derivative of the function ψ on the right-hand side of the Schrödinger equation, and integrating by parts, we obtain the Wigner–Liouville integrodifferential equation[30]

$$\frac{\partial W}{\partial t}+\frac{p}{m}\frac{\partial W}{\partial q}+F(q)\frac{\partial W}{\partial p}=\int_{-\infty}^{\infty} ds\, W(p-s,q,t)\omega(s,q). \tag{4}$$

In this equation

$$\omega(s,q)=\frac{2}{\pi\hbar^2}\int dq' V(q-q') \times\sin\left(\frac{2sq'}{\hbar}\right)+F(q)\frac{d\delta(s)}{ds} \tag{5}$$

takes account of the nonlocal contribution of the potential, and $F(q)=-\partial V(q)/\partial q$ is a classical force. In the classical limit, $\hbar\to 0$, Eq. (4) becomes the classical Liouville equation

$$\frac{\partial W}{\partial t}+\frac{p}{m}\frac{\partial W}{\partial q}=-F(q)\frac{\partial W}{\partial p}. \tag{6}$$

Equation (4) can be written in integral form. For this purpose one introduces the dynamical trajectories $\{\overline{q}_\tau(\tau;p,q,t),\overline{p}_\tau(\tau;p,q,t)\}$, $\tau\in[0,t]$, starting from the point (p,q) at time $\tau=t$:

$$\begin{aligned} &d\overline{p}/d\tau=F(\overline{p}(\tau)), \quad \overline{p}_t(t;p,q,t)=p,\\ &d\overline{q}/d\tau=\overline{q}(\tau)/m, \quad \overline{q}_t(t;p,q,t)=q. \end{aligned} \tag{7}$$

An integral equation is obtained by substituting the right-hand sides of these equations into the Wigner–Liouville equation, whose left-hand side becomes a total differential, and integrating over time

$$W(p,q,t)=W^0(\overline{p}_0,\overline{q}_0)+\int_0^t d\tau\int_{-\infty}^{\infty} ds\, W \times(\overline{p}_\tau-s,\overline{q}_\tau,\tau)\omega(s,\overline{q}_\tau). \tag{8}$$

Here $W^0(\overline{p}_0,\overline{q}_0)=W(p,q,0)$ is the Wigner distribution function at zero time. The solution of Eq. (8) can be represented as an iterative series. For this we introduce the notation $\tilde{W}^{\tau_1}$ for the distribution function, which evolves classically in the interval $[0,\tau_1]$, and the integral operator $K_{\tau_i}^{\tau_{i+1}}$ describing the evolution between the times τ_i and τ_{i+1}. Now Eq. (8) can be represented in the form

$$W^t=\tilde{W}^t+K_\tau^t W^\tau, \tag{9}$$

where $\tilde{W}^t=W^0(\overline{p}_0,\overline{q}_0)$. The corresponding iterative series solving this equation can be written as

$$W^t=\tilde{W}^t+K_{\tau_1}^t\tilde{W}^{\tau_1}+K_{\tau_2}^t K_{\tau_1}^{\tau_2}\tilde{W}^{\tau_1}+K_{\tau_3}^t K_{\tau_2}^{\tau_3}K_{\tau_1}^{\tau_2}\tilde{W}^{\tau_1}+\ldots \tag{10}$$

Now, to calculate the quantum-mechanical average (1) it is necessary to calculate a linear functional of the Wigner distribution function

$$\begin{aligned} A(t)&=\int\int dqdp A(q,p)W(q,p,t)\\ &=(A|\tilde{W}^t)+(A|K_{\tau_1}^t\tilde{W}^{\tau_1})+(A|K_{\tau_2}^t K_{\tau_1}^{\tau_2}\tilde{W}^{\tau_1})\\ &\quad+(A|K_{\tau_3}^t K_{\tau_2}^{\tau_3}K_{\tau_1}^{\tau_2}\tilde{W}^{\tau_1})+\ldots \end{aligned} \tag{11}$$

Here the brackets $(\ldots|\ldots)$ for the functions $A=A(p,q)$ and $\tilde{W}^t$ or $K_{\tau_i}^t K_{\tau_{i-1}}^{\tau_i}\ldots K_{\tau_1}^{\tau_2}\tilde{W}^{\tau_1}$ indicate averaging over the entire phase space $\{p,q\}$.

The first term on the right-hand side of Eq. (10) gives the classically evolving initial distribution $W^0(\overline{p}_0,\overline{q}_0)$, i.e., the evolution of the distribution function without quantum corrections. However, even this first term of the iterative series describes not classical but rather quantum effects and can contain arbitrary powers of the Planck constant, since a quantum initial state of the system is taken as the initial data for Eq. (10). The rest of the terms in the iterative series describe quantum corrections to evolution. Each term of the iterative series (10) is a multiple integral. This multiple integral can be replaced by an integral sum, and each term of the integral sum can be represented as a contribution of trajectories of a definite topological type. These trajectories consist of segments of classical trajectories—solutions of Eqs. (7)—separated from one another by random perturbations of the momentum.

All terms of the iterative series can be calculated in accordance with the theory of Monte Carlo methods for solving linear integral equations. For this, a Monte Carlo scheme generating a large sample of the terms making the main contribution to the series (10) has been derived. This sample also decreases the computer time required to calculate the rest of the integrals appearing in each term of the iterative series. Let us consider the second term of the series (10). This term can be rewritten as

$$\begin{aligned} K_{\tau_1}^t\tilde{W}^{\tau_1}&=\int_0^1 d\tau_1\int ds_1\omega(s_1,\overline{q}_1)W^0(\overline{p}_0^1,\overline{q}_0^1)\\ &=\int_0^1 d\tau_1[B(\overline{q}_2)(1+Q(\overline{q}_2))]\theta(1-\tau_2)r(\tau_2)\\ &\quad\times\int ds_1 P(s_1,\overline{q}_1^2)C(\overline{q}_1^2)r(\tau_1) \end{aligned}$$

$$\{\sigma(s_1,\overline{q_1^2})t\tilde{Q}(\overline{q_1^2})\theta(\tau_2-\tau_1)/C(\overline{q_1^2})r(\tau_1)\}W^0(\overline{p_0^1},\overline{q_0^1}),$$

where the substitution of variables $\tau\to\tau t(\tau\in[0,1])$ was made for all terms of the iterative series (10). The quantity $r(\tau_1)$ is the probability of choosing a random time τ_1 and θ is the step function.

Once the second term of the series (10) is written in the form (12), it can be given the following probabilistic interpretation. We shall employ the time-reversibility of the equations of classical dynamics (7) and start the construction of a trajectory at time $\tau=0$. At time τ_1 for a trajectory representing an arbitrary term in the iterative series a perturbation of the momentum of the trajectory by an amount s_1 can occur with probability $C(\overline{q_1^2})$, and the probability of rejecting a momentum perturbation is $B(\overline{q_2})$ $(C(\overline{q_1^2})+B(\overline{q_2})=1)$. The probability B for rejecting momentum jumps was introduced to make the algorithm more flexible, so that, depending on the degree of quantization of the system, a transition from quantum to classical trajectories would occur automatically.

Since we are considering a trajectory representing the second term in the iterative series, a perturbation of the momentum at the time τ_1 was accepted. Now it is necessary to choose in the time interval $[\tau_1,1]$ a random value τ_2, which is the time of the next attempt to perturb the momentum. After a perturbation of the momentum by an amount s we must continue the generation of the trajectory up to the time τ_2 in accordance with Hamilton's equations. At this time an attempt to perturb the momentum for the second term of the iterative series must be rejected, and we continue the generation of the trajectory up to the time $\tau=1$. The rejected attempt at disturbing the momentum must be taken into account by multiplying the weighting function of the trajectory by a compensating factor, which stands in the braces on the right-hand side of the expression (12). The product of the Weyl symbol of the operator under consideration and the weighting function at different points along the trajectory gives the time dependence of the computed quantities. Averaging over a large ensemble of trajectories of this type gives the contribution of the second term of the iterative series.

Similar expressions but with a large number of intermediate times on classical trajectories when a perturbation of the momentum occurs can also be written for the other terms in the series (10). The number of the term in the iterative series (10) described by the given trajectory determines the number of momentum perturbations along the trajectory.

The final expression used to calculate the linear functional (11) is

$$\begin{aligned} A(t)&=M\{\alpha(A;T_i)\}\\ &=\sum_{p,q}(\triangle p\triangle q)\sum_{i=0}^{\infty}\sum_{j=0}^{i}\sum_{\tau_j}\sum_{s_j}\alpha(A;T_i)P(T_i),\\ \alpha(A;T_i)&=A(p,q)W^0(\overline{p_0^1},\overline{q_0^1})\Omega(T_i), \end{aligned} \tag{13}$$

where the functions P and Ω are, respectively, the probability of generating a quantum trajectory T_i and the weighting function of this trajectory.

3. MEASURED QUANTITIES

The study of the evolution of a wave packet can be taken as the starting point for studying the temporal aspects of tunneling. The probability of observing a wave packet or particle at an arbitrary point X is determined by the squared absolute value $|\psi(X,t)|^2$ of the wave function. In a time-dependent problem this probability depends on the time and determines the characteristic times of the wave-packet dynamics. If an ideal detector (i.e., measurement by the detector does not disturb the wave function), responding to the presence of particles, is used in the experiment, then the average residence time measured by the detector at the point X is

$$\tilde{t}_X=\frac{\int_0^\infty dt\,t|\psi(X,t)|^2}{\int_0^\infty dt|\psi(X,t)|^2}. \tag{14}$$

A description of these times can be found in Refs. 31–33. The distribution of residence times at the point X is

$$\tilde{P}(t_X)=\frac{|\psi(X,t)|^2}{\int_0^\infty dt|\psi(X,t)|^2}. \tag{15}$$

To find the squared wave function $|\psi(X,t)|^2$ it is sufficient to calculate a quantum-mechanical average of an operator

$$\begin{aligned} \langle\psi(t)|\delta(\hat{q}-X)|\psi(t)\rangle&=\int dq\,\delta(q-X)|\psi(q,t)|^2\\ &=|\psi(X,t)|^2. \end{aligned}$$

In the language of the Wigner formalism this is equivalent to calculating the integral

$$\begin{aligned} \langle\psi(t)|\delta(\hat{q}-X)|\psi(t)\rangle&=\int\int dq\,dp\,\delta(q-X)W(q,p,t)\\ &=\int dp\,W(X,p,t). \end{aligned} \tag{16}$$

If the point X is chosen to the right of the barrier, then this integral makes it possible to calculate the squared wave function which has tunneled through the barrier. The distribution of the residence times can be rewritten, in accordance with Eq. (16), as

$$\begin{aligned} P_X(t)&=\frac{|\psi(X,t)|^2}{\int_0^\infty dt|\psi(X,t)|^2}\\ &=\frac{\int dpW(X,p,t)}{\int_0^\infty}dt\int dpW(X,p,t). \end{aligned} \tag{17}$$

To determine the average time when the wave packet passes through a detector at the point X it is necessary to calculate the integral

$$\langle t(X)\rangle = \int_0^\infty dt\, tP_X(t), \tag{18}$$

and the average transmission time of a packet from the point X_i to the point X_f will be

$$\langle t_T(X_i,X_f)\rangle = \langle t(X_f)\rangle - \langle t(X_i)\rangle. \tag{19}$$

If the points X_i and X_f are chosen on different sides of the potential barrier, then the expression (19) can be used to estimate the tunneling time.

The chief drawback of the definition (17) is that, as a rule, detectors responding to a flux density and not a probability density are used in physical experiments. Therefore a different quantity must be considered in order to compare theory and experiment. For this, the distribution of arrival times of a wave packet at a prescribed point in terms of the probability flux density was introduced:[34]

$$P_X(t) = \frac{\langle \psi(t)|\hat{J}(X)|\psi(t)\rangle}{\int_0^\infty dt\langle \psi(t)|\hat{J}(X)|\psi(t)\rangle}, \tag{20}$$

where

$$\hat{J}(X) = \frac{1}{2}[\hat{p}\,\delta(\hat{q}-X) + \delta(\hat{q}-X)\hat{p}]. \tag{21}$$

Of course, the definition (20) is not a real distribution function from probability theory, since this function can assume negative values at some points. Nonetheless, the definition (20) will be a distribution function if there is no reverse flux through the point X or the flux is negligibly small. For this the point X is chosen sufficiently far from the barrier. Measuring the distribution of the arrival times of a packet in front of and beyond the barrier, the transmission time through a region much larger than the region of the potential barrier can be calculated. This time is analogous to the asymptotic phase times[35] and in addition to the tunneling time and the packet–barrier interaction time it also contains the transmission time through the region where the potential barrier is zero. These two times cannot be resolved. Despite continuing discussions, this tunneling-time problem has still not been finally solved.[19–24,36,37]

Another problem concerns the physical implementation of an experiment in which simultaneous detection of a packet in front of and beyond a barrier would not substantially reduce the wave function. For this reason, ordinarily, a different quantity—the time delay—is measured in experiments.[38–42] A time delay arises because of the presence of a barrier and is defined as the difference of the average arrival times of the tunneling and free packets:

$$\Delta\tau_{\text{arrival}}(X) = \langle t_X\rangle^{\text{tun}} - \langle t_X\rangle^{\text{free}}. \tag{22}$$

The definition (20) for calculating the average arrival times gives a reasonable estimate of the time delays measured in an experiment.

The distribution of arrival times (20) can be rewritten in the Wigner formulation of quantum mechanics as

$$P_X(t) = \frac{\int\int dqdpJ_X(q,p)W(q,p,t)}{\int_0^\infty dt\int\int dqdpJ_X(q,p)W(q,p,t)}, \tag{23}$$

where the Weyl symbol of the current operator $\hat{J}(X)$ is

$$J_X(q,p) = \frac{\hbar}{2}\sin\left(\frac{2p(X-q)}{\hbar}\right)\frac{\partial}{\partial q}\delta(q-X). \tag{24}$$

Substituting into Eq. (20) the expression (24) and calculating the integral over the variable q by parts we obtain the expression

$$P_X(t) = \frac{\int dppW(X,p,t)}{\int_0^\infty dt\int dppW(X,p,t)}. \tag{25}$$

Comparing the expressions (17) and (25), it is easy to see that they differ by the fact that the momentum p appears in the numerator and denominator in Eq. (25). This momentum appeared in the last expression because the probability flux density is measured there.

4. SIMULATION RESULTS

We shall examine a series of experiments on the tunneling of an electron with the wave function

$$\psi(x,0) = \frac{1}{(2\pi\sigma_x)^{1/4}}\exp\left[-\left(\frac{x-x_0}{2\sigma_x}\right)^2 + ik_0x\right] \tag{26}$$

through a Gaussian potential barrier

$$V(x) = V_0\exp\left[-\frac{(x-d)^2}{\sigma^2}\right].$$

The Wigner distribution function (3) corresponding to the initial wave function of the electron can be written as

$$W(p,q,0) = 2\exp\left[-\frac{(q-x_0)^2}{2\sigma^2}\right]\exp\left[-\frac{2\sigma^2(p-\hbar k_0)^2}{\hbar^2}\right]. \tag{27}$$

The center $x_0 = \langle\psi(x,0)|\hat{x}|\psi(x,0)\rangle$ of the wave packet at zero time was chosen far enough from the left-hand boundary of the barrier so that the probability density beyond the barrier would be negligibly small compared with the transmission probability $|T|^2$ through the barrier. Tunneling occurred through a wide ($\sigma=2.5$ nm, typical of $Al_xGa_{1-x}As$ structures) and narrow ($\sigma=0.5$ nm) Gaussian barriers of height $V_0=0.3$ eV centered at $d=0$. The electron kinetic energy was $E_0=\hbar^2k_0^2/2m=V_0/2=0.15$ eV. We employ a system of units with $\hbar=m=V_0=1$. Distances were measured in units of the reduced de Broglie wavelength $\lambda=1/k_0$. In this system the parameters of the wave packet and barrier are $E_0=0.5$, $\Delta k=0.04$ (0.125), $\sigma_x=1/2\Delta k=12.5$ (4), $x_0=-92.5$ (-43), $\sigma=5$ (2.5 nm), and $\sigma=1$ (0.5 nm).

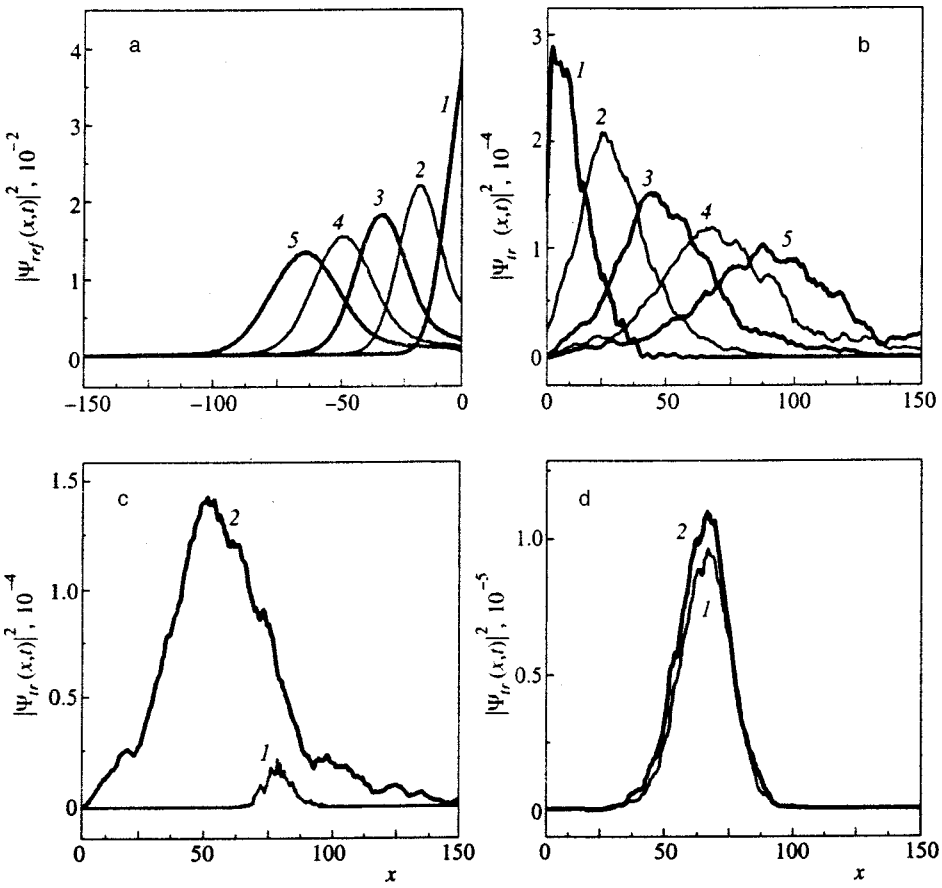

FIG. 1. Probability density (a) $|\psi_{\text{ref}}(x,t)|^2$ of the reflected wave packet and (b) $|\psi_{tr}(x,t)|^2$ of the tunneled wave packet at successive times $t_i = 144-239$ fs (curves *1–5*) with $\Delta k = 0.125$ and barrier thickness $\sigma = 1$ (0.5 nm); (c, d) $|\psi_{tr}(x,t)|^2$ at time $t = 187$ fs, $\Delta k = 0.125$ with barrier thickness $\sigma = 1$ (0.5 nm) (c) and $\sigma = 5$ (2.5 nm) (d): curve *1* — calculation using classical trajectories, curve *2* — calculation using quantum trajectories.

4.1. Evolution of the wave packet

The interaction of a wave packet ($\hbar\Delta k = 0.125$) with a narrow potential barrier [$\sigma = 1$ (0.5 nm)] is shown in Figs. 1a and 1b. These figures show the probability density $|\psi(x,t)|^2$ (curves *1–5*) of reflected (Fig. 1a) and tunneled (Fig. 1b) wave packets at successive times $t = 114-239$ fs. The probability density was calculated using Eq. (16), i.e., in terms of the Wigner distribution function. This integral was calculated along quantum and classical trajectories. In the calculation over classical trajectories only the high-energy components of a packet could pass classically above the barrier. This calculation corresponds to the curve *1* in Fig. 1c, and the evolution of the Wigner function can be described only by the first term of the series (10). In the formalism of quantum trajectories the passage of the components of a packet beyond the barrier is associated with random perturbations of the momentum, i.e., with a virtual change in energy. The results of this calculation correspond to the curve *2* in Fig. 1c. Now the quantum corrections introduced by all terms in the series (10) are taken into account in the evolution of the Wigner function.

Of course, the calculation over quantum trajectories also takes account of the high-energy components that pass above the barrier, since they describe the contribution of the first term in the series (10). However, comparing the curves *1* and *2* in Fig. 1c shows that their role is negligible for a narrow barrier and most of the packet passes above the barrier on account of the virtual change in energy, described as random perturbations of the momentum of the quantum trajectories. A study of tunneling through a wide barrier leads to the opposite conclusion. The curves *1* and *2* in Fig. 1d are almost coincident. This means that most of the packet has passed above the barrier, and the contribution of all terms in the series (10), except for the first term, is negligibly small. To avoid such a situation and to restore the importance of quantum effects, it is necessary to decrease the uncertainty of the momentum of the initial wave packet. In what follows all calculations for a wide barrier are presented for momentum uncertainty $\hbar\Delta k = 0.04$.

4.2. Average position, average momentum, and their spreads

Figure 2a shows the evolution of the average position $\langle\psi(t)|\hat{X}|\psi(t)\rangle$ of the wave packet for calculation according to classical (curve *1*) and quantum (curve *2*) trajectories. In these two methods for calculating the average position $\overline{X}$ no differences are observed before interaction with the barrier (curves *1* and *2* are coincident). This result can be explained as follows. In the method under discussion the quantum-mechanical properties appear at two points: in the properties of the initial state of a wave packet and in the evolution of the packet. Since the same initial data were chosen for the quantum and classical trajectories, the fact that $\overline{X}$ is the same must be explained by the evolution of the wave packet. Specifically, while the packet moves freely in front of the barrier, it is correctly described by classical trajectories also. In this case the first term in the series (10) suffices to describe

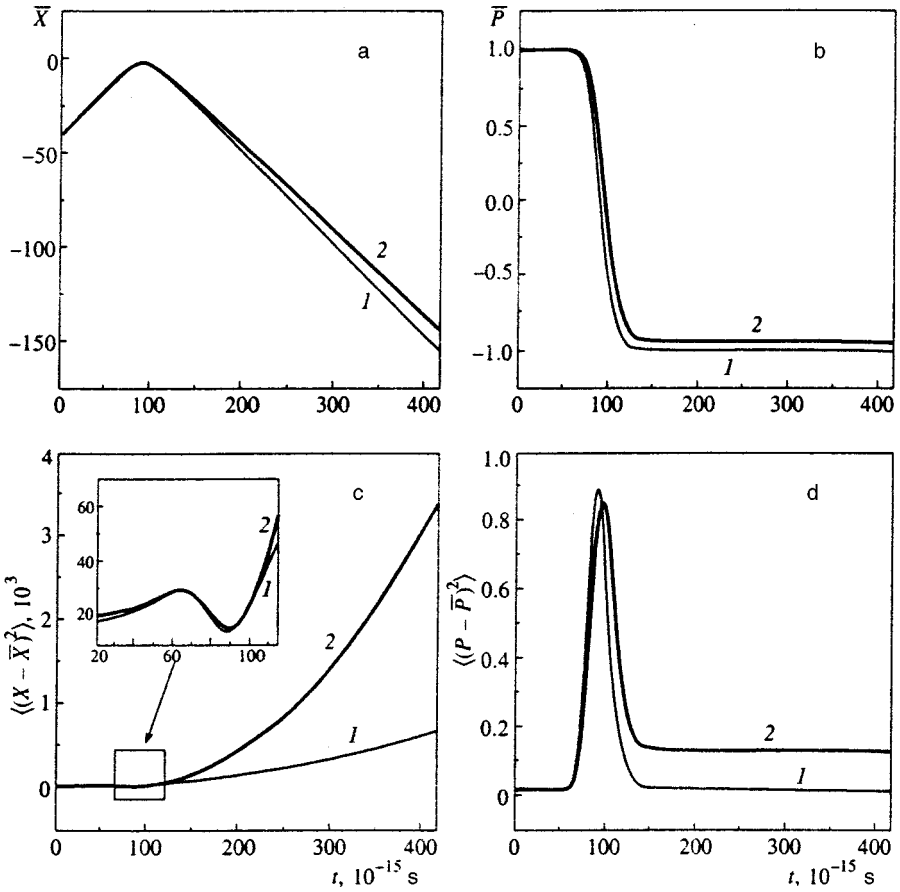

FIG. 2. Average position $\overline{X}$ (a), average momentum $\overline{P}$ (b), position spread $\langle (X-\overline{X})^2\rangle$ (c) and momentum spread $\langle (P-\overline{P})^2\rangle$ (d): *1* — calculation using classical trajectories; *2* — calculation using quantum trajectories.

the evolution of the Wigner function. This result can also be obtained analytically, estimating the right-hand side of the Wigner–Liouville equation (4). For the initial Wigner function (27) and Gaussian barrier which we have chosen it is easy to show that the integral on the right-hand side of Eq. (4) decays exponentially as a function of distance from the barrier. In this case Eq. (4) becomes the classical Liouville equation, whose characteristics are ordinary classical trajectories.

A difference in the behavior of the curves *1* and *2* appears after the packet interacts with a barrier. Now the classical trajectories are no longer characteristics and do not describe the evolution of the wave packet correctly. In Figs. 2a and 2b the average position and momentum of the calculation over quantum trajectories (curve *2*) are greater than for classical trajectories (curve *1*). This is due to the following circumstances. In the first place, since most of the packet is reflected, as one can see from Fig. 2b the average momentum changes sign after being scattering by the barrier. In the second place, the classical trajectories (curve *1*) do not take account of tunneling; they only take account of the negligible above-barrier transmission, arising because of the uncertainty in the momentum of a Gaussian wave packet. At the same time it is obvious that the tunneling part of the packet has positive momentum and moves in the opposite direction relative to the reflected part. Therefore its contribution to $\overline{X}$ and $\overline{P}$ has a different sign. This is the explanation of the difference between the curves *1* and *2*.

In addition, the motion of the tunneling and reflected packets on different sides of the barrier also explains the more rapid increase of the position spread in the quantum calculation (curve *2*, Fig. 2c) as compared with the classical calculation (curve *1*), which takes only the spreading of the wave packet into account. The behavior of the packet width on scattering by a barrier is shown in greater detail in the upper left-hand part of Fig. 2c.

The interaction of a packet with the barrier also gives rise to interesting behavior of the momentum spread in Fig. 2d. The constant values (curve *1*) on the initial and final sections show the momentum spread in the incident and reflected wave packets, i.e., before and after interaction with the barrier. The observed peak is due to the change in the sign of the momentum of the packet and to the fact that different components reach the barrier and are reflected from it at different times. The increase in the momentum spread (curve *2*) on the final section is explained by the appearance of a tunneling packet with positive momentum in the quantum computational method, while the total average momentum is negative.

4.3. Distribution of arrival and residence times: momentum distribution function

The results of the calculation of the unnormalized distribution of residence times (17) at different points in front of the barrier, inside the barrier, and beyond barrier are presented in Figs. 3a and 3b (curves *1*–*5*). Figures 4a and 4b show the analogous results for the unnormalized distribution

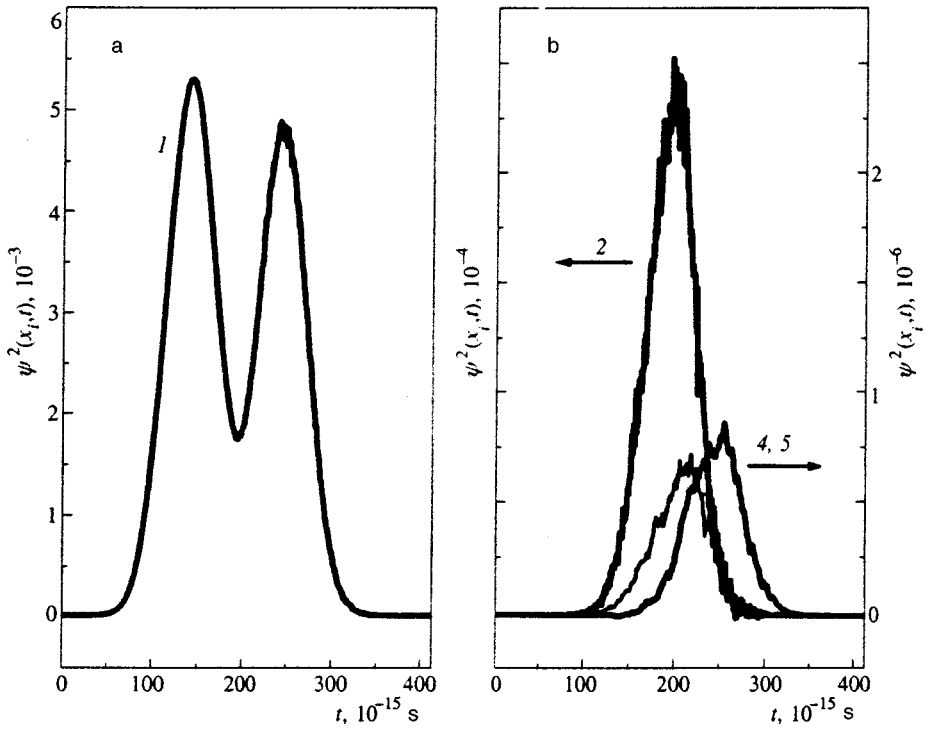

FIG. 3. Probability density or unnormalized distributions of the residence times (17): a — $|\psi(x_i,t)|^2$ at the point $x_1=-5\sigma$ (curve *1*); b — $|\psi(x_i,t)|^2$ at the point $x_2=-0.67\sigma$ (curve *2*), at $x_4=0.67\sigma$ (curve *4*), and $x_5=5\sigma$ (curve *5*); center of the barrier located at $x_3=0$, barrier thickness $\sigma=5$ (2.5 nm).

(20) of the arrival times. The curves *1* in Figs. 3a and 4a show the behavior of the probability density and flux, corresponding to the fact that the incident and reflected wave packets pass through the detector at different times. Curve *2* in Fig. 4a shows the behavior of the flux measured at a certain point to the left of barrier center. The tunneling and high-energy components present in the initial packet reaching the point classically pass through this point. An interesting result is obtained for the probability flux density in Fig. 4b (curves *3*–*5*). The flux measured at barrier center (curve *3*) is much lower than the flux on the right-hand boundary of the barrier (curve *4*) and far to the right of the barrier (curve *5*). This means that the tunneling components of the wave packet which move in oppposite directions interfere inside the barrier. Some of these components pass completely through the barrier, while others are reflected inside the barrier and do not reach its right-hand boundary. Interference of the reflected and transmitted components yields the observed decrease in the flux amplitude at the barrier center (curve *3*) and at the right-hand boundary (compare curves *4* and *5*).

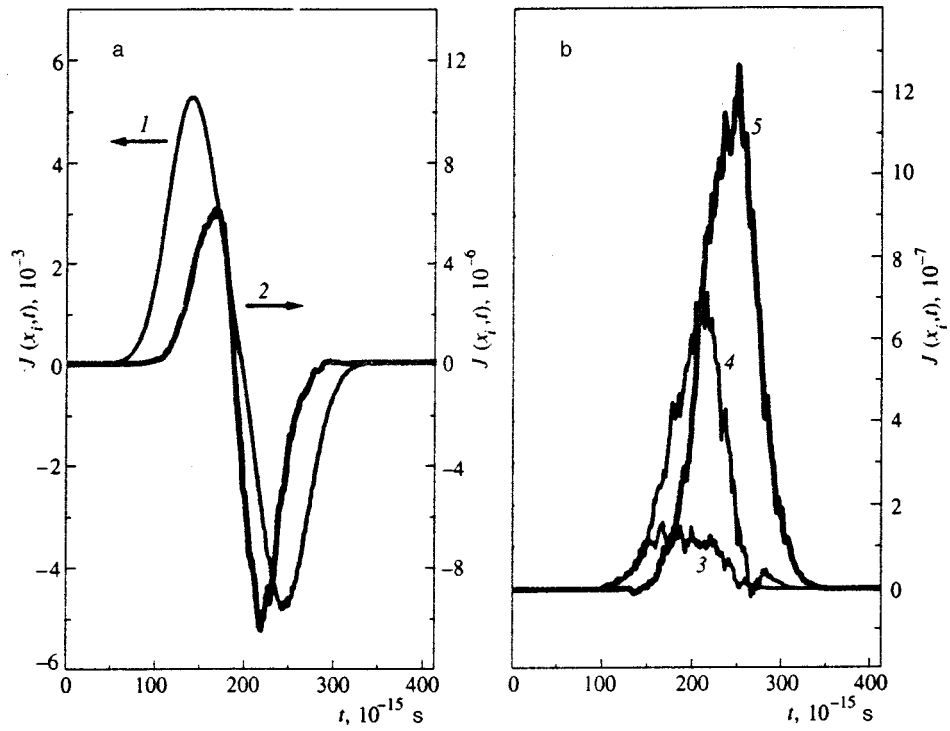

FIG. 4. Probability flux density or unnormalized distributions of the arrival times (20): a — $J(x_i,t)$ at the points $x_1=-5\sigma$ (curve *1*) and point $x_2=-0.67\sigma$ (curve *2*); b — $J(x_i,t)$ at $x_3=0$ (curve *3*), $x_4=0.67\sigma$ (curve *4*), and $x_5=5\sigma$ (curve *5*); center of the barrier located at $x_3=0$, barrier thickness $\sigma=5$ (2.5 nm).

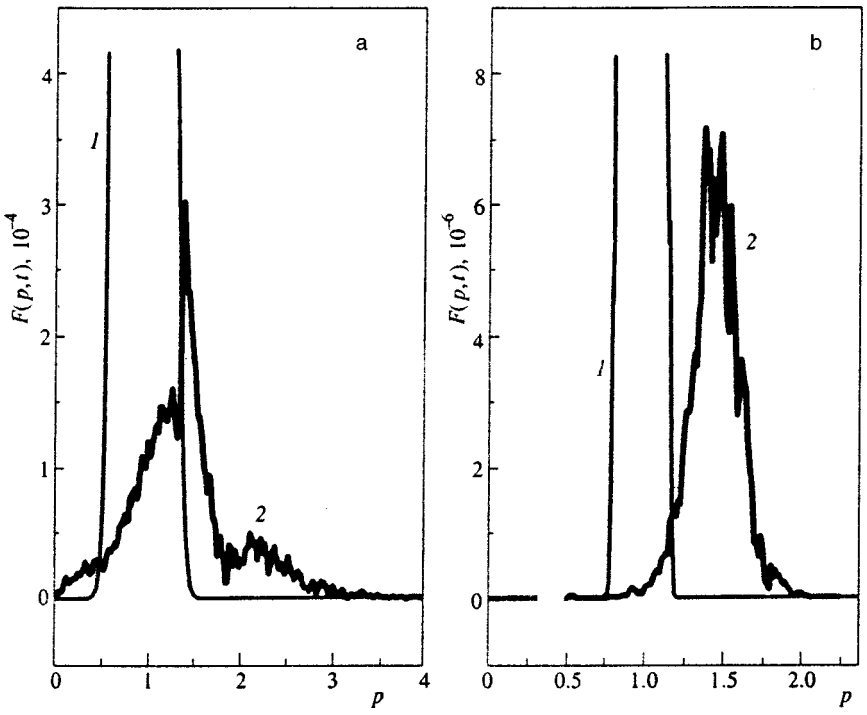

FIG. 5. Momentum distribution in a packet at $t=0$ (curve *1*) and in a packet transmitted through a potential barrier (curve *2*): a — $\Delta k=0.125$, barrier thickness $\sigma=1$ (0.5 nm), $t=218$ fs; b — $\Delta k=0.04$, barrier thickness $\sigma=5$ (2.5 nm), $t=385$ fs.

Interestingly, the investigation of tunneling using classical trajectories in complex time also shows a similar effect.[37] It is found that transmission through a barrier occurs as a series of attempts, many of which are unsuccessful because of reflections in different regions below the barrier.

Comparing the distributions of the residence and arrival times in Figs. 3b and 4b shows that they are almost identical. The computed average residence and arrival times (18) are also identical (the difference is less than 1 fs). As we have already stated, the distribution of the arrival times (20) is not a true distribution function and, as one can see from Fig. 4a (curve *2*), it is not suitable for calculating the average arrival time of a packet in front of the barrier. This makes it impossible to calculate the tunneling time as the difference (19) of the average arrival times of the packet in front of and beyond the barrier. Nonetheless, the expression (19) can be used to estimate the tunneling time if the average residence time (14) is used instead of the average arrival time in front of the barrier. Then the tunneling time through the potential barrier is $\tau_T(-0.67\sigma,+0.67\sigma)=12$ fs, i.e., it is almost equal to the transmission time of a free packet through a similar region $\tau_T^{\text{class}}(-0.67\sigma,+0.67\sigma)=13.4$ fs.

The time delays were measured at the points $x_4=0.67\sigma$ (1.6 nm) and $x_5=5\sigma$ (12 nm) and were found to be $\Delta\tau_{\text{arrival}}(x_4)=8$ fs and $\Delta\tau_{\text{arrival}}(x_5)\leqslant 0.5$ fs. If these measurements are performed even farther to the right of the barrier, then $\Delta\tau_{\text{arrival}}(x)$ becomes negative. Thus an interesting behavior is discerned: Even though the tunneling wave packet is delayed by the barrier ($\Delta\tau_{\text{arrival}}(x_4)=8$ fs) and passes through the barrier approximately in the same time as a free packet, it appears earlier at a definite distance to the right of the barrier. This effect can be explained by the fact that the transmission probability through a Gaussian barrier increases with energy, so that packet components with a larger momentum have a higher probability of ending up beyond the barrier. These components move more rapidly than a free packet and eventually overtake a free packet. Then the time delays can only be negative. This confirms the momentum distribution function

$$\frac{\langle\psi(t)|\,\delta(\hat{p}-p)\,|\psi(t)\rangle}{\langle\psi(t)|\psi(t)\rangle}, \tag{28}$$

calculated for narrow (Fig. 5a) and wide (Fig. 5b) barriers, respectively, at times $t=218$ and 385 fs. At these characteristic times the distribution function no longer changes, since the interaction with the barrier has ceased. It is clear from Fig. 5 that the average momentum of the tunneled wave packet (curve *2*) is greater than the average momentum of the wave packet initially (curve *1*). The peak observed in the momentum distribution function (curve *2* in Fig. 5a) is due to the packet components that had a large momentum and passed above the barrier. It is evident that tunneling through a narrow potential barrier increases the spread in the distribution function, while tunneling through a wide barrier substantially shifts the center of the distribution in the direction of large momenta (curve *2* in Fig. 5b).

5. CONCLUSIONS

The quantum generalization of classical molecular dynamics was used to solve the Wigner–Liouville integral equation in the Wigner formulation of quantum mechanics. The method discussed for solving this equation does not require a large increase of computer time and makes it possible to avoid the computational difficulties that arise when solving the time-dependent Schrödinger equation.

This approach was used to solve the time-dependent problem of tunneling of a finite wave packet, i.e., a problem in which it is important to take account of exponentially small quantum effects. The evolution of a wave packet, the

behavior of the average and the spreads in the position and momentum together with the distributions of the residence and arrival times for the wave packet at different positions of an ideal detector were analyzed. The following results were obtained: 1) The tunneling time through a potential barrier is of approximately the same order of magnitude as the transmission time of a free wave packet over a similar distance; 2) the tunneling wave packet is delayed by the potential barrier, so that after the barrier the time delay should be positive; 3) measurement of negative time delays is possible only at sufficiently large distances from the barrier and is associated with a shift of the momentum distribution function; 4) a Gaussian barrier transmits predominantly the high-energy components of a packet, interaction with the barrier shifts the center of the momentum distribution function so that the average momentum of the transmitted packet is larger than the initial average momentum of the entire packet; 5) tunneling through a narrow potential barrier increases the spread of the momenta of the tunneled components, while tunneling through a wide barrier appreciably increases the average momentum; and 6) the computational results for the probability flux density showed that the tunneling wave packet does not pass completely through the barrier; instead, a portion of the packet is reflected and does not reach the barrier boundary.

This work was partically supported by grants from the Russian Fund for Fundamental Research and the program ''Physics of Solid-State Nanostructures.''

*)E-mail: lozovik@isan.troitsk.ru
†)E-mail: alex@vovan.msk.ru

[1] Yu. E. Lozovik, S. P. Merkulova, D. V. Lisin *et al.*, submitted to Phys. Lett. A; see also *Proceedings of International Symposium ''Nanostructures'97: Physics and Technology,''* St. Petersburg (1997), p. 352.
[2] L. Eisenbud, PhD Thesis, Princeton University (1948).
[3] D. Bohm, *Quantum Theory* (Prentice-Hall, New York, 1951, p. 257).
[4] E. P. Wigner, Phys. Rev. **98**, 145 (1955).
[5] M. Buttiker and R. Landauer, Phys. Rev. Lett. **49**, 1739 (1982).
[6] M. Buttiker and R. Landauer, Phys. Scr. **32**, 429 (1985).
[7] R. Landauer and Th. Martin, Solid State Commun. **84**, 115 (1992).
[8] D. Sokolovski and L. Baskin, Phys. Rev. A **36**, 4604 (1987).
[9] D. Sokolovski and J. Connor, Phys. Rev. A **47**, 4677 (1993).
[10] A. I. Baz', Yad. Fiz. **5**, 229 (1967) [Sov. J. Nucl. Phys. **5**, 161 (1967)].
[11] V. F. Rybachenko, Yad. Fiz. **5**, 895 (1967) [Sov. J. Nucl. Phys. **5**, 635 (1967)].
[12] M. Buttiker and R. Landauer, Phys. Rev. A **27**, 6178 (1983).
[13] C. R. Leavens and G. C. Aers, Solid State Commun. **63**, 1101 (1987).
[14] M. Buttiker and R. Landauer, IBM J. Res. Dev. **30**, 451 (1986).
[15] Th. Martin and R. Landauer, Phys. Rev. A **47**, 2023 (1993).
[16] C. R. Leavens and G. C. Aers, J. Vac. Sci. Technol. A **6**, 305 (1988).
[17] R. Landauer, Phys. Chem. **95**, 404 (1991).
[18] R. Landauer and Th. Martin, Rev. Mod. Phys. **66**, 217 (1994).
[19] J. Kijowski, Rep. Math. Phys. **6**, 362 (1974).
[20] P. Busch, M. Grabowski, and P. J. Lahti, Phys. Lett. A **191**, 357 (1994).
[21] N. Grot, C. Rovelli, and R. S. Tate, Phys. Rev. A **54**, 4676 (1996).
[22] V. Delgado and J. G. Muga, Phys. Rev. A **56**, 3425 (1997).
[23] Y. Aharonov and D. Bohm, Phys. Rev. **122**, 1649 (1961).
[24] J. Léon, J. Phys. A **30**, 4791 (1997).
[25] W. Pauli, in *Encyclopedia of Physics*, 5/1, edited by S. Flugge (Springer, Berlin, 1958, p. 60).
[26] V. S. Filinov, Mol. Phys. **85**, 711 (1995); V. S. Filinov, Mol. Phys. **88**, 1517, 1529 (1996).
[27] V. S. Filinov, Yu. E. Lozovik, A. V. Filinov, I. E. Zakharov, and A. M. Oparin, Izv. Ross. Akad. Nauk, Ser. Fiz. **62**, 1179 (1998); V. S. Filinov, Y. E. Lozovik, A. V. Filinov, I. Zacharov, and A. Oparin, Phys. Scr. **58**, 297, 304 (1998).
[28] V. S. Filinov, S. Bonella, Y. E. Lozovik, A. V. Filinov, and I. Zacharov, in *Classical and Quantum Dynamics in Condensed Phase Simulations* (World Scientific Publishing Company, Singapore, 1998, p. 671).
[29] E. P. Wigner, Phys. Rev. **40**, 749 (1932).
[30] V. I. Tatarskiĭ, Usp. Fiz. Nauk **139**, 587 (1983) [Sov. Phys. Usp. **26**, 311 (1983)].
[31] E. P. Wigner, in *Aspects of Quantum Theory*, edited by A. Salam and E. P. Wigner (Cambridge, London, 1972, p. 237).
[32] M. Bauer, P. A. Mello, and K. W. Mc Voy, Z. Phys. A **293**, 151 (1979).
[33] V. S. Olkhovsky, E. Recami, and A. J. Gerasimchuk, Nuovo Cimento **22**, 263 (1974).
[34] R. S. Dumont and T. L. Marchioro II, Phys. Rev. A **47**, 85 (1993).
[35] E. H. Hauge and J. A. Stovneng, Rev. Mod. Phys. **61**, 917 (1989).
[36] W. R. McKinnon and C. R. Leavens, Phys. Rev. A **51**, 2748 (1995).
[37] D. Mugnai and A. Ranfagni, Nuovo Cimento D **14**, 541 (1992).
[38] A. Ranfagni, D. Mugnai, P. Fabeni, and G. P. Pazzi, Appl. Phys. Lett. **58**, 774 (1991).
[39] A. M. Steinberg, P. G. Kwiat, and R. Y. Chiao, Phys. Rev. Lett. **71**, 708 (1993).
[40] A. Enders and G. Nimtz, Phys. Rev. B **47**, 9605 (1993).
[41] Ch. Spielmann, R. Szipocs, A. Sting, and F. Krausz, Phys. Rev. Lett. **73**, 2308 (1994).
[42] N. C. Kluksdahl, A. M. Kriman, and D. K. Ferry, Phys. Rev. B **39**, 7720 (1989).

Translated by M. E. Alferieff

Kinetics of indirect photoluminescence in GaAs/Al_xGa_{1-x}As double quantum wells in a random potential with a large amplitude

L. V. Butov,*) A. V. Mintsev, and A. I. Filin

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

K. Eberl

Max-Planck-Institut für Festkörperforschung, 70569 Stuttgart, Germany

(Submitted 10 November 1998)
Zh. Éksp. Teor. Fiz. **115**, 1890–1905 (May 1999)

The kinetics of indirect photoluminescence of GaAs/Al_xGa_{1-x}As double quantum wells, characterized by a random potential with a large amplitude (the linewidth of the indirect photoluminescence is comparable to the binding energy of an indirect exciton) in magnetic fields $B \leqslant 12$ T at low temperatures $T \geqslant 1.3$ K is investigated. It is found that the indirect-recombination time increases with the magnetic field and decreases with increasing temperature. It is shown that the kinetics of indirect photoluminescence corresponds to single-exciton recombination in the presence of a random potential in the plane of the double quantum wells. The variation of the nonradiative recombination time is discussed in terms of the variation of the transport of indirect excitons to nonradiative recombination centers, and the variation of the radiative recombination time is discussed in terms of the variation of the population of optically active excitonic states and the localization radius of indirect excitons. The photoluminescence kinetics of indirect excitons, which is observed in the studied GaAs/Al_xGa_{1-x}As double quantum wells for which the random potential has a large amplitude, is qualitatively different from the photoluminescence kinetics of indirect excitons in AlAs/GaAs wells and GaAs/Al_xGa_{1-x}As double quantum wells with a random potential having a small amplitude. The temporal evolution of the photoluminescence spectra in the direct and indirect regimes is studied. It is shown that the evolution of the photoluminescence spectra corresponds to excitonic recombination in a random potential. © *1999 American Institute of Physics.* [S1063-7761(99)02605-0]

1. INTRODUCTION

The neutral system consisting of spatially separated two-dimensional layers of electrons and holes in double quantum wells has been widely studied in recent years.[1–12] This system is of interest primarily because of the possibility of constructing structures with the required architecture and a low rate of indirect (interwell) recombination. Since the effective carrier temperature is determined by the ratio of the relaxation and recombination times, the low indirect-recombination rate makes it possible to obtain a neutral electron–hole system of a high density with a low effective temperature. A number of theoretical treatments have shown that in a system of spatially separated layers of electrons and holes in double quantum wells at low temperatures collective states can be observed, including a condensate of indirect excitons similar to the Bose–Einstein condensate bosons.[13–20] An interesting particular case is a system of spatially separated layers of electrons and holes in a strong magnetic field perpendicular to the plane of the well. A number of theoretical studies have shown that the critical conditions for condensation of excitons are improved in a strong magnetic field as a result of complete quantization of the energy spectrum of electrons and holes[15,16] and as a result of the lifting of spin degeneracy. Theory predicts that when the distance between the electron and hole layers is small, $d \lesssim l_B$ ($l_B = \sqrt{\hbar c/eB}$ is the magnetic length), the ground state of the system is determined by the electron–hole interaction and is an exciton condensate, while for large distances, $d \gtrsim l_B$, the ground state is determined by electron–electron and hole–hole interaction and is an incompressible Fermi liquid or Wigner crystal of electrons and holes.[18,19]

An inevitable property of semiconductor quantum wells and double quantum wells is the existence of a random potential produced in the plane of a well by irregularities of the interfaces, composition fluctuations, defects, and impurities. A random potential qualitatively affects the properties of the system. Specifically, a strong random potential destroys possible collective states (see Ref. 12 and citations there). No theory of a system of spatially separated electron and hole layers in the presence of a random potential is yet available. We shall parametrize the magnitude of the potential by the ratio of the binding energy of an indirect exciton to the width of the indirect luminescence line, determined by the amplitude of the random potential, E_I/Δ_{PL}. In terms of the parameters d, l_B, E_I, and Δ_{PL} (in a zero magnetic field the analog of l_B is the Bohr radius of an indirect exciton), four

classes of spatially separated electron–hole systems in double quantum wells can be distinguished. Class B1 consists of double quantum wells with a small effective distance between the layers and a weak disorder ($d \lesssim l_B$, $E_I \gg \Delta_{PL}$). According to theoretical studies,[13–20] for this class of double quantum wells the ground state of the system at low temperatures should be an excitonic condensate, and the critical conditions for condensation of excitons are improved in a strong magnetic field.[15,16] Class B2 consists of double quantum wells with a small distance between the layers and strong disorder ($d \lesssim l_B$, $E_I \lesssim \Delta_{PL}$). Class A1 consists of double quantum wells with a large distance between the layers and weak disorder ($d \gtrsim l_B$, $E_I \gg \Delta_{PL}$). According to theory,[18,19] for this class of double quantum wells the ground state of the system at low temperatures should be an incompressible Fermi liquid or a Wigner crystal of electrons and holes. Class A2 consists of double quantum wells with a large distance between the layers and strong disorder ($d \gtrsim l_B$, $E_I \lesssim \Delta_{PL}$). This classification must be supplemented by the carrier density and the temperature, which determine the phase boundaries. Interclass transitions between the classes are probably, continuous. Thus a transition between the classes B1 and A1 and a transition between the classes B2 and A2 can be accomplished by increasing the magnetic field, while a transition between the classes B1 and B2 and between the classes A1 and A2 can be followed by studying double quantum wells with various degrees of disorder.

At experimental investigation of $\Gamma - X_z$ AlAs/GaAs double quantum wells, characterized by a small distance between the electron and hole layers, $d \approx 3$–4 nm, $E_I \sim 10$ meV, and Δ_{PL} from 3 to 6 meV, which therefore belong to the classes B1 and B2, implying condensation of indirect excitons in strong magnetic fields at low temperatures have been observed: an anomalous increase of the diffusion coefficient and radiative recombination rate of excitons, interpreted as the appearance of superfluidity of excitons and superradiance of an excitonic condensate,[12] and anomalously large fluctuations of the total intensity of the photoluminescence of excitons, interpreted as critical fluctuations near a phase transition, which are associated with instability of condensate domains.[8] As the disorder in the experimental AlAs/GaAs wells increased, these anomalies became weaker and disappeared, which corresponded to a transition from class B1 to class B2.[12]

The kinetics of photoluminescence in double quantum wells belonging to the class A1 has been investigated in a zero magnetic field. Specifically, double quantum wells GaAs/Al_xGa_{1-x}As with $d \approx 12$ nm, $E_I \sim 5$ meV, and $\Delta_{PL} = 1.3$ meV have been investigated.[21] A sharp increase of intensity and narrowing of the photoluminescence line of indirect excitons were found after the pulsed laser excitation was switched off at low temperatures and high exciton densities. The effect was described by a rapid increase in the population of optically active excitonic 2D states.[21]

In the present work we investigate the optical properties of a system consisting of spatially separated layers of electrons and holes in class A2 wells. Specifically, we investigate GaAs/Al_xGa_{1-x}As double quantum wells with $d \sim 11$ nm and $E_I \sim \Delta_{PL} \sim 6$ meV. The experimental data are compared with the analogous dependences for class B1, B2, and A1 double quantum wells.

2. SAMPLE AND EXPERIMENTAL PROCEDURE

In gate voltage tunable $n^+ - i - n^+$ heterostructure with a single GaAs/Al_xGa_{1-x}As double quantum well, which was adjusted by varying the gate voltage, was grown by molecular-beam epitaxy on an n^+-GaAs substrate. The i layer consists of two, 5 nm wide, GaAs quantum wells separated by a 5.5 nm thick $Al_{0.35}Ga_{0.65}As$ barrier and surrounded by 55 nm thick $Al_{0.35}Ga_{0.65}As$ barriers. The band diagram of the i layer of the structure in the indirect regime is shown in Fig. 1. The 1100 nm thick n^+ layers on the substrate side and the 110 nm thick layer on the surface side were doped with Si to a density $N_{Si} = 5 \times 10^{17}$ cm^{-3}. To improve the electrical contact, δ doping with $N_{Si} = 10^{13}$ cm^{-2} was performed at 10 nm from the surface. As a result of the high doping density the n^+ layers are of a metallic character, and the gate voltage V_g applied between the substrate and the surface decreases in the i layer. The front gate consisted of a frame around a mesa with a 200×200 μm^2 window.

The sample was placed in a helium cryostat with a superconducting solenoid. Excitation and detection were performed through a 200 μm in diameter optical light guide, placed 300 μm from the mesa surface. The carriers were excited by a pulse semiconductor laser ($\hbar\omega = 1.85$ eV). The laser pulse was approximately square with ~ 50 ns duration and ~ 1 ns edges. The temporal resolution of the detection system was 0.5 ns. A double grating monochromator, a photomultiplier, and a time-correlated photon counting system with time resolution were used to detect the signal.

3. PHOTOLUMINESCENCE KINETICS IN THE INDIRECT REGIME

The indirect regime in the $n^+ - i - n^+$ structure of the GaAs/Al_xGa_{1-x}As double quantum wells occurs for finite values of V_g. The V_g dependence of the photoluminescence spectra and kinetics are shown in Figs. 1a and 1c. For $V_g \lesssim 0.3$ V the energy of the photoluminescence line is essentially independent of V_g, and the decay of this line is characterized by a short lifetime. Therefore the ground state of the system for $V_g \lesssim 0.3$ V is a direct exciton. For $V_g \gtrsim 0.4$ V, increasing V_g produces an approximately linear energy shift of the principal photoluminescence line and increases the decay time of the line. Therefore for $V_g \gtrsim 0.4$ V electrons and holes in the ground state of the system occupy different quantum wells; this corresponds to an indirect regime. The magnitude of the shift of the indirect photoluminescence line is determined by the electrostatic energy eFd, where e is the electron charge and F is the electric field in the z direction. A transition from the direct regime into the indirect regime occurs in a nonzero electric field F_{D-I}. This corresponds to excitonic recombination with direct and indirect exciton energies $\mathscr{E}_D = E_g - E_D$ and $\mathscr{E}_I = E_g - E_I - eFd$, where E_g is the energy gap, including the electron and hole quantization energy in the double quantum well, E_D and E_I are the binding energies of the direct and indirect excitons. For $F = F_{D-I}$ the energy difference between the single-particle direct and in-

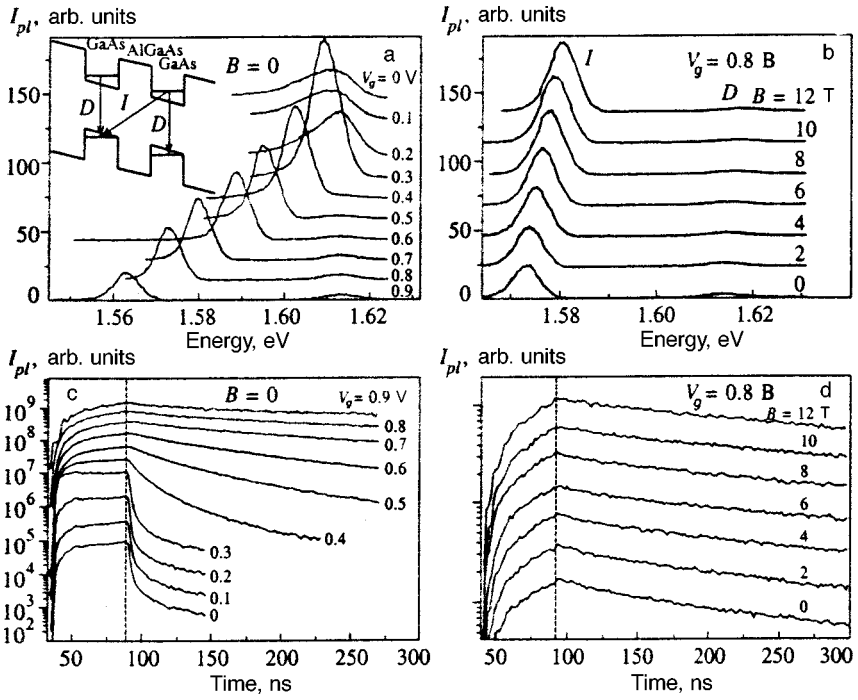

FIG. 1. Time-integrated photoluminescence spectrum (a, b) and photoluminescence kinetics, measured at the maximum of the principal line (c, d), as a function of the gate voltage (a, c) and magnetic field (b, d) at $T=1.3$ K and $W_{ex}=10$ W/cm^2. The dashed line corresponds to the trailing edge of the 50 ns laser excitation pulse. The spectra and kinetics are shifted along the ordinate for clarity. The direct and indirect photoluminescence lines are labeled by *D* and *I*, respectively. The band diagram of the GaAs/Al_xGa_{1-x}As double quantum well is shown in the inset.

direct pair states is equal to the difference between the binding energies of the direct and indirect excitons: $eF_{D-I}d = E_D - E_I$ (see Refs. 7 and 9 and the references cited there).

The width of the indirect-exciton line is determined by the random potential in the plane of the double quantum well. Several types of disorder, making the main contribution to the inhomogeneous broadening of the indirect photoluminescence line, can be distinguished: a) interfacial fluctuations; b) electric-field fluctuations in the *z* direction; c) charged impurities (other types of disorder, such as composition fluctuations, neutral impurities, and defects also exist, but their contribution to disorder is, as a rule, smaller). Fluctuations of the electric field in the *z* direction give rise to in-phase fluctuations of the potential for an electron and for a hole, so that they can be treated as fluctuations of the potential for the indirect exciton center of mass. Charged impurities give rise to antiphase fluctuations of the potential for an electron and for a hole. Strong fluctuations due to charged impurities can result in breakup of the exciton and independent localization of an electron and a hole in a local minimum of the random potential.[11] Interfacial fluctuations give rise to in-phase fluctuations of the potential for an electron and a hole in single quantum wells; for an indirect exciton (electron–hole pair) in GaAs/Al_xGa_{1-x}As wells interfacial fluctuations produce independent fluctuations of the potential for an electron and a hole. Fluctuations of the electric field in the *z* direction are determined primarily by fluctuations of the extent of the section where gate voltage drops. To reduce such fluctuations to a minimm the n^+ layers should possess good conductivity, and the *i* layer should be a good insulator. Then the region where gate voltage drops is clearly determined—it is the *i* layer.

Since fluctuations of the electric field in the *z* direction are specific to indirect excitons in double quantum wells, their relative contribution to the inhomogeneous broadening of the indirect photoluminescence line can be estimated by comparing the widths of the direct and indirect photoluminescence lines. In the GaAs/Al_xGa_{1-x}As double quantum well studied the indirect photoluminescence line (6.5 meV for $V_g=0.8$ V) is even narrower than that of the direct photoluminescence line (14.7 meV in the direct regime with $V_g=0$ and 9.4 meV in the indirect regime with $V_g=0.8$ V), indicating that fluctuations of the electric field make a negligibly small contribution to the broadening of the line. The quantum wells in the experimental structure are narrow: 5 nm corresponds to 18 monolayers. In narrow wells interface fluctuations make the main contribution to inhomogeneous broadening of the line. Thus, in the experimental double quantum well a fluctuation of the well width by one monolayer with an infinite terrace produces a change $\delta_m \sim 5$ meV in the electron energy and ~ 2 meV in the hole energy. The finiteness of the terraces results in quantization of the electron and hole energies in the plane. This produces states in the entire energy interval from 0 to δ_m. The observed width of the indirect photoluminescence line corresponds to δ_m (Fig. 1a). This confirms that interfacial fluctuations make the dominant contribution to the broadening of the line. Therefore the large magnitude of the random potential in the experimental GaAs/Al_xGa_{1-x}As double quantum wells is due primarily to the small width of the quantum wells. The shape of the photoluminescence line reflects the energy distribution of excitons over local energy minima in the random potential. The direct-photoluminescence line is probably broadened in part because the widths of the two quantum wells are different and the direct photoluminescence line includes two spectrally unresolved lines from the two quantum wells.

In the experimental structure in the indirect regime both radiative and nonradiative recombination contribute to the recombination of indirect excitons. The observed decrease of the recombination rate with delay time (see, for example,

Fig. 1) is characteristic for both radiative and nonradiative recombination of indirect excitons. The radiative recombination rate of an exciton is proportional to the population of the optically active $2D$ excitonic states (with quasimomenta $k \leqslant k_0 = \mathscr{E}/\hbar c$, where c is the velocity of light in the medium) and it increases with the extent of the wave function of the exciton center of mass in the plane, called the coherent area of the exciton (the radiative recombination rate saturates when the coherent length of the reciprocal of the wavelength of the emitted light is reached).[22–26] As a result of the spread of the localization radius in a random potential, the radiative recombination time of excitons is nonuniform over the plane of the double quantum well. As a result, the radiative recombination rate decreases with increasing delay time, since it is excitons with a large localization radius that are the first to recombine in the photoluminescence decay process.

Moreover, as the delay time increases, electrons and holes independently localized in the local minima of the random potential and having, as a result of the spatial separation in the plane (in addition to separation in the z direction for indirect electron–hole pairs), a low radiative-recombination rate make an increasingly larger contribution to the intensity of photoluminescence. Since the independently localized electrons and holes with the smallest separation in the plane are the first to recombine in the process of photoluminescence decay, the radiative recombination rate of independently localized electrons and holes also decreases with increasing delay time.[11] In narrow double quantum wells, characterized by a low diffusion coefficient of indirect excitons, nonradiative recombination is determined by exciton transport toward nonradiative recombination centers.[12,27–29] The decrease of the nonradiative recombination rate with the delay time is due to the monotonic decrease of the exciton diffusion coefficient. As a result, more and more localized excitons, which have not had enough time to reach the centers of nonradiative recombination and to recombine, predominate in the spectrum.

The magnetic-field and temperature dependences of the indirect-photoluminescence spectra and kinetics are displayed in Figs. 1b, 1d and 2. The temperature dependence of the indirect-photoluminescence spectra and kinetics in strong magnetic fields, $B \leqslant 12$ T, is qualitatively the same as the analogous dependence for $B = 0$. The corresponding initial decay time of photoluminescence, τ, and the integrated intensity of indirect photoluminescence, I_{PL}, are shown in Fig. 3. The time τ increases with magnetic field (Figs. 1d and 3a) and decreases with increasing temperature (Fig. 2b, 3b, and 3c).

The observed photoluminescence kinetics in GaAs/Al_xGa_{1-x}As double quantum wells is qualitatively different from the kinetics in the B1-, B2-, and A1-class double quantum wells investigated: in contrast to GaAs/Al_xGa_{1-x}As double quantum wells with weak disorder (A1 class),[21] in the experimental GaAs/Al_xGa_{1-x}As double quantum wells a sharp increase in the intensity of photoluminescence after the laser excitation pulse ceases is not observed; in contrast to AlAs/GaAs double quantum wells (B1 class),[8,12] in the experimental GaAs/Al_xGa_{1-x}As wells a sharp decrease of τ in strong magnetic fields is not observed. The intensity of photoluminescence does not exhibit a sharp increase in the studied double quantum well with a large random potential after the laser excitation pulse ceases because the large potential smears the boundary between the optically active (with $k \leqslant k_0$) and passive ($k > k_0$) excitonic states[30], so that the possible sharp increase in the population of optically active states does not result in a higher photoluminescence intensity. The absence of a sharp decrease of τ in strong magnetic fields, which is observed in AlAs/GaAs double quantum wells (B1 class) and indicates the appearance of superfluidity of excitons,[12] indicates the absence of collective effects in the studied double quantum wells with a large random potential and a low binding energy of the indirect exciton (A2 class), as expected (see Sec. 1).

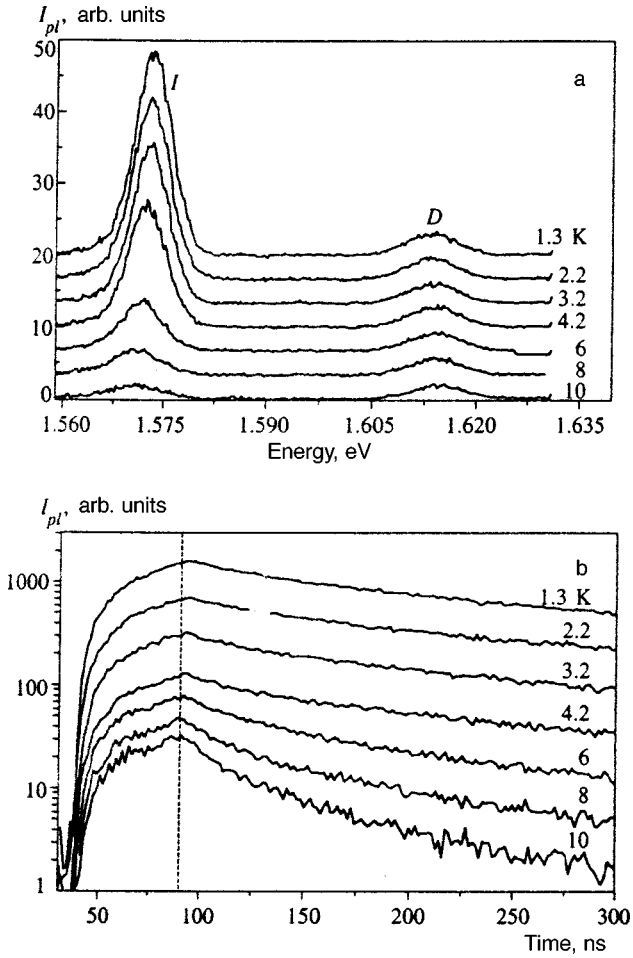

FIG. 2. Time-integrated photoluminescence spectrum (a) and indirect photoluminescence kinetics, measured at the maximum of the indirect line (b), as a function of temperature with $V_g = 0.8$ V, $W_{ex} = 10$ W/cm^2, and $B = 0$. The spectra and kinetics are shifted along the ordinate for clarity. The dashed line corresponds to the trailing edge of the 50 ns laser excitation pulse. The direct and indirect photoluminescence lines are labeled by D and I, respectively.

The observed monotonic increase of the lifetime with the magnetic field is characteristic for radiative and nonradiative single-exciton recombination in a random potential. The change in the radiative lifetime of excitons with increasing magnetic field is determined by the ratio of the increase in the oscillator strength of the exciton as a result of a decrease in the exciton radius[31] and the decrease in the oscillator

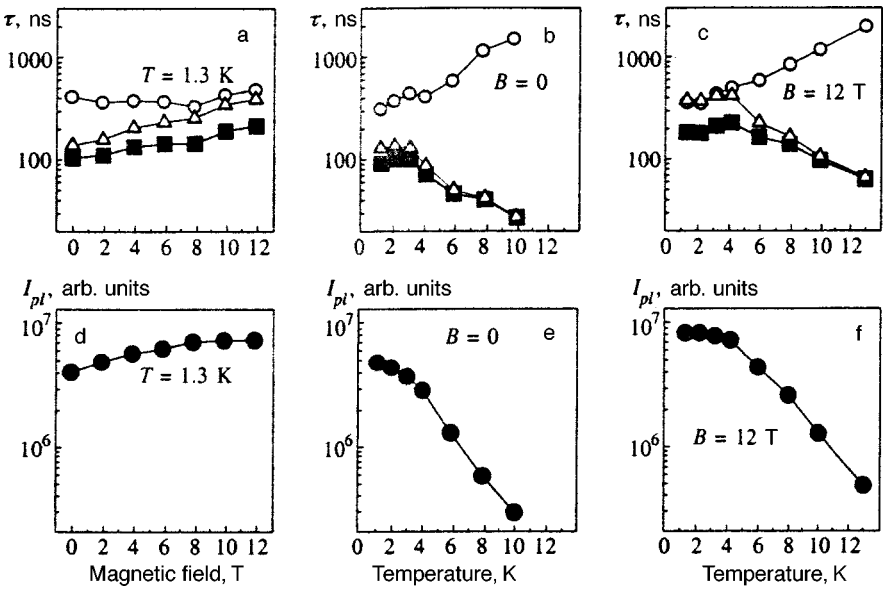

FIG. 3. Measured indirect-recombination times τ (■) and the integrated intensity of indirect photoluminescence together with the deduced radiative and nonradiative indirect recombination times τ_r (○) and τ_{nr} (△) versus the magnetic field with $V_g = 0.8$ V and $W_{ex} = 10$ W/cm^2.

strength of an exciton as a result of a decrease in the localization radius of the exciton[22–26] in a magnetic field as a result of an increase in the mass of the magnetoexciton.[31,32] The increase in the nonradiative lifetime of an exciton (determined by transport toward nonradiative recombination centers) with the magnetic field is due to a decrease in the diffusion coefficient caused by an increase in the magnetoexciton mass.[33] A decrease of the diffusion coefficient with increasing magnetic field has also been observed for indirect excitons in AlAs/GaAs double quantum wells in weak magnetic fields[12] and for direct excitons in single GaAs/Al_xGa_{1-x}As double quantum wells.[34] We note that a monotonic increase of the radiative and nonradiative lifetimes with increasing magnetic field is also characteristic for independently localized electrons and holes and is due to a decrease of the carrier localization radius.

The method described in Ref. 12 was used to distinguish the radiative and nonradiative lifetimes. The radiative lifetime τ_r can be directly extracted from the measured total lifetime τ and the time-integrated photoluminescence intensity I_{PL}. For single-exponential decay of photoluminescence $\tau_r = (G/I_{PL})\tau$, where G is the generation rate of electron–hole pairs in double quantum wells (weak nonexponentiality introduces negligible quantitative corrections[12]). The quantity G is unknown; to estimate it the quantum yield with parameters corresponding to maximum I_{PL} was taken to be 1. Then $G = I_{\max}$ and $\tau_r = (I_{\max}/I_{PL})\tau$, where $I_{\max}$ is the maximum integrated photoluminescence intensity observed in the experimental double quantum well with $V_g = 0.3$ V (Fig. 1a). The formula $\tau^{-1} = \tau_r^{-1} + \tau_{nr}^{-1}$ was used to find the nonradiative lifetime τ_{nr} using the measured value of τ and the calculated value of τ_r. The values of τ_r and τ_{nr} found in this manner are shown in Fig. 3.

Note that the parameter dependence of τ_r found by the method indicated above is correct if G does not depend on the given parameter. This condition was satisfied for the AlAs/GaAs double quantum wells studied in Ref. 12. However, it is not satisfied in the present investigations of GaAs/Al_xGa_{1-x}As double quantum wells. The photon energy for photoexcitation is only 100 meV lower than the energy of the $Al_{0.35}Ga_{0.65}As$ barrier, so that as a result of the Franz–Keldysh effect and the tails of the density of states in $Al_xGa_{1-x}As$, the absorption in the barrier layers was substantial. The collection of carriers generated in the barrier layers to double quantum wells made an appreciable contribution to G, comparable to the generation in GaAs layers. Thus the increase in the total photoluminescence intensity with a small applied gate voltage (Fig. 1) is due to an increase in G as a result of the drift of carriers generated in the barrier layers (this effect is absent for photoexcitation with photon energy much less than the gap width in the barrier layers). The drift of carriers generated in the barrier layers in double quantum wells can depend on the magnetic field and temperature. For this reason G is not independent of the parameters, and the method described above for finding the dependence of τ_r on the magnetic field and temperature from the measured values of τ and I_{PL} is not entirely correct for the present experiment. Moreover, an error in estimating the quantum yield for parameters corresponding to maximum I_{PL} will enter in the absolute value of τ_r and τ_{nr} as well as in the magnetic-field and temperature dependences of τ_{nr}. Nonetheless the method employed makes it possible to follow the qualitative variations of τ_r and τ_{nr} as a function of magnetic field and temperature.

It is evident in Fig. 3 that τ_r is virtually independent of the magnetic field and τ_{nr} increases monotonically with the field. This corresponds to the single-exciton behavior described above. As temperature increases, τ_r increases and τ_{nr} decreases (Fig. 3; the opposing behavior of τ_r and τ_{nr} could cause a weak nonmonotonicity of the temperature variations of τ, and the possible increase of τ at low temperatures falls within the experimental error, Figs. 2 and 3). The decrease in τ_{nr} with increasing temperature is characteristic for both single-exciton recombination and recombination of independently localized electrons and holes and is due to the increase in the diffusion of excitons (electrons and holes) toward nonradiative recombination centers as a result of their thermal activation from local minima of the random poten-

tial. The increase in τ_r with temperature is characteristic for excitonic recombination because the population of optically active excitonic states decreases (for a Boltzmann distribution of excitons the fraction of optically active excitons with $k \leqslant k_0$ is $1-\exp(-E_0/kT)$, where $E_0=\hbar^2k_0^2/2m\sim 1$ K; for $T \gg E_0$ the Boltzmann distribution leads to a linear increase of the radiative lifetime of excitons with increasing temperature).[22–26] However, the observed increase of τ_r with temperature contradicts the model of recombination of independently localized electrons and holes, on the basis of which an increase in temperature should enhance the overlapping of the electron and hole wave functions due to their delocalization. Hence it follows that the random potential in the double quantum well under study is not strong enough to break up excitons, and the fraction of independently localized electrons and holes is low. Note that the magnetic-field and temperature dependences of τ_r and τ_{nr} in the studied $GaAs/Al_xGa_{1-x}As$ double quantum well correspond qualitatively to the analogous dependences for AlAs/GaAs wells (B1 class) in weak magnetic fields.[12] In both cases these dependences correspond to single-exciton recombination.

The photoluminescence spectra and kinetics of indirect excitons are virtually independent of the exciton density, fixed by the laser excitation density W_{ex} (Fig. 4). For $W_{ex}=10$ W/cm^2 the density of indirect excitons with lifetime $\tau=100$ ns is estimated to be several times 10^{10} cm^{-2}. As W_{ex} increases, a shift of the indirect exciton line in the direction of high energies is observed. This behavior corresponds to the theoretically predicted increase of the energy of indirect excitons with increasing density[18,20] and is explained by the repulsive dipole–dipole interaction between indirect excitons for low exciton densities and by the energy shift for high electron–hole densities due to the electric field between the separated electron and hole layers. Moreover, since the degeneracy of the zero-dimensional excitonic state in a local minimum of the random potential is finite (neglecting the exciton-exciton interaction, the degeneracy is $\sim S/a_B^2$, where S is the area of the local minimum and a_B is the Bohr radius of an indirect exciton), an increase of the exciton density results in an increase of the average exciton energy. This effect should also contribute to the observed increase in the indirect-exciton energy with density.

A small decrease of the recombination time of indirect excitons is observed with increasing W_{ex} (Fig. 4b). This behavior is characteristic for single-exciton recombination in a random potential: the exciton localization radius increases with exciton density (since at low density excitons are strongly localized in deep local minima of the potential). This decreases both the radiative-recombination time of excitations[22–26] and the nonradiative-recombination time of excitons due to transport toward nonradiative-recombination centers.

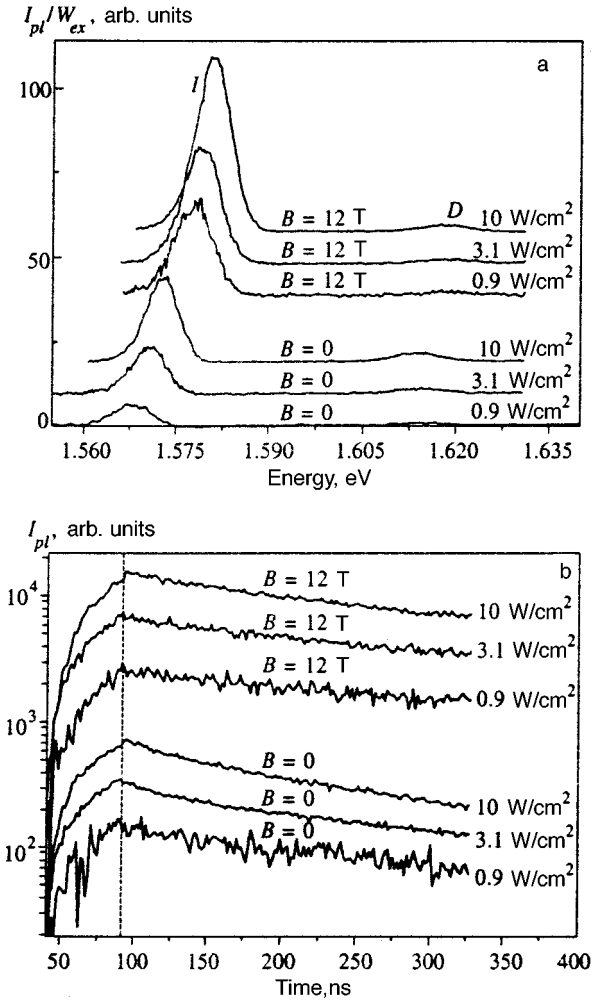

FIG. 4. Time-integrated photoluminescence spectrum normalized to the laser excitation density (a) and indirect-photoluminescence kinetics measured at the maximum of the indirect line (b) as a function of the laser excitation density with $V_g=0.8$ V, $T=1.3$ K, and $B=0$ and 12 T. The spectra and kinetics are shifted along the ordinate for clarity. The dashed line corresponds to the trailing edge of the 50 ns laser excitation pulse. The direct and indirect photoluminescence lines are labeled by *D* and *I*, respectively.

4. EVOLUTION OF PHOTOLUMINESCENCE SPECTRA IN DIRECT AND INDIRECT REGIMES

In this section we examine the temporal revolution of the photoluminescence spectra. This is the evolution of the energy distribution of excitons (electron–hole pairs) with weight proportional to the radiative-recombination probability. The dependence of the photoluminescence spectra on the time delay is shown in Fig. 5a and 5b, respectively, for the direct and indirect regimes. After the laser excitation pulse ends, both the direct-photoluminescence line in the direct regime and the indirect-photoluminescence line in the indirect regime shift monotonically in the direction of low energies as the time delay increases (Fig. 5). Similar behavior is also observed in strong magnetic fields. The shift of the photoluminescence line with increasing time delay is typical both for excitonic recombination and for recombination of independently localized electron–hole pairs as well as for radiative and nonradiative recombinations. In the first place, as the delay increases, the average energy of photoexcited excitons (electron–hole pairs) in a random potential decreases as a result of the energy relaxation of the carriers: excitons (electrons and holes) migrate in the plane of the double quantum well in search for lower-energy local minima of the potential with the emission of acoustic

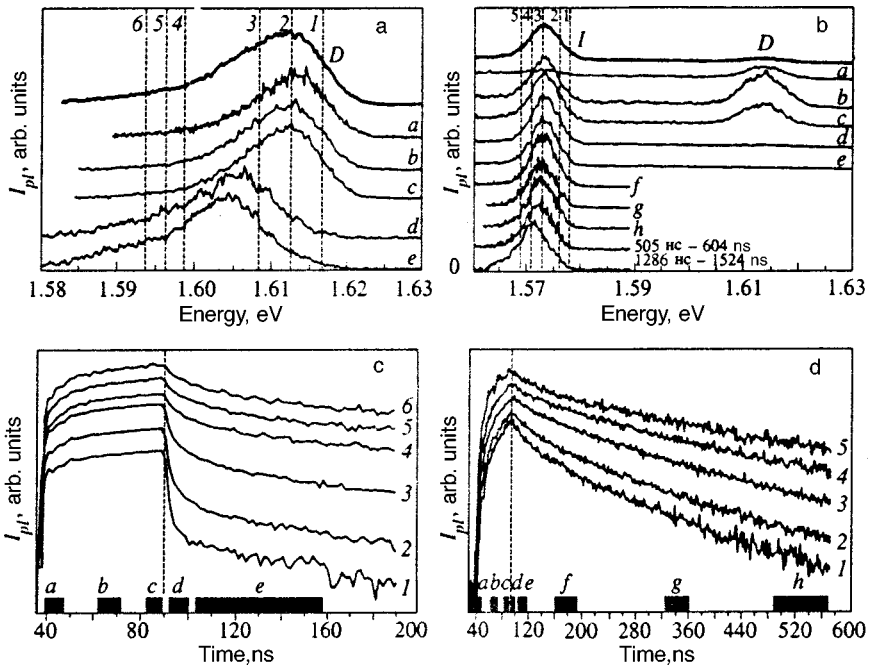

FIG. 5. Direct-photoluminescence spectrum in the direct regime with $V_g=0$ (a) and the indirect-photoluminescence spectrum in the indirect regime with $V_g=0.8$ V (b) versus the delay time, $T=1.3$ K, $B=0$, and $W_{ex}=10$ W/cm^2. The spectra were measured in the time intervals shown in panels c and d. The time-integrated spectra are shown in the top. All spectra are scaled to roughly the same intensity. The curves of photoluminescence kinetics versus energy are shown for the same parameters (c, d); the energies of the detected signal are shown in panels a and b by dashed lines. The spectra and kinetics are shifted along the ordinate for clarity. The dashed line in panels c and d corresponds to the trailing edge of the 50 ns laser excitation pulse. The direct and indirect photoluminescence lines are labeled by *D* and *I*, respectively.

phonons. This mechanism of energy relaxation of excitons in a random potential was theoretically examined in Ref. 35 and has been observed for indirect excitons in a double quantum well.[10,36] In the second place, as the delay time increases, the average energy of the excitons (electron–hole pairs) in a random potential decreases because the higher-energy excitons (independently localized electrons and holes) have higher radiative and nonradiative recombination rates. As the energy of the excitons in a random potential increases, their localization radius increases,[37] which results in a higher radiative recombination rate[22–26] and higher nonradiative recombination rate, due to transport of excitons toward nonradiative-recombination centers. For independently localized electron–hole pairs, the higher-energy pairs also have a higher recombination rate because of their larger localization radius and the corresponding larger overlap between the electron and hole in the plane.[11]

Another aspect of the dependence of the photoluminescence spectra on the time delay is the energy dependence of the photoluminescence kinetics, shown in Figs. 5c and 5d. As energy decreases, the photoluminescence decay becomes slower and slower, which corresponds to the time dependence of the spectra (Fig. 5a and 5b) and was discussed above.

As a result of the inhomogeneous broadening of the direct and indirect photoluminescence lines, for the appropriate gate voltages a mixed regime in which the direct and indirect photoluminescence energies overlap can be obtained. Such a regime of energy resonance between direct and indirect excitons was considered in Ref. 38 in a study of the photoluminescence of zero-dimensional excitonic states in the local minima of the random potential (natural quantum dots). In our double quantum well the mixed regime appears for $V_g \sim 0.2-0.5$ V; this is evident from the V_g dependence of the photoluminescence spectra (Fig. 1). In the studied double quantum well, characterized by a larger width of the direct-photoluminescence line than the width of the indirect line, a specific energy dependence of the photoluminescence kinetics is observed in the mixed regime (Fig. 6). This dependence is different from the monotonic decrease of the recombination rate with decreasing energy, as is observed in the direct and indirect regimes. Specifically, in the mixed regime the recombination rate on the initial times of photoluminescence decay depends nonmonotonically on the energy, reaching a minimum at energies corresponding to indirect photoluminescence (kinetics *4–6* in Fig. 6b). The high recombination rate of direct photoluminescence at energies below the indirect-photoluminescence energy (*8*, *9* in Fig. 6b) indicates that the electron–hole distance in the plane is less than the distance in the *z* direction. Since the latter distance is ~ 11 nm, the electron–hole distance in the plane is less than the radius of a direct exciton (≈ 10 nm) and especially an indirect exciton, which has a larger radius as a result of the lower binding energy (≈ 20 nm).[7] A small electron–hole distance in the plane indicates that excitonic recombination predominates over recombination of independently localized electrons and holes.

5. CONCLUSIONS

We have investigated the kinetics of indirect photoluminescence in GaAs/Al_xGa_{1-x}As double quantum wells characterized by a random potential with a large amplitude (the width of the photoluminescence line is comparable to the binding energy of an indirect exciton) and a large distance between the electron and hole layers ($d \sim 11$ nm, which is greater than the magnetic length for $B \gtrsim 5.5$ T) in magnetic fields $B \leqslant 12$ T at low temperatures $T \geqslant 1.3$ K. It was found that the indirect recombination time τ increases with the magnetic field and decreases with temperature. Analysis of the variation of τ and the total intensity of photoluminescence gave the radiative and nonradiative indirect recombi-

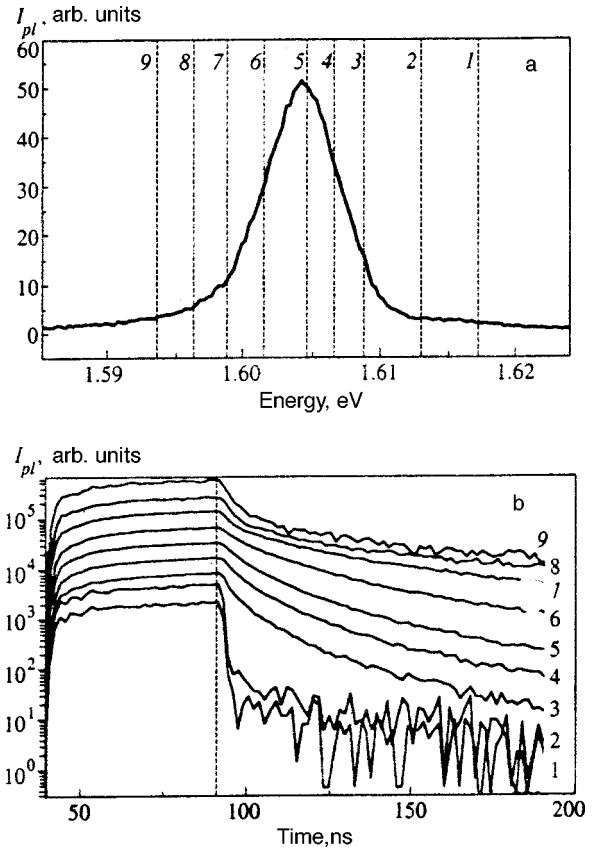

FIG. 6. Energy dependence of the photoluminescence kinetics in the mixed regime corresponding to overlapping of the direct and indirect photoluminescence lines ($V_g=0.4$ V) at $T=1.3$ K, $B=0$, and $W_{ex}=10$ W/cm^2 (b). The energies of the detected signal are shown in panel a, which shows the time-integrated spectrum. The spectra are shifted along the ordinate for clarity. The dashed line in panel b corresponds to the trailing edge of the 50 ns laser excitation pulse.

nation times, τ_r and τ_{nr}. It was found that τ_{nr} increases with the magnetic field and decreases with increasing temperature, while τ_r is essentially independent of the magnetic field and increases with temperature. It was shown that the kinetics of indirect photoluminescence corresponds to the single-exciton recombination in the presence of a random potential in the plane of the double quantum well. The variation of the nonradiative recombination time was discussed in terms of the variation of the transport of indirect excitons toward nonradiative recombination centers, and the variation of the radiative recombination times was discussed in terms of the variation of the population of optically active excitonic states and the localization radius of indirect excitons. The photoluminescence kinetics of indirect excitons, which was observed in the studied GaAs/Al_xGa_{1-x}As double quantum wells with a random potential having a large amplitude, is qualitatively different from the photoluminescence kinetics of indirect excitons in AlAs/GaAs and GaAs/Al_xGa_{1-x}As double quantum wells when the amplitude of the random potential is small.

The temporal evolution of the photoluminescence spectra in the direct and indirect regimes was examined. It was found that after the laser excitation pulse ceases both the direct photoluminescence line in the direct regime and the indirect photoluminescence line in the indirect regime shift monotonically in the direct of lower energies as the delay time increases. It was shown that the evolution of the photoluminescence spectra corresponds to excitonic recombination in a random potential and is determined by the energy relaxation of excitons and by the energy dependence of the recombination rate.

We thank V. D. Kulakovskiĭ and S. G. Tikhodeev for a discussion of the results obtained in this work as well as the Russian Fund for Fundamental Research (Project 98-02-18371) and the program ''Physics of Solid-State Nanostructures'' (Project 97-1067) for financial support.

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

[1] Y. J. Chen, Emil. S. Koteles, B. S. Elman, and C. A. Armiento, Phys. Rev. B **36**, 4562 (1987).
[2] M. N. Islam, R. L. Hillman, D. A. B. Miller, D. S. Chemla, A. C. Gossard, and J. H. English, Appl. Phys. Lett. **50**, 1098 (1987).
[3] S. R. Andrews, C. M. Murray, R. A. Davies, and T. M. Kerr, Phys. Rev. B **37**, 8198 (1988).
[4] S. Charbonneau, M. L. W. Thewalt, E. S. Koteles, and B. Elman, Phys. Rev. B **38**, 6287 (1988).
[5] C. C. Phillips, R. Eccleston, and S. R. Andrews, Phys. Rev. B **40**, 9760 (1989).
[6] T. Kukuzawa, E. E. Mendez, and J. M. Hong, Phys. Rev. Lett. **64**, 3066 (1990); A. Alexandrou, J. A. Kash, E. E. Mendez, M. Zachau, J. N. Hong, T. Kukuzawa, and Y. Hase, Phys. Rev. B **42**, 9225 (1990); J. A. Kash, M. Zachau, E. E. Mendez, J. M. Hong, and T. Fukuzawa, Phys. Rev. Lett. **66**, 2247 (1991).
[7] M. M. Dignamn and J. E. Sipe, Phys. Rev. B **43**, 4084 (1991).
[8] L. V. Butov, A. Zrenner, G. Abstreiter, G. Böhm, and G. Weimann, Phys. Rev. Lett. **73**, 304 (1994).
[9] L. V. Butov, A. Zrenner, G. Abstreiter, A. V. Petinova, and K. Eberl, Phys. Rev. **52**, 12153 (1995); A. B. Dzyubenko and A. L. Yablocnskii, Phys. Rev. B **53**, 16355 (1996).
[10] J. E. Golub, S. D. Baranovskii, and P. Thomas, Phys. Rev. Lett. **78**, 4261 (1997).
[11] V. B. Timofeev, A. I. Filin, A. V. Larionov, J. Zeman, G. Martinez, J. M. Hvam, D. Birkedal, and C. B. Sorensen, Europhys. Lett. **41**, 535 (1998); V. B. Timofeev, A. V. Larionov, A. S. Iozelevich, J. Zeman, G. Martinex, J. Hvam, and K. Sorensen, JETP Lett. **67**, 613 (1998).
[12] L. V. Butov and A. I. Filin, Phys. Rev. B **58**, 1980 (1998).
[13] Yu. E. Lozovik and V. I. Yudson, Zh. Éksp. Teor. Fiz. **71**, 738 (1976) [Sov. Phys. JETP **44**, 389 (1976)].
[14] S. I. Sevchenko, Fiz. Nizk. Temp. **2**, 505 (1976) [Sov. J. Low Temp. Phys. **2**, 251 (1976)].
[15] I. V. Lerner and Yu. E. Lozovik, JETP Lett. **27**, 467 (1978); I. V. Lerner and Yu. E. Lozovik, J. Low Temp. Phys. **38**, 333 (1980); I. V. Lerner and Yu. E. Lozovik, Zh. Éksp. Teor. Fiz. **80**, 1488 (1981) [Sov. Phys. JETP **53**, 763 (1981)].
[16] Y. Kuramoto and C. Horie, Solid State Commun. **25**, 713 (1978).
[17] T. Fukuzawa, S. S. Kano, T. K. Gustafson, and T. Ogawa, Surf. Sci. **228**, 482 (1990).
[18] D. Yoshioka and A. H. MacDonald, J. Phys. Soc. Jpn. **59**, 4211 (1990).
[19] X. M. Chen and J. J. Quinn, Phys. Rev. Lett. **67**, 8895 (1991).
[20] X. Zhu, P. B. Littlewood, M. S. Hybersten, and T. M. Rice, Phys. Rev. Lett. **74**, 1633 (1995).
[21] L. V. Butov, A. Imamoglu, A. V. Minstev, K. L. Campman, and S. C. Gossard, Phys. Rev. B **59**, 1625 (1999).
[22] J. Feldman, G. Peter, E. O. Göbel, P. Dawson, K. Moore, C. Foxon, and R. J. Elliot, Phys. Rev. Lett. **59**, 2337 (1987).
[23] E. Hanamura, Phys. Rev. B **38**, 1228 (1988).
[24] L. C. Andreani, F. Tassone, and F. Bassani, Solid State Commun. **77**, 641 (1991).
[25] B. Deveaud, F. Clerot, N. Roy, K. Satzke, B. Sermage, and D. S. Katzer, Phys. Rev. Lett. **67**, 2355 (1991).
[26] D. S. Citrin, Phys. Rev. B **47**, 3832 (1993).

[27] F. Minami, K. Hirata, K. Era, T. Yao, and Y. Matsumoto, Phys. Rev. B **36**, 2875 (1987).
[28] M. Maaref, F. F. Charfi, D. Scalbert, C. Benoir a la Guillaume, and R. Planet, Phys. Status Solidi B **170**, 637 (1992).
[29] G. D. Gilliland, A. Antonelli, D. J. Wolford, K. K. Bajaj, J. Klem, and J. A. Bradley, Phys. Rev. Lett. **71**, 3717 (1993).
[30] W. Zhao, P. Stenius, and A. Imamoglu, Phys. Rev. B **56**, 5306 (1997).
[31] I. V. Lerner and Yu. E. Lozovik, Zh. Éksp. Teor. Fiz. **78**, 1167 (1980) [Sov. Phys. JETP **51**, 588 (1980)].
[32] Yu. E. Lozovik and A. M. Ruvinskiĭ, Zh. Éksp. Teor. Fiz. **112**, 1791 (1997) [JETP **85**, 979 (1997)].
[33] A. V. Dzyubenko and G. E. W. Bauer, Phys. Rev. B **51**, 14524 (1995).
[34] M. Jiang, H. Wang, R. Merlin, D. G. Steel and M. Cardona, Phys. Rev. B **48**, 15476 (1993).
[35] T. Takagahara, Phys. Rev. B **31**, 6552 (1985).
[36] L. V. Butov and A. I. Filin, Zh. Éksp. Teor. Fiz. **114**, 1115 (1998) [JETP 87, 608 (1998)].
[37] Zh. S. Gevorkyan and Yu. E. Lozovik, Fiz. Tverd. Tela (Leningrad) **27**, 1800 (1985) [Sov. Phys. Solid State **27**, 1079 (1985)].
[38] A. Zrenner, L. V. Butov, M. Hagn, G. Abstreiter, G. Böhm, and G. Weimann, Phys. Rev. Lett. **72**, 3382 (1994).

Translated by M. E. Alferieff

Inhomogeneous strains in semiconducting nanostructures

N. A. Gippius*) and S. G. Tikhodeev†)

Institute of General Physics, Russian Academy of Sciences, 117942 Moscow, Russia
(Submitted 29 December 1998)
Zh. Éksp. Teor. Fiz. **115**, 1906–1914 (May 1999)

We have developed a numerical technique for calculating inhomogeneous strains in stressed semiconducting nanostructures, such as quantum wires and dots manufactured by nanolithography from stressed InGaAs/GaAs quantum wells. The technique is based on solving a linear problem of elasticity theory by the Green's function method and presumes a lack of defects and dislocations in nanostructure heterojunctions. Spatial distributions of strain tensor components and shifts of electron and hole potentials in a nanostructure due to the strain have been calculated. © *1999 American Institute of Physics.* [S1063-7761(99)02705-5]

1. INTRODUCTION

In open nanostructures, such as quantum wires and dots which are manufactured using nanolithography on the surfaces of samples containing InGaAs/GaAs quantum wells at small depths, stressed InGaAs layers give rise to large inhomogeneous strains determined by shapes of lithographically defined nanostructures. Partial elastic relaxation of stressed layers has considerable effect on spectra and wave functions of electrons and holes, and this offers fresh opportunities for band-gap engineering and control over wave functions of carriers.[1–3] The present paper reports a technique developed for calculating inhomogeneous strain fields in stressed nanostructures, and calculations of strain distributions and resulting shifts of electron and hole levels in InGaAs/GaAs quantum wires and cylindrical quantum dots.

2. THEORETICAL MODEL

In what follows, we will assume that the system under consideration is dislocation-free, and the equations of elasticity theory apply at nanometer scales. We suggest an extremely efficient approach to strain calculations based on Green's functions of elasticity theory.[4] This method explicitly takes into account the piecewise-constant form of elastic parameters of semiconducting nanostructures, which usually consist of homogeneous, epitaxially connected pieces of a variety of semiconductors. We will prove that the strain tensor can be expressed in regions of constant elastic parameters as a functional of the forces at the boundaries of these regions, which allows one to reduce the problem of elasticity theory to an integral equation at the interfaces between the various fragments of a nanostructure.

We model elastic properties of fragments in the approximation of an isotropic elastic medium, whose Green's functions can be expressed in explicit form.[4] Luckily, the anisotropy of semiconductors in III-V nanostructures, for example, InGaAs/GaAs, is fairly low. In particular, the difference between elastic moduli C_{12} and C_{44} in both GaAs and InGaAs is small:

$$\frac{C_{12}-C_{44}}{C_{12}}\approx 0.057 \ll 1.$$

In the approximation of an isotropic medium, the elasticity problem is fully characterized by two elastic constants, namely, Young's modulus E and the Poisson coefficient σ. In fact, we only require ratios of Young's moduli.

The source of stress in a nanostructure is the difference between the lattice constants of GaAs and InGaAs. In our calculations we use the following low-temperature formulas for the lattice constants (measured in nanometers):[5]

$$a_{\mathrm{GaAs}}(T(\mathrm{K}))=0.565325+6.86\times 10^{-6}\ \ T,$$

$$a_{\mathrm{InAs}}(T(\mathrm{K}))=0.60583+4.52\times 10^{-6}\ \ T,$$

and linear interpolation for InGaAs.

For example, in an unbounded two-dimensional $In_{0.1}Ga_{0.9}As$ quantum well on a GaAs substrate, the difference between the lattice constants leads to the following homogeneous uniaxial strain in the InGaAs well layer, provided that the epitaxial interfaces are matched at the atomic level:

$$u_{xx}^{(0)}=u_{yy}^{(0)}=\frac{a_{\mathrm{GaAs}}-a_{\mathrm{InGaAs}}}{a_{\mathrm{InGaAs}}}\equiv -\delta,$$

$$u_{zz}^{(0)}=-\frac{2\sigma}{1-\sigma}u_{xx}^{(0)}=\frac{2\sigma}{1-\sigma}\delta,$$

$$u_{xy}^{(0)}=u_{yz}^{(0)}=u_{zx}^{(0)}=0. \qquad (1)$$

The GaAs barrier layers are unstrained. For finite fragment sizes, compression is partially relaxed in InGaAs, and the strain becomes inhomogeneous and encompasses GaAs regions.

Consider the simplest case of dislocation-free matching between two fragments of a semiconducting nanostructure V_1 and V_2 along some surface S_{12}. If the equilibrium lattice constants a_1 and a_2 of the two fragments are different, both are strained owing to their mutual interaction on the inter-

face. We denote the surface force on the first fragment due to the second by $\mathbf{g}_{12}\equiv\mathbf{g}$. By Newton's third law, $\mathbf{g}_{12}=-\mathbf{g}_{21}\equiv-\mathbf{g}$.

In the general case, strains also lead to displacements and rotations of fragments as a whole. A fragment's arbitrary displacement (due to strains) much smaller than its size can be expressed as

$$\mathbf{u}(P)=\mathbf{u}_0(P)+\mathbf{t}_{\rm cm}+\mathbf{\Omega}\times(\mathbf{r}_P-\mathbf{r}_{\rm cm})+\mathbf{u}^d(P), \tag{2}$$

where $\mathbf{u}_0(P)$ is the initial equilibrium displacement of cells in the unstrained fragment, $\mathbf{t}_{\rm cm}$ and $\mathbf{\Omega}$ are the shift and rotation angle of the fragment as a whole, $\mathbf{r}_{\rm cm}$ is the center-of-mass coordinate of the fragment, $\mathbf{u}^d(P)$ is the purely deformation displacement, which is a functional of the surface forces $\mathbf{g}$:

$$u_i^d(P)=\int_S U_{ij}^g(P,P')g_j(P')dS_{P'}. \tag{3}$$

In this paper, we show how to calculate $U_{ij}^g(P,P')$, the kernel of the functional that relates strain displacements in a homogeneous fragment to the forces applied to its surface S.

The condition of the defect-free matching between the fragments at each point of surface S_{12} can be written

$$\mathbf{u}^{(1)}(P)=\mathbf{u}^{(2)}(P), \tag{4}$$

where

$$\mathbf{u}^{(1)}(P)\equiv\mathbf{u}_0^{(1)}(P)+\mathbf{t}_{\rm cm}^{(1)}+\mathbf{\Omega}^{(1)}\times(\mathbf{r}_P-\mathbf{r}_{\rm cm}^{(1)})+\int_S U^{g(1)}(P,P')\mathbf{g}(P')dS_{P'},$$

$$\mathbf{u}^{(2)}(P)\equiv\mathbf{u}_0^{(2)}(P)+\mathbf{t}_{\rm cm}^{(2)}+\mathbf{\Omega}^{(2)}\times(\mathbf{r}_P-\mathbf{r}_{\rm cm}^{(2)})-\int_S U^{g(2)}(P,P')\mathbf{g}(P')dS_{P'}.$$

Note that the different signs in front of the integrals are due to the definition $\mathbf{g}_{12}\equiv\mathbf{g}$ and $\mathbf{g}_{21}=-\mathbf{g}_{12}\equiv-\mathbf{g}$.

For a time-independent stress constant, the total surface force and torque on each fragment should be zero:

$$\int_S g_i(P')dS_{P'}=0,\quad \int_S \varepsilon_{ij\gamma}r_j(P')g_\gamma(P')dS_{P'}=0. \tag{5}$$

In numerical calculations, it is convenient to replace integrals with sums over small surface elements.

Without a loss of generality, one can single out one fragment, for example the first, and set $\mathbf{t}_{\rm cm}^{(0)}=0$ and $\mathbf{\Omega}^{(1)}=0$. Suppose that the surface S_{12} is broken up into N_S elements; then Eqs. (4) and (5) yield $3N_S+6$ equations for the forces $\mathbf{g}(P')$ ($3N_S$ unknown quantities), $\mathbf{t}_{\rm cm}^{(2)}$, and $\mathbf{\Omega}^{(2)}$ (another six unknowns). Solving these equations, one can determine the surface forces $\mathbf{g}(P')$, and given these quantities and $U_{ij}^{g(n)}(P,P')$, $n=1,2$, for arbitrary fragment points P and P' on the surface S_{12}, one can calculate the displacement at an arbitrary point inside the fragment.

We now describe the technique for constructing the kernel $U_{ij}^g(P,P')$. We fill the space outside the fragment under consideration with the same material and apply the forces $\mathbf{f}(P')$ to points P' on the surface S in this homogeneous elastic medium. The displacements caused by these forces can be calculated using the known expressions for the Green's functions for an unbounded, homogeneous elastic medium:[4]

$$u_i(P)=\int_S A_{ij}(P,P')f_j(P')dS_{P'},$$

$$A_{ij}(P,P')\equiv A_{ij}(\mathbf{R}=\mathbf{r}_P-\mathbf{r}_{P'})=\frac{1}{8\pi E}\frac{1+\sigma}{1-\sigma}\frac{(3-4\sigma)\delta_{ij}R^2+R_iR_j}{R^3}, \tag{6}$$

where E is Young's modulus and σ is the Poisson coefficient. The strain tensor

$$u_{ij}(\mathbf{r})=\frac{1}{2}\left(\frac{\partial u_i}{\partial r_j}+\frac{\partial u_j}{\partial r_i}\right)$$

and the stress tensor

$$\sigma_{ij}=\frac{E}{1+\sigma}\left(u_{ij}+\frac{\sigma}{1-2\sigma}u_{ll}\delta_{ij}\right),$$

are obviously also known functionals of $\mathbf{f}(P')$.

The forces acting at points P' on a fragment (defined by surface S) of an unbounded elastic medium equal, by definition,

$$g_i^{(in)}(P')=\sigma_{ij}(P)n_j(P'),\quad P\to P', \tag{7}$$

where $n_j(P')$ are the components of the unit normal to the surface S at point P', and P is an interior point of the fragment. Just like u_{ij} and σ_{ij}, the distribution of forces $\mathbf{g}^{(\rm in)}$ also becomes a known functional of $\mathbf{f}(P')$:

$$g_i^{(in)}(P')=\int_S B_{ij}(P',P'')f_j(P'')dS_{P''}. \tag{8}$$

We emphasize that the forces $\mathbf{g}^{(in)}(P')$ are different from $\mathbf{f}(P')$, which can easily be verified by considering the special case in which the applied forces $\mathbf{f}$ are nonvanishing only at certain points on the surface S. Nonetheless, the strains generated by these forces, and hence $\mathbf{g}^{(in)}$, are generally nonvanishing over the entire surface.

Since the elastic interaction is short-range, the strain in the interior of an unbounded elastic medium due to forces $\mathbf{f}(P')$ is obviously the same as that in a free body of the same shape and elastic properties acted upon by forces $\mathbf{g}^{(in)}$.

Thus, the problem of the strain in a free body acted upon by surface forces $\mathbf{g}(P')$ can be solved in two stages. First, we calculate the distribution of forces $\mathbf{f}(P')$ that satisfies the equation

$$\int_S B_{ij}(P',P'')f_j(P'')dS_{P''}=g_i(P'), \tag{9}$$

then, given Eq. (6), we calculate the strain in the interior of the elastic medium, which is identical to the desired strain in the free body.

Since we are interested in time-independent strains, the distributions of forces $\mathbf{f}$ and $\mathbf{g}$ should yield vanishing total forces and torques. Therefore, in addition to Eq. (5), a similar condition for $\mathbf{f}$ should be satisfied:

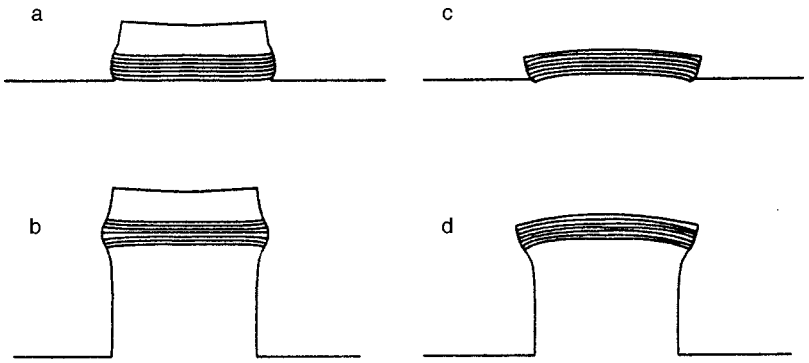

FIG. 1. Cross sections of InGaAs/GaAs wires of width $L_x=45$ nm with strain fields calculated for four different combinations of the thicknesses of the cap barrier layer, $L_{\rm cap}$, quantum well layer, L_z, and the etch depth in the substrate barrier layer, $L_{\rm b}$. The quantities $L_{\rm cap}$, L_z, and $L_{\rm b}$ equal, respectively, (a) 10, 5, and 0 nm; (b) 10, 5, and 35 nm; (c) 0, 5, and 0 nm; (d) 0, 5, and 35 nm. In order to visualize the calculated strain tensor components, the vertical displacements are multiplied by a factor of 200.

$$\int_S f_i(P')dS_{P'}=0, \quad \int_S \varepsilon_{ijk}(r_j(P')-r_j^{\rm cm})f_k(P')dS_{P'}=0. \tag{10}$$

Equations (5) and (10) can be treated as scalar products of vectors in the configuration space of surface force distributions:

$$(\mathbf{g}\cdot\mathbf{F}^{(i)})=0, \quad (\mathbf{g}\cdot\mathbf{M}^{(i)})=0, \quad i=1,\ 2,\ 3, \tag{11}$$

where the vectors $\mathbf{F}^{(i)}$ and $\mathbf{M}^{(i)}$ have the components $F_j^{(i)}(P')=\delta_{ij}$, $M_k^{(i)}(P')=\varepsilon_{ijk}(r_j(P')-r_j^{\rm cm})$, and the scalar product of vectors **a** and **b** is defined by

$$(\mathbf{a}\cdot\mathbf{b})=\int_S a_k(P')b_k(P')dS_{P'}. \tag{12}$$

These conditions indicate that out of the entire configuration space of surface force distributions, only the subspace orthogonal to the hyperplane *FM* defined by the vectors $\mathbf{F}^{(i)}$ and $\mathbf{M}^{(i)}$, $i=1,\ 2,\ 3$, is interesting from the standpoint of physics.

Note that for an arbitrary distribution of **f**, even if the total forces and torques are zero, the resulting distributions of **g** should be orthogonal to the hyperplane *FM*, since these are by definition the forces on an immobile fragment. This means that operator $B_{ij}(P',P'')$ in Eq. (8) transforms the entire configuration space to the orthogonal complement of the hyperplane *FM*; hence, it is degenerate. Therefore, the matrix $B_{ij}(P',P'')$ cannot be directly inverted.

This difficulty, however, can easily be eliminated if B in Eq. (8) is replaced with an operator equivalent over the subspace of vanishing total forces and torques:

$$B_0=\Pi_{FM}+(1-\Pi_{FM})B(1-\Pi_{FM}), \tag{13}$$

where Π_{FM} is the projection operator onto the hyperplane *FM*. The operator B_0 is nondegenerate. When integration over the surface is replaced with summation over N_S small areas, B_0 becomes a nondegenerate $N_S\times N_S$ matrix, which can be inverted using standard techniques. The matrix obtained by this method yields a solution of Eq. (8) in the form $\mathbf{f}=B_0^{-1}\mathbf{g}$. Substituting this expression for **f** into $\mathbf{u}=A\mathbf{f}$ in Eq. (6), we obtain the desired relationship between the strain field **u** in an elastic body and the forces **g** on its surface:

$$\mathbf{u}=A\mathbf{f}=AB_0^{-1}\mathbf{g}. \tag{14}$$

A comparison of this expression and Eq. (3) yields the final expression for $U_{ij}^g(P,P')$:

$$U_{ij}^g(P,P')=\sum_{k,P''} A_{ik}(P,P'')(B_0^{-1})_{kj}(P'',P'). \tag{15}$$

At the beginning of this section, we discussed the case of two fragments abutting one another. This approach can easily be generalized to an arbitrary structure composed of N_r fragments joined at N_s interfaces.

3. RESULTS FOR InGaAs/GaAs STRUCTURES; DISCUSSION

We now demonstrate the opportunities offered by the suggested technique by calculating strain fields in specific semiconducting structures, namely quantum wires and cylindrical quantum dots. The cross sections of these structures are shown in Fig. 1. The strain tensor components in quantum wires are independent of axial location, and in cylindrical quantum dots they are independent of the azimuthal angle. Under these conditions, the formulas derived in the previous section can be simplified. The resulting integral equations become one-dimensional in this case and can easily be solved using numerical techniques.

Figure 1 illustrates our calculations for cross sections of quantum wires based on $In_{0.1}Ga_{0.9}As/GaAs$ heterostructures with quantum wells near their surfaces. In order to visualize the calculated strain fields, the vertical displacements shown in the graph are magnified by a factor of 200. The values of elastic parameters used in our calculations (the Poisson coefficient σ and the relative change in Young's modulus E) are given in Table I. We have analyzed structures with various combinations of the upper cap GaAs layer thickness $L_{\rm cap}$, InGaAs quantum well thickness L_z, and the depths $L_{\rm b}$ of etched grooves in GaAs under the quantum well, given a common lateral size of the wire of $L_x=45$ nm. The diagrams

TABLE I. Material parameters of GaAs and $In_{0.1}Ga_{0.9}As$ at low temperatures used in calculations.

	GaAs	$In_{0.1}Ga_{0.9}As$
σ	0.31	0.31
E, arb. units	1	0.95
a, nm	0.56534	0.56939
$-\delta\times10^3$	0	7.11295
a_c, eV	−7.2	−7.0
a_v, eV	1.2	1.2
d_v, eV	1.9	1.8
b_v, eV	4.5	4.4

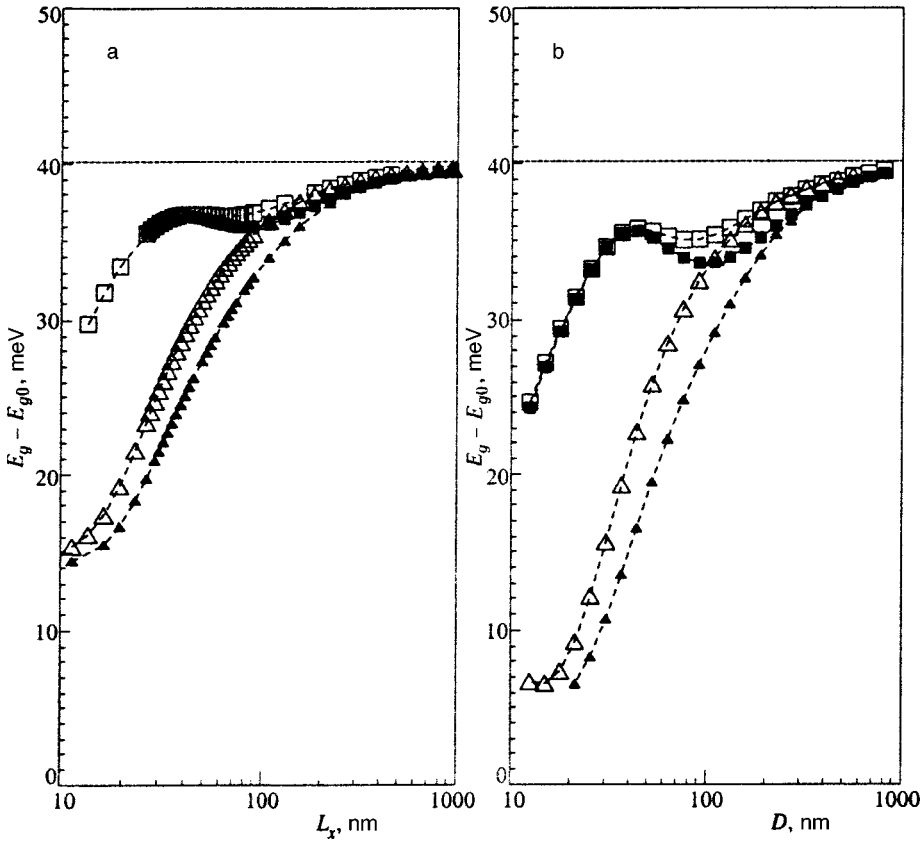

FIG. 2. Changes in the band gap width at the centers of the quantum well layers in strained $In_{0.1}Ga_{0.9}As/GaAs$ heterostructure (a) quantum wires and (b) quantum dots, plotted as functions of lateral sizes of nanostructures calculated for four different configurations shown in Fig. 1 (squares correspond to the configuration in Fig. 1a, black squares to Fig. 1b, triangles to Fig. 1c, and black triangles to Fig. 1d).

clearly show that the strain relaxation in the InGaAs layers peaks near the vertical walls defining the wires. Moreover, at small thicknesses of the cap barrier layer, the quantum well layer as a whole is bent (Figs. 1c and 1d). One can see that the GaAs cap layer leads to additional compression of the InGaAs quantum-well layer. As will be shown below, this can lead to extremely abrupt changes in band parameters of structures under discussion. The solutions for quantum dots are similar to those for quantum wires, the only difference being larger relaxation amplitudes in InGaAs quantum dots than in the wires. This is to be expected, due to the greater relative surface exposure of the quantum dots.

The shift of the semiconductor band gap,

$$\Delta_g = \Delta_c - \Delta_v, \tag{16}$$

can be expressed in terms of the strain tensor components. The shifts of the conduction band bottom and valence band top are determined by the formulas

$$\Delta_c = a_c \epsilon, \tag{17}$$

$$\Delta_v = a_v \epsilon + \sqrt{\frac{b_v^2}{2}\epsilon_1^2 + d_v^2 \epsilon_2^2}, \tag{18}$$

respectively, where

$$\epsilon = \mathrm{Tr}\, u_{ij},$$

$$\epsilon_1^2 = (u_{xx}-u_{yy})^2 + (u_{yy}-u_{zz})^2 + (u_{zz}-u_{xx})^2,$$

$$\epsilon_2^2 = u_{xy}^2 + u_{yz}^2 + u_{zx}^2.$$

The deformation potential constants a_c, a_v, b_v, and d_v (Ref. 5) are listed in Table 1.

Figure 2 plots calculations of renormalization of the band gap width Δ_g at the Γ-point in the middle of the quantum well InGaAs layer versus nanostructure lateral size (quantum wire width L_x and cylindrical dot diameter D, respectively). For each quantum object, the renormalization of the band gap width is given for the four nanostructure cross sections shown in Fig. 1.

The value $\Delta_g = 0$ in Fig. 2 corresponds to complete elastic relaxation in an InGaAs layer, whereas the value

$$\Delta_g^{(0)} = (a_c - a_v)\mathrm{Tr}\, u^{(0)} - |b_v(u_{xx}^{(0)} - u_{zz}^{(0)})| \tag{19}$$

[see also Eq. (2)], which equals approximately 40 meV in $In_{0.1}Ga_{0.9}As/GaAs$ structures at $T = 2$ K and is shown in Fig. 2 as a dashed horizontal line, corresponds to the total uniaxial strain in a stressed $In_{0.1}Ga_{0.9}As$ quantum well on an unbounded GaAs substrate. The closer the calculated change in the gap width to zero, the greater the elastic relaxation of the InGaAs layer; on the contrary, the closer the change is to $\Delta_g^{(0)}$, the greater the elastic stress in the InGaAs layer.

Of special note is the difference between the curves for structures with and without a cap GaAs layer on top of the InGaAs quantum well (squares and triangles in Fig. 2, respectively). This difference is considerably greater at L_x, $D < 100$ nm and increases to 15–20 meV. This is due to greater elastic relaxation in structures without the GaAs cap layer owing to the larger bend, which can be seen in Fig. 1. As expected, the amplitudes of the gap width variations in

the quantum dots are much greater than in the quantum wires.

A comparison of our calculations and data derived from measurements of photoluminescence spectra of quantum wires and dots fabricated lithographically from InGaAs/GaAs heterostructures[6] indicates that the red shift of the excitonic luminescence spectrum with lateral sizes decreasing to L_x, $D \approx 100$ nm can be attributed to the partial elastic relaxation of the quantum well layer. In structures with smaller lateral sizes, the red shift related with the strain can be high, but is offset by a large blue shift due to the size quantization of carriers. Note that a correct quantum-mechanical description of the system should take into account nonuniform variations in the positions of band edges due to the relaxation of elastic strains in semiconducting nanostructures, as well as the large difference between the dielectric constants of semiconductors and vacuum.

The authors are indebted to M. Bayer, A. Forchel, and V. D. Kulakovskiĭ for useful discussions. This work was supported by the Russian Fund for Fundamental Research (Grant 97-02-17872) and the Nanostructure program sponsored by the Ministry of Science and Technology of Russia (Grant 97-1072). One of the authors (N.A.G.) is currently the A. V. Humboldt fellow at the University of Würzburg, Germany.

*)E-mail: gippius@wpax13.physik.uni-wuerzburg.de
†)E-mail: tikh@gpi.ru

[1] J. P. Reithmaier, R. Höger, H. Riechert, A. Heberle, G. Abstreiter, and G. Weimann, Appl. Phys. Lett. **56**, 536 (1990).
[2] K. Kash, R. Bhat, Derek D. Mahoney, P. S. D. Lin, A. Scherer, J. M. Worlock, B. P. Van der Graag, M. Koza, and P. Grabbe, Appl. Phys. Lett. **55**, 681 (1989).
[3] K. Kash, B. P. Van der Graag, Derek D. Mahoney, A. S. Gozdz, L. T. Florez, and J. P. Harbison, Phys. Rev. Lett. **67**, 681 (1989).
[4] L. D. Landau and E. M. Lifshitz, *Theory of Elasticity*, 3rd ed., Pergamon Press, New York (1986).
[5] Landolt-Börnstein, *Numerical Data and Functional Relationships in Science and Technology*, New Series, III/17a, O. Madelung (ed.), Springer-Verlag, Berlin (1982).
[6] R. Steffen, Th. Koch, J. Oshiniwo, F. Faller, and A. Forchel, Appl. Phys. Lett. **68**, 223 (1996); A. Forchel, Private communication.

Translation provided by the Russian Editorial office

Emission of charged particles by a charged black hole

I. B. Khriplovich*)

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

(Submitted 23 December 1998)
Zh. Éksp. Teor. Fiz. **115**, 1539–1548 (May 1999)

The probability of charged particle production by the electric field of a charged black hole strongly depends on the particle energy. This probability is found in the nonrelativistic and ultrarelativistic limits. The range of values for the mass and charge of a black hole in which the mechanism of emission discussed here dominates over the Hawking mechanism is indicated. © *1999 American Institute of Physics.* [S1063-7761(99)00105-5]

1. The problem of particle production by the electric field of a black hole has been repeatedly discussed.[1–6] The probability of this process was estimated in these treatments using, in one way or another, earlier results[7–9] for the case of an electric field constant over all space. This approximation would seem to be quite natural in relation to sufficiently large black holes, for which the gravitational radius significantly exceeds the Compton wavelength of the particle, $\lambda = 1/m$. (In the present paper we employ a system of units in which $\hbar = 1$ and $c = 1$; Newton's gravitational constant k is written explicitly.) However, as we will shortly see, generally the constant-field approximation is inadequate here, since it does not reflect several essential features of the problem.

It is convenient to begin the discussion with the problem of particle production by a constant electric field. Here and in what follows we will limit ourselves to the case of electron and positron production, primarily because the emission probability is maximized for these lightest charged particles. Moreover, the concept of the Dirac sea makes it possible in the case of fermions to do without second quantization, thus making our reasoning transparent.

To calculate the principal exponential dependence of the effect, it is sufficient to limit ourselves to a simple approach originating in Ref. 7 (see also the textbook in Ref. 10). In the potential $-eEz$ of a constant electric field E, the ordinary Dirac gap (Fig. 1a) tilts (see Fig. 1b). As a result, a particle that had a negative energy in the absence of the field can now tunnel through the gap (see the dashed line in Fig. 1b) and go to infinity as an ordinary particle. The hole created in this way is nothing but an antiparticle. Elementary calculations lead to the well-known result for the particle production probability:

$$W \sim \exp\left\{-\frac{\pi m^2}{eE}\right\}. \tag{1}$$

This simple derivation clearly explains important properties of the phenomenon. First, the action inside the barrier does not change when the dashed line in Fig. 1b is shifted up or down, and it is for this reason that Eq. (1) is independent of the energy of the particles produced. Moreover, for the external field to be considered constant, it must vary only slightly along the path inside the barrier. However, the length of this path is not directly related to the Compton wavelength of the particle. In particular, for an arbitrarily weak field the path inside the barrier becomes arbitrarily long.

Thus, one can expect that the constant-field approximation generally cannot be used in the problem of particle production by a charged black hole and that the probability of particle production in this problem is strongly energy-dependent. The explicit form of this dependence is found below. In this paper we will limit ourselves to the case of a nonrotating black hole.

2. We start the solution of the problem by calculating the action inside the barrier. The metric of a charged black hole is well known:

$$ds^2 = f\,dt^2 - f^{-1}dr^2 - r^2(d\theta^2 + \sin^2\theta\, d\phi^2), \tag{2}$$

where

$$f = 1 - \frac{2kM}{r} + \frac{kQ^2}{r^2}, \tag{3}$$

with M and Q the mass and charge of the black hole, respectively. The equation for the particle 4-momentum in these coordinates is

$$f^{-1}\left(\epsilon - \frac{eQ}{r}\right)^2 - fp^2 - \frac{l^2}{r^2} = m^2, \tag{4}$$

where e and p are the energy and radial momentum of the particle, respectively. We assume that the particle charge e is of the same sign as the black-hole charge Q, attributing the charge $-e$ to an antiparticle.

Clearly, the action inside the barrier is minimized when the orbital angular momentum l is zero. It is evident therefore (and this will be demonstrated in the next section explicitly) that as a result of summation over l the s state determines the exponential factor in the total probability of the process. Thus, for the time being we limit ourselves to the case of purely radial motion. The equation for the Dirac gap at $l = 0$ is

$$\epsilon_{\pm}(r) = \frac{eQ}{r} \pm m\sqrt{f}. \tag{5}$$

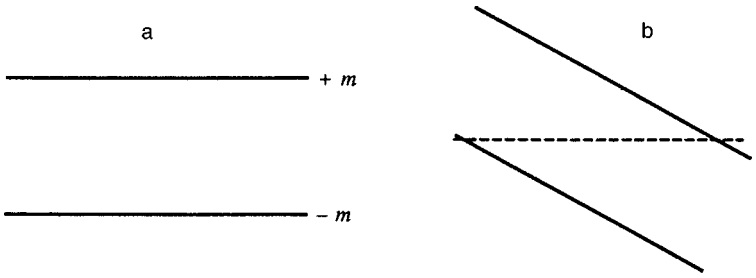

FIG. 1.

Figure 2 depicts the curve representing this dependence. As is well known,[11] at the horizon of a black hole, i.e., at $r=r_+=kM+\sqrt{k^2M^2-kQ^2}$, the gap vanishes. Then, as r increases, the lower boundary of the gap $\epsilon_-(r)$ decreases monotonically, tending asymptotically to $-m$. The upper branch $\epsilon_+(r)$ at first generally increases, and then decreases, tending asymptotically to m.

Figure 2 shows that the particles of the Dirac sea whose coordinate r exceeds the gravitational radius r_+ and whose energy ϵ lies between m and $\epsilon_-(r)$ tunnel through the gap to infinity. In other words, a black hole loses its charge as a result of this effect by emitting particles with the same sign of the charge e as that of Q. Clearly, the phenomenon takes place only if

$$\frac{eQ}{r_+}>m. \tag{6}$$

For an extreme black hole, with $Q^2=kM^2$, the Dirac gap looks somewhat different (see Fig. 3): as Q^2 tends to kM^2, the position of the maximum of the curve $\epsilon_+(r)$ tends to r_+, and the value of the maximum tends to eQ/r_+. It is obvious, however, that qualitatively the situation is the same. Thus, although an extreme black hole has a zero Hawking temperature and, correspondingly, gives off no thermal radiation, it still produces charged particles due to the effect in question.

In the general case $Q^2\leqslant kM^2$ the doubled action inside the barrier in the exponent for the particle production probability is

$$2S=2\int_{r_1}^{r_2}dr\,|p(r,\epsilon)|$$

$$=2\int_{r_1}^{r_2}\frac{dr\,r}{r^2-2kMr+kQ^2}$$

$$\times\sqrt{-p_0^2+2(\epsilon eQ-km^2M)r-(e^2-km^2)Q^2}\,, \tag{7}$$

where $p_0=\sqrt{\epsilon^2-m^2}$ is the momentum of the emitted particle at infinity, and the turning points $r_{1,2}$ are, as usual, the roots of the quadratic polynomial under the radical; here we are interested in the energy interval $m\leqslant\epsilon\leqslant eQ/r_+$. Of course, the integral can be found explicitly, although this requires doing tedious calculations. The result, however, is simple:

$$2S=2\pi\frac{m^2}{(\epsilon+p_0)p_0}[eQ-(\epsilon-p_0)kM]. \tag{8}$$

Of course, in contrast to the exponent in Eq. (1), this expression is strongly energy-dependent.

Note that the action inside the barrier does not vanish even at the energy limit $\epsilon_m=eQ/r_+$. Figure 2 clearly shows that this is the case for a non-extreme black hole. For an extreme black hole this is not as obvious. However, due to the singularity of $|p(r,\epsilon)|$, the action inside the barrier is finite at $\epsilon=\epsilon_m=eQ/r_+$ for an extreme black hole as well. In this case the exponential factor in the probability is

$$\exp\left(-\pi\frac{\sqrt{k}\,m}{e}kmM\right). \tag{9}$$

Due to the extreme smallness of the ratio

$$\frac{\sqrt{k}\,m}{e}\sim 10^{-21}, \tag{10}$$

the exponent here is large only for a very heavy black hole, with a mass M exceeding that of the Sun by more than five orders of magnitude. And since the total probability, integrated over energy, is dominated by the energy region $\epsilon\sim\epsilon_m$, the semiclassical approach is applicable in the case of extreme black holes only for these very heavy objects. Note, finally, that for particles emitted by an extreme black hole, the typical values of ϵ/m are very large:

$$\frac{\epsilon}{m}\sim\frac{\epsilon_m}{m}=\frac{eQ}{kmM}=\frac{e}{\sqrt{k}\,m}\sim 10^{21}.$$

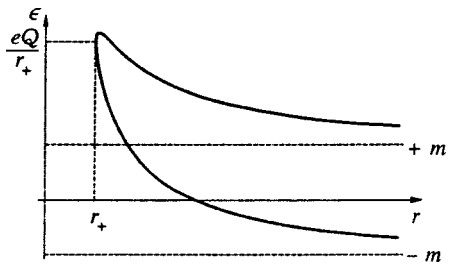

FIG. 2.

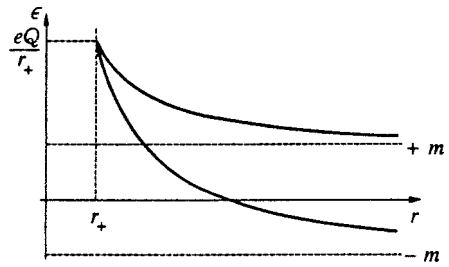

FIG. 3.

In other words, in all cases an extreme black hole emits mainly ultrarelativistic particles.

Let us now go back to non-extreme holes. In the nonrelativistic limit, when we have $eQ/r_+ \to m$ and, correspondingly, the particle velocity v tends to zero, the exponential factor is of course very small:

$$\exp\left(-\frac{2\pi kmM}{v}\right). \tag{11}$$

Therefore, we will consider mainly the opposite (ultrarelativistic) limit, where the exponential factor is

$$\exp\left(-\pi\frac{m^2}{\epsilon^2}eQ\right). \tag{12}$$

Here too the energy range $\epsilon \sim \epsilon_m \sim eQ/kM$ plays an important role, so that the ultrarelativistic limit corresponds to the condition

$$eQ \gg kmM. \tag{13}$$

But then the semiclassical result (12) is applicable (i.e., the action inside the barrier is large) only if

$$kmM \gg 1. \tag{14}$$

Note that this last condition means that the gravitational radius of the black hole ($r_+ \sim kM$) is much larger than the electron Compton wavelength $1/m$. In other words, the result (12) refers to macroscopic black holes. Combining (13) with (14), we arrive at one more condition for the applicability of formula (12):

$$eQ \gg 1. \tag{15}$$

We will study this relationship later.

Note that Gibbons[4] calculated the action inside the barrier under the assumptions used in deriving (12). However, the result presented in Ref. 4, $2S = \pi m^2 r_+^2/eQ$, is totally energy-independent (and corresponds to Eq. (1), which refers to the case of a constant electric field). There is no explanation of how such a result can be obtained for the integral in question in the general case $\epsilon \neq \epsilon_m$.

3. The obtained exponential yields the probability of a particle that approaches the turning point r_1 (see Figs. 2 and 3) from the left passing through the potential barrier. It should be recalled that generally the position of a turning point depends not only on the particle energy ϵ but also on its orbital angular momentum l. The total number of particles with given ϵ and l approaching a spherical surface of radius r_1 per unit time is equal to the product of the area of this surface

$$S = 4\pi r_1^2(\epsilon, l) \tag{16}$$

by the particle current density

$$j^r(\epsilon, l) = \frac{\rho}{\sqrt{g_{00}}}\frac{dr}{dt} \tag{17}$$

(see, e.g., §90 in Ref. 12). As usual, the particle velocity is

$$v^r = \frac{dr}{dt} = \frac{\partial\epsilon}{\partial p} \tag{18}$$

(note that we have again omitted the subscript r on the radial momentum). To obtain an explicit expression for the particle density ρ, we will use the semiclassical approximation (the conditions of its applicability for the region $r_+ \leqslant r \leqslant r_1$ will be discussed later). Note that the phase-space volume element

$$2\frac{dp_x\,dp_y\,dp_z\,dx dy dz}{(2\pi)^3} \tag{19}$$

is a scalar. (Here the factor 2 is due, as usual, to the two possible orientations of the electron spin.) On the other hand, the number of particles in the elementary cell $dx\,dy\,dz$ is (see §90 in Ref. 12)

$$\rho\sqrt{\gamma}\,dx\,dy dz, \tag{20}$$

where γ is the determinant of the space metric tensor. Since all the states of the Dirac sea are occupied, by comparing Eqs. (19) and (20) we find that

$$\frac{\rho}{\sqrt{g_{00}}} = \frac{2}{\sqrt{g_{00}\gamma}}\sum\frac{dp_x\,dp_y\,dp_z}{(2\pi)^3} = \frac{2}{\sqrt{-g}}\sum\frac{dp_x\,dp_y\,dp_z}{(2\pi)^3}$$

must be plugged into Eq. (17) for the current density [the summation here and below is done with fixed ϵ and l; see Eq. (17)]. In our case the determinant g of the four-dimensional metric tensor does not differ from the flat metric tensor, so that the radial current density of the Dirac-sea particles is

$$j^r(\epsilon, l) = 2\sum\frac{d^3p}{(2\pi)^3}\frac{\partial\epsilon}{\partial p}. \tag{21}$$

The summation on the right hand side actually reduces to multiplication by $2l+1$ of the possible projections of the orbital angular momentum $\mathbf{l}$ on the z axis and to integration over the azimuthal angle of the vector $\mathbf{l}$, which yields 2π. If we now allow for the identity

$$\frac{\partial\epsilon}{\partial p_r}dp_r = d\epsilon,$$

we arrive at

$$j^r(\epsilon, l) = 2\frac{2\pi(2l+1)}{(2\pi)^3 r_1^2(\epsilon, l)}. \tag{22}$$

Finally, the pre-exponential factor in the probability, which is differential in energy and orbital angular momentum, is

$$\frac{2(2l+1)}{\pi}. \tag{23}$$

Accordingly, the number of particles emitted per unit time is

$$\frac{dN}{dt} = \frac{2}{\pi}\int d\epsilon\sum_l (2l+1)\exp[-2S(\epsilon, l)]. \tag{24}$$

In the most interesting (ultrarelativistic) case, dN/dt can be calculated explicitly. Consider the expression for the momentum in the region inside the barrier for $l \neq 0$:

$$|p(\epsilon,l,r)|=f^{-1}\sqrt{\left(m^2+\frac{l^2}{r^2}\right)f-\left(\epsilon-\frac{eQ}{r}\right)^2}. \qquad (25)$$

The main contribution to the integral over energies in formula (24) is provided by the region $\epsilon\to\epsilon_m$. In this region the functions $f(r)$ and $\epsilon-eQ/r$ entering into expression (25) are small and vary rapidly. In the quantity

$$\mu^2(r,l)=m^2+\frac{l^2}{r^2}, \qquad (26)$$

we can replace r by the average value, which lies between the turning points r_1 and r_2. Obviously, in the limit $\epsilon\to\epsilon_m$ under discussion, the nearest turning point coincides with the horizon radius, $r_1=r_+$. In this limit the expression for the farthest turning point is

$$r_2=r_+\left[1+\frac{2\mu^2}{\epsilon_m^2-\mu^2}\,\frac{\sqrt{k^2M^2-kQ^2}}{r_+}\right]. \qquad (27)$$

If for the sake of making an estimate we assume $r\sim r_+$ in (26), we can easily show that the correction to 1 in the square brackets is limited by the ratio $l^2/(eQ)^2$. If we now assume that this ratio is small (below we will see that this assumption is self-consistent), we conclude that $r_2\approx r_+$, so that μ^2 can be considered independent of r, i.e., $\mu^2(r,l)=m^2+l^2/r_+^2$. As a result we obtain

$$2S(\epsilon,l)\approx\pi eQ\left(\frac{m^2}{\epsilon^2}+\frac{l^2}{r_+^2\epsilon^2}\right). \qquad (28)$$

Now we can easily find that

$$\frac{dN}{dt}=m\left(\frac{eQ}{\pi mr_+}\right)^3\exp\left(-\frac{\pi m^2r_+^2}{eQ}\right). \qquad (29)$$

Note that the range of orbital angular momenta contributing to the total probability (29) is effectively limited by the condition $l^2\leqslant eQ$. Since $eQ\gg 1$ holds, this condition makes it possible to pass from summation with respect to l in (24) to integration. On the other hand, this condition justifies the approximation $\mu^2(r,l)=m^2+l^2/r_+^2$.

However, up to now we have not considered one more condition necessary for the derivation of formula (29), the applicability of the semiclassical approximation to the left of the barrier, for $r_+\leqslant r\leqslant r_1$. This condition has the usual form

$$\frac{d}{dr}\,\frac{1}{p(r)}<1. \qquad (30)$$

In other words, the minimum size of the initial wave packet must not exceed the distance from the horizon to the turning point. Using the estimate

$$p(r)\sim\frac{r_+(eQ-\epsilon r_+)}{(r-r_+)(r-r_-)}$$

for the momentum in the most important region, we can easily verify that for an extreme black hole the condition (30) is valid because $eQ\gg 1$. In the non-extreme case, for $r_+-r_-\sim r_+$, the situation is different: the condition (30) reduces to

$$\epsilon<\frac{eQ-1}{r_+}\sim\frac{eQ}{r_+}. \qquad (31)$$

Thus, for a non-extreme black hole in the most important region $\epsilon\to\epsilon_m$, the condition for the applicability of the semiclassical approximation fails to hold. Nevertheless, the semiclassical result (24) remains valid qualitatively to within a factor of order unity in the coefficient.

We conclude this section with a few words about the emission of particles by light charged black holes, for which $kmM<1$ holds, i.e., for which the gravitational radius is smaller than the Compton wavelength of the electron. In this case the first part of inequality (31),

$$\epsilon<\frac{eQ-1}{r_+},$$

which guarantees the localization of the initial wave packet in the strong-field region, means in particular that

$$eQ=Z\alpha>1 \qquad (32)$$

(here we have introduced $Z=Q/e$). It is well-known (see, e.g., Refs. 13 and 14) that the vacuum near a point charge with $Z\alpha>1$ is unstable, so that such an object loses its charge by emitting charged particles. It is quite natural that for a black hole whose gravitational radius is smaller than the electron Compton wavelength the condition for emission of a charge is the same as in pure quantum electrodynamics. (Note that the unity in all these conditions should not be taken literally: even in quantum electrodynamics, where the instability condition for the vacuum of spin-1/2 particles is $Z\alpha>1$ (for a point nucleus), for a finite nucleus it becomes[13,14] $Z\alpha>1.24$. On the other hand, for the vacuum of scalar particles in the field of a point nucleus the instability condition is $Z\alpha>1/2$ (see, e.g., Refs. 15 and 16). As noted earlier, for a light black hole ($kmM<1$ the condition $eQ>1$ gives rise to a small action inside the barrier and to the inapplicability of the semiclassical approximation adopted in the present paper. The problem of emission of particles by a charged black hole with $kmM<1$ has been studied numerically by Page.[17]

4. The exponential factor

$$\exp\left(-\frac{\pi m^2r_+^2}{eQ}\right)$$

in our formula (29) coincides with the expression arising from Eq. (1), which refers to the case of a constant electric field E, if we replace this field by its value Q/r_+^2 at the black hole horizon. As mentioned earlier, an approach based on the formulas for a constant electric field was used in Refs. 1–6. Thus, our result for the main (exponential) dependence of the probability integrated over energy coincides with the corresponding result of these papers. Moreover, our final formula (29) agrees with the corresponding result of Ref. 6 to within a common factor 1/2. (This difference is of no interest by itself: as noted earlier, for a non-extreme black hole the semiclassical approximation cannot guarantee an exact value for the common numerical factor.)

Nevertheless, we believe that the analysis of the phenomenon performed in the present paper, which demon-

strates the essential difference between this phenomenon and particle production by a constant external field, is useful. First, the analysis shows that the probability of particle production by a charged black hole has a nontrivial energy spectrum. Moreover, real particles are produced by a charged black hole not in the entire space: for a given energy ϵ particles are emitted by a spherical surface of radius $r_2(\epsilon)$, a surface that is close to the horizon at maximum energy eQ/r_+. (For instance, it follows that the derivation of the result of Ref. 6 for dN/dt has no physical meaning: it amounts to substituting $E=Q/r^2$ into the well-known Schwinger formula,[9] obtained for a constant field, and then integrating over the entire space outside the horizon.)

We will now compare the intensity I of particle emission due to the mechanism discussed in this paper with the Hawking intensity I_H. Introducing an additional weight ϵ into the integrand of formula (24), we obtain

$$I=\pi m^2\left(\frac{eQ}{\pi m r_+}\right)^4 \exp\left(-\frac{\pi m^2 r_+^2}{eQ}\right). \tag{33}$$

The simplest way to estimate the Hawking intensity it is to use dimensional arguments: we divide the Hawking temperature $T_H=1/4\pi r_+$ by the typical classical time of the problem, r_+ (recall that we use a system of units in which $c=1$). Thus,

$$I_H\sim\frac{1}{4\pi r_+^2}. \tag{34}$$

A more accurate result for I_H differs from this estimate by a small numerical factor $\simeq 2\times 10^{-2}$, but for qualitative estimates this difference can be neglected. The intensities (33) and (34) become equal at

$$eQ\sim\frac{\pi}{6}\frac{(mr_+)^2}{\ln mr_+}\sim\frac{\pi}{6}\frac{(kmM)^2}{\ln kmM}. \tag{35}$$

We believe that the condition $eQ\sim 1/(4\pi)$ for the equality of these intensities, which was obtained by Novikov and Starobinskiĭ[6] from a comparison of $\epsilon_m=eQ/r_+$ and $T_H=1/(4\pi r_+)$, is not justified.

In conclusion, consider the change of the horizon surface of a black hole (and hence of the hole's entropy) due to this nonthermal emission. To this end it is convenient to introduce, following Ref. 18, the so-called irreducible mass M_0 of a black hole:

$$2M_0=M+\sqrt{M^2+Q^2} \tag{36}$$

(here and below we assume that $k=1$). It is convenient to write this relationship as

$$M=M_0+\frac{Q^2}{4M_0}. \tag{37}$$

Obviously, $r_+=2M_0$, so that the horizon surface and the black-hole entropy are proportional to M_0^2.

When a charged particle is emitted, the charge of the black hole changes by $\Delta Q=-e$ and its mass, by $\Delta M=-eQ/r_++\xi$, where ξ is the deviation of the particle energy from its maximum value. Using Eq. (37), we can easily see that as a result of particle emission the irreducible mass M_0 and hence the horizon surface and entropy of a nonextreme black hole do not change, provided that the particle energy is at its maximum eQ/r_+. In other words, such a process, which is the most probable one, is adiabatic. For $\xi>0$, the irreducible mass, the horizon surface, and the entropy increase.

As usual, an extreme black hole, with $M=Q=2M_0$, is a special case. Here at the maximum energy of the emitted particle, $\epsilon_m=e$, we have $\Delta M=\Delta Q=-e$, so that the black hole remains extreme after particle emission. In this case we have $\Delta M_0=e/2$, so that the irreducible mass and the horizon surface decrease. In the more general case $\Delta M=-e+\xi$, the irreducible mass changes as follows:

$$\Delta M_0=-\frac{e-\xi}{2}+\sqrt{\left(M_0-\frac{e}{2}+\frac{\xi}{4}\right)\xi}. \tag{38}$$

Clearly, for a heavy extreme black hole, even for a small deviation ξ of the emitted energy from the maximum value, the square root dominates in this expression, so that the horizon surface increases.

The author is grateful to I. V. Kolokolov, A. I. Mil'shteĭn, V. V. Sokolov, and O. V. Zhirov for the interest in the work and for useful comments. The work was supported by the Russian Fund for Fundamental Research (Grant No. 98-02-17797), the Russian Ministry of Education (Grant 3H-224-98), and the Federal Program ''Integration 1998'' (Project No. 274).

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

[1] M. A. Markov and V. P. Frolov, Teor. Mat. Fiz. **3**, 3 (1970).
[2] W. T. Zaumen, Nature (London) **247**, 531 (1974).
[3] B. Carter, Phys. Rev. Lett. **33**, 558 (1974).
[4] G. W. Gibbons, Commun. Math. Phys. **44**, 245 (1975).
[5] T. Damour and R. Ruffini, Phys. Rev. Lett. **35**, 463 (1975).
[6] I. D. Novikov and A. A. Starobinskiĭ, Zh. Éksp. Teor. Fiz. **78**, 3 (1980) [Sov. Phys. JETP **51**, 1 (1980)].
[7] F. Sauter, Z. Phys. **69**, 742 (1931).
[8] W. Heisenberg and H. Euler, Z. Phys. **98**, 714 (1936).
[9] J. Schwinger, Phys. Rev. **82**, 664 (1951).
[10] V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ, *Quantum Electrodynamics*, 2nd ed., Butterworth–Heinemann, Boston (1996).
[11] N. Deruelle and R. Ruffini, Phys. Lett. B **52**, 437 (1974).
[12] L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, 4th ed., Pergamon, Oxford (1975).
[13] Ya. B. Zel'dovich and V. S. Popov, Usp. Fiz. Nauk **105**, 403 (1971) [Sov. Phys. Usp. **14**, 673 (1972)].
[14] A. B. Migdal, Usp. Fiz. Nauk **123**, 369 (1977) [Sov. Phys. Usp. **20**, 879 (1977)].
[15] A. Sommerfeld, *Wave Mechanics*, 6th ed., Methuen, London (1944).
[16] H. A. Bethe, *Intermediate Quantum Mechanics*, W. A. Benjamin, New York (1964).
[17] D. N. Page, Phys. Rev. D **16**, 2402 (1977).
[18] D. Christodoulou, Phys. Rev. Lett. **25**, 1596 (1970); D. Christodoulou and R. Ruffini, Phys. Rev. D **4**, 3552 (1971).

Translated by Eugene Yankovsky

Neutralization of negative hydrogen ions in collisions with multicharged fast ions

A. B. Voĭtkiv

U. A. Arifov Institute of Electronics, Uzbek Academy of Sciences, 700143 Tashkent, Uzbekistan

N. Grün and W. Scheid

Institute for Theoretical Physics (Theorie II) at the Justus-Liebig-University, Giessen, Heinrich-Buff-Ring 16, D-35392 Giessen, Germany

(Submitted 17 April 1998)
Zh. Éksp. Teor. Fiz. **115**, 1549–1560 (May 1999)

We study the neutralization of negative hydrogen ions in collisions with multicharged fast ions (including relativistic ions) by using an approach that allows a simple expression for the neutralization cross section to be derived over a range of collision parameters where the standard Born approximation breaks down. © *1999 American Institute of Physics.*
[S1063-7761(99)00205-X]

1. INTRODUCTION

Atomic collisions involving negative ions that result in a change in the charge composition of the colliding particles have been thoroughly studied in the past several decades (see, e.g., Refs. 1 and 2 and the literature cited therein). The study of such collisions may lead to important practical applications (e.g., the production of beams of fast neutral particles[3]). A theoretical investigation of the neutralization of H^- ions in collisions with fast particles that carry a relatively small charge, i.e., in the limit $Z \ll v$ (here Z is the charge of the particle, and v is the collision velocity; in this paper we use, if not stated otherwise, the atomic system of units), can be carried out either in the Born approximation (see, e.g., Ref. 4) or by using the impact-parameter method, when for all impact parameters the probability of electron transitions is calculated to first order in the interaction between the electron and the field of the fast particle. For collisions with heavy fast particles the two approaches yield equivalent results for the cross sections (see, e.g., Ref. 5), and for the sake of brevity we call these approaches the standard Born approximation. Calculations in this approximation yield a simple expression for the neutralization cross section:[6]

$$\sigma_{-0} = 183.7 \frac{Z^2}{v^2} \ln 5.44v. \tag{1}$$

In this paper we examine the neutralization of negative hydrogen ions in collisions with multicharged fast ions,

$$H^- + A^{Z+} \rightarrow H^- + A^{Z+} + e^-, \tag{2}$$

in the collision parameter range $Z \gtrsim v \gg v_0$ (here v_0 is the characteristic velocity of the loosely bound electron in H^-), where the standard Born approximation cannot be used. At such values of charges and velocities the process of electron capture by the multicharged ion has a very low probability (see, e.g., Refs. 7 and 8). Hence the reaction (2) almost fully determines the total cross section of neutralization for H^- ions in such collisions. Experimentally, the cross sections of neutralization for H^- ions were studied by Melchert *et al.*[2] in collisions with the multicharged ions Ne^{Z+} ($Z \leqslant 4$), Ar^{Z+} and Xe^{Z+} ($Z \leqslant 8$) at center-of-mass collision energies $E_{\text{c.m.}} \lesssim 200$ keV, while Tawara *et al.*[9] have started an experimental investigation of the neutralization of H^- ions by multicharged fast ions at much higher collision energies, on the order of several MeV/amu. The problem of neutralization of H^- ions in collisions with multicharged fast ions was investigated theoretically in Refs. 1,2,8–10, where the following theoretical approaches were used to calculate the neutralization cross section: (1) the method of classical Monte Carlo paths,[1] (2) the method developed in Ref. 11 as a generalization of the Keldysh theory[12] for the photoionization in a strong field and applied by Melchert *et al.*[2] to the problem of neutralization of H^- ions, (3) the two-state model,[10] and (4) the close-coupling method.[8]

In the present paper we study the problem of neutralization of H^- ions in collisions with multicharged fast ions (including relativistic ions) by using an approach that, in contrast to the methods just mentioned, makes it possible to solve the problem of finding the cross sections analytically, with results applicable over a broad range of values of the parameters of the problem, $v \lesssim Z$ and $v_0 \ll v < c$ (here $c = 137$ is the speed of light; the limits of this approach are specified below).

2. GENERAL FORMULAS

In accordance with the description of H^- ions in the model of a split shell $(1s,1s')$, we assume that one of the two electrons is on an almost hydrogenlike $1s$-orbital, while the other, loosely bound, electron occupied a diffuse orbital with radius $\simeq 4$. The large difference in the binding energies of the electrons in H^- (0.5 and 0.0275, respectively) makes it possible to examine the problem of neutralization of the negative hydrogen ion by the one-electron approach, in which it is assumed that the outer, loosely bound, active electron moves in the effective field of the ''frozen'' core (the proton plus the tightly bound inner electron). Following one of the methods of selecting the effective one-electron

wave function of the bound state of the negative ion and the valence *s*-electron (see, e.g., Ref. 13), we can write the wave function in the form

$$\psi_0(r)=N\frac{\exp\{-\kappa r\}-\exp\{-\beta r\}}{r}, \tag{3}$$

where r is the distance between the active electron and the nucleus of the H^- ion, $N=\sqrt{\kappa\beta(\kappa+\beta)/2\pi(\beta-\kappa)^2}$ is the normalization factor, and $\kappa=0.235$ is determined from the known value of the affinity energy: $\kappa^2/2=0.0275$. For the value of parameter β we take $\beta=0.913$, which was defined by Sidis *et al.*[14] so as to correctly describe the value of the singlet scattering length of an electron (in the *s*-wave) scattered by a hydrogen atom at energies lower than the first excitation threshold of the hydrogen atom. The wave function (3) behaves properly at large distances,

$$\psi_0(\kappa r\gg 1)\simeq G\sqrt{\frac{\kappa}{2\pi}}\frac{\exp\{-\kappa r\}}{r}, \tag{4}$$

with $G=1.51$,[1)] and remains finite as $r\to 0$. Below we use the wave function (3) and the corresponding functions of the continuum spectrum (assuming, as usual, that the scattering phase shift of the electron scattered by the hydrogen atom at low energies is finite only for the *s*-wave) to describe the eigenstates of the discrete and continuous spectrum of the H^- ion.

Now we discuss the collision. We assume that prior to the collision the negative hydrogen ion was at the origin and the multicharged ion was moving along a classical straight path $\mathbf{S}(t)=\mathbf{b}+\mathbf{v}t$, where $\mathbf{b}$ is the impact-parameter vector. Recent calculations by Lin *et al.*[8] have shown that even for moderate values of Z the cross section of neutralization for H^- ions in collisions with multicharged fast ions depends only on the total charge of the ions and not on the internal structure of the ions. Hence in what follows we assume that a multicharged fast ion is a point charge.

To date the experimental results with respect to neutralization of H^- ions by multicharged fast ions have been obtained in studies of collisions where the velocities are much smaller than the speed of light, so that relativistic effects can be neglected. However, to obtain a more general solution, we will not use the condition $v\ll c$ in deriving an expression for the neutralization cross section. Accordingly, the field of a multicharged point ion moving, in general, with a relativistic speed, is described by the scalar and vector potentials (see, e.g., Ref. 16)

$$\varphi=\frac{Z}{R},\quad \mathbf{A}=\frac{\mathbf{v}}{c}\varphi,$$
$$R=\sqrt{(z-vt)^2+(1-v^2/c^2)(\boldsymbol{\rho}-\mathbf{b})^2}, \tag{5}$$

where $(x,y,z)=(\boldsymbol{\rho},z)=\mathbf{r}$ are the coordinates of the active electron, the z axis is directed along the velocity of the multicharged ion, and $\boldsymbol{\rho}\cdot\mathbf{v}=0$. As is known (see, e.g., Ref. 2), collisions with impact parameters $b>r_0$ (where $r_0\approx\kappa^{-1}$ is the characteristic size of the H^- ion) provide the principal contribution to the cross section of neutralization for H^- ions in collisions with multicharged fast (but nonrelativistic) ions. Below we show that for collisions with $b>r_0$ the electron velocity in the course of the entire process remains much smaller than the speed of light. Hence in calculating the transitions of the active electron in collisions with multicharged ions with $b>r_0$ we can use the Schrödinger equation

$$i\frac{\partial\Psi}{\partial t}=[H_0+W(t)]\Psi,\quad b>r_0, \tag{6}$$

where H_0 is the effective one-electron Hamiltonian for the free hydrogen ion, and $W(t)$ describes the interaction of the nonrelativistic electron and the field of the relativistic particle:

$$W(t)=\frac{1}{2c}(\mathbf{p}\cdot\mathbf{A}+\mathbf{A}\cdot\mathbf{p})+\frac{\mathbf{A}^2}{2c^2}-\varphi, \tag{7}$$

with $\mathbf{p}$ the electron momentum operator. Since the negative hydrogen ion is a nonrelativistic system, we can expect the magnetic field generated by the multicharged ions to affect only slightly the probability of electron detachment from H^- in collisions with $b>r_0$. By examining the expression for the classical Lorentz force acting on an electron in the electric and magnetic fields generated by the multicharged ions,

$$\mathbf{F}=\mathbf{E}+\frac{1}{c}\mathbf{v}_{\rm el}\times\mathbf{H},$$

with $\mathbf{v}_{\rm el}$ the electron velocity, we can make an estimate of the relative importance of these fields. For the fields generated by relativistic particles we can generally assume that $H\simeq E$. However, since $v_{\rm el}/c\ll 1$, for the ratio of the magnetic component of the Lorentz force to the electric component we have $v_{\rm el}/v/c^2\ll 1$. Then, when examining collisions with $b>r_0$, we can ignore the magnetic field of the multicharged ions for all values of the velocities in the collisions. Hence the interaction (7) can be expressed to the first approximation in terms of only the electric field. To this end we use the gauge transformation

$$\mathbf{A}_1(\mathbf{r},t)=\mathbf{A}(\mathbf{r},t)+\operatorname{grad}\chi(\mathbf{r},t),$$
$$\varphi_1(\mathbf{r},t)=\varphi(\mathbf{r},t)-\frac{1}{c}\frac{\partial}{\partial t}\chi(\mathbf{r},t),$$
$$\Psi_1(\mathbf{r},t)=\exp[-i\chi(\mathbf{r},t)]\,\Psi(\mathbf{r},t), \tag{8}$$

where χ is the gauge function. Using the Göppert-Mayer gauge function[17]

$$\chi(\mathbf{r},t)=-\mathbf{r}\cdot\mathbf{A}(0,t), \tag{9}$$

and Eqs. (6)–(9), expanding the scalar and vector potential in powers of the electron position up to first-order terms inclusive, and keeping only the principal term in the expression of the interaction in the new Schrödinger equation, we find that

$$i\frac{\partial\Psi_1}{\partial t}=[H_0+W_1(t)]\Psi_1,\quad b>r_0, \tag{10}$$

where $W_1(\mathbf{r},t)=\mathbf{r}\cdot\mathbf{E}(0,t)$, and

$$\mathbf{E}(0,t)=\mathbf{E}=-\operatorname{grad}\varphi(0,t)-\frac{1}{c}\frac{\partial\mathbf{A}(0,t)}{\partial t}=\frac{Z\mathbf{R}_1}{R_0^3\gamma^2},\quad b>r_0,$$

$$\mathbf{R}_1=(-\mathbf{b},vt),\quad R_0=\sqrt{v^2t^2+\left(\frac{b}{\gamma}\right)^2},\quad \gamma=\frac{1}{\sqrt{1-(v/c)^2}}. \tag{11}$$

Here **E** is the electric field calculated in the dipole approximation. The interaction W_1 can be interpreted as acting over a finite time interval. Indeed, we have

$$\int_{-\infty}^{\infty}W_1(t)\,dt=W_1(t=0)T, \tag{12}$$

where $T=T(b)=2b/\gamma v$ has the meaning of the effective collision time (see, e.g., Ref. 18). Depending on the magnitude of the impact parameter, this time may be either shorter or longer than the characteristic time of revolution, $\tau\approx r_0/v_0$, of the active electron in the bound state of a free H^- ion. Using the wave function (3), we can estimate the characteristic velocity v_0 of the electron in the bound state of the H^- ion at $v_0\simeq\sqrt{\beta\kappa}\simeq\sqrt{\kappa}$, i.e., $\tau\approx\kappa^{-3/2}$. Following Ref. 19, we divide the entire range $b>r_0$ of the impact parameter into two subranges: (1) $r_0<b\ll\gamma v\tau$, and (2) $b\gg Z/v\kappa$. These subranges overlap if $Z/\kappa\ll\gamma v^2\tau$ holds, and below we assume this condition to be true.

When the effective collision time is short compared to the characteristic internal time of the atomic system, in calculating the transitions of the atomic system we can use the sudden approximation; see, e.g., Refs. 20 and 21. In the impact-parameter range $r_0<b\ll\gamma v^2\tau$, where $T(b)\ll\tau$, to calculate the neutralization probability we use the zeroth-order sudden approximation, within which the neutralization probability $w_{-0}(b)$ can be written

$$w_{-0}(b)\simeq w^s_{-0}(b)=\int d\mathbf{k}\left|\langle\mathbf{k}|\exp\left[-i\int_{-\infty}^{\infty}W_1(t)\,dt\right]|\psi_0\rangle\right|^2, \tag{13}$$

where $|\mathbf{k}\rangle$ is a state of the continuous spectrum of the effective one-electron Hamiltonian H_0 for H^-. Allowing for the fact that $\int_{-\infty}^{\infty}W_1(t)\,dt=\mathbf{q}\cdot\mathbf{r}$, where $\mathbf{q}=2Z\mathbf{b}/vb^2$ has the meaning of the average momentum transferred by the field of the multicharged ion to the active electron, we can write (13) as follows:

$$w^s_{-0}(b)=\int d\mathbf{k}\,|\langle\mathbf{k}|\exp\{-i\mathbf{q}\cdot\mathbf{r}\}|\psi_0\rangle|^2. \tag{14}$$

The value of q is negligible compared to $m_ec=137$, which justifies our assumption about the essentially nonrelativistic electron velocities in collisions with $b>r_0$. The spatial shift of the electron over the collision time, ξ, a shift that can be estimated at $\xi(b)\sim q(b)T(b)\sim Z/\gamma v^2\ll r_0$, is small compared to the characteristic size of the H^- ion. The conditions $T(b)\ll\tau$ and $\xi(b)\ll r_0$ justify the use of the zeroth-order sudden approximation when investigating the problem in question in the impact-parameter range $r_0<b\ll\gamma v\tau$.

In our case, when calculating the probability (14), it is convenient to use the completeness condition for the states $|\psi_0\rangle$ and $\{|\mathbf{k}\rangle\}$ of the negative ion:

$$|\psi_0\rangle\langle\psi_0|+\int d\mathbf{k}\,|\mathbf{k}\rangle\langle\mathbf{k}|=\mathbf{I}.$$

As a result,

$$w^s_{-0}(b(q))=1-\frac{16\pi^2N^4}{q^2}\left(\arctan\frac{q}{2\kappa}+\arctan\frac{q}{2\beta}-2\arctan\frac{q}{\kappa+\beta}\right)^2. \tag{15}$$

For the contribution to the neutralization cross section σ_{-0} of collisions with impact parameters $b_1\leqslant b\leqslant b_2$ (where $b_2\ll\gamma v\tau$, and b_1 is defined below) we have

$$\Delta\sigma_{-0}(b_1\leqslant b\leqslant b_2)=2\pi\int_{b_1}^{b_2}db\,bw^s_{-0}(b)=8\pi\frac{Z^2}{v^2}\int_{q_2}^{q_1}\frac{dq}{q^3}p(q), \tag{16}$$

where $p(q)=w^s_{-0}(b(q))$, and $q_{1,2}=2Z/vb_{1,2}$.

Collisions with small impact parameters, $b\lesssim r_0$, are characterized (provided that $v\lesssim Z$) by large energy transfers to the active electron in comparison to the electron binding energy in H^-. The average energy transferred to the electron in a collision with an impact parameter b can be estimated at $\epsilon_{\rm tr}(b)\simeq 2Z^2/b^2v^2$ for $b\gtrsim r_0$ (see, e.g., Ref. 18). This means that the energy is high even at $b\simeq r_0$ ($\epsilon_{\rm tr}\simeq 2Z^2\kappa^2/v^2\gg\kappa^2/2$) and, correspondingly, becomes even higher for $b\lesssim r_0$. Such a large energy transfer to the active electron results in electron detachment from the negative ion with a probability $w^s_{-0}(b)\simeq 1$. What is important is that the condition $b>r_0$ plays an important role in deriving Eq. (14) (since we use the dipole expansion for the interaction of the electron and the field of the multicharged ion) and, correspondingly, the use of this expression in calculating the neutralization probability is justified only if $b>r_0$. Calculations show, however, that for $v\lesssim Z$ Eq. (14) yields reasonable values (close to unity) of the detachment probability also in the impact-parameter range $b\lesssim r_0(w^s_{-0}(b)\lesssim 1)$. Hence below we use Eq. (14) to estimate the electron detachment probability in the range of small impact parameters, $b\lesssim\kappa^{-1}$.

Our calculations show that the contribution (16) to the neutralization cross section is almost independent of q_1 for $q_1\gtrsim 1$. Hence below we simply put $q_1=\infty$ ($b_1=0$):

$$\Delta\sigma_{-0}(0\leqslant b\leqslant b_2)=8\pi\frac{Z^2}{v^2}\int_{q_2}^{\infty}\frac{dq}{q^3}p(q). \tag{17}$$

The integral in (17) can be evaluated in the following way. We write

$$\int_{q_2}^{\infty}\frac{dq}{q^3}p(q)=\int_{q_0}^{\infty}\frac{dq}{q^3}p(q)+\int_{q_2}^{q_0}\frac{dq}{q^3}p(q), \tag{18}$$

where $q_0\ll\kappa$. Equation (15) suggests that $p(q)=6.055q^2$ for $q\ll\kappa$. Moreover, $q_2\ll\kappa$ for $Z/v\kappa\ll b_2\ll\gamma v\tau$. Hence we can write

$$\int_{q_2}^{\infty}\frac{dq}{q^3}p(q)=\int_{q_0}^{\infty}\frac{dq}{q^3}p(q)+6.055\ln\frac{q_0b_2v}{2Z}, \tag{19}$$

where $Z/v\kappa \ll b_2 \ll \gamma v\tau$. For small values of q_0, the integral on the right-hand side of Eq. (19) depends on q_0 as $\ln(1/q_0)$. Indeed, numerical calculations with different (but small) values of q_0 show that this integral can written

$$\int_{q_0}^{\infty} \frac{dq}{q^3} p(q) = 6.055 \ln \frac{C}{q_0}, \tag{20}$$

where $C=0.46$ is almost independent of q_0. Thus, using Eqs. (17)–(20), we arrive at an expression for the contribution to the detachment cross section of collisions with impact parameters from the range $0 \leqslant b_0 \leqslant b_2$, with $Z/v\kappa \ll b_2 \ll \gamma v\tau$:

$$\Delta\sigma_{-0}(0 \leqslant b \leqslant b_2) = 152.2 \frac{Z^2}{v^2} \ln \frac{0.46 v b_2}{2Z}. \tag{21}$$

For large impact parameters, $b \gtrsim \gamma v\tau$, we cannot use the sudden approximation. However, in collisions with large impact parameters one can expect that the field of the multicharged ion (notwithstanding the large ion charge) is only a small perturbation even for a loosely bound active electron in H^-. Calculations show that in collisions with $b \gg Z/v\kappa$ the electron detachment probability is low compared to unity. Indeed, for $Z/v\kappa \ll b \ll \gamma v\tau$ we can also use Eq. (14), which yields

$$w_{-0}(b) \simeq \int d\mathbf{k}\, |\langle \mathbf{k}|\mathbf{q}\cdot\mathbf{r}|\psi_0\rangle|^2 \simeq 1.34 \frac{Z^2}{v^2\kappa^2 b^2} \ll 1. \tag{22}$$

Hence to calculate the electron detachment probability for collisions with $b \gg Z/v\kappa$ we can use perturbation theory in the interaction $W_1(t)$. In first-order perturbation theory, the expression for the electron detachment probability is (see the Appendix)

$$w_{-0}(b) \simeq w^p_{-0}(b) = \frac{4Z^2}{\gamma^4 v^4} \int_0^{\infty} dk k^2 \omega_{k1}^2 y_{k1}^2 \left[K_0^2\left(\frac{\omega_{k1} b}{\gamma v}\right) + \gamma^2 K_1^2\left(\frac{\omega_{k1} b}{\gamma v}\right)\right], \quad b \gg \frac{Z}{v\kappa}, \tag{23}$$

where $\omega_{k1} = (\mathbf{k}^2 + \kappa^2)/2$ are the electron transition frequencies, $y_{k1}^2 (z_{k1}^2 = x_{k1}^2 = y_{k1}^2 = r_{k1}^2/3)$ are the averaged squares of the components of the (one-electron) dipole matrix elements for H^-, and K_0 and K_1 are the modified Bessel functions. The contribution to the neutralization cross section of collisions with impact parameters b from the range $b_3 \leqslant b < \infty$ has a simple form if the point b_3 is selected so that $Z/v\kappa \ll b_3 \ll \gamma v\tau$:

$$\Delta\sigma_{-0}(b_3 \leqslant b < \infty) = 2\pi \int_{b_3}^{\infty} db\, b w^p_{-0} = 152.2 \frac{Z^2}{v^2} \left[\ln \frac{1.123 v\gamma}{\omega_{\text{eff}} b_3} - \frac{v^2}{2c^2}\right], \tag{24}$$

where

$$\omega_{\text{eff}} = \exp \frac{\int_0^{\infty} dk\, k^2 y_{k1}^2 \ln \omega_{k1}}{\int_0^{\infty} dk\, k^2 y_{k1}^2} = 0.081.$$

Using (21) and (24) and putting $b_2 = b_3$ (which is possible, since the two impact-parameter ranges in question overlap), we arrive at a simple formula for the total cross section of neutralization for H^- ions in collisions with multicharged fast ions:

$$\sigma_{-0} = 152.2 \frac{Z^2}{v^2} \left[\ln \frac{3.2 v^2 \gamma}{Z} - \frac{v^2}{2c^2}\right]. \tag{25}$$

The contribution of the impact-parameter range $b \lesssim r_0$ does not exceed πr_0^2 (in order of magnitude). As Eq. (25) implies, $\sigma_{-0} \gg \pi r_0^2$ for $Z \gtrsim v$. This means that our initial assumption about the relative contribution of the impact-parameter range $b > r_0$ to the cross section was correct.

We now discuss the limits of our approach. First, to use the sudden approximation we must ascertain that $v \gg v_0 \simeq \sqrt{\kappa}$ holds. Second, the condition $Z \gtrsim v$ is needed to minimize the error introduced by using Eq. (14) in calculations in the impact-parameter range $b \lesssim r_0$. Third, an important requirement is that the two impact-parameter ranges in question overlap, which is the case if $Z/\kappa v \ll \gamma v\tau \simeq \gamma v \kappa^{-3/2}$. Hence Eq. (25) can be used to calculate the cross section in a parameter range defined by the conditions $z \lesssim Z \ll \gamma v^2/\sqrt{\kappa}$ and $v \gg v_0 \simeq \sqrt{\kappa}$. In the relativistic case the applicability range reduces to $Z \sim v \sim c$. For nonrelativistic collisions, when $(v/c)^2 \ll 1$, the nonrelativistic limit of Eq. (25), [2)]

$$\sigma_{-0} = 152.2 \frac{Z^2}{v^2} \ln \frac{3.2 v^2}{Z}, \tag{26}$$

can be used to calculate cross sections if both conditions, $v \lesssim Z \ll v^2/\sqrt{\kappa}$ and $v \gg v_0 \simeq \sqrt{\kappa}$, are met. A remark is in order. Since the active electron in H^- has (on the atomic scale) a very low binding energy and a low orbital velocity, even collisions with particles with $Z \sim 1$ and $v \sim 1$ can be considered collisions with ''multicharged'' fast ions, and the values of the cross sections of neutralization for H^- ions can be estimated by Eq. (26).

3. DISCUSSION

In Fig. 1 we compare the results of calculations done using Eq. (26) and the experimental data on the cross sections of neutralization for H^- ions in collisions with ions of Ne ($Z \leqslant 4$), of Ar ($Z \leqslant 8$), and of Xe ($Z \leqslant 8$) at a center-of-mass energy $E_{\text{c.m.}}$ of the colliding particles equal to 200 keV (Ref. 2). If we take into account the assertion of Melchert *et al.*[2] that, within experimental error, the values of the cross sections depend only on the total charge of the ion and not on the details of the ion internal structure, we can conclude that there is good agreement between our results and the experimental data. Figure 1 also depicts the results of calculations in the standard Born approximation [Eq. (1)] and of calculations that use the analytical Presnyakov–Uskov approximation (Eq. (9) of Ref. 2). For these values of the col-

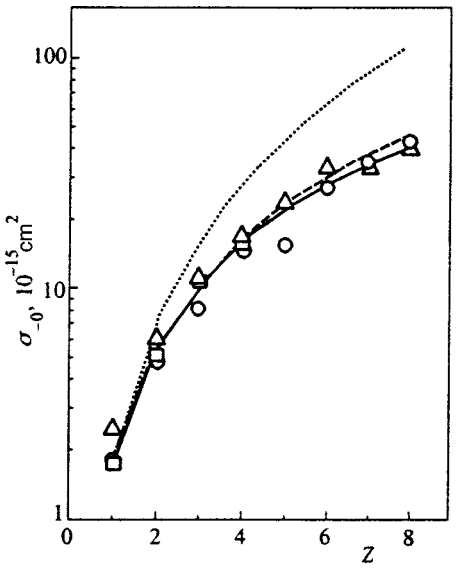

FIG. 1. Cross sections of neutralization for H^- ions at a (center-of-mass) collision energy $E_{c.m.}=200$ keV. The experimental data are taken from Ref. 2: □, Ne^{Z+} ($Z=1$–4); ○, Ar^{Z+} ($Z=1$–8); and, Xe^{Z+} ($Z=1$–8). The solid, dashed, and dotted curves represent the results of calculations by Eq. (26), by the Presnyakov–Uskov formula, and by Eq. (1), respectively.

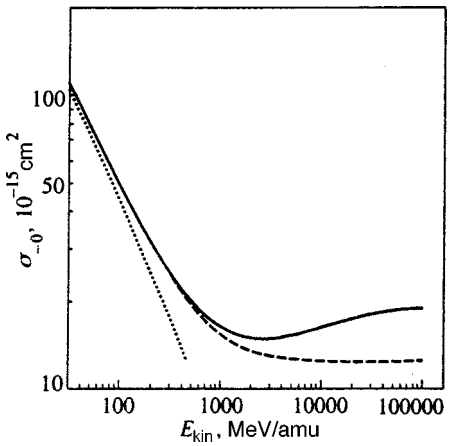

FIG. 2. Cross section of neutralization for H^- ions in relativistic collisions with U^{92+} ions as a function of the kinetic energy (per atomic mass unit) of the multicharged incoming ion. The solid curve represents the results of calculations by Eq. (25), while the dashed and dotted curves represent the results of calculations by Eq. (26) that use the relativistic and nonrelativistic relationship between the velocity and kinetic energy of the multicharged ion.

lision parameters, Eq. (26) and the Presnyakov–Uskov formula yield very similar results (similar to each other and with respect to the experimental data) for all $Z=1$–8, while for large values of Z the standard Born approximation yields overvalued results.

Note that for other values of the parameters Z and v lying within the applicability range of our approach, the calculation of cross sections by Eq. (26) also yields results that are in good agreement both with the experimental data[2] and with the results of other calculations.[2,8,10]

Tawara *et al.*[9] have started an experimental investigation of ion neutralization at high collision energies E (of order of several MeV/amu). At the 9th International Conference on the Physics of Highly Charged Ions, Tawara *et al.*[9] reported on their (preliminary) experimental results for the cross section of neutralization for H^- ions by Ar^{18+} ions at $E=2$ MeV/amu. The value of the cross section they obtained was $\sigma_{-0}^{expt}=3.8\times10^{-14}\,cm^2$. For such collision parameters Eq. (26) yields a value of $4.55\times10^{-14}\,cm^2$ for the cross section, the Presnyakov–Uskov formula yields $4.8\times10^{-14}\,cm^2$, and Eq. (1) yields $8\times10^{-14}\,cm^2$. The substantial overvaluation of the cross sections for large Z stemming from the use of Eq. (1) is the result of the nonunitarity of the standard Born approximation. In this approximation the neutralization probability is proportional to Z^2 and for collisions involving multicharged ions with small impact parameters may exceed unity. The approach used in the present paper is free from this drawback, since it uses the sudden approximation (which is unitary) in calculations of the neutralization probability with moderate values of b. Interestingly, the Presnyakov–Uskov formula and Eq. (26), derived by very different approaches, yield extremely close values of cross sections when $v^2/Z>1$. For instance, when the parameter $x=v^2/Z$ varies within the range $1<x<100$, for the ratio of the cross section (26) to the Presnyakov–Uskov cross section, $f=\sigma_{-0}/\sigma_{Pr-Us}$, we have $0.9<f(x)<0.96$.

As noted earlier, our approach can also be used to describe the neutralization of H^- ions in relativistic collisions with ions whose charge is excessively large. As an example, Fig. 2 depicts the results of calculations of the cross section of neutralization for the H^- ion in relativistic collisions with U^{92+}. The relativistic effects, which influence the dependence of the neutralization cross section on the kinetic energy of the multicharged ions, can be divided into two types: (a) the effect related to an increase in the velocity of the multicharged ions with the ion kinetic energy that is slower than predicted by nonrelativistic mechanics, and (b) the effect related to the ''flattening'' of the electric field a relativistic particle experiences in the direction of its motion (see, e.g., Ref. 16). Figure 2 suggests that effect (a) becomes appreciable at $E_{kin}\sim100$ MeV/amu, while effect (b) begins to manifest itself at $E_{kin}\sim1$ GeV/amu. While the cross section specified by Eq. (26) tends to a finite limit in the range of ultrarelativistic collisions (provided that we use the relativistic relationship between velocity and energy), the relativistic Eq. (25) describes in this range an increase in cross section. The physical meaning of this increase is simple. In the ultrarelativistic limit, an increase in energy has almost no effect of the collision velocity, but γ increases and hence the flattening of the electric field generated by the relativistic particle becomes more pronounced. This reduces the effective collision time $T(b)$ and extends the impact-parameter range $b\lesssim\gamma v\tau$ where the collisions are sudden for the electron and where the detachment probability decreases, as b increases, much more slowly [by a power law; see Eq. (22)] than it does in the range $b\gtrsim\gamma v\tau$, where the external perturbation is adiabatically slow for the electron and where the detachment probability decreases exponentially with increasing b. As a result, the flattening of the electric field gives rise to a cross section that diverges with increasing γ as $\ln\gamma$ (see, e.g., Ref. 23 and the literature cited therein).

4. CONCLUSION

We have examined the process of neutralization of negative hydrogen ions in collisions with multicharged fast ions. In this approach, the H^- ion is described in the one-electron approximation, while the expression for the wave function of the active electron is selected in a form that guarantees correct asymptotic behavior of this function for $r \gg \kappa^{-1}$ and finite results for $r \lesssim \kappa^{-1}$. Collisions with multicharged fast ions are described by dividing the entire impact-parameter range into two overlapping subranges. To describe neutralization in collisions with $b \ll \gamma v \tau$, when the effective collision time $T(b) \sim b/\gamma v$ is short, we have used the zeroth-order sudden approximation. This approximation is unitary and within $v_0 \ll v \lesssim Z$ yields meaningful results for the detachment probability even for collisions with $b \lesssim r_0$, i.e., in the impact-parameter range where $w_{-0}(b) \simeq 1$. To describe collisions with $b \gg Z/v\kappa$, i.e., in the impact-parameter range where the interaction of an electron with a multicharged ion is already weak and the electron detachment probability is low, we have used first-order perturbation theory in the field of the multicharged ion. The resulting expression for the neutralization cross section can be used in a wide range of the parameters Z of v of the problem discussed above.

APPENDIX

In first-order perturbation theory, the electron transition amplitude has the form

$$a(\mathbf{k}) = -i \int_{-\infty}^{\infty} dt \exp(i\omega_{k1} t) \langle \mathbf{k} | W_1 | \psi_0 \rangle, \tag{27}$$

where $\omega_{k1} = (k^2 + \kappa^2)/2$ is the transition frequency. Integration with respect to time in (27) yields

$$\begin{aligned} a(\mathbf{k}) &= \frac{2iZ}{\gamma^2 v} \left\{ K_0\left(\frac{\omega_{k1} b}{\gamma v}\right) \left\langle \mathbf{k} \middle| \exp \frac{i\omega_{k1} z}{v} \middle| \psi_0 \right\rangle \right. \\ &\quad \left. + \frac{\gamma \omega_{k1}}{v} K_1\left(\frac{\omega_{k1} b}{\gamma v}\right) \langle \mathbf{k} | y | \psi_0 \rangle \right\} \\ &= \frac{2iZ\omega_{k1}}{\gamma^2 v^2} \left\{ \gamma K_1\left(\frac{\omega_{k1} b}{\gamma v}\right) \langle \mathbf{k} | y | \psi_0 \rangle \right. \\ &\quad \left. + iK_0\left(\frac{\omega_{k1} b}{\gamma b} \langle \mathbf{k} | z | \psi_0 \rangle\right) \right\}, \end{aligned} \tag{28}$$

where the y axis is directed along the impact-parameter vector, and K_0 and K_1 are the modified Hankel functions. Using the standard expansion in spherical functions for the wave function of the continuous spectrum (see, e.g., Ref. 4),

$$|\mathbf{k}\rangle = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} i^l \exp(-i\delta_l) R_{kl}(r) Y_{lm}^*\left(\frac{\mathbf{r}}{r}\right) Y_{lm}\left(\frac{\mathbf{k}}{k}\right), \tag{29}$$

from (28) we find that

$$|a(\mathbf{k})|^2 = \frac{4Z^2 \omega_{k1}^2}{v^4 \gamma^4} r_{k1}^2 \left[K_0^2\left(\frac{\omega_{k1} b}{\gamma v}\right) \cos^2\theta + \gamma^2 K_1^2\left(\frac{\omega_{k1} b}{\gamma v}\right) \sin^2\theta \sin^2\varphi \right], \tag{30}$$

where θ ($0 \leqslant \theta \leqslant \pi$) is the angle between $\mathbf{k}$ and $\mathbf{v}$, φ ($0 \leqslant \varphi < 2\pi$) is the angle between $\mathbf{b}$ and the projection of the electron momentum $\mathbf{k}$ in the final state on the impact-parameter plane, and

$$r_{k1} = \sqrt{4\pi} \int_0^{\infty} dr\, r^3 R_{k1}(r) \psi_0(r).$$

From Eq. (30) we finally arrive at an expression for the detachment probability:

$$\begin{aligned} w_{-0}(b) &= w_{-0}^p(b) = \int d\mathbf{k} \, |a(\mathbf{k})|^2 \\ &= \frac{4Z^2}{3\gamma^4 v^4} \int_0^{\infty} dk\, k^2 \omega_{k1}^2 r_{k1}^2 \left[K_0^2\left(\frac{\omega_{k1} b}{\gamma v}\right) \right. \\ &\quad \left. + \gamma^2 K_1^2\left(\frac{\omega_{k1} b}{\gamma v}\right) \right]. \end{aligned} \tag{31}$$

[1)] The value $G = 1.51$ is very close to $G_{\text{Pek}} = 1.56$, obtained from calculations with a multiparameter two-electron wave function.[15].

[2)] Interestingly, at $Z = 1$ Eq. (26) coincides, with an insignificant discrepancy in the numerical factors in front of the logarithm and under the logarithm sign, with the formula for the cross section of neutralization of H^- ions by electron impact obtained by Smirnov and Chibisov[22] under the assumption that the incoming electron moves along a classical path.

[1] F. Melchert, W. Debus, M. Liehr *et al.*, Europhys. Lett. **9**, 433 (1989).
[2] F. Melchert, M. Benner, S. Krüdener, R. Schulze, S. Meuser, K. Huber, E. Salzborn, D. B. Uskov, and L. P. Presnyakov, Phys. Rev. Lett. **74**, 888 (1995).
[3] D. B. Uskov, in *The Physics of Electronic and Atomic Collisions*, Whistler, Canada, July–August 1995, AIP Conf. Proc. **360**(1), 687 (1996).
[4] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-relativistic Theory*, 3rd ed., Pergamon Press, Oxford (1977).
[5] M. R. C. McDowell and J. P. Coleman, *Introduction to the Theory of Ion–Atom Collisions*, North-Holland, Amsterdam (1970).
[6] Y. K. Kim and M. Inokuti, Phys. Rev. A **3**, 665 (1971).
[7] R. K. Janev, L. P. Presnyakov, and V. P. Shevel'ko, *Physics of Highly Charged Ions*, Springer, New York (1985).
[8] J. T. Lin, T. F. Jiang, J. Y. Kuang, and C. D. Lin, Phys. Rev. A **56**, 2020 (1997).
[9] H. Tawara, T. Tonima, H. Kumagai, T. Imai, D. B. Uskov, and L. P. Presnyakov, in *Abstracts 9th Int. Conf. on the Physics of Highly Charged Ions* (HCI-98), Bensheim, September 14–18 (1998), p. 107.
[10] M. H. Cherkani, D. Fussen, M. I. Chibisov, and F. Brouillard, Phys. Rev. A **54**, 1445 (1996).
[11] L. P. Presnyakov and D. B. Uskov, Zh. Éksp. Teor. Fiz. **86**, 882 (1984) [Sov. Phys. JETP **59**, 515 (1984)].
[12] L. V. Keldysh, Zh. Éksp. Teor. Fiz. **47**, 1945 (1964) [Sov. Phys. JETP **20**, 1307 (1965)].
[13] B. M. Smirnov, *Negative Ions*, McGraw-Hill, New York (1982).
[14] V. Sidis, C. Kubach, and D. Fussen, Phys. Rev. A **27**, 2431 (1983).
[15] C. L. Pekeris, Phys. Rev. **126**, 1470 (1962).
[16] L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, 4th ed., Pergamon Press, Oxford (1975).
[17] M. Göppert-Mayer, Ann. Phys. **9**, 273 (1931).

[18] N. Bohr, ''The penetration of atomic particles through matter,'' Kgl. Danske Videnskab. Selskab. Mat.-Fys. Medd. **18**, No. 8 (1948).
[19] A. B. Voitkiv and A. V. Koval, J. Phys. B **31**, 499 (1998).
[20] J. H. Eichler, Phys. Rev. A **15**, 1856 (1977).
[21] A. M. Dykhne and D. L. Yudin, Usp. Fiz. Nauk **125**, 377 (1978) [Sov. Phys. Usp. **21**, 549 (1978)].
[22] B. M. Smirnov and M. I. Chibisov, Zh. Éksp. Teor. Fiz. **49**, 841 (1965) [Sov. Phys. JETP **22**, 585 (1966)].
[23] J. Eichler and W. E. Meyerhof, *Relativistic Atomic Collisions*, Academic Press, New York (1995).

Translated by Eugene Yankovsky

Self- and cross-phase modulation accompanying third-harmonic generation in a hollow waveguide

A. M. Zheltikov,*) N. I. Koroteev,†) and A. N. Naumov

International Laser Center, M. V. Lomonosov Moscow State University, 119899 Moscow, Russia
(Submitted 23 September 1998)
Zh. Éksp. Teor. Fiz. **115**, 1561–1579 (May 1999)

The influence of self- and cross-phase modulation on third-harmonic generation in a hollow waveguide is investigated. Analytic solutions of the coupled equations for the slowly varying amplitudes of the pump pulse and the third harmonic in a gas-filled hollow waveguide are obtained with consideration of self- and cross-phase modulation and first-order dispersion effects. The possibility of controlling the nonlinear phase trajectory of the third harmonic by cross-phase modulation is demonstrated. © *1999 American Institute of Physics.* [S1063-7761(99)00305-4]

1. INTRODUCTION

The coherent generation of optical harmonics in a field of short high-power laser pulses has been actively investigated for a fairly long time (see, for example, Refs. 1–16). Despite the impressive results achieved in the area of generating high-order harmonics and obtaining coherent short-wavelength radiation—with a wavelength shorter than 2.4 nm (Ref. 17)—in an inert-gas jet, this experimental approach, which is based on the use of focused laser beams, does not permit the achievement of high frequency-conversion efficiencies even for low-order harmonics (in contrast to low-order harmonic generation in a laser plasma[18–26]). The way to solve the problem of increasing the nonlinear-optical interaction length in the gas medium was pointed out by Miles *et al.*,[27] who demonstrated the possibility of significantly (by three orders of magnitude) increasing the efficiency of the nonlinear-optical interaction in a hollow dielectric waveguide for a four-photon process coherent anti-Stokes scattering process back in 1977. Nisoli *et al.*[28] demonstrated that the use of a hollow optical waveguide permits effective broadening of the spectrum of an ultrashort laser pulse due to self-phase modulation. In particular, it was shown that pulses with a duration of 20 fs propagating in a hollow optical waveguide filled with an inert gas experience broadening of the spectrum as a consequence of self-phase modulation, which is sufficient for subsequent compression of the pulse to a duration of 4.5 fs.

Since the optical breakdown threshold for a gas filling a waveguide significantly exceeds the characteristic values of the breakdown threshold for ordinary optical fibers, the approach developed in Refs. 27 and 28 is especially promising for the generation of high-power ultrashort laser pulses covering several light-field periods. Since hollow-waveguide technology permits the use of high-power laser pulses, this approach also seems very promising for optical frequency conversion using parametric mixing of light waves and harmonic generation. When femtosecond pulses are used, the frequency-conversion efficiency in such processes is restricted because of phase mismatching and group delay. The first experiments on third-harmonic generation in hollow waveguides filled with inert gases using pulses with a duration of 20 fs from a Ti:sapphire laser was reported in Ref. 29. However, no quantitative data regarding the parameters of the third-harmonic beam were presented in Ref. 29. The experiments in Ref. 30 showed that the use of hollow-waveguide technology permits the achievement of high laser-radiation frequency-conversion efficiencies for third-harmonic generation (0.2%) and parametric four-wave mixing (13%). Durfee *et al.*[30] also experimentally demonstrated that the phase mismatch accompanying a nonlinear-optical interaction due to the dispersion of the gas can be compensated by phase mismatch of the waveguide modes. The efficiency of the nonlinear-optical interaction then increases significantly.

The results of the experiments in Ref. 30 stimulated intensive research on nonlinear-optical frequency-conversion processes in hollow waveguides.[31–33] In particular, it was shown that compensation of the phase mismatch in hollow waveguides permits a 100–1000-fold increase in the frequency-conversion efficiency for harmonic generation up to the 45th order in comparison to the frequency-conversion efficiencies achieved in experiments with gas jets.[5–10]

Thus, the result of the experiments in Ref. 30 provide evidence that phase matching and temporal overlap of the fundamental and third-harmonic pulses can be ensured for a fairly long interaction length in the gas filling a hollow optical waveguide. Since the fundamental pulse effectively interacts with the third-harmonic pulse and, at the same time, according to Ref. 28, experiences fairly strong self-phase modulation, it can be presumed that the third harmonic should also be phase-modulated owing to cross-phase modulation.[34–39] In particular, the influence of cross-phase modulation can lead to the significant broadening detected in Ref. 30 of the spectrum of the UV signal appearing as a result of parametric four-wave mixing in a hollow optical waveguide. Similar phenomena associated with self- and cross-phase modulation were observed in experiments on

third-harmonic generation in a field of ultrashort laser pulses in connection with the optical breakdown of air[40,41] under conditions such that the temporal self-interaction of the light pulses was accompanied by self-defocusing of the fundamental light due to the induced plasma-electron density profile. According to the results of the theoretical analysis performed by Tempea and Brabec,[42] the self-phase modulation due to plasma nonlinearity under ionization conditions in a hollow waveguide permits the production of pulses with linear chirp, which can be effectively compensated using a dispersion delay line.

This paper shows that cross-phase modulation can be an important factor, which has a significant influence on the efficiency of frequency conversion upon the generation of optical harmonics in gas-filled hollow waveguides and, at the same time, permits control of the nonlinear (with respect to the intensity of the laser radiation) phase trajectory of the harmonic pulse. In Sec. 2 we obtain analytic solutions of the coupled equations for the slowly varying amplitudes of the pump pulse and the third harmonic in hollow waveguides with consideration of self- and cross-phase modulation and the first-order dispersion effects in a medium with a noninertial Kerr nonlinearity. Unlike the plane-wave approximation, the equations obtained permit allowance for the influence of the waveguide through the propagation constants, the group velocities of the pump and third-harmonic pulses, and nonlinear coefficients written with consideration of the transverse pump and third-harmonic fields for the corresponding waveguide modes. Section 3 describes a method for performing numerical calculations of the parameters of the pump and third-harmonic pulses in an argon-filled hollow waveguide. Section 4 discusses the results of the numerical calculations for a regime in which the effects of the group delay of the pump and third-harmonic pulses are insignificant, as well as for the case where the difference between the group velocities of the pump and third-harmonic pulses has a significant influence on the efficiency and properties of the process of third-harmonic generation in a hollow waveguide. The main results of this investigation are briefly summarized in the Conclusion.

2. BASIC RELATIONS FOR THIRD-HARMONIC GENERATION IN A HOLLOW WAVEGUIDE WITH CONSIDERATION OF SELF- AND CROSS-PHASE MODULATION

Let us consider the process of third-harmonic generation in a hollow optical waveguide filled with a gas medium having third-order nonlinearity. In analyzing this phenomenon we shall use the approximation of slowly varying amplitudes, presuming that the duration of the light pulses is large in comparison to the period of the light field (see, for example, Ref. 43). Readily interpreted analytical expressions describing third-harmonic generation in a hollow waveguide with allowance for the effects of self- and cross-modulation can be obtained, in analogy to Refs. 44 and 45, in the case where we can confine ourselves to consideration of the first-order dispersion effects.

2.1. Solutions of the equations for third-harmonic generation in a hollow waveguide within the approximation of slowly varying amplitudes

Let us represent the fundamental (pump) and third-harmonic pulses propagating along the z axis in a hollow waveguide filled by a gas with a noninertial Kerr nonlinearity in the following form

$$\mathbf{E}_p=\frac{1}{2}f_p^{n'n}(\boldsymbol{\rho})\mathbf{e}_p^{n'n}A^{n'n}(t,z)\exp[-i(\omega t-K_p^{n'n}z)]+\text{c.c.}, \quad (1)$$

$$\mathbf{E}_h=\frac{1}{2}f_h^{m'm}(\boldsymbol{\rho})\mathbf{e}_h^{m'm}B^{m'm}(t,z)\exp[-i(3\omega t-K_h^{m'm}z)]+\text{c.c.}, \quad (2)$$

where ω is the central frequency of the fundamental light; the subscripts p and h refer to the pump and third-harmonic pulses, respectively; $f_p^{n'n}(\boldsymbol{\rho})$ and $f_h^{m'm}(\boldsymbol{\rho})$ are the transverse distributions of the fields of the fundamental light and the third harmonic in the hollow waveguide corresponding to the eigenmodes of the hollow waveguide with the indices n' and n and with m' and m, respectively; $K_p^{n'n}$ and $K_h^{m'm}$ are the propagation constants of the pump and third-harmonic pulses corresponding to the same eigenmodes of the hollow waveguide;[46] $A^{n'n}(t,z)$ is the slowly varying amplitude of the fundamental pulse (it is assumed that a definite waveguide mode is excited at the frequency of the fundamental light); $B^{m'm}(t,z)$ is the slowly varying amplitude of the third-harmonic pulse (the indices corresponding to the transverse mode of the pump wave have been omitted for simplicity); and $\mathbf{e}_p^{n'n}$ and $\mathbf{e}_h^{m'm}$ are the unit vectors of the polarizations of the fundamental and third-harmonic pulses, respectively.

When the inequalities

$$\frac{\omega_l a}{c}\gg 1, \quad (3)$$

$$\left|\frac{K_l^{m'm}c}{\omega_l n_1(\omega_l)}-1\right|\ll 1, \quad (4)$$

hold, where $l=p$, h, and $n_1(\omega_l)$ is the refractive index of the gas in the hollow waveguide for radiation with the frequency ω_l, we can use approximate analytic solutions for the transverse distribution of the field and the propagation constants of the electromagnetic field in the hollow waveguide.[46] In particular, for the TEM_{1m} modes of a hollow waveguide we have

$$f_l^{1m}(\boldsymbol{\rho})\equiv f_l^m(\boldsymbol{\rho})=J_0\left(\frac{u_l^m\rho}{a}\right). \quad (5)$$

Here $J_0(x)$ is the zeroth-order Bessel function, u_l^m is the eigenvalue of the TEM_{1m} mode, a is the internal radius of the hollow waveguide, and

$$K_l^{1m}\equiv K_l^m\approx\frac{\omega_l n_1(\omega_l)}{c}\left[1-\left(\frac{u_l^m c}{a\omega_l n_1(\omega_l)}\right)^2\times\left(\frac{1}{2}+\frac{\text{Im}(\mu(\omega_l))}{a\omega_l}c\right)\right], \quad (6)$$

where

$$\mu(\omega_l)=\frac{\varepsilon_2(\omega_l)+n_1^2(\omega_l)}{2n_1^2(\omega_l)(\varepsilon_2(\omega_l)-n_1^2(\omega_l))^{1/2}}$$

for TEM modes and $\varepsilon_2(\omega_l)$ is the dielectric constant of the waveguide walls at the frequency ω_l.

Using a procedure similar to the one described in Ref. 47, we obtain equations for the slowly varying amplitudes of the pump $[A^n(z,t)\equiv A^{1n}(z,t)]$ and third-harmonic $[B^m(z,t)\equiv B^{1m}(z,t)]$ pulses:

$$\left(\frac{\partial}{\partial t}+\frac{1}{\nu_p^n}\frac{\partial}{\partial z}\right)A^n=i\tilde{\gamma}_1^n A^n|A^n|^2, \tag{7}$$

$$\left(\frac{\partial}{\partial t}+\frac{1}{\nu_h^m}\frac{\partial}{\partial z}\right)B^m=i\tilde{\beta}^{mn}(A^n)^3\times\exp(-i\Delta k^{mn}z)+2i\tilde{\gamma}_2^{mn}B^m|A^n|. \tag{8}$$

Here ν_p^n and ν_h^m are the group velocities of the pump and third-harmonic pulses, and

$$\Delta k^{mn}=K_h^m-3K_p^n\approx\Delta k_0+\Delta k_w^{mn} \tag{9}$$

is the phase mismatch taking into account the waveguide dispersion. In this equation

$$\Delta k_0=\frac{\omega_h}{c}[n_1(\omega_h)-n_1(\omega_p)],$$

$$\Delta k_w^{mn}=\frac{c}{2\omega_p}\left[3\left(\frac{u_p^n}{a}\right)^2-\frac{1}{3}\left(\frac{u_h^m}{a}\right)^2\right]$$

are the components of the phase mismatch due to the dispersion of the gas and the waveguide, respectively (the total phase mismatch can be represented as a sum of two components if the inequality $n_1(\omega_l)-1\ll 1$ holds). The nonlinear coefficients $\tilde{\gamma}_1^m$, $\tilde{\gamma}_2^{mn}$, and $\tilde{\beta}^{mn}$ can be expressed in terms of the nonlinear-optical cubic susceptibilities with the corresponding frequency arguments:

$$\tilde{\gamma}_1^n=\frac{3\pi\omega^2}{2K_p^nc^2}\mathbf{e}_p^{n*}\hat{\chi}^{(3)}(\omega;\omega,-\omega,\omega)\mathbf{e}_p^n\mathbf{e}_p^{n*}\mathbf{e}_p^n\times\frac{\int\int[f_p^n(\boldsymbol{\rho})]^4\rho\,d\rho\,d\theta}{\int\int[f_p^n(\boldsymbol{\rho})]^2\rho\,d\rho\,d\theta}, \tag{10}$$

$$\tilde{\gamma}_2^{mn}=\frac{27\pi\omega^2}{2K_h^mc^2}\mathbf{e}_h^{m*}\hat{\chi}^{(3)}(3\omega;3\omega,-\omega,\omega)\mathbf{e}_h^m\mathbf{e}_p^{n*}\mathbf{e}_p^n\times\frac{\int\int[f_h^m(\boldsymbol{\rho})]^2[f_p^n(\boldsymbol{\rho})]^2\rho\,d\rho\,d\theta}{\int\int[f_h^m(\boldsymbol{\rho})]^2\rho\,d\rho\,d\theta}, \tag{11}$$

$$\tilde{\beta}^{mn}=\frac{9\pi\omega^2}{2K_h^mc^2}\mathbf{e}_h^*\hat{\chi}^{(3)}(3\omega;\omega,\omega,\omega)\mathbf{e}_p\mathbf{e}_p\mathbf{e}_p\times\frac{\int\int f_h^m(\boldsymbol{\rho})[f_p^n(\boldsymbol{\rho})]^3\rho\,d\rho\,d\theta}{\int\int[f_h^m(\boldsymbol{\rho})]^2\rho\,d\rho\,d\theta}. \tag{12}$$

Equations (7) and (8) are similar to the equations describing third-harmonic generation in a gas medium with allowance for self- and cross-phase modulation in the plane-wave approximation (see, for example, Ref. 47). The nonlinear term on the right-hand side of Eq. (7) describes self-phase modulation. The first term on the right-hand side of Eq. (8) describes the nonlinear polarization of the third-order medium, which is responsible for third-harmonic generation and the cross-phase modulation of the third harmonic due to the self-phase modulation of the fundamental pulse (XPM1). The second term on the right-hand side of this equation corresponds to the cross-phase modulation due to the modulation of the refractive index at the frequency of the third harmonic by the fundamental pulse (XPM2).

We assumed in writing Eq. (8) that the third-harmonic pulse has a fairly low intensity and that the self-phase modulation of this pulse can be neglected. However, unlike the plane-wave approximation, Eqs. (7) and (8) take into account the influence of the waveguide through the propagation constants (6), the group velocities of the pump radiation and the third harmonic, and the nonlinear coefficients (10)–(12), which were written with allowance for the transverse distributions of the pump and third-harmonic fields for the corresponding waveguide modes. In particular, the mismatch of the wave vectors, which appears in Eq. (8) and specifies the efficiency of third-harmonic generation, depends not only on the dispersion of the gas, but also on the dispersion of the waveguide modes. As was noted in Ref. 30, this circumstance allows the phase-matching conditions for a particular pair of transverse modes of the pump radiation and the third harmonic to be improved.

Equations similar to (7) and (8) were analyzed in detail for the case of second-harmonic generation in fiber waveguides in Ref. 48 and for the combined propagation of optical pulses of different frequency in a waveguide in Refs. 49–51. In particular, Ho *et al.*[48] obtained an analytic solution for the process of second-harmonic generation with consideration of cross-phase modulation.

Solving the system of equations (7) and (8) after Ref. 45, we obtain the following expressions for the amplitudes of the fundamental and third-harmonic pulses:

$$A^n(\eta_p^n,z)=A_0^n(\eta_p^n)\exp[i\tilde{\gamma}_1^n|A_0^n(\eta_p^n)|^2z], \tag{13}$$

$$B^m(\eta_h^m,z)=i\tilde{\beta}^{mn}\exp\left[2i\tilde{\gamma}_2^{mn}\int_0^z|A_0^n(\eta_h^m+\zeta^{mn}z')|^2\,dz'\right]\times\int_0^z dz'\,(A_0^n(\eta_h^m+\zeta^{mn}z'))^3\exp\Big[-i\Delta k^{mn}z'+3i\tilde{\gamma}_1^m|A_0^n(\eta_h^m+\zeta^{mn}z')|^2z'$$

TABLE I. Characteristic spatial scales for the generation of the TEM_{1n} modes of the third harmonic in an argon-filled hollow waveguide with a diameter equal to 80 μm and in the free gas for pump radiation with a wavelength equal to 0.78 μm, a pulse duration equal to 25 fs, and a power equal to 2 GW at an argon pressure equal to 0.5 atm.

n	l_d, m	l_{d0}, m	l_w^{1n}, cm	l_{w0}^{1n}, cm	l_{ph}^{1n}, cm	l_{ph0}^{1n}, cm	l_{xpm1}, cm	l_{spm}, cm
1	16.2	15.9	27	22	0.25	0.39	4	12
2	17.6	15.9	24	22	0.31	0.39	4	12
3	20.8	15.9	20	22	0.55	0.39	4	12

[a]*Note*. Here l_d and l_{d0} are the dispersive spreading lengths in the waveguide and in the free gas, respectively; l_w^{1n} and l_{w0}^{1n} are the characteristic spatial scales of group delay in the waveguide and in the free gas, respectively; l_{ph}^{1n} and l_{ph0}^{1n} are the characteristic spatial scales of phase mismatch in the waveguide and in the free gas, respectively; and l_{xpm1} and l_{spm} are the characteristic spatial scales of XPM1 and self-phase modulation, respectively.

$$-2i\tilde{\gamma}_2^{mn}\int_0^{z'}|A_0^n(\eta_h^m+\zeta^{mn}z'')|^2\,dz''\Bigg], \qquad (14)$$

where $\eta_l^m=(t-z/\nu_l^m)/\tau$ is the time in the coordinate frame traveling with one of the pulses ($l=p,h$), normalized to the duration τ of the pulse of incident radiation, and $\zeta^{mn}=(1/\nu_h^m-1/\nu_p^n)/\tau$. Because we confined our treatment to first-order dispersion theory, the main pulse (13) propagates in the medium with no change in the shape of the envelope, i.e., $A^n(\eta_p^n,z)=A_0^n(\eta_p^n)$ [$A_0^n(\eta_p^n)$ is the shape of the envelope at the entrance to the medium]. The nonlinear phase trajectory of the fundamental pulse due to self-phase modulation can be written

$$\Phi_{\mathrm{spm}}=\tilde{\gamma}_1^n|A_0^n(\eta_p^n)|^2z. \qquad (15)$$

Like Eqs. (7) and (8), the solutions for the pump and third-harmonic pulses are formally similar to the expressions obtained in the plane-wave approximation,[45] but, unlike the equations presented in Ref. 45, they take into account the dispersion of the waveguide and the transverse intensity distributions of the pump and third-harmonic pulses.

2.2. Estimation of the characteristic spatial scales

As can be seen from Eq. (14), the significant restrictions on the efficiency of harmonic generation in the presence of ultrashort laser pulses are associated with the phase mismatch and group delay of the fundamental and third-harmonic pulses due to the dispersion of the phase and group velocities inherent in a nonlinear medium. As can be seen from Eqs. (6)–(9), in the case of third-harmonic generation in a hollow waveguide, it is possible to partially compensate the mismatch of the phase and group velocities of the pump and third-harmonic pulses associated with the dispersion of the phase and group velocities in the gas by utilizing the dispersion of the waveguide modes. The characteristic spatial scales of the phase mismatch and group delay of the pulses with respect to time in a waveguide are specified in the following manner:

$$l_{\mathrm{ph}}^{mn}=\frac{1}{\Delta k^{mn}}, \quad l_w^{mn}=(\zeta^{mn})^{-1}=\frac{\tau}{|(1/\nu_h^m-1/\nu_p^n)|}.$$

Table I presents estimates of the characteristic spatial scales of the phase mismatch and group delay in a waveguide and in a free gas for fundamental pulses of a Ti:sapphire laser having a duration of 25 fs and the third harmonic in an argon-filled hollow waveguide at a pressure of 0.5 atm. The values presented were obtained using the data on the dispersion of inert gases from Ref. 52. The table also presents estimates for the dispersive spreading length of the pump pulse in the hollow waveguide and the free gas and the characteristic spatial scales of self-phase modulation (l_{spm}) and XPM1 ($l_{\mathrm{xpm1}}\approx l_{\mathrm{spm}}/3$).

As can be seen from the estimates presented, the characteristic scales of the nonlinear-optical interactions for the conditions considered are significantly smaller than the dispersive spreading length of the pump pulse and the group-delay length of the pump radiation and the third harmonic. At the same time, in the general case the phase-mismatching effects have a significant influence on the frequency-conversion efficiency. Phase mismatch can be partially compensated only in a narrow range of pressures for individual pairs of spatial modes of the fundamental light and the third harmonic (the TEM_{11} mode of the fundamental light and the TEM_{13} mode of the third harmonic in the case under consideration; see also Secs. 4.1 and 4.2).

2.3. Case of combined propagation of the pump and third-harmonic pulses

Taking into account that under the conditions specified above $l_{\mathrm{spm}},l_{\mathrm{xpm1}}\ll l_w^{mn}$ (see Table I), we shall illustrate some important features of cross-phase modulation by examining the initial stage of third-harmonic generation for $z\ll l_w^{mn}$, where the group delay of the fundamental and third-harmonic pulses can be neglected. In other words, we shall examine the case of combined propagation of the pump and third-harmonic pulses. In this approximation we have $\eta_p^n=\eta_h^m=\eta$, and Eq. (14) reduces to

$$B^m(\eta,z)=i\tilde{\beta}^{mn}A_0^3\exp\left\{iz\left[\left(\tilde{\gamma}_2^{mn}+\frac{3}{2}\tilde{\gamma}_1^n\right)|A_0^n(\eta)|^2-\frac{\Delta k^{mn}}{2}\right]\right\}\frac{\sin(\xi^{mn}z)}{\xi^{mn}}, \qquad (16)$$

where

$$\xi^{mn}=\frac{(3\tilde{\gamma}_1^n-2\tilde{\gamma}_2^{mn})|A_0^n(\eta)|^2-\Delta k^{mn}}{2}. \quad (17)$$

As can be seen from Eqs. (13), (14), and (16), self- and cross-phase modulation give rise to an additional phase shift between the fundamental and third-harmonic pulses, which depends on the pump intensity and can partially compensate the intensity-independent phase mismatch $-\Delta kz$ for a certain part of the pump pulse. The characteristic length of the synchronous interaction between the fundamental and third-harmonic pulses is given by $L_{in}=1/\xi^{mn}$. Thus, when the condition

$$\Delta k^{mn}=(3\tilde{\gamma}_1^n-2\tilde{\gamma}_2^{mn})|A_0^n(\eta)|^2 \quad (18)$$

is satisfied (i.e., when the nonlinear medium, fundamental frequency, and pump intensity are appropriately selected), the combined action of the XPM1 and XPM2 effects can partially or completely compensate the intensity-independent phase mismatch for third-harmonic generation in a certain part of the pump pulse. In particular, if the condition (18) holds near the maximum of the pump pulse, the efficiency of third-harmonic generation is higher in this region than on the leading and tailing edges of the pump pulse, where the phase mismatch remains. Shortening of the third-harmonic pulse can also occur. This phenomenon is illustrated by the results of the numerical calculations presented in Sec. 4.1.

As can be seen from Eq. (14), the phase trajectory and variation of the frequency of the third harmonic depend on the amplitude of the fundamental pulse. Thus, the spectrum of the third harmonic at the exit from the nonlinear medium can be controlled and the conditions for subsequent pulse compression can be optimized by varying the amplitude of the fundamental pulse. Physically, such cross-phase control of the chirp of optical harmonics is possible because the harmonics are generated in the pump field, which experiences self-phase modulation, and the phase of the third harmonic is modulated by the correction to the refractive index of the medium induced by the fundamental pulse.

To conclude this section we present an estimate for the efficiency of third-harmonic generation in a hollow waveguide when the group delay of the fundamental pulse and the third harmonic can be neglected for the experimental parameters realized in Ref. 30. If we use the estimate for the third-order nonlinear-optical susceptibility per krypton atom given in Ref. 47, $\chi^{(3)}(3\omega;\omega,\omega,\omega)\approx 3.2\times 10^{-37}$ esu, assuming that the phase mismatches for the TEM_{11} mode of the fundamental light and the TEM_{13} mode of the third harmonic are completely compensated, and setting the gas pressure equal to $p=61$ Torr, the energy of the fundamental light injected into the waveguide equal to 145 μJ, the internal diameter of the hollow waveguide equal to 153 μm, and the length of the waveguide equal to 30 cm, then we find that the intensity of the third harmonic estimated from Eq. (16) for these conditions is roughly 5×10^{18} erg/cm$^2\cdot$s. For the TEM_{13} mode of the third harmonic with an effective area of 1.4×10^{-5} cm^2 this energy corresponds to an efficiency of the conversion of the energy of the fundamental light into the third harmonic on the 0.1% level, which is consistent in order of magnitude with the results of experimental measurements (0.2%) obtained in Ref. 30.

3. NUMERICAL SIMULATION

We studied the process of third-harmonic generation in a gas-filled hollow optical waveguide, treating the effects of self- and cross-phase modulation,via numerical calculations using Eqs. (13), (14), and (16). It was assumed that the pump pulse has an envelope of Gaussian shape,

$$A_0^1(\eta_p)=A_0(\eta_p)=\exp\left(-\frac{\eta_p^2}{2}\right), \quad (19)$$

where $\eta_p\equiv\eta_p^1$, and a transverse intensity distribution which corresponds to the TEM_{11} mode of the hollow waveguide and can be described by Eq. (5). Third-harmonic generation with a transverse intensity distribution corresponding to the lowest TEM_{11}, TEM_{12}, and TEM_{13} modes of a hollow waveguide with the eigenvalues $u_l^m\approx 2.4$, 5.5, and 8.7, respectively, was considered. As will be shown below, the bulk of the energy of the third harmonic generated in the case considered is contained in the three lowest modes, so that the effects associated with the generation of higher modes can be neglected.

The calculations were performed for an argon-filled hollow waveguide with dielectric walls. The internal radius a was set equal to 80 μm, and the wavelength of the pump radiation was 780 nm. The waveguide length was set equal to $L=40$ cm. The power attenuation constants estimated according to Ref. 46 for a pump pulse of wavelength 780 nm and the third harmonic for the modes considered are less than 7×10^{-3} cm^{-1}, permitting neglect of the absorption effects in the hollow waveguide for the wavelengths considered.

The coefficient γ_1^n, which is responsible for self-phase modulation of the pump pulse, was calculated using the experimental data for argon in Ref. 52. Far from the resonances of the third-order nonlinear-optical susceptibilities appearing in (10)–(12), γ_2^{mn} can be set equal to γ_1^n to within a multiplier specified by the mode structure of the pump radiation and the third harmonic in the waveguide:

$$\gamma_2^{mn}=\gamma_1^n\frac{\int\int(f^p(\boldsymbol{\rho}))^2(f^h(\boldsymbol{\rho}))^2\rho\,d\rho\,d\theta}{\int\int(f^p(\boldsymbol{\rho}))^4\rho\,d\rho\,d\theta}. \quad (20)$$

In estimating the minimum pulse duration to which the phase-modulated third-harmonic pulse can be compressed it was assumed that chirp compensation is described by[53]

$$B^m(L_c,\eta)=\frac{1}{\sqrt{-2\pi i\Gamma^m L_c}}\int_{-\infty}^{\infty}B^m(\eta',L)\times\exp\left(-\frac{i}{2\Gamma^m L_c}(\eta'-\eta)^2\right)d\eta', \quad (21)$$

where L_c is the interaction length with a dispersive medium which provides for chirp compensation, and

$$\Gamma^m=\frac{1}{\tau^2}\left(\frac{\partial^2 K_h^m}{\partial\omega'^2}\right)_{\omega'=3\omega}$$

is a parameter which characterizes the dispersion of the group velocity in that medium.

The second-order dispersion Γ^m and the interaction length L_c were chosen so as to ensure maximum compression of the phase-modulated third-harmonic pulse.

4. CALCULATION RESULTS AND DISCUSSION

In this section we discuss the results of the numerical calculations performed for the process of third-harmonic generation in an argon-filled hollow optical waveguide under the conditions specified above. Depending on the relationship between the group-delay length of the fundamental and third-harmonic pulses l_w^{mn} and the length of the hollow optical waveguide L, third-harmonic generation can take place in a regime in which the effects associated with group delay are insignificant or in a regime in which the difference between the group velocities of the pump and third-harmonic pulses has a significant influence on the efficiency of the process, as well as on the duration and spectrum of the third harmonic. Let us evaluate the conditions under which the effects associated with group delay of the pump radiation and the third harmonic in a hollow waveguide can be neglected. When the argon pressure is $p=0.5$ atm, the group-delay length l_w^{mn} of the pulses of the fundamental light with wavelength 0.78 μm and of the third harmonic becomes of order the length $L=40$ cm of the waveguide under consideration for a pump-pulse duration $\tau\sim 50$ fs. Thus, the group delay of the pump and third-harmonic pulses does not significantly affect the process of third-harmonic generation under these conditions for pump-pulse durations exceeding 50 fs.

The case of weak group delay is of special interest in connection with ensuring highly efficient third-harmonic generation. We begin our treatment with this case and then move on to explore the influence of group delay on the parameters of the third-harmonic pulse.

4.1. Cross-phase modulation during third-harmonic generation in the absence of group delay

Let the total energy of the third-harmonic pulse at the waveguide exit be

$$W_h=\sum_m W_h^m,$$

where

$$W_h^m=\int P_h^m(\eta)\,d\eta$$

is the energy of the third-harmonic pulse in the TEM_{1m} eigenmode of the waveguide and

$$P_h^m(\eta)=|B^m(\eta,L)|^2\int\int |f_h^m(\rho)|^2\rho\,d\varphi\,d\rho$$

is the power of the third-harmonic pulse corresponding to the TEM_{1m} eigenmode of the waveguide. Figure 1 presents plots of the total energy of the third-harmonic pulse W_h at the exit from a hollow waveguide as a function of argon pressure in the absence of spatial separation of the pump and third-harmonic pulses, which were calculated from Eq. (16) for various values of the peak pump power P_{p0}. The pressure dependence of the total energy of the third-harmonic pulse oscillates because the phase mismatch varies as the pressure is increased. The energy of the third-harmonic pulse reaches a constant level because at sufficiently high pressures the increase in the nonlinear susceptibility responsible for generation of the third harmonic in proportion to the pressure is compensated by worsening of the phase-matching conditions [$\Delta k^{mn}\propto p$, so that at sufficiently high pressures the envelope of the function $\sin(\xi^{mn}L)/\xi^{mn}$ falls off as $(\xi^{mn})^{-1}\propto 1/p$]. The oscillation amplitude decreases with increasing pressure, because, as follows from Eq. (17), an increase in pressure increases in the difference in phase mismatch for the process of third-harmonic generation at the maximum and on the tails of the pump pulse. As a result, the integration over time in the expression for W_h gives less pronounced oscillations than in the case of low pressures.

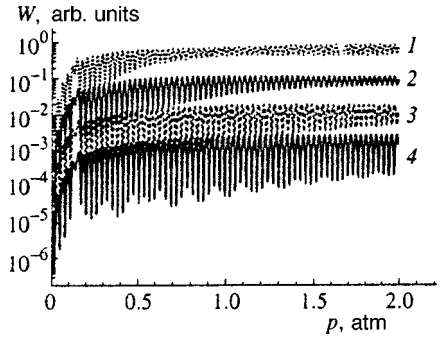

FIG. 1. Total energy of the third-harmonic pulse at the exit from a hollow waveguide W_h versus argon pressure in the absence of group delay for a waveguide length $L=40$ cm and various peak pump powers: 4 (*1*), 2 (*2*), 1 (*3*), and 0.5 GW (*4*).

As can be seen from Fig. 2, a large part of the energy of the third harmonic belongs to the lowest TEM_{11} mode of the hollow waveguide. The TEM_{12} mode also contains significant energy. The energy belonging to the TEM_{13} mode is comparable to the energy contained in the two lower modes only in a narrow range of argon pressures ($p\sim 0.14$ atm),

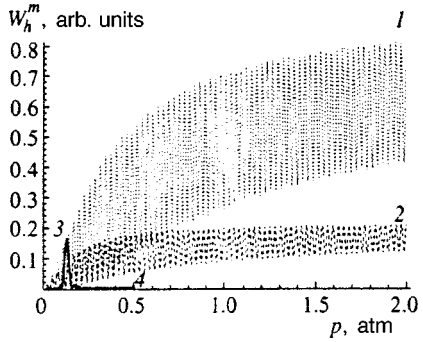

FIG. 2. Energy of the third-harmonic pulse at the exit from a hollow waveguide W_h^m versus argon pressure for the TEM_{11} (*1*), TEM_{12} (*2*), TEM_{13} (*3*), and TEM_{14} (*4*) modes in the absence of group delay for $L=40$ cm and $P_{p0}=4$ GW.

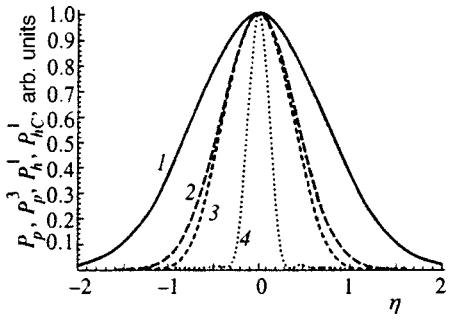

FIG. 3. Time dependence of the energy of the normalized pump-pulse power P_p (*1*) and P_p^3 (*2*) and the normalized power of the TEM_{11} mode of the third-harmonic pulse at the exit from a hollow waveguide P_h^1 (*3*) and the normalized power after compensation of the linear chirp P_{hC}^1 (*4*) in the traveling coordinate frame η for $p=0.536$ atm and $P_{p0}=2$ GW.

where the phase-matching conditions are satisfied for the TEM_{13} mode. As can be seen from Fig. 2, the energy of the higher modes is negligibly small compared to the energy of the first three modes in the pressure range considered.

Figure 3 presents the envelopes of the power (*1*) and the cubed power (*2*) of the pump pulse, as well as the envelopes of the power of the lowest mode of the third-harmonic pulse at the exit from the hollow waveguide (*3*) and after compensation of the linear chirp (*4*), calculated from Eqs. (16) and (21) for $p=0.536$ atm and a peak pump power $P_{p0}=2$ GW. Figure 4 shows the absolute value of the spectrum

$$S_h(\Omega)=\left|\frac{1}{2\pi}\int B^1(\eta,L)\exp(i\Omega\eta)\,d\eta\right|$$

of the third-harmonic pulse for the TEM_{11} mode (*1*) and of the pump pulse (*2*) at the waveguide exit, as well as the absolute value of the spectrum of the pump pulse at the waveguide entrance (*3*) for $p=0.536$ atm and $P_{p0}=2$ GW. The argon pressure was chosen so as to satisfy the conditions for a local phase-matching optimum for the TEM_{11} mode of the third harmonic [$\xi^{11}=(N+0.5)\pi$, where N is an integer] at a fairly high pressure. As can be seen from Fig. 2, under these conditions a large part of the energy of the third harmonic is contained in the TEM_{11} mode. The results presented in Figs. 3 and 4 provide evidence that the third-harmonic pulse is phase-modulated due to cross-phase modulation effects and that its spectrum is broadened significantly even in comparison to the pump pulse, which experiences the influence of self-phase modulation. For this reason, the third-harmonic pulse can be effectively compressed by compensating the linear chirp.

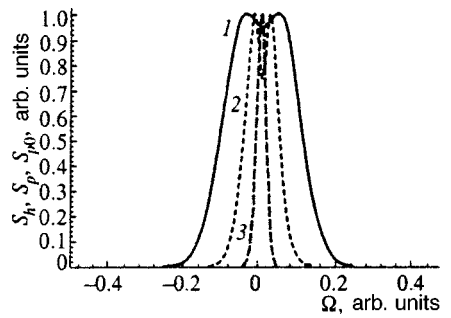

FIG. 4. Absolute value of the spectrum of the amplitude of the TEM_{11} mode of the third-harmonic pulse $S_h=|(1/2\pi)\int B^1(\eta,L)\exp(i\Omega\eta)\,d\eta|$ (*1*) and the pump pulse at the exit from the waveguide $S_p=|(1/2\pi)\int A^1(\eta,L)\times\exp(i\Omega\eta)\,d\eta|$ (*2*) and at the entrance to the waveguide $S_{p0}=|(1/2\pi)\int A^1(\eta,0)\exp(i\Omega\eta)\,d\eta|$ (*3*) for $p=0.536$ atm and $P_{p0}=2$ GW.

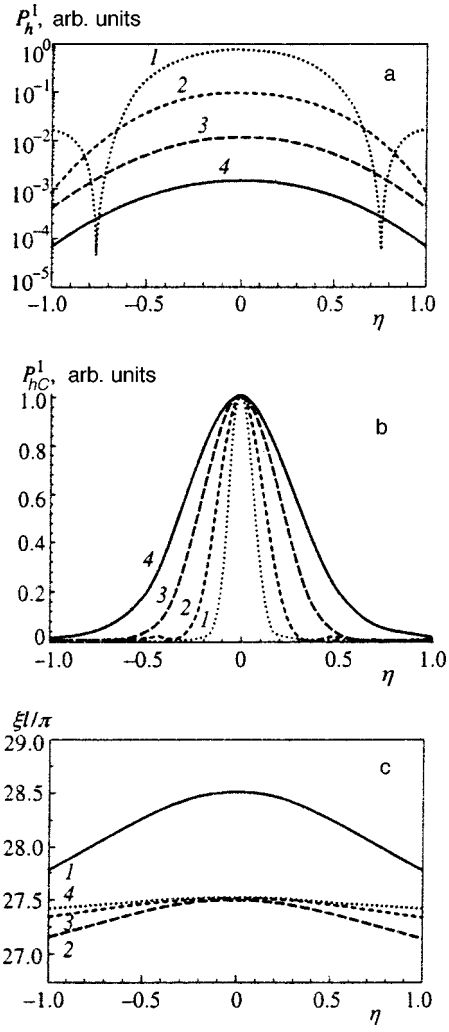

FIG. 5. Time dependence of the power of the TEM_{11} mode of the third-harmonic pulse at the exit from the waveguide P_h^1 (a) and after compensation of the linear chirp P_{hC}^1 (b) and of $\xi L/\pi$ (c) in the traveling coordinate frame η for $P_{p0}=4$ GW and $p=0.549$ atm (*1*), $P_{p0}=2$ GW and $p=0.536$ atm (*2*), $P_{p0}=1$ GW and $p=0.545$ atm (*3*), and $P_{p0}=0.5$ GW and $p=0.549$ atm (*4*).

Figure 5, which shows the power envelopes of the third-harmonic pulse corresponding to the TEM_{11} waveguide mode at the waveguide exit (Fig. 5a) and after compensation of the linear chirp (Fig. 5b), as well as a plot of $\xi^{11}(\eta)$ (Fig. 5c), illustrates the sharpening of the third-harmonic pulse due to the intensity-dependent correction to the phase mismatch [see Eqs. (16) and (17)] when the gas pressure and the pump power are selected to achieve the best phase-matching conditions near the maximum of the pump pulse. It can be seen from a comparison of curves *1* in Figs. 5a and 5c that the third-harmonic signal corresponding to the TEM_{11} waveguide mode vanishes at the waveguide exit for values of η such that $\xi^{11}=N\pi$, where N is an integer.

As can be seen from Figs. 5a and 5b, the chirp of the

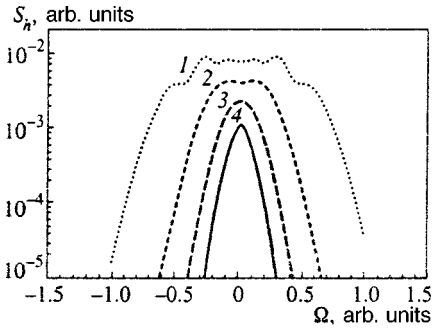

FIG. 6. Absolute value of the spectrum of the amplitude of the TEM_{11} mode of the third-harmonic pulse at the exit from the waveguide S_h for $P_{p0}=4$ GW and $p=0.549$ atm (*1*), $P_{p0}=2$ GW and $p=0.536$ atm (*2*), $P_{p0}=1$ GW and $p=0.545$ atm (*3*), and $P_{p0}=0.5$ GW and $p=0.549$ atm (*4*).

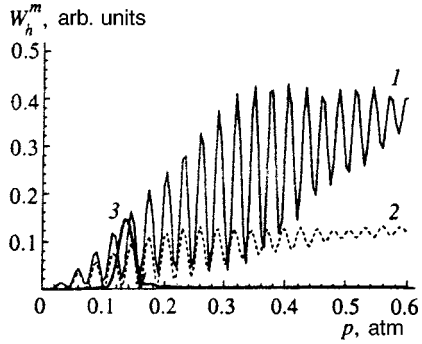

FIG. 8. Energy of the third-harmonic pulse for the TEM_{11} (*1*), TEM_{12} (*2*), and TEM_{13} (*3*) waveguide modes at the exit from a hollow waveguide versus argon pressure for $L=40$ cm, $P_{p0}=4$ GW, and $2\tau=25$ fs.

third-harmonic pulse (Fig. 5a) and its spectral width (Fig. 6) increase with increasing pump power, making it possible to obtain fairly short pulses at the output of the compressor. Note, however, that increasing the pump power can distort the pulseform of the third harmonic.

Thus, the results of these calculations show that the phase cross-modulation accompanying third-harmonic generation makes it possible to control the chirp of the third harmonic by varying the amplitude of the pump pulse and the parameters (pressure and dispersion) of the gas filling the hollow waveguide and thereby permitting the formation of pulses of tripled frequency and regulated duration.

4.2. Cross-phase modulation during third-harmonic generation with group delay of the pump and third-harmonic pulses

When dispersive spreading of the pulses can be neglected (the dispersion length l_d for the parameters presented above significantly exceeds the length of the hollow optical waveguide), but the difference between the group velocities at the frequencies of the pump radiation and the third harmonic must be taken into account, third-harmonic generation in a hollow waveguide can be analyzed with allowance for the effects of self- and cross-phase modulation using Eq. (14). Figure 7, showing how the total energy of the third-harmonic pulse W_h at the exit from the hollow waveguide

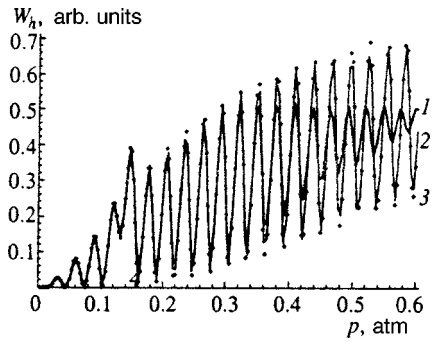

FIG. 7. Normalized total energy of the third-harmonic pulse at the exit from a hollow waveguide versus argon pressure for two values of the pump-pulse duration $2\tau=25$ fs (*1*) and 200 fs (*2*) and with neglect of the group delay of the pump pulse and the third harmonic (*3*) for $L=40$ cm and $P_{p0}=4$ GW.

varies with argon pressure, was obtained from Eq. (14) with allowance for the group delay of the pump and third-harmonic pulses for various values of the pump-pulse duration and a peak pump power $P_{p0}=4$ GW. In analogy to the case of small group delay of the pump and third-harmonic pulses considered in Sec. 4.1, the total energy of the third-harmonic pulse oscillates as a function of gas pressure due to the variation of the phase mismatch with increasing pressure. As in the case of small group delay, the energy of the third-harmonic pulse reaches a constant energy because at sufficiently high pressures the increase in the nonlinear susceptibility responsible for third-harmonic generation is compensated by the worsening of the phase-matching conditions.

The dependence of the energy contained in the transverse TEM_{11} (*1*), TEM_{12} (*2*), and TEM_{13} (*3*) modes of the third-harmonic pulse at the exit from the hollow waveguide on argon pressure calculated with consideration of the spatial separation of the pump and third-harmonic pulses from Eq. (14) for $L=40$ cm, $P_{p0}=4$ GW, and $\tau=25$ fs is presented in Fig. 8. In analogy to the case of small group delay, a large part of the energy of the third harmonic belongs to the lowest TEM_{11} mode of the hollow waveguide, a smaller part of the energy is contained in the TEM_{12} mode, and the energy of the TEM_{13} mode is comparable to the energy contained in the two lowest modes only in a narrow range of argon pressures ($p\sim 0.14$ atm), where the phase-matching conditions for the TEM_{13} mode are satisfied.

As can be seen from the envelopes of the power of the pump pulse and the power of the lowest mode of the third-harmonic pulse at the waveguide exit in Fig. 9a, the group delay of the pump and third-harmonic pulses at first causes an increase in the duration of the third-harmonic pulse (curve *2* in Fig. 9a, the pump and third-harmonic pulses spatially separate by a distance $\eta_h\equiv\eta_h^1=0.8$ over the length of the waveguide) and then (as the pump pulse duration further decreases) the third harmonic splits into two pulses (curve *1* in Fig. 9a, and the pump and third-harmonic pulses separate by a distance $\eta_h=3$ over the length of the waveguide). The right-hand pulse in Fig. 9a forms as a result of third-harmonic generation over a distance of order the coherence length $l_{\rm ph}^{11}=1/\Delta k^{11}$ near the entrance endplate of the waveguide, while the left-hand pulse corresponds to third-

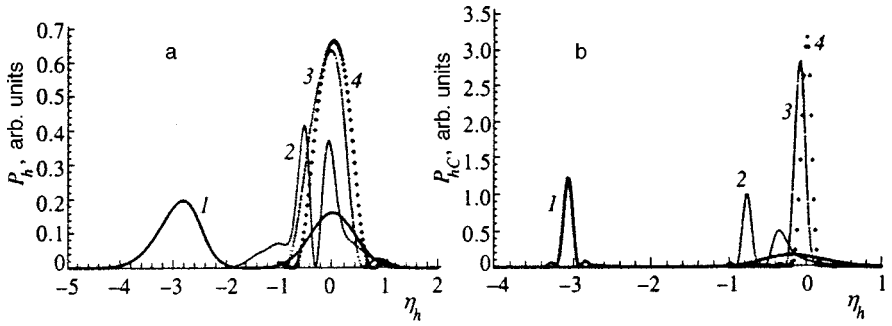

FIG. 9. Time dependence of the power of the lowest mode of the third-harmonic pulse at the exit from the waveguide $P_h \equiv P_h^1$ (a) and after compensation of the linear chirp (b) in the traveling coordinate frame η_h for $p=0.52$ atm, $P_{p0}=4$ GW, and durations of the pump pulse $2\tau=25$ fs (*1*), 100 fs (*2*), and 800 fs (*3*), as well as with neglect of the effects of group delay (*4*).

harmonic generation in the analogous region near the exit endplate. The power of the third-harmonic pulse generated in the central part of the waveguide is negligibly small compared to the third-harmonic signal formed near the waveguide endplates. This phenomenon can easily be explained using Eq. (14). In fact, assuming that $l_w \ll L$, letting the upper integration limit in (14) tend to infinity, and noting that at low pump powers the second and third terms in the argument of the exponential function are small compared to the first, we find that the power of the third-harmonic signal is specified by the Fourier transform of the pump-pulse amplitude in the Δk representation:

$$B^m(\eta_h^m, z) \approx i \frac{\tilde{\beta}^m}{\zeta^{mn}} \int_0^{z/\zeta^{mn}} d\eta'' \, (A_0^n(\eta_h + \eta''))^3 \times \exp\left[-i \frac{\Delta k^{mn}}{\zeta^{mn}} \eta'' \right]. \quad (22)$$

Physically, the low efficiency of third-harmonic generation in the central portion of the waveguide is associated with the significant phase mismatch between the pump radiation and the third harmonic ($\Delta k^{11}/\zeta^{11} \sim 100$ for pulses of 25 fs duration at an argon pressure of 0.5 atm).

It is noteworthy that the third-harmonic pulse appearing near the exit endplate of the waveguide is phase-modulated far more strongly and, accordingly, can be compressed with a far higher efficiency (Fig. 9b) than the third-harmonic pulse formed near the entrance endplate of the waveguide. This is because under conditions for which the pump and third-harmonic pulses separate by a fairly large distance over the length of the waveguide (as occurs in Fig. 10), the XPM2 effect in the noninertial Kerr medium has a fairly weak influence on the phase of the third harmonic. As a consequence of self-phase modulation, the pump pulse is phase-modulated far more strongly near the exit endplate of the waveguide than near the entrance endplate. For this reason, the third-harmonic pulse formed at the exit endplate of the waveguide is characterized by significant phase modulation and can be effectively compressed by chirp compensation (Figs. 9b and 10b). The corresponding spectra of the pump pulse and the third harmonic are shown in Fig. 11. As in the case of small group delay of the pump and third-harmonic pulses, an increase in the pump power broadens the spectrum of the third harmonic (Fig. 12).

Thus, the picture of the nonlinear-optical interaction of short laser pulses in a gas-filled hollow waveguide is considerably more complicated with a significant group delay of the pump and third-harmonic pulses than in the absence of group delay of the pulses. However, under these conditions, too, the cross-phase modulation makes it possible to control the parameters of the third-harmonic pulses (Figs. 10 and 12) and permits the formation of ultrashort pulses of radiation in the ultraviolet range.

5. CONCLUSION

An analysis of the solutions obtained for coupled equations describing the slowly varying amplitudes of the pump pulse and third harmonic with allowance for self- and cross-phase modulation and first-order dispersion effects has shown that self- and cross-phase modulation can be important factors, which have a significant influence on the efficiency of frequency conversion in third-harmonic generation in hollow waveguides. Unlike the plane-wave approximation, these equations include the influence of the waveguide through the propagation constants, the group velocities of the pump radiation and the third harmonic, and the nonlinear coefficients written with consideration of the transverse distributions of the pump and third-harmonic fields for the corresponding waveguide modes. It has been shown the phase trajectory of the third harmonic can partially compensate the

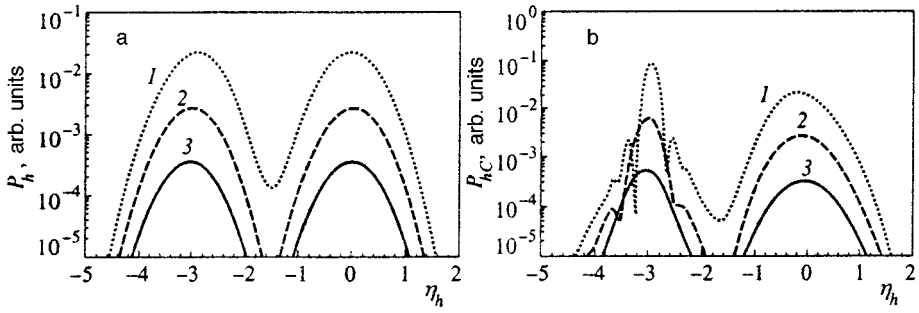

FIG. 10. Time dependence of the power of the TEM_{11} mode of the third-harmonic pulse at the exit from the waveguide $P_h \equiv P_h^1$ (a) and after compensation of the linear chirp P_{hC} (b) in the traveling coordinate frame η_h for $P_{p0}=2$ GW and $p=0.506$ atm (*1*), $P_{p0}=1$ GW and $p=0.514$ atm (*2*), and $P_{p0}=0.5$ GW and $p=0.518$ atm (*3*) with $2\tau=25$ fs.

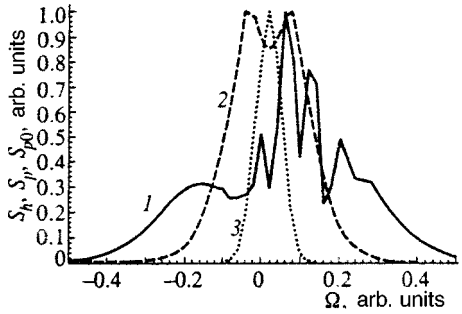

FIG. 11. Absolute value of the spectrum of the amplitude of the TEM_{11} mode of the third-harmonic pulse S_h (*1*), of the pump pulse at the exit from the waveguide S_p (*2*), and of the pump pulse at the entrance to the waveguide S_{p0} (*3*) for $p=0.506$ atm, $P_{p0}=2$ GW, and $2\tau=50$ fs.

constant phase mismatch between the third-harmonic pulse and the fundamental pulse as a result of cross-phase modulation in a certain part of the pulse. The influence of phase mismatch on the efficiency of third-harmonic generation can be minimized by selecting media with favorable dispersion properties of the refractive index and fiber modes, as well as the frequency dependences of the corresponding nonlinear-optical susceptibilities.

This study has shown that the chirp of optical harmonics at the exit from a nonlinear medium can be controlled by varying the amplitude of the pump pulse. The physical mechanism for controlling the spectrum and phase of optical harmonics is associated with the fact that the harmonics are generated under the conditions of self-phase modulation of the fundamental pulse and the phase of the third-harmonic pulse is modulated by a correction to the refractive index of the medium induced by the fundamental pulse. Such a method of controlling phase modulation seems promising for obtaining ultrashort pulses in the vacuum-ultraviolet and x-ray ranges by high-order harmonic generation and parametric wave mixing in hollow optical waveguides and a laser plasma.

This research was carried out with support from the Russian Fund for Fundamental Research (Project No. 97-02-17351) and the U.S. Civilian Research and Development Foundation for the Independent States of the Former Soviet Union (CRDF) (Award No. RP1-255).

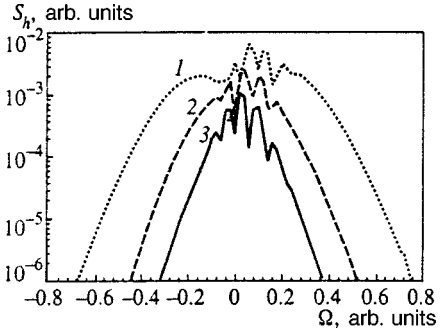

FIG. 12. Absolute value of the spectrum of the amplitude of the TEM_{11} mode of the third-harmonic pulse at the exit from the waveguide for $2\tau=25$ fs, $P_{p0}=2$ GW, and $p=0.506$ atm (*1*), $P_{p0}=1$ GW and $p=0.514$ atm (*2*), and $P_{p0}=0.5$ GW and $p=0.518$ atm (*3*).

*)E-mail: zhelt@ilc.phys.msu.su
†)Deceased

[1] J. Reintjes, C.-Y. She, R. C. Eckardt, IEEE J. Quantum Electron. **QE-14**, 581 (1978).
[2] J. Reintjes, Appl. Opt. **19**, 3889 (1980).
[3] J. Bokor, P. H. Bucksbaum, R. R. Freeman, Opt. Lett. **8**, 217 (1983).
[4] A. McPherson, G. Gibson, H. Jara *et al.*, J. Opt. Soc. Am. B **4**, 595 (1987).
[5] M. Ferray, A. L'Huillier, X. F. Li *et al.*, J. Phys. B **21**, L31 (1988).
[6] X. F. Li, A. L. Huillier, M. Ferray *et al.*, Phys. Rev. A **39**, 5751 (1989).
[7] P. Balcou, A. S. L. Gomes, C. Coruaggia *et al.*, J. Phys. B **25**, 4467 (1992).
[8] N. Sarukura, K. Hata, T. Adachi *et al.*, Phys. Rev. A **43**, 1669 (1991).
[9] A. L'Huillier and P. Balcou, Phys. Rev. Lett. **70**, 774 (1993).
[10] J. J. Macklin, J. D. Kmetec, and C. L. Gordon III, Phys. Rev. Lett. **70**, 766 (1993).
[11] J. K. Crane, M. D. Perry, S. Herman, and R. W. Falcone, Opt. Lett. **17**, 1256 (1992).
[12] K. Kondo, N. Sarukura, K. Sajiki, and S. Watanabe, Phys. Rev. A **47**, 2480 (1993).
[13] C.-G. Wahlstrom, J. Larsson, A. Persson *et al.*, Phys. Rev. A **48**, 4709 (1993).
[14] D. M. Mittleman, D. C. Douglass, Z. Henis *et al.*, J. Opt. Soc. Am. B **13**, 170 (1996).
[15] C. Altucci, T. Starczewski, E. Mevel *et al.*, J. Opt. Soc. Am. B **13**, 148 (1996).
[16] C. Spielmann, N. H. Burnett, S. Sartania *et al.*, Science **278**, 661 (1997).
[17] C. Spielmann, M. Schnuerer, N. H. Burnett *et al.*, in *Proceedings of the Conference on Lasers and Electro-Optics (CLEO'98), San Francisco, CA, 1998, (OSA Technical Digest Series, Vol. 6)*, Optical Society of America, Washington, DC (1998), p. 81.
[18] S. M. Gladkov, A. M. Zheltikov, N. I. Koroteev, and A. B. Fedotov, Pis'ma Zh. Tekh. Fiz. **14**, 1399 (1988) [Sov. Tech. Phys. Lett. **14**, 610 (1988)].
[19] A. B. Fedotov, S. M. Gladkov, N. I. Koroteev, and A. M. Zheltikov, J. Opt. Soc. Am. B **8**, 373 (1991).
[20] A. M. Zheltikov, N. I. Koroteev, and A. B. Fedotov, Opt. Spektrosk **72**, 971 (1992) [Opt. Spectrosc. **72**, 527 (1992)].
[21] S. Kubodera, Y. Nagata, Y. Akiyama *et al.*, Phys. Rev. A **48**, 4576 (1993).
[22] A. M. Zheltikov, N. I. Koroteev, and A. B. Fedotov, Proc. SPIE **2097**, 308 (1993).
[23] A. M. Zheltikov, N. I. Koroteev, and A. B. Fedotov, Laser Phys. **4**, 569 (1994).
[24] A. B. Fedotov, N. I. Koroteev, and A. M. Zheltikov, Laser Phys. **5**, 835 (1995).
[25] W. Theobald, C. Wulker, F. P. Schafer, and B. N. Chichkov, Opt. Commun. **120**, 177 (1995).
[26] D. von der Linde, T. Engers, G. Jenke *et al.*, Phys. Rev. A **52**, R25 (1995).
[27] R. B. Miles, G. Laufer, and G. C. Bjorklund, Appl. Phys. Lett. **30**, 417 (1977).
[28] M. Nisoli, S. De Silvestri, and O. Svelto, Appl. Phys. Lett. **68**, 2793 (1996); M. Nisoli, S. De Silvestri, O. Svelto *et al.*, Opt. Lett. **22**, 522 (1997).
[29] F. Giammanco, in *Second International Symposium on Modern Problems of Laser Physics, Akademgorodok, Novosibirsk, 1997*, Institute of Laser Physics, Russian Academy of Sciences, Akademgorodok, Novosibirsk (1997).
[30] C. G. Durfee III, S. Backus, M. M. Murnane, and H. C. Kapteyn, Opt. Lett. **22**, 1565 (1997).
[31] N. Tamaki, O. Maya, K. Midorikawa, and M. Obara, in *Proceedings of the Conference on Lasers and Electro-Optics (CLEO'98), San Francisco, CA, 1998, (OSA Technical Digest Series, Vol. 6)*, Optical Society of America, Washington, DC (1998), p. 83.
[32] Z. Chang, A. Rundquist, H. Wang *et al.*, Phys. Rev. Lett. **79**, 2967 (1997).
[33] C. G. Durfee III, A. Rundquist, Z. Chang *et al.*, in *International Quantum Electronic Conference (IQEC'98), San Francisco, CA, 1998, Postdeadline Papers*, Optical Society of America, Washington, DC (1998), QPD5.
[34] J. Gersten, R. Alfano, and M. Belic, Phys. Rev. A **21**, 1222 (1980).
[35] A. R. Chraplyvy, D. Marcuse, and P. S. Henry, IEEE J. Lightwave Technol. **LT-2**, 6 (1984).
[36] R. Alfano, T. Jimbo, J. Manassah, and P. Ho, Opt. Lett. **11**, 626 (1986).

[37] R. Alfano, Q. Wang, T. Jimbo, and P. Ho, Phys. Rev. A **35**, 459 (1987).
[38] R. Alfano, P. Baldeck, F. Raccah, and P. Ho, Appl. Opt. **26**, 3491 (1987).
[39] G. P. Agrawal, *Nonlinear Fiber Optics*, Academic Press, Boston (1989).
[40] C. W. Siders, N. C. Turner III, M. C. Downer *et al.*, J. Opt. Soc. Am. B **13**, 330 (1996).
[41] A. B. Fedotov, N. I. Koroteev, M. M. T. Loy *et al.*, Opt. Commun. **133**, 587 (1997).
[42] G. Tempea and T. Brabec, Opt. Lett. **23**, 1286 (1998).
[43] Y. R. Shen, *The Principles of Nonlinear Optics*, Wiley, New York (1984).
[44] N. I. Koroteev and A. M. Zheltikov, Laser Phys. **8**, 512 (1998).
[45] N. I. Koroteev and A. M. Zheltikov, Appl. Phys. B **67**, 53 (1998).
[46] E. A. J. Marcatili and R. A. Schmeltzer, Bell Syst. Tech. J. **43**, 1783 (1964).
[47] J. F. Reintjes, *Nonlinear Optical Parametric Processes in Liquids and Gases*, Academic Press, Orlando (1984).
[48] P. P. Ho, D. Ji, Q. Z. Wang, and R. R. Alfano, J. Opt. Soc. Am. B **7**, 276 (1990).
[49] E. M. Dianov, P. V. Mamyshev, A. M. Prokhorov, and S. V. Chernikov, Kvantovaya Elektron. **15**, 1941 (1988) [Sov. J. Quantum Electron. **18**, 1211 (1988)].
[50] G. P. Agrawal, P. L. Baldeck, and R. R. Alfano, Phys. Rev. A **40**, 5063 (1989).
[51] R. F. de Souza, E. J. S. Fonseca, J. Miguel Hickmann, and A. S. Gouveia-Neto, Opt. Commun. **124**, 79 (1996).
[52] *Landolt Bornstein Physikalisch-Chemische Tabellen*, W. A. Roth and K. Scheel (Eds.), Verlag von Julius Springer, Berlin, Vol. 2 (1931), Vol. 3 (1935).
[53] S. A. Akhmanov, V. A. Vysloukh, and A. S. Chirkin, *Optics of Femtosecond Laser Pulses*, AIP, New York (1992).

Translated by P. Shelnitz

Transient NQR signals in the effective field of multipulse trains

D. Ya. Osokin

Kazan Physicotechnical Institute, Kazan Science Center of the Russian Academy of Sciences, 420029 Kazan, Russia

(Submitted 4 September 1998)
Zh. Éksp. Teor. Fiz. **115**, 1580–1592 (May 1999)

This paper is a theoretical analysis of time-dependent nutation and echo signals in the effective field of multipulse trains in NQR. The results of experimental investigations of the same aspects were reported in earlier papers. The various features of dipole–dipole interactions in three-level quadrupole spin systems are discussed. It is shown that, in contrast to NMR, the dipole Hamiltonian in the interaction representation determined by the quadrupole Hamiltonian contains only three diagonal components of the dipole–dipole interaction tensor. On the other hand, the strong inhomogeneous broadening characteristic of NQR hinders exact measurement of these components by ordinary methods. The theoretical analysis suggests that the decay of the echo-signal envelope in the effective field of multipulse trains is determined solely by the dipole relaxation time, which serves as justification of a new experimental method used in measuring the characteristics of the dipole–dipole interaction tensor in spin systems with an inhomogeneously broadened spectrum. © *1999 American Institute of Physics.* [S1063-7761(99)00405-9]

1. INTRODUCTION

Multipulse trains were introduced into the realm of nuclear magnetic resonance studies by J. Waugh and collaborators[1,2] in order to average dipole–dipole interactions, which made it possible to observe chemical shifts in solids. In NQR, the use of multipulse trains was also initially introduced for practical reasons, since such trains made it possible to dramatically increase the number of events per unit time, which is especially important in the search for weak NQR signals (especially NQR in ^{14}N). However, the successful use in NQR of multipulse trains in studies of polycrystalline samples, where the principal axes of the tensor of the electric-field gradient are oriented at random in relation to the vector of the radio-frequency (rf) field so that there can be no 90° and 180° pulses in the strict sense of the word, required reviewing several aspects of the theoretical explanation of multipulse regimes. The strict requirement (obligatory in NMR) that the mean dipole–dipole interaction Hamiltonian be zero, which is impossible to meet in NQR for polycrystalline samples, was replaced by a less stringent requirement that the initial density matrix commute with the mean Hamiltonian, provided that the initial density matrix contains the transverse magnetization operator.[3]

Because of the practical problems mentioned above, the heightened interest of researchers in multipulse spin-locking in NQR caused the part of the initial density matrix that does not commute with the mean Hamiltonian (is orthogonal to it) to be neglected. It is obvious, however, that the evolution of that term may be of interest from the practical viewpoint as well as from the theoretical. Previous papers[4–6] reported the results of experimental studies of transient NQR signals in ^{14}N in the effective field of multipulse trains, i.e., signals that appear immediately after the spin system has been subjected to a multipulse train or the structure or parameters of the train have changed. In these papers several new results are listed: echo in the effective field (echo on the echo-signal envelope) that arises after the phase in the multipulse train has been inverted or an additional pulse has been applied, and a time-dependent nutation signal in the effective field that has a pronounced oscillatory nature and coincides in shape with the NMR induction signal in a solid (the ordinary induction signal in NQR has no oscillations and resembles a Gaussian curve).

In Ref. 6 it was noted that the experimental results could be interpreted as stemming from time-dependent nutation in the effective field of multipulse trains under conditions of stroboscopic observation, which means that the information is gathered once per cycle or ''supercycle'' of the train. In the same paper it was assumed that one application of the method could be the study of dipole–dipole interactions in a system with an inhomogeneously broadened spectrum, which is characteristic of NQR.

Goldman[7] calls such a situation ''dirty'' and advises avoiding it. However, much research[8,9] has gone into the problem of a spin temperature setting in systems with large inhomogeneous broadening (primarily in EPR). Moreover, inhomogeneous broadening was taken into account unconditionally in research devoted to calculations of spin echo in EPR (see Refs. 10 and 11). The spin-packet approximation[12] proved to be fairly effective in such calculations. However, all these approaches did not involve transient processes, which precede the setting in of quasistationary states in multipulse regimes.

The decay time of echo signals in the effective field equals several ordinary echo (Hahn echo) decay times in the same samples.[6] This may occur because of partial averaging of the dipole–dipole interactions by a multipulse train or

because of the effect of large inhomogeneous broadening in the decay time of Hahn echo. Hence before we analyze transient process in the effective field of multipulse trains we must examine the features of dipole–dipole interactions in NQR and their effect on formation of ordinary (Hahn) echo.

2. DIPOLE–DIPOLE INTERACTIONS IN A THREE-LEVEL QUADRUPOLE SPIN SYSTEM

The quadrupole Hamiltonian[13]

$$H'_Q = A[3I_z^2 - I^2 + \eta(I_x^2 - I_y^2)/2] \tag{1}$$

for $I=1$ (NQR of ^{14}N, which corresponds to the experiments described in Refs. 3–5) can be represented in terms of single-transition operators:[14]

$$H'_Q = \omega_{pq} S_z^{pq} + (\omega_{qr} - \omega_{rp})(S_z^{qr} - S_z^{rp}). \tag{2}$$

Here A is the nuclear quadrupole coupling constant; η is the asymmetry parameter of the tensor of the electric-field gradient; I_x, I_y, and I_z are the spin-operator projections; $\omega_{pq} = E_p - E_q$, $\omega_{qr} = E_q - E_r$, and $\omega_{rp} = E_r - E_p$ are the transition frequencies, with E_p, E_q, and E_r the energy levels; and I is replaced with S in the notation for single-transition operators used in Ref. 14 so as to emphasize the difference between these operators and true spin operators.

The first feature of NQR we would like to stress and allow for in our analysis is the large inhomogeneous broadening of lines due to the presence of impurities, imperfections in the lattice, etc. Ordinarily, the inhomogeneous width of NQR lines for ^{14}N, denoted by $1/T_2^*$, is about 1 kHz in molecular crystals and large compared to the dipole width $1/T_2$ (of order 100 Hz). At the same time, the width of the spectrum of rf pulses, $1/t_p$, is within 10 kHz, i.e., $T_2 > T_2^* > t_p$. Bearing all this in mind, we can write the Hamiltonian of pairwise interacting spins as a sum of two-spin Hamiltonians:

$$\begin{aligned} H' &= \omega_{pq1} S_{z1}^{pq} \mathbf{1}_2 + \omega_{pq2} \mathbf{1}_1 S_{z2}^{pq} + H'_D \\ &= \omega_{pq}(S_{z1}^{pq}\mathbf{1}_2 + \mathbf{1}_1 S_{z2}^{pq}) + \delta(S_{z1}^{pq}\mathbf{1}_2 - \mathbf{1}_1 S_{z2}^{pq}) + H'_D. \end{aligned} \tag{3}$$

The second term on the right-hand side commutes with all the operators of the pq subspace and does not take part in the evolution process if the rf pulse is applied in this transition, so that it can dropped. Here $\mathbf{1}_1$ and $\mathbf{1}_2$ are the identity matrices for the first and second spins, $\omega_{pq} = (\omega_{pq1} + \omega_{pq2})/2$ is the median quadrupole frequency of these spins for the pq transition, $\delta = (\omega_{pq1} - \omega_{pq2})/2$ is the ''difference'' of these frequencies, and H'_D is the dipole–dipole interaction Hamiltonian.

The dipole Hamiltonian in the interaction representation determined by the quadrupole Hamiltonian can be expressed in terms of single-transition operators:[15]

$$\begin{aligned} H_D = {} & 2\Omega_{pq}(S_x^{pq}S_x^{pq} + S_y^{pq}S_y^{pq}) - \Omega_{pq}(S_x^{qr}S_x^{qr} + S_y^{qr}S_y^{qr} \\ & + S_x^{rp}S_x^{rp} + S_y^{rp}S_y^{rp}) + (\Omega_{qr} - \Omega_{rp})(S_x^{qr}S_x^{qr} + S_y^{qr}S_y^{qr} \\ & - S_x^{rp}S_x^{rp} - S_y^{rp}S_y^{rp}), \end{aligned} \tag{4}$$

where Ω_{pq}, Ω_{qr}, and Ω_{rp} are the diagonal components of the dipole–dipole interaction tensor (note that the sum of these components is zero). Here and below we drop the numbering of the spin operators and use the definition by position, i.e., the operator of the first spin is in the first position and that of the second spin, in the second. While the initial dipole Hamiltonian has the same form in NQR and NMR, the ''truncated'' dipole Hamiltonian in NQR differs significantly from the similar Hamiltonian in NMR. First, the truncated dipole Hamiltonian in NQR has no A-term of the ''dipole alphabet'' because the Z-component of the nuclear spin is not diagonal in the representation where the quadrupole Hamiltonian is diagonal. Second, the three remaining terms describe flip–flop processes, and each term corresponds to an exchange of photons of the three allowed transition frequencies. The first term of the truncated dipole Hamiltonian (4) describes flip–flop processes at a transition frequency equal to the frequency of the external rf field, while the second and third terms represent the sum and difference of similar terms for the other two processes. All commute with each other; the first transforms as a tensor of rank 2, i.e., is rotated by a rf pulse through a double angle, the third as a tensor of rank 1, and the second as a scalar. Thus, in NQR the dipole reservoir splits into two constants of motion. This fact was first noted in Refs. 15 and 17. As a result, NQR provides the possibility of measuring the asymmetry parameter of the dipole–dipole interaction tensor in experiments, although in the laboratory reference frame the dipole–dipole interaction tensor is axisymmetric,[2] in the same way as the tensor of interest in the truncated dipole Hamiltonian in NMR is axisymmetric.

In the interaction representation we can write the quadrupole Hamiltonian as follows:

$$H_Q = \Delta(S_z^{pq}\mathbf{1} + \mathbf{1}S_z^{pq}) + \delta(S_z^{pq}\mathbf{1} - \mathbf{1}S_z^{pq}), \tag{5}$$

where $\Delta = \omega_{pq} - \omega_0$ is the deviation of the median spin-interaction frequency from the frequency of the rf field, which is usually assumed to be the median frequency of the resonance line. The first term on the right-hand side of Eq. (5) commutes with all the terms in H_D and with the second term in H_Q. Thus, it is the ''quadrupole reservoir,'' corresponding to the Zeeman reservoir in NMR. The second term in (5) does not commute with the truncated dipole Hamiltonian, and hence no purely dipole reservoir can be specified in NQR. The difference of quadrupole frequencies of the interacting spins bring us close to such a reservoir.

Another approach to the transition to the interaction representation involving the total quadrupole Hamiltonian of two spins is possible. In this case, the interaction representation lacks ''quadrupole'' terms but the dipole Hamiltonian becomes time dependent:

$$\begin{aligned} H_D = {} & 2\Omega_{pq}[(S_x^{pq}S_x^{pq} + S_y^{pq}S_y^{pq})\cos 2\delta_{pq}t + (S_x^{pq}S_y^{pq} - S_y^{pq}S_x^{pq}) \\ & \times \sin 2\delta_{pq}t] + 2\Omega_{qr}[(S_x^{qr}S_x^{qr} + S_y^{qr}S_y^{qr})\cos 2\delta_{qr}t \\ & + (S_x^{qr}S_y^{qr} - S_y^{qr}S_x^{qr})\sin 2\delta_{qr}t] + 2\Omega_{rp}[(S_x^{rp}S_x^{rp} \\ & + S_y^{rp}S_y^{rp})\cos 2\delta_{rp}t + (S_x^{rp}S_y^{rp} - S_y^{rp}S_x^{rp})\sin 2\delta_{rp}t], \end{aligned} \tag{6}$$

where δ_{pq}, δ_{qr}, and δ_{rp} are the differences of the quadrupole frequencies of the interacting spins in the pq-, qr-, and rp-transitions.

It is now convenient to base our reasoning on examining the evolution of an isolated pair of spins (the two-spin model). This model makes it possible to solve the quantum Liouville equation exactly, which can produce useful information about the effect of inhomogeneous broadening of the dipole–dipole interaction and can be generalized to a multispin system by various approximation schemes.

To obtain a solution we must introduce the initial density matrix. In the high-temperature approximation, the equilibrium density matrix is determined by the Boltzmann distribution. When a 90° pulse is applied along the x axis, the initial density matrix is transformed (if we drop the first term and a constant factor) into

$$\begin{aligned}\rho_1 &= \omega_{pq1} S_y^{pq}\mathbf{1} + \omega_{pq2}\mathbf{1}S_y^{pq} \\ &= \omega_{pq}(S_y^{pq}\mathbf{1} + \mathbf{1}S_y^{pq}) + \delta(S_y^{pq}\mathbf{1} - \mathbf{1}S_y^{pq}) \\ &\simeq \omega_{pq}(S_y^{pq}\mathbf{1} + \mathbf{1}S_y^{pq}),\end{aligned} \tag{7}$$

since $\omega_{pq} \gg \delta$.

Further evolution due to the internal interaction Hamiltonian is determined by the quantum Liouville equation. Its formal solution,

$$\rho_2 = \exp(-iHt)\,\rho_1 \exp(iHt),$$

can be reduced to nonoperator form by successive differentiation and calculation of the resulting commutators.[3,15,16] In this process of such calculations the initial density matrix splits into two parts:

$$\begin{aligned}&\rho_1 = \rho_{1pq} + \rho_{1r}, \\ &\rho_{1pq} = \omega_{pq}(S_y^{pq}\mathbf{1}^{pq} + \mathbf{1}^{pq}S_y^{pq}), \\ &\rho_{1r} = \omega_{pq}(S_y^{pq}\mathbf{1}^{r} + \mathbf{1}^{r}S_y^{pq}),\end{aligned} \tag{8}$$

where $\mathbf{1}^{pq} = 2S_e^{pq}$ is the identity matrix in the pq subspace, and $\mathbf{1}^r = \mathbf{1} - \mathbf{1}^{pq}$. Accordingly, the spin space splits into two subspaces, for each of which we can put together a set of spin operators that commute with the operators of the other subspace.[18] The matrix ρ_{pq} commutes with the second and third terms of the dipole Hamiltonian (4), while ρ_r commutes with the first and second terms. Formally, the first subspace is entirely analogous to spin-$\frac{1}{2}$ NMR, although the physical meaning of the operators in these two cases do not coincide. The operators of the second subspace describe two-photon transitions through a third level and have no direct analogs in NMR.[12] The spin evolution in this subspace requires a separate investigation. The experiments described in Refs. 4–6 used 180° multipulse trains that averaged the internal interactions in this subspace,[1–3] which requires using other trains to study this subspace in experiments and makes it possible to discard the second part of the density matrix. We will also drop the first term of the quadrupole Hamiltonian in (5), which commutes with the other terms of the total Hamiltonian. The effect of this term will be determined in the last stage, although it is obvious from the start that it manifests itself in the divergence of the isochromatic curves as the detuning frequency increases. For the remaining terms in the Hamiltonian and the part of the density matrix referring to the first spin, $\rho_1 = \omega_{pq} S_y^{pq}\mathbf{1}$, the solution of the quantum Liouville equation for the pq-subspacehas the form

$$\rho_2 \sim \omega_{pq1}[S_y S_e \cos\,\omega_e t + (-S_x S_e c_\delta + S_z S_x c_\Omega)\sin\,\omega_e t], \tag{9}$$

where $\omega_e = (\Omega_{pq}^2 + \delta^2)^{1/2}$ is the effective precession frequency, and $c_\delta = \delta/\omega_e$ and $c_\Omega = \Omega_{pq}/\omega_e$ are the direction cosines of the rotation axis. Here and below we place no upper indices on the operators and assume, if not stated otherwise, that the operators belong to the pq-subspace.

Applying a second 180° pulse to the spin system leads to formation of spin echo. The density matrix describing the evolution of the observable $S_y S_e$ after the second pulse has been applied can be written

$$\begin{aligned}\rho_3 = \omega_{pq1} S_y S_e \{-[c_\delta^2 \cos\,\omega_e(t-\tau) + c_\Omega^2 \cos\,\omega_e(t+\tau)] \\ \times \cos\,\Delta(t-\tau) + c_\delta \sin\,\omega_e(t-\tau)\,\sin\,\Delta(t-\tau)\}.\end{aligned} \tag{10}$$

Thus, at $\delta = 0$, which corresponds to the spin-packet approximation,[11] formula (10) yield an echo signal with a decay time $T_2 = 1/\Omega_{pq}$. At $\Omega_{pq} = 0$ the initial transverse magnetization is fully restored, but in the general case there is partial refocusing of the isochromatic curves, which is determined by the ratio of the direction cosines, and the echo signal decays with a decay time depending on both Ω_{pq} and δ. Hence, measuring the spin–spin relaxation time in systems with an inhomogeneously broadened spectrum by the usual methods (the Hahn two-pulse train method) may lead to incorrect results. For the other part of the density matrix, ρ_{2r}, the explicit form of the solution of the quantum Liouville equation is more complicated and describes two rotations with frequencies

$$\omega_{e1,2} = \frac{1}{\sqrt{2}}\sqrt{\delta^2 + \frac{\Omega_{qr}^2 + \Omega_{rp}^2}{2} \pm \sqrt{\delta^4 + \delta^2(\Omega_{qr}^2 + \Omega_{rp}^2) + \Omega_{qr}^2\Omega_{rp}^2}}. \tag{11}$$

Obviously, the difference of these two frequencies is determined by the asymmetry of the dipole–dipole interaction tensor. At $\Omega_{qr}^2 = \Omega_{rp}^2$ one frequency vanishes and the other becomes equal to the frequency ω_e in (9).

3. TIME-DEPENDENT NUTATION IN THE EFFECTIVE FIELD

The experiments described in Refs. 4–6 used a multipulse train with alternating phases,

$$\varphi_y(\tau_1 - \psi_x - 2\tau_1 - \psi_{-x} - \tau_1)_n, \tag{12}$$

where φ_y is the angle of the rotation of the nuclear magnetization vector initiated by a priming pulse directed along the y axis of a rotating reference frame, ψ_x (ψ_{-x}) is the angle of the rotation initiated by a pulse from the train applied along the x ($-x$) axis, τ_1 ($2\tau_1$) is the pulse spacing, and n is the number of cycles.

In the first approximation (in which the pulses are approximated by delta functions) we can use the ordinary

method of calculating mean Hamiltonians.[1–3] Since the experiments used a 180°-pulse train, the inhomogeneous-broadening Hamiltonian is averaged to zero, and the dipole Hamiltonian, which a tensor of rank 2 and is transformed by a double angle by the pulses, remains unchanged. As a result the time-dependent nutation in the effective field is an induction signal whose decay is determined by the dipole–dipole interaction Hamiltonian. However, this approximation describes neither the dependence of the period of oscillations of the time-dependent nutation on pulse length and spacing nor the formation of echo signals in the effective field after the phase of rf pulses in the train has been inverted. Thus, in describing the experimental results discussed in Refs. 4–6 one must not neglect the inhomogeneous-broadening Hamiltonian during the action of the pulses. In this case the common method used in calculating the mean Hamiltonian[1] is inapplicable, since the pulse propagators do not represent a unitary transformation. Hence in what follows we will use a transition to the interaction representation in each spin [see Eq. (6)]. In this representation there is no inhomogeneous-broadening Hamiltonian, and the dipole–dipole interaction Hamiltonian and the Hamiltonian representing the interaction with the rf field become time-dependent.

For a 180°-pulse train we may leave only the first term in the expression for the dipole Hamiltonian (6). The Hamiltonian representing the interaction with the rf field in this case can be written

$$H_1=\omega_1(S_x\mathbf{1}\cos\Delta_1 t+\mathbf{1}S_x\cos\Delta_2 t-S_y\mathbf{1}\sin\Delta_1 t - \mathbf{1}S_y\sin\Delta_2 t), \tag{13}$$

where $\Delta_1=\omega_{pq1}-\omega_0$, $\Delta_2=\omega_{pq2}-\omega_0$, ω_0 is the frequency of the rf field, and ω_1 is amplitude of the Hamiltonian.

The cycle (t_c) averages of the dipole Hamiltonian and of the Hamiltonian representing the interaction with the rf field are given by the formulas

$$\begin{aligned}\tilde H_D^{(0)}&=\frac{2\Omega\sin\delta t_c}{\delta t_c}[(S_yS_y+S_xS_x)\cos\delta t_c \\ &\quad+(S_xS_y-S_yS_x)\sin\delta t_c] \\ &=\frac{2\Omega\sin\delta t_c}{\delta t_c}\exp\{i(S_zS_e\Delta_1+S_eS_z\Delta_2)t_c\}(S_yS_y+S_xS_x) \\ &\quad\times\exp\{-i(S_eS_z\Delta_2+S_zS_e\Delta_1)t_c\},\end{aligned} \tag{14}$$

$$\begin{aligned}\tilde H_1^{(0)}&=\frac{4\omega_1 t_p}{t_c}[\sin\Delta_1\tau(S_yS_e\cos 2\Delta_1\tau+S_xS_e\sin 2\Delta_1\tau) \\ &\quad+\sin\Delta_2\tau(S_eS_y\cos 2\Delta_2\tau+S_eS_x\sin 2\Delta_2\tau)] \\ &=\frac{4\omega_1 t_p}{t_c}\exp\{i(S_zS_e\Delta_1+S_eS_z\Delta_2)t_c\}(S_yS_e\sin\Delta_1\tau \\ &\quad+S_eS_y\sin\Delta_2\tau)\exp\{-i(S_zS_e\Delta_1+S_eS_z\Delta_2)t_c\},\end{aligned} \tag{15}$$

If we pass to a new reference frame defined by the propagator

$$P=\exp\{-i(S_zS_e\Delta_1+S_eS_z\Delta_2)t_c\},$$

we obtain

$$\begin{aligned}\tilde H_D&=\frac{2\Omega\sin\delta t_c}{\delta t_c}(S_xS_x+S_yS_y)=\Omega_1(S_xS_x+S_yS_y), \\ \tilde H_1&=\frac{4\omega_1 t_p}{t_c}(S_yS_e\sin\Delta_1\tau+S_eS_y\sin\Delta_2\tau) \\ &=\omega_{1a}(S_yS_e+S_eS_y)+\omega_{1b}(S_yS_e-S_eS_y),\end{aligned} \tag{16}$$

where $\tau=t_c/4$.

The total Hamiltonian consists of two commuting terms, $S_yS_e+S_eS_y$ and $S_yS_e-S_eS_y$, each of which does not commute with the third term, $S_xS_x+S_yS_y$. If we perform another transformation with the propagator

$$P_1=\exp\{i(S_zS_x+S_xS_z)\alpha\}\exp\{i(S_zS_x-S_xS_z)\beta\},$$

where

$$\sin\alpha=\frac{2\omega_{1a}}{\omega_{e1}},\quad \omega_{e1}=2\sqrt{\omega_{1a}^2+\frac{\Omega_1^2}{4}},$$

$$\sin\beta=\frac{2\omega_{1b}}{\omega_{e2}},\quad \omega_{e2}=2\sqrt{\omega_{1b}^2+\frac{\Omega_1^2}{4}},$$

the Hamiltonian (16) becomes a sum of commuting terms:

$$H=2[\Omega_1S_yS_y+S_xS_x(\omega_{e1}+\omega_{e2})+S_zS_z(\omega_{e2}-\omega_{e1})]. \tag{17}$$

Thus, to obtain a solution, we must transform the initial density matrix via the propagators P and P_1, find the solution of the quantum Liouville equation with the Hamiltonian (17), and perform the inverse transformation. All these operations are extremely involved and cannot be described in detail in a journal article (it is best to do all calculations on a computer using a package with computer algebra capabilities).

A. Pulse train (12) without a priming pulse

The initial density matrix (the Boltzmann matrix density in the high-temperature approximation) for the first spin is proportional to $\omega_{pq1}S_zS_e$ and for the second spin, to $\omega_{pq2}S_eS_z$. The contribution of the first spin to the observable S_xS_e is

$$\begin{aligned}\rho_1\sim S_xS_e[&(\sin\alpha+\sin\beta)\sin(\omega_{e1}+\omega_{e2})t_e+(\sin\alpha \\ &-\sin\beta)\sin(\omega_{e1}-\omega_{e2})t_e]\cos\Omega_1t_e\cos\frac{\Delta_1t_c}{2}.\end{aligned} \tag{18}$$

The contribution of the second spin is

$$\rho_2\sim S_xS_e\sin\Omega_1t_e\cos\frac{\Delta_2t_c}{2}\sin\omega_{e1}t_e\sin\omega_{e2}t_e\sin(\alpha+\beta), \tag{19}$$

where $t_e=nt_c$ is the effective discrete time. These expressions imply that there is exchange of coherence between the spins. The contribution of the second spin to the transverse magnetization of the first spin is zero if $\Omega=0$. The experimental curve for time-dependent nutation obtained by the pulse train (12) without a priming pulse[6] is depicted in Fig. 1. The curve represents damped sinusoidal oscillations and qualitatively agrees with the expressions (18) and (19). The damping is determined by the dipole–dipole interaction Hamiltonian (the factor $\cos\Omega_1t_e$) and the mixing of oscillations with combinations of the frequencies ω_{e1} and ω_{e2}.

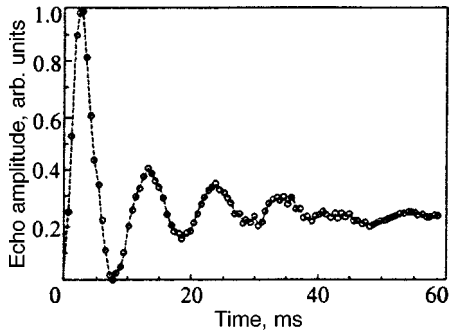

FIG. 1. Envelope of an echo signal in a multipulse train with alternating phases [see (12)] for the case $\varphi=0$, $\psi=\pi$, $\tau=0.3$ ms, and $t_p=50$ μs.

B. Pulse train (12) with a 90° priming pulse

After a 90° pulse has acted on the system, the initial density matrix becomes proportional to S_xS_e and S_eS_x for the first and second spins, respectively. The first spin creates the observed coherence

$$\rho_1 \sim S_xS_e\Big\{\cos\,\Omega_1 t_e[\cos\,\omega_{e1}t_e\,\cos\,\omega_{e2}t_e+\sin\,\omega_{e1}t_e \times\sin\,\omega_{e2}t_e\,\cos\,(\alpha+\beta)]\cos^2\frac{\Delta_1 t_c}{2}+(\cos^2\alpha\,\cos\,\omega_{e1}t_e +\cos^2\beta\,\cos\,\omega_{e2}t_e)\,\sin^2\frac{\Delta_1 t_c}{2}\Big\}. \quad (20)$$

The contribution of the second spin is

$$\rho_2 \sim S_xS_e\Big\{\sin\,\Omega_1 t_e(-\sin\,\omega_{e1}t_e\,\cos\,\omega_{e2}t_e\,\cos\,\alpha+\cos \times\omega_{e1}t_e\,\sin\,\omega_{e2}t_e\,\cos\,\beta)\cos\frac{\Delta_1 t_c}{2}\cos\frac{\Delta_2 t_c}{2} +(-\cos^2\alpha\,\cos\,\omega_{e1}t_e+\cos^2\beta\,\cos\,\omega_{e2}t_e-\sin^2\alpha +\sin^2\beta)\,\sin\frac{\Delta_1 t_c}{2}\sin\frac{\Delta_2 t_c}{2}\Big\}. \quad (21)$$

The experimental curve[6] for this case is depicted in Fig. 2 and, according to the above expressions, represents damped cosinusoidal oscillations.

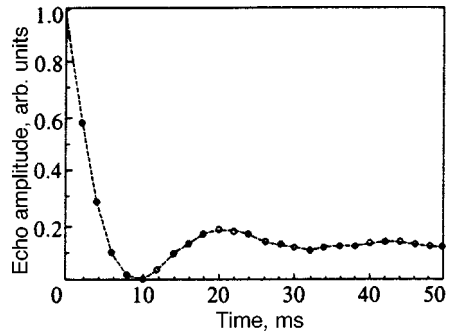

FIG. 2. Envelope of an echo signal in a multipulse train with alternating phases with $\varphi=\pi/2$, $\psi=\pi$, $\tau=0.5$ ms, and $t_p=50$ μs. Signal sampling was done once per cycle.

4. ECHO IN THE EFFECTIVE FIELD

Echo signals in the effective field are formed by the multipulse train[4–6]

$$\varphi_y-(\tau_1-\psi_x-2\tau_1-\psi_{-x}-\tau_1)_n -(\tau_1-\psi_{-x}-2\tau_1-\psi_x-\tau_1)_N. \quad (22)$$

Hence analysis requires that we examine all coherences that occur (during the time $\tau_e=nt_c$) due the mean Hamiltonians (14) and (15) and the reverse evolution of these coherences after the phase of the rf field has been inverted in the train. For this part of the train the mean Hamiltonian (15) changes sign. The priming pulse φ_y excites a transverse magnetization S_xS_e. Then, due to the action of the sum of Hamiltonians (14) and (15), this magnetization transforms into the orthogonal cophased coherence S_yS_e, the two antiphased coherences S_xS_z and S_yS_z, and the two-photon coherences S_xS_y and S_yS_x. After the phase has been inverted, additional contributions to the observed coherence S_xS_e are provided by the antiphased coherence S_xS_z and the orthogonal cophased coherence S_yS_e. According to (14) and (15), in the first stage there are rotations with combinations of the three frequencies Ω_1, ω_{e1}, and ω_{e2}. After phase inversion is completed, there is convergence of coherence at one frequency and divergence at another. The observed coherence S_xS_e and the antiphased coherence S_xS_z are produced by terns that diverge at the frequency Ω_1 and converge at the frequencies $\omega_{e1}-\omega_{e2}$ and $\omega_{e1}+\omega_{e2}$, respectively:

$$\rho(S_xS_e)\sim\cos^2\frac{\Delta_1 t_c}{2}\cos^2\frac{\alpha+\beta}{2}\sin\,\alpha\,\sin\,\beta \times\cos\,\Omega_1(t_e+\tau_e)\cos(\omega_{e1}-\omega_{e2})(t_e-\tau_e), \quad (23)$$

$$\rho(S_xS_z)\sim\sin^2\frac{\Delta_1 t_c}{2}\cos^2\frac{\alpha+\beta}{2}\sin\,\alpha\,\sin\,\beta \times\cos\,\Omega_1(t_e+\tau_e)\,\cos\,(\omega_{e1}+\omega_{e2})(t_e-\tau_e). \quad (24)$$

This signal decays with an effective spin–spin relaxation time $T_{2e}=1/\Omega_1$. The orthogonal cophased coherence provides a contribution which is totally convergent at the frequencies ω_{e1} and ω_{e2}. Thus, in the given experiment, the echo-signal decay time is independent of inhomogeneous broadening, which makes it possible to recommend this method for measuring dipole–dipole interaction times in systems with an inhomogeneously broadened spectrum. The decay of the echo-signal envelope obtained via the pulse train (22) is depicted in Fig. 4 of Ref. 6. The decay time amounts to 20 μs, while T_2 in the same sample measured by the Hahn two-pulse method is 7 ms.

Another approach to analyzing echo signals in the effective field amounts to calculating the mean Hamiltonian for a ''supercycle'' consisting of two cycles with inverse phases. With such a train,

$$\varphi_y-(\tau-\psi_x-2\tau-\psi_{-x}-\tau)_n-[(\tau-\psi_{-x}-2\tau -\psi_x-\tau)_{2n}-(\tau-\psi_x-2\tau-\psi_{-x}-\tau)_{2n}]_N, \quad (25)$$

multiple echo signals in the effective field have been obtained.[5,6] The mean Hamiltonian calculated in this way

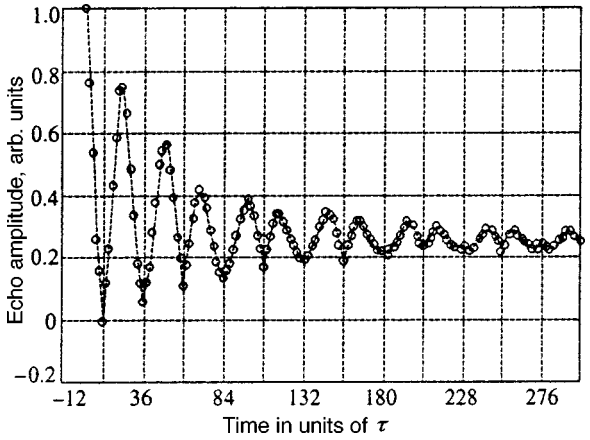

FIG. 3. Multiple echo signals in the effective field obtained via train (25) at $\tau=0.3$, $t_p=50$ μs, and $n=3$. The grid denotes the points of phase inversion.

describes the echo-signal envelope in such a train. The dipole Hamiltonian (14) in this case remains unchanged, while the Hamiltonian (15) representing the interaction with the rf field becomes

$$\tilde{H}_1^{(0)}=\frac{4\omega_1 t_p}{t_c}\left[\frac{\sin^2(\Delta_1 t_c/4)}{\cos\Delta_1\tau}\left(S_xS_e\cos\frac{\Delta_1 t_c}{2}+S_yS_e\sin\frac{\Delta_1 t_c}{2}\right)\right.$$
$$\left.+\frac{\sin^2(\Delta_2 t_c/2)}{\cos\Delta_2\tau}\left(S_eS_x\cos\frac{\Delta_2 t_c}{2}+S_eS_y\sin\frac{\Delta_2 t_c}{2}\right)\right]$$
$$=\frac{4\omega_1 t_p}{t_c}\exp\{i(S_zS_e\Delta_1+S_eS_z\Delta_2)t_c\}$$
$$\times\left(S_xS_e\frac{\sin^2(\Delta_1 t_c/4)}{\cos\Delta_1\tau}+S_eS_x\frac{\sin^2(\Delta_2 t_c/4)}{\cos\Delta_2\tau}\right)$$
$$\times\exp\{-i(S_zS_e\Delta_1+S_eS_z\Delta_2)t_c\},\qquad(26)$$

where t_c now denotes the duration of the entire ''supercycle.'' This Hamiltonian describes time-dependent nutation in the effective field of the multipulse train (22), provided that signal strobing is done once per ''supercycle,'' and resembles the Hamiltonian (15), differing only in the coefficients of the operators and the operators proper, in which S_y is replaced by S_x.

Accordingly, the solution of the quantum Liouville equation is close to (21):

$$\rho\sim S_zS_e\{\cos\Omega_1 t_e[\cos\omega_{e1}t_e\cos\omega_{e2}t_e+\sin\omega_{e1}t_e$$
$$\times\sin\omega_{e2}t_e\cos(\alpha+\beta)]+\cos^2\alpha\cos\omega_{e1}t_e$$
$$+\cos^2\beta\cos\omega_{e2}t_e+\sin^2\alpha+\sin^2\beta\}\sin\Delta_1\tau.\qquad(27)$$

These patterns are reflected in the experimental curve in Fig. 3 (see Ref. 6), whose envelope decays according to a double-exponential law. The fast exponential is probably determined by the time T_{2e} and the slow exponential by the formation of multispin coherence.

The contribution of the second spin exhibits the same features, but here expressions of the type $\cos\omega(t_e-\tau_e)$ are replaced by $\sin\omega(t_e-\tau_e)$. This probably explains the ''pairing'' of echo signals depicted in Fig. 3.

To go over to the many-body problem, we must average the above expressions for the density matrix over the dipole and quadrupole frequencies with appropriate distribution functions. Of course, in this case the solution does not become ''many-body.'' However, the approximations used here are no rougher than those employed in widely accepted methods that use second-moment calculations or an iteration procedure (if the latter is limited to two iterations).[7] In both cases one is forced to calculate a double commutator, which corresponds to decoupling the equations before three-spin coherence is excited.[16]

A detailed review of the methods used in describing time-dependent nutation in EPR in solids, which large inhomogeneous broadening must be taken into account, can be found in Ref. 19. Here preference is given to a phenomenological description that the Bloch equation instead of the Liouville equation. In the spin-packet approximation, time-dependent nutation represents damped oscillations (with a decay time T_2) with a frequency equal to the amplitude of the rf field, ω_1:

$$M_y(t)=M_z(0)\sin\omega_1 t\exp\left\{-\frac{t}{T_2}\right\},$$

provided that the initial magnetization is directed along the z axis. This approach can also be used to interpret time-dependent nutation in the effective field of multipulse trains, since the mean Hamiltonians (14), (15), and (26) formally (i.e., in the constituent operators) are similar to the Hamiltonian describing time-dependent nutation in a steady-state rf field in EPR. For the sake of comparison, it is also useful to consider the signals of time-dependent nutation and echo in the effective field of the single-phase multipulse train

$$\varphi_x-(\tau-\psi_x-\tau)_n.\qquad(28)$$

The mean Hamiltonian for this train can be written

$$\tilde{H}_1^{(0)}=\frac{2\omega_1}{t_c}\left[\frac{\sin\Delta_1 t_p}{\Delta_1}(S_xS_e\cos\Delta_1\tau+S_yS_e\sin\Delta_1\tau)\right.$$
$$\left.-\frac{\sin\Delta_2 t_p}{\Delta_2}(S_eS_x\cos\Delta_2\tau+S_eS_x\sin\Delta_2\tau)\right]$$
$$\simeq\frac{2\omega_1 t_p}{t_c}\exp\{i(S_zS_e\Delta_1+S_eS_z\Delta_2)t_c\}(S_xS_e$$
$$+S_eS_x)\exp\{-i(S_zS_e\Delta_1+S_eS_z\Delta_2)t_c\}.\qquad(29)$$

The experimental curves representing time-dependent nutation and echo signals that arise after the phase has been inverted are depicted in Figs. 1 and 2 of Ref. 4. The oscillation frequency in this case ($2\omega_1 t_p/t_c$) is at its maximum, is lower for the train (12) with alternating phases ($4\omega_1 t_p\sin(\Delta_{1,2}\tau)/t_c$) (Fig. 2), and is still lower for the supercycle of the train (22)($4\omega_1 t_p\sin^2(\Delta_{1,2}t_c/4)/t_c\cos\Delta_{1,2}\tau$) (Fig. 3); oscillations are essentially unobservable in the last case.

All the experiments referred to in this paper were carried out using single-crystal samples of sodium nitrite $NaNO_2$

(NQR in ^{14}N) at 77 K and the $+\leftrightarrow 0$ transition. A more thorough description of the experimental methods can be found in Refs. 4–6.

5. CONCLUSION

The theoretical investigation of echo signals and time-dependent nutation in the effective field of multipulse trains obtained in experiments and described in the previous papers cited in Refs. 4–6 has shown that the proposed method can be used to find in experiments the true spin–spin relaxation time for spin systems with an inhomogeneously broadened spectrum. The results of the experiments imply that the decay times of the envelopes of echo signals and of the time-dependent nutation in the effective field are several times longer than the decay time of the envelope of ordinary-echo signals measured by the Hahn two-pulse method. The reason is that in systems with an inhomogeneously broadened spectrum the dipole reservoir ceases to be a constant of motion, its place being taken by the sum of the dipole Hamiltonian and the Hamiltonian describing the splitting of the levels of the interacting spins. One consequence of this is that the decay time of the transverse magnetization, measured by the Hahn method, is determined by the values of the two Hamiltonians. A multipulse train separates these Hamiltonians, shifting the inhomogeneous broadening to the subspace of the external rf field. As a result, we can select a representation in which the dipole Hamiltonian (scaled, to be exact) becomes a constant of motion.

Thus, the oscillations of time-dependent nutation are determined by the mean Hamiltonian of the rf field (with allowance for inhomogeneous broadening), and the damping of these oscillations is determined by the dipole–dipole interaction Hamiltonian. Phase inversion of the rf field leads to echo formation, and the isochromatic curves that generate the oscillations are refocused.

Quantitative comparison of the results of the present investigation and the experimental data asks for a separate study after the time-dependent nutation in the effective field in the other subspace has been analyzed. This requires averaging over dipole and quadrupole frequencies. The experimentally observed time-dependent nutation in a 180°-pulse train and a zero priming pulse is of the greatest interest. Probably, these signals are due to mutual correlation of dipole and quadrupole frequencies over the bulk of the sample. If this is the case, then the proposed method can serve as a useful instrument for studying the inner structure (homogeneity) of crystals.

ACKNOWLEDGMENTS

This work was made possible by a grant from the Russian Fund for Fundament Research (Grant No. 96-03-33911).

[1] J. Waugh, *New NMR Methods in Solid State Physics*, Cambridge Univ. Press, Cambridge (1978).
[2] U. Haeberlen and M. Mehring, *High Resolution NMR in Solids. High Resolution NMR Spectroscopy in Solids*, Springer, New York (1976).
[3] D. Ya. Osokin, Phys. Status Solidi B **102**, 681 (1980).
[4] V. L. Ermakov, R. Kh. Kurbanov, D. Ya. Osokin, and V. A. Shagalov, JETP Lett. **54**, 466 (1991).
[5] D. Ya. Osokin, R. Kh. Kurbanov, and V. A. Shagalov, JETP Lett. **62**, 309 (1995).
[6] D. Ya. Osokin and V. A. Shagalov, Solid State Nucl. Magn. Reson. **10**, 63 (1997).
[7] M. Goldman, *Spin Temperature and Nuclear Magnetic Resonance in Solids* [Russian translation], Mir Publishers, Moscow (1972), pp. 301, 113, and 304 [English orig.: Clarendon Press, Oxford (1970)].
[8] V. A. Atsarkin, *Dynamic Polarization of Nuclei in Solid Insulators* [in Russian], Nauka, Moscow (1980).
[9] G. R. Khutsishvili, Zh. Éksp. Teor. Fiz. **50**, 1641 (1966) [Sov. Phys. JETP **23**, 1092 (1966)].
[10] K. M. Salikhov, A. G. Semenov, and Yu. D. Tsvetkov, *Spin Echo and Its Application* [in Russian], Nauka, Novosibirsk (1976).
[11] Ya. S. Lebedev and V. I. Muromtsev, *EPR and Relaxation of Stabilized Radicals* [in Russian], Khimiya, Moscow (1972).
[12] A. M. Portis, Phys. Rev. **91**, 1071 (1953).
[13] I. A. Safin and D. Ya. Osokin, *Nuclear Quadrupole Resonance in Nitrogen Compounds* [in Russian], Nauka, Moscow (1977), p. 25.
[14] R. Ernst, G. Bodenhausen, and A. Wokaun, *Principles of NMR in One and Two Dimensions* [Russian translation], Mir Publishers, Moscow (1990), p. 57 [English orig.: Clarendon Press, Oxford (1987)].
[15] D. Ya. Osokin, J. Mol. Struct. **83**, 243 (1982).
[16] M. Munowitz, *Coherence and NMR*, Wiley, New York (1988), p. 39.
[17] M. Matti Maricq, Phys. Rev. B **33**, 4501 (1986).
[18] V. L. Ermakov and D. Ya. Osokin, Mol. Phys. **53**, 1335 (1984).
[19] D. Stehlik, C. H. Bock, and M. C. Thurnauer, in *Advanced EPR: Applications in Biology and Biochemistry*, A. J. Hoff (Ed.), Elsevier, Amsterdam (1989).

Translated by Eugene Yankovsky

Synchronizing a chaotic laser by injecting a chaotic signal with a frequency offset

A. P. Napartovich*) and A. G. Sukharev

Troitsk Institute of Innovative and Thermonuclear Research, 142092 Troitsk, Moscow Region, Russia
(Submitted 29 September 1998)
Zh. Éksp. Teor. Fiz. **115**, 1593–1604 (May 1999)

It has been theoretically proven that a chaotic laser driven by an injected signal derived from a similar chaotic master oscillator working at a slight frequency offset operates in a new mode of synchronization. There is a certain relationship between average laser fields, which has been approximately calculated in an explicit form, and instantaneous fields form an attractor about the point defined by these average values, with a relatively small spread in the phase trajectories around it. It has been shown that such a configuration can be used in transmitting confidential information.
[S1063-7761(99)00505-3]

1. INTRODUCTION

From the standpoint of oscillation dynamics, a laser is a nonlinear dynamic system that operates, depending on the conditions, in various modes, including regimes of dynamic chaos, which have been studied extensively in recent time. Since lasers are used in communication systems, the issue of transmission of confidential information through commonly accessible channels has been brought up.[1,2] The underlying idea is utilization of chaotic generation regimes for coding transmitted information. The development of this idea required meticulous research into operating regimes of chaotic lasers and feasibility of their synchronization by coupling radiation from one laser to another.

The issue of synchronization of oscillating systems dates from several centuries ago. However, since the synchronization effects in chaotic nonlinear oscillators are various and sundry, the theory of these effects is far from complete. It has become clear by this time that the concept of ''synchronization of coupled nonlinear systems'' is still in the process of development, because a lot of new synchronization modes are being discovered. The simplest form is synchronization of identical systems when they are coupled using an appropriate technique. In this case, the state vectors x_1 and x_2 of these systems in the phase space approach one another as closely as one pleases as the time tends to infinity. One example taken from the laser physics is the synchronization of optically coupled lasers: the experiment was reported in Ref. 3, and the theoretical aspects in Ref. 4.

The concept of generalized synchronization has also been introduced. It refers to the situation when certain functional asymptotic relations are established between the state vectors of two lasers. For example, two chaotic lasers with a symmetric coupling and an upper limit imposed on the frequency offset between their cavities are locked in such a manner that the field amplitudes are identical, whereas the phase difference takes a certain value depending on the offset and coupling constant.[4]

Weaker forms of synchronization, namely, the so-called ''phase synchronization''[6] and delayed synchronization,[7] have been discovered in dynamic systems described by Rössler's equations.[5] The ''phase'' in Ref. 6 is the phase of a point on the trajectory of Rössler's attractor, which corresponds to the rotation on average, if the origin is chosen appropriately. In the case of the delayed synchronization, the following asymptotic relation between state vectors applies: $x_2(t)=x_1(t+t_0)$, where t_0 is the delay between the evolution of the ''slave'' system with respect to the master (here the case of unidirectional coupling is considered). As in the case of the transition to turbulence in hydrodynamics, which proceeds via sequential changes of oscillation modes, it turns out that at certain system parameters of two coupled chaotic lasers, essentially perfect synchronization is fully disrupted from time to time, and then restored.[8]

The brief review of papers on the synchronization of chaotic systems given above indicates that the effects in such generators are many and varied, and this variety can hamper the development of practicable optical communication lines for transmission of confidential information based on chaotic lasers. Nonetheless, successful attempts[1,2] to test in experiment a system of optical communication based on synchronization of a chaotic laser (receiver) by injecting into the latter the radiation from a master oscillator have demonstrated the feasibility of this scheme. The technique for coding/decoding information using a random sequence of laser pulses is based on a nontrivial effect: it turns out[1] that a slave laser fed with an optical signal carrying coded information restores the initial signal and eliminates modulation. By comparing the input beam carrying information with the output signal, one can retrieve the coded information. It is obvious that the effect of information losses in the receiver with conservation of the carrier chaotic signal does not take place under all conditions. There are limitations on the density of coded information and degree of identity between the two lasers. The admissible bandwidth for information transmission was calculated numerically[9] in the case of two identical lasers. The crucial parameter of two nonidentical lasers is the difference between effective lengths of their optical cavities if it is not a multiple of the light wavelength. The frequency offset between the cavities raises the issue of syn-

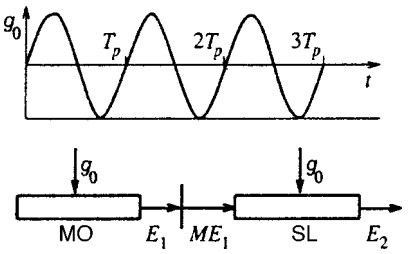

FIG. 1. Diagram of the laser system: (MO) master oscillator; (SL) slave laser; $g_0(t)$ is the pumping power.

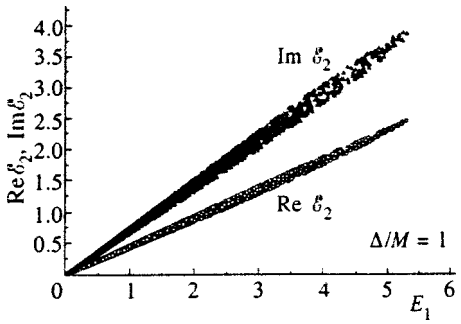

FIG. 2. Phase portrait of the real and imaginary components of field $\mathscr{E}_2$ at $\Delta = M$.

chronization of two chaotic lasers and erasure of information in the output beam of the slave laser.

This paper addresses the synchronization of chaotic lasers at a finite frequency offset between them and the possibility of decoding information transmitted by such a system.

2. MATHEMATICAL MODEL

The model considers the chaotic operation of lasers owing to a periodic modulation of the rate of pumping to the upper level with period T_p. When T_p is close to the period of relaxation oscillations[10] in a laser under constant pumping at a level equal to the amplitude of the periodic pumping, the laser generates a random sequence of pulses. A diagram of the studied laser system is given in Fig. 1.

The differential equations describing the system shown in Fig. 1 are given below in the simplest possible form:

$$\tau\dot{g}_1 = g_0(t) - g_1(1+|\mathscr{E}_1|^2),$$

$$\tau\dot{g}_2 = g_0(t) - g_2(1+|\mathscr{E}_2|^2),$$

$$\dot{\mathscr{E}}_1 = \frac{1}{2}(g_1 - g_{\text{th}} + i\Delta)\mathscr{E}_1,$$

$$\dot{\mathscr{E}}_2 = \frac{1}{2}(g_2 - g_{\text{th}} - i\Delta)\mathscr{E}_2 + M(\mathscr{E}_1 - \mathscr{E}_2). \tag{1}$$

The kinetic properties of the gain $g_{1,2}$ in the medium are described in the simplest approximation of a fixed pumping $g_0(t)$ and a single effective relaxation time τ of the population inversion. The field amplitudes $\mathscr{E}_1$ and $\mathscr{E}_2$ are complex parameters, and the squares of their absolute values are normalized to the saturation intensity I_s. The time is expressed in dimensionless form, and the natural selection of the time unit is the inherent time characteristic of the laser, namely, the photon round-trip time in the cavity, $2L/c$. The effective dimensionless time of the medium relaxation $\tau \gg 1$.

The two chaotic lasers are coupled by injecting a fraction of the first laser's output into the second. The coupling efficiency is determined by the coupling factor M, which takes account of the propagation and coupling losses (the parameter M is assumed to be a real value). The difference between characteristic laser frequencies is determined by the cavity offset $\Delta = \Delta\omega \cdot 2L/c$, where $\Delta\omega$ is the difference between the eigenfrequencies, and L is the cavity length of the master oscillator. For identical laser cavities $\Delta = 0$. In order to obtain identical fields in the two lasers at zero offset, we set the threshold gain g_{th} of the slave laser to equal that of the master oscillator when they are optically coupled, so the dynamic equations for the two laser are identical at $\mathscr{E}_1 = \mathscr{E}_2$.

Replacing the complex fields $\mathscr{E}_1 = E_1 \exp(i\varphi_1 t)$ and $\mathscr{E}_2 = E_2 \exp(i\varphi_2 t)$ with the amplitudes and phases $\varphi = \varphi_1 - \varphi_2$, we rewrite Eq. (1) in the form more convenient for further analysis:

$$\tau\dot{g}_1 = g_0 - g_1(1+E_1^2),$$

$$\tau\dot{g}_2 = g_0 - g_2(1+E_2^2),$$

$$\dot{E}_1 = \frac{1}{2}(g_1 - g_{\text{th}})E_1,$$

$$\dot{E}_2 = \frac{1}{2}(g_2 - g_{\text{th}})E_2 + M(E_1\cos\varphi - E_2),$$

$$\dot{\varphi} = \Delta - M(E_1/E_2)\sin\varphi. \tag{2}$$

In the numerical calculations described below, the dimensionless parameters of the cavities and media were selected after a preliminary analysis of the bifurcation diagram of a laser under periodic pumping,[11] so that the master oscillator should operate in the regime of ''maximal chaos,'' which corresponds to the maximum positive Lyapunov exponent (Lyapunov dimensionality $\Lambda \approx 1.5$). The system parameters were $g_{\text{th}}/g_0 = 0.5$, $M = 0.1$, and $g_{\text{th}} = 0.5$, where g_0 is the pumping intensity averaged over time.

3. CHARACTERIZATION OF THE SYNCHRONIZATION MODE

Based on previous studies of synchronization of two coupled chaotic lasers with a certain frequency offset between their cavities,[4] one might expect a generalized synchronization at small offset parameters in the scheme under discussion (Fig. 1). It follows, however, from numerical calculations that even at a small frequency offset between the cavities, there is no asymptotically exact relation between fields $\mathscr{E}_1$ and $\mathscr{E}_2$. Figure 2 shows a phase diagram with the field amplitude E_1 plotted as the abscissa, whereas the ordinate represents two parameters, namely, Re $\mathscr{E}_2$ and Im $\mathscr{E}_2$. One can see that the set of phase points for each parameter is concentrated near a straight line, but remains, nonetheless, chaotically spread about it.[1)] Strictly speaking, this behavior might simply be treated as a lack of synchronization between

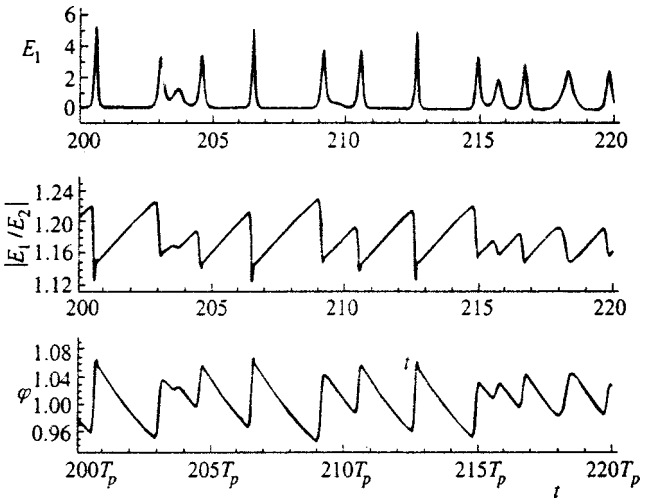

FIG. 3. The amplitude of the master oscillator output, ratio between field amplitudes generated by the two lasers, and their phase difference as functions of time; $\Delta = M$.

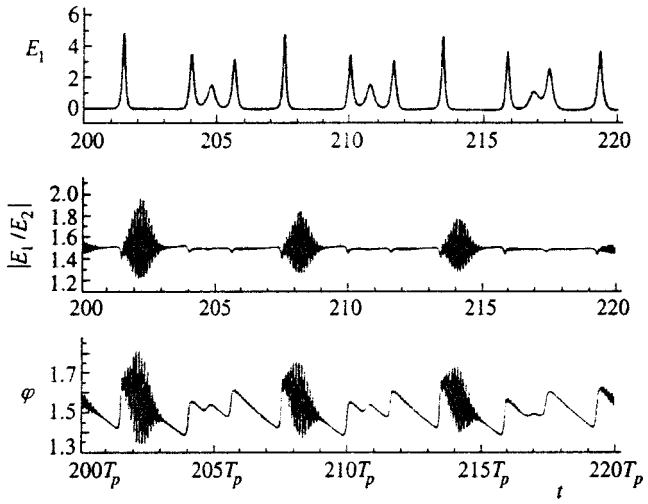

FIG. 4. The same as in Fig. 3, but at $\Delta = 1.5M$.

the two lasers. It is noteworthy, however, that the spread of phase points about the average straight lines is fairly small.

Thus, there is no generalized synchronization in its ideal sense, but the correlation between the two fields is fairly strong. This effect is also illustrated by Fig. 3, which shows the chaotic dynamics of the field amplitude E_1 generated by the master oscillator, the ratio between field amplitudes $|E_1/E_2|$, and the phase difference at frequency offset $\Delta = M$. It is also noteworthy that although $|E_1/E_2|$ and φ are random functions of time, the amplitude of their deviations from the mean is small (no greater than 5%). At lower Δ the rms amplitude of deviation from nonvanishing means is smaller. Therefore, the resulting process can be described as a generalized synchronization ''on average''with fluctuations whose amplitude increases with the ratio Δ/M. This description is inconsistent with all suggested scenarios of synchronization between chaotic oscillators that have been known to date. Our calculations indicate that the suggested mode of synchronization is realized in the region of parameters extending to $\Delta/M = 1.5$, and calculations corresponding to the latter condition are plotted in Fig. 4. It clearly shows that deviations from the mean values are up to tens of percent. At higher Δ/M (see Fig. 5 for $\Delta/M = 1.6$) even the synchronization on average is destroyed, and spreads of parameters are higher than their mean values.

Based on considerations of physical dimensionality, it is natural to suppose that, in the case of the generalized synchronization, a proportionality relation between the fields should exist:

$$\mathscr{E}_1 = \mathscr{E}_2 f e^{i\varphi_0}, \tag{3}$$

where f and φ_0 are real constants. In the case of undamped small fluctuations, one can calculate the constants minimizing the functional

$$\sigma^2 = \left(\int_0^t |\mathscr{E}_1|^2 dt \right)^{-1} \int_0^t |\mathscr{E}_1 - f_m \exp(i\varphi_m)\mathscr{E}_2|^2 dt, \tag{4}$$

where t is a sufficiently long time interval. By minimizing this functional, one can derive the following expressions for optimal f_m and φ_m in terms of $\mathscr{E}_1$ and $\mathscr{E}_2$, which can be calculated as functions of time numerically:

$$f_m = \left(\int E_2^2 dt \right)^{-1} \times \sqrt{\left(\int E_1 E_2 \cos\varphi \, dt \right)^2 + \left(\int E_1 E_2 \sin\varphi \, dt \right)^2}, \tag{5}$$

$$\varphi_m = \arg\left(\int E_1 E_2 \exp(i\varphi) dt \right). \tag{6}$$

The parameters f_m and φ_m calculated by Eqs. (5) and (6) using the fields derived from Eq. (2) are plotted in Fig. 6 and 7, respectively, as functions of the frequency offset divided by the coupling constant M. Note that in the case of the real generalized synchronization, the minimum of functional (4) tends to zero as $1/t$, whereas in our calculations it remains nonzero as $t \to \infty$.

In the case of small spreads of calculations about the straight lines, as in Fig. 2, there is good reason to suppose

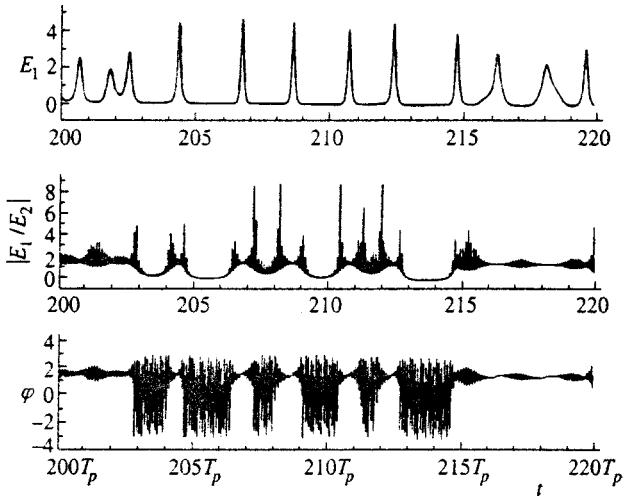

FIG. 5. The same as in Figs. 3 and 4, but at $\Delta = 1.6M$.

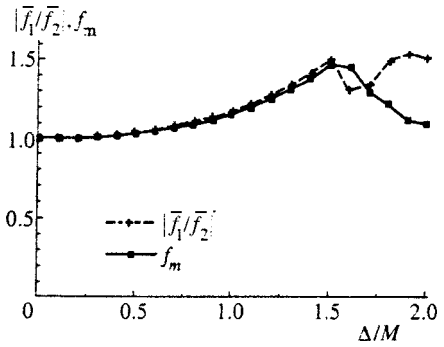

FIG. 6. Parameter f_m and time average $\bar{f}$ as functions of the normalized frequency offset.

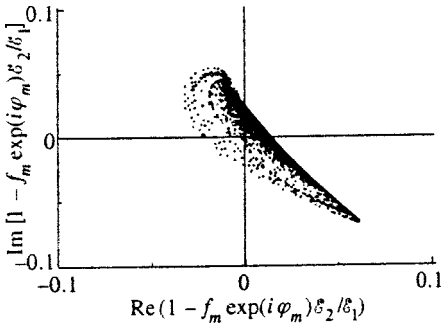

FIG. 8. Phase diagram of the relative spread of states plotted on the complex plane about the state of generalized synchronization with parameters f_m and φ_m for $\Delta=M$.

that directly calculated time averages of the ratio f between field amplitudes and phase differences φ should be close to the optimal values f_m and φ_m. This is supported by direct calculations. These parameters are very close to their optimal values over the interval $\Delta/M\leqslant 1.5$ (Figs. 6 and 7). When Δ/M is more than the critical value $\Delta/M=1.5$, the curves diverge, which is accounted for by the disruption of synchronization even in the sense of average. Figure 5 shows that at $\Delta/M=1.6$ the phase difference between fields $\mathscr{E}_1$ and $\mathscr{E}_2$ grows indefinitely (note that the phase in this graph is defined modulo 2π). Thus, in the synchronization mode under discussion, the average difference between field frequencies is zero, whereas beyond the synchronization interval the frequencies of fields $\mathscr{E}_1$ and $\mathscr{E}_2$ on average are different.

The proximity between constants defined in different manners, which is demonstrated by Figs. 6 and 7, indicates that the deviations of fields from their mean values are small at each moment of time when Δ/M ranges up to the critical value. The spread of points in the strange attractor for the ratios between the field amplitudes is shown in Fig. 8, which plots the phase diagram of the complex ratio between the fields at $\Delta/M=1$. The emergence of the strange attractor allows us to speak of the synchronization mode under consideration as ''strange.'' The spread of points in the phase diagram can be characterized by the parameter σ^2, which has the sense of a variance. The curve of σ as a function of Δ/M is plotted in Fig. 9.

The small amplitude of fluctuations justifies the application of perturbation theory to integrate Eq. (2) in explicit form.

Under the conditions of the ''strange'' synchronization discussed in this paper, the derivatives on the left-hand sides of Eq. (2) averaged over a large time interval are zero, whereas the variations of all parameters, including the phase difference φ, are finite. By averaging the first and third lines of Eq. (2) over time, one can easily find that $\overline{g_1}=g_{\mathrm{th}}$ and $\overline{E_1^2}=\overline{g_0}/g_{\mathrm{th}}-1$, where an overbar denote time averaging of variables. After introducing a new variable $F=E_1/E_2$ instead of E_2, we obtain a differential equation

$$\dot{F}=MF(1-F\cos\varphi)-\frac{1}{2}(g_2-g_1)F. \tag{7}$$

By averaging this equation divided by F over time, we obtain

$$\overline{g_2}=g_{\mathrm{th}}+2M(1-\overline{F\cos\varphi}). \tag{8}$$

Another relationship can be derived by averaging the difference between the kinetic equations for gains g_1 and g_2:

$$\overline{g_2}+\overline{g_2E_2^2}=\overline{g_1}+\overline{g_1E_1^2}=\overline{g_0}. \tag{9}$$

Let us replace, using a rough approximation, the average of the product of two variables, g_2 and E_2^2, with the product of the averages: $\overline{g_2E_2^2}\approx\overline{g_2}\cdot\overline{E_2^2}$. Given that the spread of the points in the attractor shown in Fig. 8 is small, let us use the

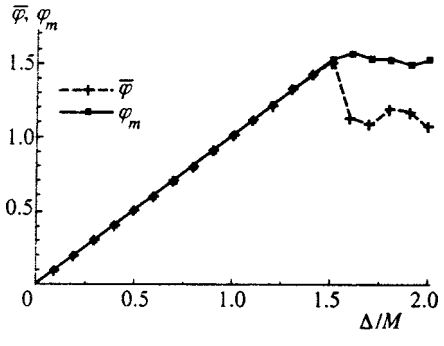

FIG. 7. Optimal phase difference φ_m and time averaged $\bar{\varphi}$ as functions of the normalized frequency offset.

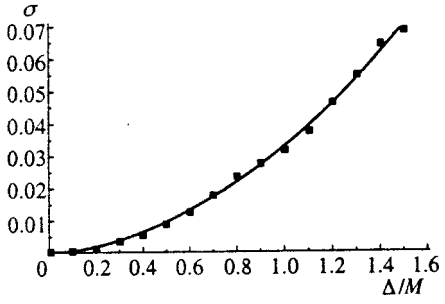

FIG. 9. Function $\sigma(\Delta/M)$, where σ^2 is the dispersion. The solid curve is an interpolation through the points.

approximate relationship $\overline{E_2^2}=\overline{E_1^2}/f_m^2$. Using Eq. (5) and the expressions for $\overline{E_1^2}$ given above, we derive another expression for $\overline{g_2}$ from Eq. (9):

$$\overline{g_2}=\frac{\overline{g_0}f_m^2}{f_m^2-1+\overline{g_0}/g_{\rm th}}. \tag{10}$$

The average value $\overline{F\cos\varphi}$ in Eq. (7) can be transformed at F close to unity and small φ, by neglecting small components of higher orders, to $\overline{F\cos\varphi}\approx\overline{F}-\overline{\varphi^2}/2$, and the data plotted in Fig. 6 indicate that $\overline{F}\approx f_m$. The combination of Eqs. (8) and (10) and the approximate expression for $\overline{F\cos\varphi}$ yield

$$\frac{f_m^2}{1+(g_{\rm th}/\overline{g_0})(f_m^2-1)}=1+\frac{2M}{g_{\rm th}}(1-f_m+\overline{\varphi^2}). \tag{11}$$

Usually $M\ll g_{\rm th}$. Hence, with due account of the smallness of the expression in the parentheses on the right of Eq. (11), we find that f_m^2 is close to unity. If we define $f_m=1+\kappa$, from Eq. (11) follows

$$\kappa=\frac{\overline{\varphi^2}/2}{1+(g_{\rm th}/M)(1-g_{\rm th}/\overline{g_0})}. \tag{12}$$

For the constants used in numerical calculations ($\overline{g_0}/g_{\rm th}=2$ and $M=0.1$, $g_{\rm th}=0.5$), we have $\kappa\approx\overline{\varphi^2}/7$. Averaging the last line in Eq. (2) and replacing the average of the product with the product of the averages, we obtain

$$\Delta/M=\overline{F\sin\varphi}\approx(1+\kappa)\overline{\varphi}(1-\overline{\varphi}^2/6)\approx\overline{\varphi}(1+\alpha\overline{\varphi}^2), \tag{13}$$

where the numerical factor α for the parameters of our calculations is small. Thus, the average phase difference is $\overline{\varphi}\simeq\Delta/M$, and the average ratio between the field amplitudes $\overline{F}\simeq 1+(\Delta/M)^2/6$. Comparison to numerical calculations reveals extremely good agreement between the approximate formulas and accurate calculations over an unexpectedly broad range, $\Delta/M\leqslant 1$.

Figure 8 shows that the solution is attracted to a certain compact set of parameters near the average values of field and phase differences. The small dimension of this set is the key factor for separating a signal from its chaotic carrier. The dispersion $\sigma(\Delta/M)$ defined by Eq. (4) is plotted in Fig. 9. It is small over the interval extending to $\Delta/M=1.5$, and beyond this value the ''strange'' synchronization is disrupted. Note that the dynamic chaos discussed here is equivalent to noise in an optical communication line.

4. INFORMATION DECODING

Let us consider the problem of retrieving an encoded signal in the system of two coupled chaotic lasers. A signal $f_k(t)$ is impressed on the output radiation by means of modulation at a modulation amplitude much smaller than the radiation intensity (within 10%). It turns out that such external perturbations do not destroy the effect of ''strange'' synchronization. In order to incorporate the signal into equation system (1), it suffices to modify the last line:

$$\dot{\mathscr{E}}_2=\frac{1}{2}(g_2-g_{\rm th}-i\Delta)\mathscr{E}_2+M[\mathscr{E}_1(1+f_k(t))-\mathscr{E}_2]. \tag{14}$$

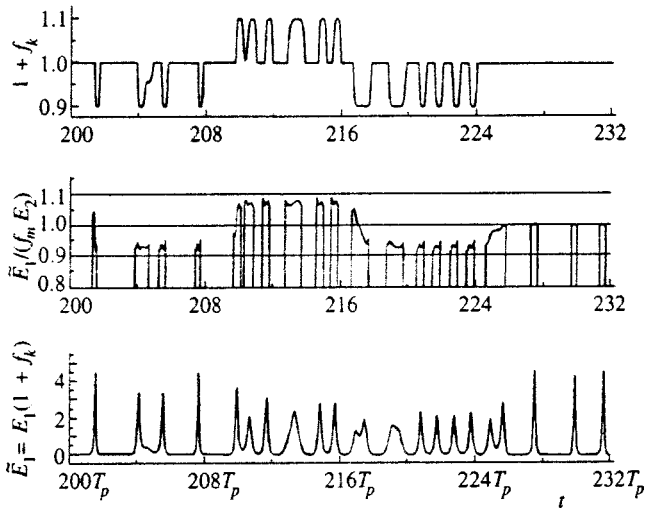

FIG. 10. The original message signal (upper curve), retrieved signal(middle curve), and signal carrying information at the slave laser input (lower curve). The ratio between the amplitudes was processed, $\Delta=0$.

The signal in our calculations is a random function of time, $f_k(t)$, which takes three values: ± 0.1, which correspond to one bit of information, and zero, which denotes the absence of the signal. The width of a step with a given value f_k comprises four to six random pulses (see the upper traces in Figs. 10–12, which show the function $1+f_k$). To avoid abrupt perturbations of the system by jumps in the signal intensity when f_k switches to another value, the signal was smoothed above a certain amplitude. As a result, the function $1+f_k$ has the shape of smooth pulses synchronized with those of field E_1 (compare the upper and lower traces in Figs. 10–12).

It was shown previously[9] that the signal bandwidth should be several times less than the frequency of relaxation oscillations, so the time interval between state changes of the function f_k is fairly long. The ten-percent amplitude modulation is too small for visual detection of the signal against the background of random laser pulses (see the third curves in Figs. 10–12). On the other hand, this modulation amplitude is higher than the dispersion characterizing the dynamic spread of the field amplitudes about the state of generalized synchronization (Fig. 9). Therefore, one can retrieve the signal $\tilde{\mathscr{E}}_1=\mathscr{E}_1(1+f_k)$ by calculating the difference between the fields with due account of the complex optimal approximation factor, $\tilde{\mathscr{E}}_1-f_m\exp(i\varphi_m)\mathscr{E}_2$.

The middle traces in Figs. 10 and 11 show numerical calculations of the retrieved signal for the cases of $\Delta=0$ and $\Delta=M$, respectively. It is clear that the accuracy of the signal transmission deteriorates as the frequency offset increases. Note that only changes in the amplitude ratio were processed.

In the case of a nonzero frequency offset, the phase difference between the fields can be used as a set of auxiliary data. The results of this processing are shown in Fig. 12 for $\Delta=M$. The data processing procedure taking into account both the amplitude ratio and phase difference yield the signal with a higher accuracy. The accuracy of the signal process-

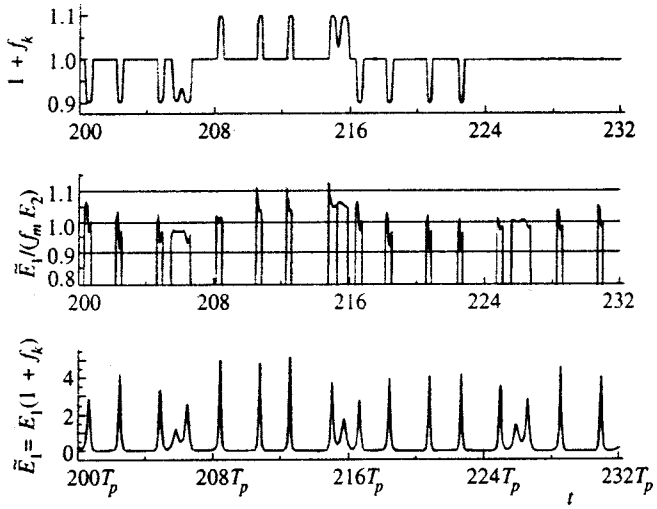

FIG. 11. The same as in Fig. 10, but at the frequency offset $\Delta = M$.

ing can be also improved by taking into account repeating distortions of the signal in the first pulses after steps in the modulation function (see the first and second traces in Figs. 10 and 11). In the general case, techniques for separation of a signal from noise[12] can be used to retrieve the signal with higher accuracy.

It is natural that the signal restoration efficiency is greatest when the frequency offset is zero. Nonetheless, the efficiency of restoration of the retrieved random signal carrying the message $[\tilde{E}_1/(f_m E_2) - 1]$ is less than 100% even in this case, coming in at about 2/3 of the original level. At the frequency offset $\Delta = M$ (Fig. 11) the restored signal is no greater than half the original signal, if the averaging is performed within the laser pulse. Note that the feasibility of retrieving information depends not on the signal amplitude, but on the possibility of identifying pulses carrying bits of information. The waveforms of retrieved signals are indented considerably, which makes identification of bit pulses more

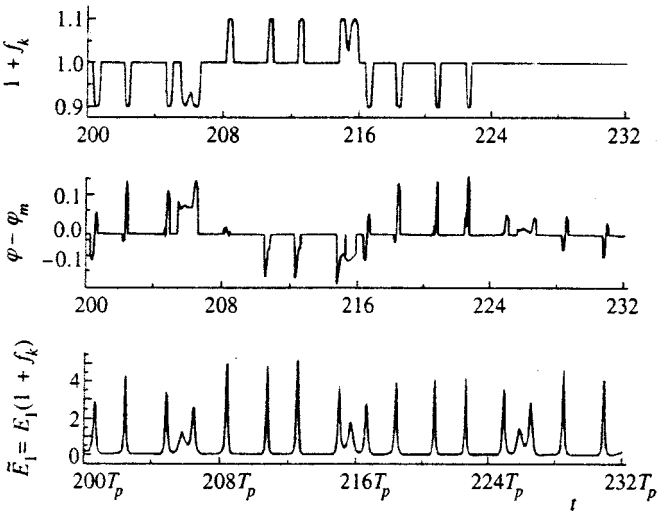

FIG. 12. The same as in Fig. 11, but the signal was retrieved by processing the difference between field phases.

difficult. Note that the phase response to an input signal is zero in the absence of a frequency offset and increases with the offset parameter, whereas the amplitude response to a modulated input signal drops (compare Figs. 10 and 11).

Let us conclude this section with a remark on the coding using phase modulation of the injection field. Numerical calculations indicate that in this case the slave laser is synchronized with the injected modulated radiation, instead of the input signal. Using Eq. (14), we derive from Eq. (1)

$$\dot{\varphi} = \Delta - MF \sin(\varphi + \eta),$$

$$\dot{F} = MF[1 - F\cos(\varphi + \eta)] - \frac{1}{2}(g_2 - g_1)F, \tag{15}$$

where η is the phase of function f_k. Using the change of variables $\tilde{\varphi} = \varphi + \eta$, we obtain the equations in their original form. Therefore, we expect that in a time of order $(MF)^{-1}$, a state with a new phase is established. Thus, one cannot retrieve encoded information in the case of phase modulation.

5. CONCLUSIONS

The study of synchronization between two chaotic lasers through unidirectional injection of a signal from the master oscillator has demonstrated that a new mode of synchronization can be implemented in the presence of frequency offset between laser cavity lengths, and this synchronization mode can be dubbed 'strange.' The output of the slave laser is almost synchronized with the injected signal, since their intensities and phases are related in a definite way, but with small errors that vary randomly with time. The relations between the average field parameters have been obtained in the reported work both numerically and analytically. The chaotic deviations of field parameters from those prescribed by deterministic equations play the role of noise when the described system of chaotic lasers is used for transmitting confidential information. At a sufficiently high amplitude of modulation of laser pulses by a signal carrying a message, the encoded signal can be retrieved by comparing the intensity of the beam carrying the information with the output of the slave laser.

This work was supported by the Russian Fund for Fundamental Research (Grant No. 98-02-17096 and 96-02-19203).

*)E-mail: apn@fly.triniti.troitsk.ru

1)The average straight lines for the two parameters are different, and the difference between their slopes depends on the frequency offset. As $\Delta \to 0$, the slope of Re $\mathscr{E}_2$ tends to 45°, whereas that of Im $\mathscr{E}_2$ tends to zero.

[1] P. Colet and R. Roy, Opt. Lett. **19**, 2056 (1994).
[2] G. D. Van Wiggeren and R. Roy, Science **279**, 1198 (1998).
[3] R. Roy and K. S. Thorburg, Jr., Phys. Rev. Lett. **72**, 2009 (1994).
[4] V. V. Likhanskii, A. P. Napartovich, and A. G. Sukharev, Quantum Electron. **25**, 40 (1995).
[5] Yu. I. Neimark and P. S. Landa, *Stochastic and Chaotic Oscillations* [in Russian], Nauka, Moscow (1987).
[6] M. G. Rosenblum, A. S. Pikovsky, and J. Kurths, Phys. Rev. Lett. **76**, 1804 (1996); U. Parlitz, L. Junge, W. Lauterborn, and L. Kocarev, Phys. Rev. E **54**, 2115 (1996).

[7] M. G. Rosenblum, A. S. Pikovsky, and J. Kurths, Phys. Rev. Lett. **78**, 4193 (1997).
[8] J. F. Heagy, T. L. Carroll, and L. M. Pecora, Phys. Rev. E **52**, R1253 (1995).
[9] A. P. Napartovich and A. G. Sukharev, Quantum Electron. **28**, 81 (1998).
[10] Ya. I. Khanin, *Dynamics of Quantum Oscillators* [in Russian], Sovetskoe Radio, Moscow (1975).
[11] V. V. Likhanskii, A. P. Napartovich, and A. G. Sukharev, Proc. SPIE **2800**, 16 (1996).
[12] S. A. Akhmanov, Yu. E. D'yakov, and A. S. Chirkin, *Introduction to Statistical Radiophysics and Optics* [in Russian], Nauka, Moscow (1981), p. 241.

Translation provided by the Russian Editorial office

A theory of two-photon light generation involving inversionless two-level atoms

V. N. Gorbachev*) and A. I. Trubilko†)

St. Petersburg Institute of the Moscow State University of Publishing, 190000 St. Petersburg, Russia
(Submitted 13 October 1998)
Zh. Éksp. Teor. Fiz. **115**, 1605–1618 (May 1999)

We propose a theory that makes it possible to examine both laser and micromaser two-photon light generation. The theory is based on the Lamb–Scully model, in which two-level atoms have initial coherence. Such an initial state allows for light generation involving atoms without inversion or even atoms with negative population inversion in the transition. We analyze the conditions needed for the emergence of squeezed states of the light field. We also show that the photon statistics in lasing is always classical but that there appears a state of the field squeezed in the phase with maximum photodetection shot-noise suppression. For a micromaser we find a regime where sub-Poisson light with a shot-noise suppression level of 60% is realized in an inversionless transition. In another regime, a squeezed state that produces 75% suppression of noise can be generated. © *1999 American Institute of Physics.* [S1063-7761(99)00605-8]

1. INTRODUCTION

Lately there has been an upsurge of interest in studies of the interaction of light and atomic systems without inversion.[1] The common approach is to develop three-level cavity models of atoms with and without relaxation and with different types of multimode interaction. Of interest are the study of properties of generation of light fields as a light source of a new type, the detection of the selective dynamics of the system, dynamic chaos, and multistable regimes[2] and the possibility of extracting (from spectroscopic measurements of probe radiation) information about coherent interactions in the system and about mode competition.[3]

Ordinarily, models of laser and micromaser light generation based on the Lamb–Scully theory examine incoherent pumping, which excites a two-level atom to the upper of lower active level.[4] In general, however, a mixed state or atomic coherence may emerge. For a maser or micromaser this may due, for example, to a spread in velocities in the atomic beam interacting with pulses of the pump field. Here, due to the Doppler shift of the signal transition frequency, it is impossible to push all the atoms into the upper state, so that some of them end up in the lasing area in a mixed state. The fact that the atoms in an atomic beam have different kinetic energies was investigated by An *et al.*,[5] who numerically analyzed the equation for the density matrix of the field with allowance for averaging over the velocities. However, by controlling the pump pulse we can create mixed states of the atom. Here the properties of the generated system prove to be different. Such lasing and masing under conditions of one-photon interaction of two-level atoms was analyzed by Lu and Bergou.[6] Scully *et al.*[7] demonstrated the possibility of squeezed field states forming in light generation involving cascade transitions, provided that there is atomic coherence between the upper and lower state of a three-level system, and Lu[8] did the same for Λ- and V-type transitions.

The aim of our investigation is to analyze optical systems where pumping creates atomic coherence involving the active levels. We examine the effect of the atomic coherence in the steady-state generation regimes of two-photon transitions, on the linewidth, and on the formation of photon statistics. Following resonant two-photon interaction between the coherence pump pulse and the active atoms, which are modeled by the two-level system, the atoms are injected into an optical cavity. The presence of atomic coherence gives rise to a steady-state lasing regime when there is negation population inversion on the active levels and when the populations of these levels are equal. Due to phase locking, the signal lineshape becomes monochromatic.

The two factors, the presence of initial coherence in the atomic beam and the nonlinearity of the two-photon transition, allows phase-squeezed laser light to form with maximum noise suppression in photodetection. The photon statistics of the laser field always remains classical. For a two-photon micromaser, the field in the ordinary case possesses quantum features, which manifest themselves in the photon statistics, with shot-noise suppression not exceeding 33% (see Ref. 9). We establish the conditions needed for stronger noise suppression in an inversionless transition. A phase-squeezed field state can also be formed in such a transition.

Ordinarily, the Lamb–Scully model is used to derive a kinetic equation for the density matrix of the field, which is the starting point in analyzing the photon statistics. This is an operator equation, which is written either in operator form or in the diagonal representation. In this paper we develop a formalism that makes it possible to immediately write the Fokker–Planck equation for the Glauber quasiprobability of a two-photon laser and micromaser. The method by which this equation is derived for the Lamb–Scully model is described in Sec. 2. In Sec. 3 we calculate the coefficients of this equation with allowance for an arbitrary state of the active atoms injected into the cavity. The model used in pumping the atomic beam is discussed in Sec. 4. In Sec. 5 we discuss lasing (in particular, involving an inversionless tran-

sition). The laser linewidth and the properties of the micromaser are elaborated on in Secs. 6 and 7.

2. THE FOKKER–PLANCK EQUATION FOR THE LAMB–SCULLY MODEL

We use the standard Lamb–Scully theory to examine the two-photon interaction of two-level atoms with a secondary signal transition that are injected with a rate r into a high-Q cavity. The starting equation is that for the density matrix F. This equation describes the interaction of a single atom and a field:

$$\frac{\partial F}{\partial t}=[V,F]+RF, \tag{1}$$

$$V=g(A_{ba}a^{\dagger 2}-a^2A_{ab}), \tag{2}$$

where $A_{ab}=|a\rangle\langle b|$ is a single-atom operator that transfers the atom from the lower active level b to the upper level a, and g is the coupling constant reflecting the interaction with a cavity mode, whose creation and annihilation operators are denoted by $a^\dagger$ and a, respectively. The exit of the field from the cavity and the departure of the atom from the active levels are described by the relaxation operator R. The equations are written for the case of exact resonance. Since the atom enters the cavity at time $t=0$ in a given state $F_A(0)$, we have

$$F(0)=F_A(0)\rho(0), \tag{3}$$

where $\rho=\mathrm{Tr}_A F$ is the density matrix of the field.

To write the kinetic equation for the density matrix ρ, we introduce the usual assumption that the field varies slowly. According to this approximation, in the time interval T during which the atom interacts with the field and the atom passes through the cavity, we have $\rho(T)\approx\rho(0)$, with $F_{mn}(T)=0$ for $m,n=a,b$. The change of the density matrix of the field brought on by this single atom, $\delta\rho=\rho(T)-\rho(0)$, can be found by solving Eq. (1) with allowance for (3):

$$F(T)-F_A(0)\rho(0)=\int_0^T([V,F]+RF)\,dt. \tag{4}$$

Now we introduce the change of the density matrix of the field due to the contribution of rT atoms, $\Delta\rho=rT\,\delta\rho$. Then the kinetic equation can be written

$$\frac{\Delta\rho}{T}=r\,\mathrm{Tr}_A\int_0^T[V,F]\,dt+R_F\rho, \tag{5}$$

where the left-hand side is a large-scale time derivative, and the operator R_F describes the exit of the field from the cavity.

A remark concerning the adopted approach is in order. If the atom–field interaction time T is short compared to characteristic atomic relaxation times, the method can be used to describe a maser or micromaser, in which case atomic relaxation should be dropped from the starting equations (1) and (4). Usually the incoherent model is sufficient for a description of pumping, which requires knowing the initial state $F_A(0)$ of the atom, i.e., it is assumed that the atom injected into the cavity is on its upper or lower active level. Here we allow for the general case where pumping produces an atomic coherence $\langle a|AF_A(0)|b\rangle$. A convenient way to do this is to employ the approach developed in Ref. 10. This makes it possible to immediately write the initial Fokker–Planck equation for the Glauber quasiprobability $P(\alpha)$, bypassing the extremely involved derivation of an operator equation for the density matrix ρ, which is actually only an intermediate stage.

For the single-atom density matrix F we introduce the quasiprobability Φ as follows:

$$F=\int\Phi|\alpha\rangle\langle\alpha|\,d^2\alpha. \tag{6}$$

This quasiprobability is an operator function of the atomic variables. Averaging it over the atoms yields the Glauber quasiprobability $\mathrm{Tr}_A\Phi=P(\alpha)$. From Eq. (1) we arrive at an equation for Φ:

$$\frac{\partial\Phi}{\partial t}=[V_0,\Phi]+\partial(d\Phi)+R\Phi, \tag{7}$$

where

$$V_0=gA_{ba}\alpha^{*2}-gA_{ab}\alpha^2,$$

$$\partial(dX)=-\left[g\frac{\partial}{\partial\alpha}2\alpha^{*2}|b\rangle\langle a|X+g\frac{\partial}{\partial\alpha^*}2\alpha^2X|a\rangle\langle b|\right]$$

$$+\left[g\frac{\partial^2}{\partial\alpha^2}|b\rangle\langle a|X+g\frac{\partial^2}{\partial\alpha^{*2}}X|a\rangle\langle b|\right]. \tag{8}$$

Here the first term in square brackets is due to the noise of the atomic system, while the second term in square brackets is due to the diagonal representation of the field employed in this paper. For Φ we use the representation $\Phi(t)=P(0)\varphi(t)+\Pi(t)$, where φ is the single-atom density matrix: $\mathrm{Tr}_A\varphi=1$. Since the field develops slowly, we can assume that $P(0)\approx P(T)$. The function Π (here it is not a correlation matrix) has the properties

$$\Pi(0)=0,\quad \mathrm{Tr}_A\Pi\neq 0,\quad \mathrm{Tr}_F\Pi\neq 0,\quad \mathrm{Tr}_{AF}\Pi=0,$$

where the subscripts A and F denote the variables of the atomic and field systems over which averaging is done. Next we denote the averaging over atoms by $\mathrm{Tr}_A=\langle\cdots\rangle$. Equation (7) yields equations for P, φ, and Π:

$$\frac{\partial P}{\partial t}P=\langle\partial(d\varphi)\rangle P(0)+\langle\partial(d\Pi)\rangle+R_F'P, \tag{9}$$

$$\frac{\partial}{\partial t}\varphi=[V_0,\varphi]+R_A'\varphi, \tag{10}$$

$$\frac{\partial}{\partial t}\Pi=[V_0,\Pi]+\partial(d\varphi)P(0)+\partial(d\Pi)+(R_A'+R_F')\Pi, \tag{11}$$

where $\partial(dX)$ has been defined in (8), and R_F' and R_A' are the relaxation operators in the representation (6).

The change of the field brought on by a single atom can be found from (9):

$$\delta P = \int_0^T (\langle \partial(d\varphi)\rangle P(0) + \langle \partial(d\Pi)\rangle)\, dt, \tag{12}$$

where the relaxation term has been discarded. To obtain a Fokker–Planck equation for the quasiprobability P, the solution for Π must contain derivatives with respect to α of order not higher than the first. This can easily be done by solving Eq. (11), say, by the iteration method. As a result, instead of Eq. (5) we arrive at an equation for P,

$$\frac{\Delta P}{T} = r \int_0^T (\langle \partial(d\varphi)\rangle P(0) + \langle \partial(d\pi)\rangle)\, dt, \tag{13}$$

where the functions φ and π can be found by solving the problem

$$\frac{\partial}{\partial t}\varphi = [V_0, \varphi] + R_A' \varphi,$$

$$\frac{\partial}{\partial t}\pi = [V_0, \pi] + \partial\langle (d\varphi)\rangle P(0) + R_A' \pi.$$

Here $\varphi(0) = \varphi^0$ is determined by the initial state of the atom injected into the cavity, and $\pi(0) = 0$. This problem for the functions φ and π immediately yields a Fokker–Planck equation for the Glauber quasiprobability.

3. THE STARTING EQUATIONS

The equations for φ and π have the same structure. They can be represented in matrix form:

$$\partial_t f_{mn} = [V_0, f]_{mn} + (R_A' f)_{mn} + \Gamma_{mn}, \tag{14}$$

where

$$\Gamma = -g\left(\frac{\partial}{\partial \alpha} 2\alpha^* |b\rangle\langle a|\varphi + \frac{\partial}{\partial \alpha^*} 2\alpha\varphi |a\rangle\langle b| \right) P(0).$$

This implies that $f = \varphi$, where $\varphi(0) = \varphi^0$, if $\Gamma = 0$ and $f = \pi$, where $\pi(0) = 0$, if $\Gamma \neq 0$. Allowing for the fact that $P(0) \approx P(T)$, we write the kinetic equation (5) in the form

$$\partial_t P = -rg \frac{\partial}{\partial \alpha} 2\alpha \left[\int_0^T \varphi_{ab}(t)\, dt + \int_0^T \pi_{ab}(t)\, dt \right] P + \text{c.c.} \tag{15}$$

To write Eq. (15), we must find the solutions of the base system (14), which describes the evolution of the atoms in the fixed-field approximation over the time interval T. The nature of this evolution strongly depends on the ratio of T and the atom relaxation time γ^{-1}. Two cases are usually distinguished. For $T\gamma \gg 1$, when the atom reaches a steady state, the atom ''leaves'' the active levels and lasing is said to be achieved. For $T\gamma \ll 1$, when atomic relaxation can be neglected, γ can be set to zero and maser or micromaser oscillation is said to be achieved. In this case, after a time T has elapsed, the atom leaves the cavity, which is equivalent to the atom departing from the active levels

For a micromaser it is convenient to solve Eq. (14), where we must put $R_A' = 0$, by applying, say, the Laplace transformation: $f(s) = \int_0^\infty \exp(-st) f(t)\, dt$. Then

$$s f_{mn}(s) - f_{mn}(0) = [V_0, f(s)]_{mn} + \Gamma_{mn}, \tag{16}$$

and the coefficients of the kinetic equation can be found by solving Eq. (16):

$$\int_0^T f(t)\, dt = \int ds \frac{\exp(sT) - 1}{s} f(s). \tag{17}$$

For a laser we can use a simple relaxation model, with equal relaxation constants, and neglect decay in the signal transition:

$$(R_A' f)_{mn} = -\gamma f_{mn}. \tag{18}$$

Then for the coefficients of the kinetic equation we have

$$\int_0^T f(t)\, dt = f(s), \tag{19}$$

where $f(s)$ is the solution of (16) at $s = \gamma$.

Now we examine the case where $T\gamma$ has an arbitrary value. Then

$$\int_0^T f(t)\, dt = \int ds \frac{\exp(sT) - 1}{s} f(s + \gamma), \tag{20}$$

where $f(s)$ is still a solution of Eq. (16), and by replacing s with $s + \gamma$ we allow for relaxation in the form (18). As a result, the Fokker–Planck equation for the Glauber quasiprobability becomes

$$\partial_t P = (\partial_\alpha 2\alpha^* \Lambda_1 + \partial^2_{\alpha\alpha} \Lambda_2 + \partial^2_{\alpha\alpha^*} 2\alpha^* \Lambda_3) P + \text{c.c.}, \tag{21}$$

where the coefficients are specified by the relationships

$$\Lambda_1(s) = -rg \left\{ \frac{\varphi^0_{ab}}{s} + g\alpha^2 \frac{s(\varphi^0_{aa} - \varphi^0_{bb}) - 2g(\alpha^{*2}\varphi^0_{ab} + \text{c.c.})}{s(s^2 + 4g^2|\alpha|^4)} \right\} + \frac{1}{2} C\alpha,$$

$$\Lambda_2(s) = -rg^2 2\alpha^* \times \left\{ \frac{4g^2\alpha^3|\alpha|^2 \varphi_{aa}(s) + 2sg|\alpha|^2 \alpha \varphi_{ab}(s)}{s(s^2 + 4g^2|\alpha|^4)} \right\} - \left\{ -\Lambda_1 + \frac{1}{2} C\alpha \right\},$$

$$\Lambda_3(s) = -rg^2 2\alpha \left\{ -\frac{\varphi_{aa}(s)}{s} + \frac{2g^2|\alpha|^4 \varphi_{aa}(s) + sg\alpha^2 \varphi_{ba}(s)}{s(s^2 + 4g^2|\alpha|^4)} \right\}.$$

Here $\varphi_{mn}(s)$ is the solution of the system (16) with the initial data φ^0_{mn}, and C is the cavity width, which characterizes the rate at which the field leaves the cavity. In the terms in $\Lambda_2(\sigma)$ we have, as usual, specified the contributions due to noise of the atomic system (the first terms in braces) and to the P-representation for the field (the second term in braces). For lasing, in accordance with (19), we must put $s = \gamma$, and for describing a micromaser, with allowance for (17), we must put

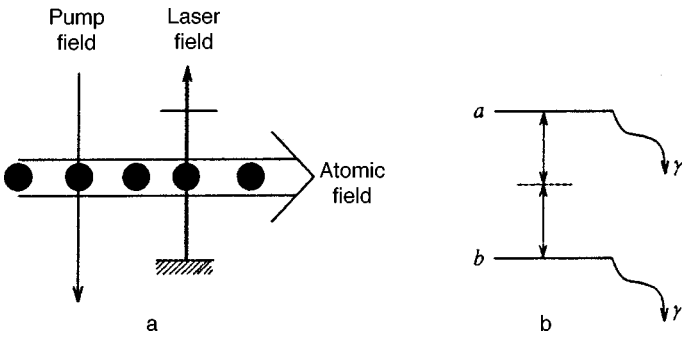

FIG. 1. (a) Optical scheme, and (b) active-level diagram.

$$\Lambda(s)\to\int ds\,\frac{\exp(sT)-1}{s}\Lambda(s).$$

Note that when $\varphi_{ab}^0=0$ holds we have the well-known equations of the micromaser and laser theories.[9,11]

4. THE PUMPING MODEL

The populations of the active levels, φ_{aa}^0 and φ_{bb}^0, and the coherence φ_{ab}^0 are determined by the mechanism of pumping and generally are not independent quantities. Suppose that the initial atoms or the atomic beam also interacts in a two-photon manner with a classical pump field, whose Rabi frequency is

$$\Omega=2g|\alpha_0|^2.$$

Suppose that the population of the lower active level is N_b. We assume (only to simplify matters) that $N_b\approx 1$. If the interaction time τ is much shorter than γ^{-1}, the state of the atoms is determined by the density matrix:

$$\varphi_{ab}^0=-\frac{1}{2}\exp\{2i\arg\alpha_0\}\sin\Omega\tau, \tag{22}$$

$$\varphi_{aa}^0-\varphi_{bb}^0=-\cos\Omega\tau, \tag{23}$$

$$\varphi_{aa}^0=\frac{1}{2}(1-\cos\Omega\tau), \tag{24}$$

where α_0 is the pump field amplitude. These solutions are well known.

Depending on the size of $\Omega\tau$, or ''pulse area,'' the following states of the atom are possible:

1. In the region $\pi/2<\Omega\tau\leqslant\pi$ the inversion is positive. At $\Omega\tau=\pi$ the atom is on the upper level, and the coherence φ_{ab}^0 is zero.

2. In the region $0<\Omega\tau\leqslant\pi/2$ the inversion is negative and the transition is inversionless. At $\Omega\tau=\pi/2$ the populations of both levels are the same and equal to 1/2, and the coherence is at its maximum: $|\varphi_{ab}^0|=1/2$.

5. LASING IN A MEDIUM WITHOUT INVERSION: PHOTON STATISTICS

In the optical scheme depicted in Fig. 1a the classical pump field produces populations φ_{aa}^0 and φ_{bb}^0 on the active levels and the coherence

$$\varphi_{ab}^0=\tilde{\varphi}_{ab}^0\exp\{2i\varphi_0\}, \tag{25}$$

which have been defined in Eqs. (22)–(24). The atoms are injected into a high-Q cavity where they interact with the cavity mode whose frequency is half the frequency of the signal transition. The rates at which the atoms leave the active levels are the same and equal to γ, as shown in Fig. 1b. Here we examine the case $T\gamma\gg 1$, which corresponds to lasing, as described by Eq. (21).

If in Eq. (21) we keep only first-order derivatives with respect to α, we find the solution $P(\alpha,t)=\delta(\alpha-z(t))$ of Eq. (21). Here z is the complex-valued amplitude of the field in the cavity (with $|z|^2=n$ the average number of photons) and satisfies the semiclassical problem

$$\partial_t n=2An^2\frac{\varphi_{aa}^0-\varphi_{bb}^0}{1+\beta n^2}+\frac{4An}{\sqrt{\beta}}\frac{\tilde{\varphi}_{ab}^0\cos\psi}{(1+\beta n^2)}-Cn, \tag{26}$$

$$\partial_t\psi=-4\frac{A}{\sqrt{\beta}}\tilde{\varphi}_{ab}^0\sin\psi, \tag{27}$$

where $A=(2g^2/\gamma)(r/\gamma)$ is the linear gain, r/γ is the number of atoms in the cavity, βn^2 is the dimensionless intensity, $\beta=4(g/\gamma)^2$ is the saturation parameter, $\psi=2(\arg z-\arg\alpha_0)$ is the phase difference of the laser field and coherence, $\tilde{\varphi}_{ab}^0$ is defined according to (25), and C is the cavity width, which characterizes the rate at which the field leaves the cavity.

The presence of initial coherence is responsible for two features of lasing. First, an additional term appears in the equation for n. Second, the phases of the laser field and of the initial coherence become locked. The latter results in a situation in which a phase difference sets in the steady-state regime, and for this difference

$$\sin\psi_0=0. \tag{28}$$

Here the value of ψ_0 is determined by the initial coherence:

$$\tilde{\varphi}_{ab}^0\cos\psi_0>0. \tag{29}$$

This condition follows from the requirement that the steady-state regime be stable. Combining (29) with (22), we get

$$\tilde{\varphi}_{ab}^0\cos\psi_0=\frac{1}{2}|\sin\Omega\tau|.$$

Then the steady-state number of photons in the cavity is given by the expression

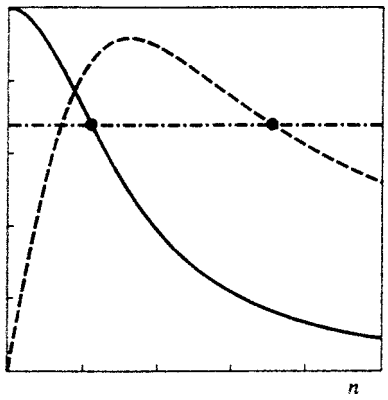

FIG. 2. Graphical solution of Eq. (26) in the steady-state case for a two-photon inversionless laser (the solid curve); the case of an ordinary two-photon laser is depicted by the dashed curve. The dash–dot straight line represents the loss level. The heavy dots represent stable lasing regimes.

$$\frac{2An}{1+\beta n^2}\left(-\cos\Omega\tau+\frac{|\sin\Omega\tau|}{\sqrt{\beta}\,n}\right)=C. \tag{30}$$

At $\Omega\tau=\pi$, Eq. (30) yields the ordinary (for a two-photon laser) condition for steady-state lasing, where the necessary threshold condition for the gain is

$$A>C\,\frac{1+\beta n^2}{2n}. \tag{31}$$

At $\Omega\tau=\pi/2$ lasing is stable and the condition (31) is not needed. For $\Omega\tau<\pi/2$ the inversion at the signal transition is negative. The steady-state regime emerges if

$$\frac{|\sin\Omega\tau|}{\sqrt{\beta}\,n}>\cos\Omega\tau. \tag{32}$$

Note that for an inversionless laser the conditions for lasing are significantly less stringent than in the case of an ordinary laser, whose stable lasing regimes are depicted by dots in Fig. 2.

To determine the photon statistics, we employ the condition of small fluctuations for the number of photons in the cavity:

$$|\alpha|^2=n+\epsilon,\quad \epsilon\ll n, \tag{33}$$

where n has been defined in Eq. (30). Under the conditions (33), we have an equation for the quasiprobability $R(\epsilon,t)=\int P(\alpha,t)\,d(\arg\alpha)$, (which describes photon-number fluctuations in the steady-state regime):

$$\partial_t R=(\Gamma_\epsilon\partial_\epsilon\epsilon+D_{\epsilon\epsilon}\partial_\epsilon^2)R, \tag{34}$$

where the coefficients are

$$\Gamma_\epsilon=\Gamma_0\frac{2\sqrt{\beta}\,n|\sin\Omega\tau|(\beta n^2-1)^{-1}-\cos\Omega\tau}{|\sin\Omega\tau|/\sqrt{\beta}\,n-\cos\Omega\tau},$$

$$D_{\epsilon\epsilon}=D_0\left(\frac{|\sin\Omega\tau|}{\sqrt{\beta}\,n}-\cos\Omega\tau\right)^{-1}\left(2\,\frac{1+\beta n^2}{5+\beta n^2}+\frac{|\sin\Omega\tau|}{\sqrt{\beta}\,n}\,\frac{1-3\beta n^2}{5+\beta n^2}-\cos\Omega\tau\,\frac{3-\beta n^2}{5+\beta n^2}\right).$$

Here

$$\Gamma_0=C\,\frac{\beta n^2-1}{\beta n^2+1},\quad D_0=\Gamma_0 n\xi,\quad \xi=\frac{5+\beta n^2}{2(\beta n^2-1)}.$$

The quantities Γ_0 and D_0 and the Mandel parameter ξ correspond to the case of an ordinary two-photon laser, for which we must put $\Omega\tau=\pi$.

The diffusion coefficient $D_{\epsilon\epsilon}$ in the Fokker–Planck equation for R always proves to be positive (for an ordinary laser and in conditions of lasing without inversion), and the photon statistics is classical; more than that, it is always super-Poisson.

6. THE SPECTRAL CONTOUR OF TWO-PHOTON LASER RADIATION; PHASE-SQUEEZED LIGHT

The spectral contour of a lasing line is determined by the Fourier transform of the correlation function $G(\tau)=\langle a^+(0)a(\tau)\rangle$:

$$G(\Omega)=\int_{-\infty}^{+\infty}\exp(i\Omega\tau)\,G(\tau)\,d\tau. \tag{35}$$

To find $G(\Omega)$, we use the small-fluctuation condition, assuming that the fluctuations μ of the field phase are small:

$$\varphi=\varphi_{\rm st}+\mu,\quad \mu\ll 1,\quad \varphi_{\rm st}=(1/2)(\psi_0+\varphi_0), \tag{36}$$

where ψ_0 and φ_0 have been defined in (29) and (25). Then the expression for $G(\tau)$ can be written in the form

$$G(\tau)=n\left[1+\frac{1}{4n^2}\langle\epsilon(0)\epsilon(\tau)\rangle+\langle\mu(0)\mu(\tau)\rangle\right]\times\exp(-i\omega\tau), \tag{37}$$

where n is the steady-state number of photons in the cavity given by (30), and $\langle\epsilon(0)\epsilon(\tau)\rangle$ is the correlation function of the photon-number fluctuations, which can be by found by solving Eq. (34). The phase-fluctuation correlation function $\langle\mu(0)\mu(\tau)\rangle$ can be found by solving an equation that follows from (21) under small-fluctuation conditions:

$$\partial_t Y=(\Gamma_\mu\partial_\mu\mu+D_{\mu\mu}\partial_\mu^2)Y,$$

$$Y(\mu,t)=\int P(\alpha,t)d(|\alpha|^2), \tag{38}$$

where

$$\Gamma_\mu=\frac{2A|\sin\Omega\tau|}{\sqrt{\beta}}=\frac{C|\sin\Omega\tau|(1+\beta n^2)}{|\sin\Omega\tau|-\sqrt{\beta}\,n\cos\Omega\tau}, \tag{39}$$

$$D_{\mu\mu}=\frac{\Gamma_\mu}{8n}\left\{-\frac{1+3\beta n^2+2\beta^2n^4}{(1+\beta n^2)^2}+\frac{\sqrt{\beta}\,n}{|\sin\Omega\tau|}\times\left(2-\frac{\cos\Omega\tau}{1+\beta n^2}\right)\right\}. \tag{40}$$

In contrast to an ordinary laser, here the lasing phase has a steady-state distribution for $\Gamma_\mu\neq 0$. This is a consequence of phase locking, which gives rise to a spectral line consisting

of a monochromatic component and two Lorentzians at the lasing frequency due to amplitude and phase fluctuations:

$$G(\Omega)=n\left[\delta(\Omega-\omega)+\frac{D_{\epsilon\epsilon}}{2n^2\Gamma_\epsilon}\frac{\Gamma_\epsilon^2}{(\Omega-\omega)^2+\Gamma_\epsilon^2}+\frac{2D_{\mu\mu}}{\Gamma_\mu}\frac{\Gamma_\mu^2}{(\Omega-\omega)^2+\Gamma_\mu^2}\right]. \quad (41)$$

We introduce the parameter

$$\eta=4n\frac{D_{\mu\mu}}{\Gamma_\mu}, \quad (42)$$

whose negative value indicates that the phase quadrature of the field is squeezed. The field in such a squeezed state may exhibit a dip below the shot-noise level in the low-frequency region of the noise spectrum in heterodyne reception:

$$i^2(\Omega)=1+Q\int_{-\infty}^{+\infty}\exp(i\Omega\tau)\langle\mu(0)\mu(\tau)\rangle\,d\tau, \quad (43)$$

where normalization has been chosen so that the shot-noise level is unity, and Q is the quantum efficiency of heterodyne reception.

The observable quantity is the relative reduction in the noise, which in terms of the introduced notation can be written

$$\delta=i^2(0)-1=2Q\frac{C}{\Gamma_\mu}\eta. \quad (44)$$

For an inversionless two-photon laser, as well as for a laser with negative population inversion, the limit of the phase-squeezed state of the lasing field is realized at low intensities $\sqrt{\beta}\,n\ll 1$, with $\eta\to -1/2$ and the noise in the region of $\Omega\approx 0$ completely suppressed ($\delta\to -1$).

7. AN INVERSIONLESS TWO-PHOTON MICROMASER

In the case of an inversionless two-photon micromaser, $T\gamma\ll 1$. The semiclassical masing equation has the form

$$\partial_t n=2r(\varphi_{aa}^0-\varphi_{bb}^0)\sin^2 I+2r\tilde{\varphi}_{ab}^0\cos\psi\sin I\cos I-Cn, \quad (45)$$

$$\partial_t\psi=-4rgT\tilde{\varphi}_{ab}^0\sin\psi, \quad (46)$$

where now the dimensionless intensity is defined as $I=gnT$, with $n=|z|^2$. As in the case of a laser, in the steady-state regime the phases of the maser field and of the initial coherence become locked, which leads to condition (29). As a result, the average number of photons in the cavity can be found from the condition

$$2r(\varphi_{aa}^0-\varphi_{bb}^0)\sin^2 I+r|\tilde{\varphi}_{ab}^0|\sin I\cos I=Cn. \quad (47)$$

To establish the photon statistics by employing the small-fluctuation approximation, we write Eq. (34) for the distribution function R, where the coefficients are given by the formulas

$$\Gamma_\epsilon=\frac{\Gamma_0}{|\sin\Omega\tau|-2\cos\Omega\tau\tan I}\left\{\cos\Omega\tau\frac{2(2I-\tan I)}{1-2I\cot I}+|\sin\Omega\tau|\left(1+\frac{I(\tan I+\cot I)}{1-2I\cot I}\right)\right\}, \quad (48)$$

$$D_{\epsilon\epsilon}=\frac{2D_0}{1+4I\cot I}\times\left\{\frac{2\tan I+\cos\Omega\tau(\tan I-4I)+(1/2)|\sin\Omega\tau|(1-4I\cot I+4I\tan I)}{|\sin\Omega\tau|-2\cos\Omega\tau\tan I}\right\}. \quad (49)$$

Here the quantities

$$\Gamma_0=C[1-2I\cot I], \quad (50)$$

$$D_0=\frac{Cn}{2}[1+4I\cot I] \quad (51)$$

correspond to the case of an ordinary two-photon maser. The masing regimes specified by (47) are stable if $\Gamma_\epsilon>0$.

Using (48) and (49), we define the intracavity Mandel parameter ξ:

$$\xi=\frac{D_{\epsilon\epsilon}}{\Gamma_\epsilon n}. \quad (52)$$

The value of ξ determines the difference between the photon statistics and Poisson statistics: $\langle n^2\rangle-\langle n\rangle^2=n(1+\xi)$. For $\xi<0$ the state of the light is nonclassical with sub-Poisson statistics. In measurements of the photocurrent spectrum or the noise spectrum $i^2(\Omega)$ near $\Omega\approx 0$, the Mandel parameter determines the level of the surplus noise in relation to the shot-noise component:

$$i^2(0)=1+\delta,\quad \delta=2q\xi\frac{C}{\Gamma}. \quad (53)$$

Here q is the quantum efficiency of the detector, and the level of shot noise is assumed to be unity.

At $\Omega\tau=\pi$, when active atoms are injected into the cavity only at the upper active level, combining (53) with (50) and (51) yields the well-known result[9]

$$\delta=q\frac{1+2y}{(1-y)^2},\quad y=2I\cot I. \quad (54)$$

The value $y=-2$ is optimal for the formation of a dip in the photocurrent noise spectrum. Here $\delta=-q/3$, which corresponds to a reduction of noise in the low-frequency region to 33% of the shot-noise level.

Numerical analysis of (53) with (48) and (49) taken into account shows that the optimal minimum value of δ is realized in the case of the initial atomic state of the active atoms of the maser without inversion at $\Omega\tau=\pi/2$, i.e., when atomic coherence is at its maximum. Under this condition, a set of almost periodic states in which the minimum value of the parameter δ is approximately $-0.6q$ can be realized. This depends on the value of the parameter n and hence on the parameters of the maser system: the solution of Eq. (47) and the simultaneous requirement that Γ_ϵ be positive. Even a slight variation in n ''destroys'' the photon statistics, and in this sense the conditions for masing are stringent.

The evolution of the linearized phase [in conditions of small fluctuations; see (36)] of the maser field is determined by Eq. (38) in which the coefficients Γ_μ and $D_{\mu\mu}$, with allowance for the steady-state regime (47), have the form

$$\Gamma_\mu=2rgT|\sin\Omega\tau| = \frac{2CI|\sin\Omega\tau|}{\sin I\cos I(|\sin\Omega\tau|-2\cos\Omega\tau)\tan I}, \tag{55}$$

$$D_{\mu\mu}=\frac{\Gamma_\mu}{4n}\left[-1+\frac{I}{|\sin\Omega\tau|}-\frac{\cos\Omega\tau}{|\sin\Omega\tau|}\frac{\sin^2 I}{2I} + \frac{\sin I\cos I}{4I}\right]. \tag{56}$$

The maximum manifestation of the effect of phase squeezing for a two-photon micromaser field can also be observed in an inversionless atomic medium, $\Omega\tau=\pi/2$. The level of noise suppression (44) in heterodyne reception near $\Omega\approx 0$ is given in this case by the formula

$$\delta=Q\frac{C}{\Gamma_{0\mu}}\left\{\frac{C}{\Gamma_{0\mu}}+\frac{\Gamma_{0\mu}}{2C}\sin 2I-2\right\}, \tag{57}$$

where Q is the quantum efficiency of heterodyne reception, and

$$\frac{C}{\Gamma_{0\mu}}=\frac{\sin I\cos I}{2I}.$$

Only with a generated maser field of low intensity, $I\ll 1$, do we have maximum suppression in the photodetection noise spectrum in the low-frequency region $\delta\to -3/4Q$. Note that these conditions are stable: both Γ_μ of (55) and Γ_ϵ of (48) are positive.

8. CONCLUSION

We have used the Lamb-Scully model to construct a theory that makes it possible to simultaneously study laser and micromaser light generation, a model that uses active atoms that initially were in a mixed state. Such an initial state allows generating light using atoms without population inversion and even a negative population inversion in the transition. Due to locking of the phase of the coherence and the phase of the generated field, the laser (maser) line is monochromatic.

We have analyzed the conditions under which squeezed states of light appear. We have found that the photon statistics in lasing is always classical but that it is also possible to generate phase-squeezed states of the field, states characterized by complete suppression of heterodyne photodetection noise. We have also found that the photon statistics of micromaser light may prove to be nonclassical and that light can be generated by an inversionless transition, with the level of photodetection-noise suppression amounting to about 60% of the relatively standard quantum limit, which is higher than in conditions of ordinary two-photon generation of light by atoms in the excited upper state. The same type of transition can be used to generate a phase-squeezed field with a level of noise suppression in heterodyne photodetection amounting to 75% of the shot-noise level.

*)E-mail: vn@vg3024.spb.edu
†)E-mail: tai@at3025.spb.edu

[1] O. A. Kocharovskaya and Ya. I. Khanin, Zh. Éksp. Teor. Fiz. **90**, 1610 (1986) [Sov. Phys. JETP **63**, 945 (1986)]; M. O. Scully, S.-Y. Zhu, and A. Gavrielides, Phys. Rev. Lett. **62**, 2813 (1989); O. Kocharovskaya and P. Mandel, Phys. Rev. A **42**, 523 (1990).
[2] A. J. Merriam and A. V. Sokolov, Phys. Rev. A **56**, 967 (1997); S.-Y. Zhu, De-Zhong Wang, and J.-Y. Gao, Phys. Rev. A **55**, 1339 (1997).
[3] P. Bogár and J. A. Bergou, Phys. Rev. A **56**, 1012 (1997); O. A. Kocharovskaya, Yu. V. Rostovtsev, and A. Imamoglu, Phys. Rev. Lett. **58**, 649 (1998).
[4] M. O. Scully and W. E. Lamb, Jr., Phys. Rev. A **159**, 208 (1967); P. Filipowicz, J. Javanainen, and P. Meystre, Phys. Rev. A **34**, 3077 (1986); J. Krause, M. O. Scully, and H. Walther, Phys. Rev. A **34**, 2032 (1986); L. A. Lugiato, M. O. Scully, and H. Walther, Phys. Rev. A **36**, 740 (1987).
[5] K. An, J. J. Childs, R. R. Dasari, and M. S. Feld, Phys. Rev. Lett. **73**, 3375 (1994).
[6] N. Lu and J. A. Bergou, Phys. Rev. A **40**, 237 (1989); N. Lu, Phys. Rev. Lett. **70**, 912 (1993).
[7] M. O. Scully, K. Wodkiewicz, and M. S. Zubairy, J. Bergou, Ning Lu, and J. Meyer ter Vehn, Phys. Rev. Lett. **60**, 1832 (1988); N. Lu, Phys. Rev. Lett. **70**, 8154 (1993).
[8] N. Lu, Phys. Rev. Lett. **45**, 5011 (1992).
[9] V. N. Gorbachev, N. V. Slepukhina, and A. I. Trubilko, Opt. Spektrosk. **85**, 155 (1998) [Opt. Spectrosc. **85**, 141 (1998)].
[10] V. N. Gorbachev and A. I. Zhiliba, Quantum Opt. **5**, 193 (1993).
[11] Yu. M. Golubev, Opt. Spektrosk. **46**, 3 (1979) [Opt. Spectrosc. **46**, 1 (1979)].

Translated by Eugene Yankovsky

Polarization-interference effects in the bremsstrahlung of quasiclassical electrons on ions with a core

V. A. Astapenko*)

Moscow Institute for Physics and Technology, 141700 Dolgoprudnyi, Moscow Region, Russia
(Submitted 11 November 1998)
Zh. Éksp. Teor. Fiz. **115**, 1619–1629 (May 1999)

The polarization dependence of the stimulated bremsstrahlung and inverse bremsstrahlung (SBIB) of quasiclassical electrons on highly charged ions with a core is calculated in the approximation of a specified Coulomb current. Emission frequencies close to an eigenfrequency of the ion core are considered. The contributions of the static and polarization channels are taken into account in the amplitude of the process. When the nondipole nature of the interaction between the incident particle and a resonant transition in the ion core is taken into account, interference between these channels causes the spectral-amplitude characteristics of the process to assume a specific dependence on the angle α between the electric field intensity vector of the electromagnetic wave and the initial velocity vector of the incident particle. This dependence, which persists after integration of the cross section over the scattering angle of the incident particle, causes interference effects, viz., asymmetry of the line shape and dips in the dependence of the SBIB cross section on electric field intensity, to appear for $\alpha=\pi/2$ and significantly reduces them for $\alpha=0$. © *1999 American Institute of Physics.* [S1063-7761(99)00705-2]

1. INTRODUCTION

Investigations of the influence of the polarization of the electron shell of a target on collisional radiative processes have been carried out in recent years.[1–6] Along with bremsstrahlung on atoms and clusters,[2,6] stimulated bremsstrahlung (including the multiphoton effect) on highly charged ions[1,4,5] and recombination on ions[3] were examined.

The dynamic polarization of the electron core of a target opens up a new polarization channel for the radiative process along with the traditionally considered static channel.[7] As a result, the cross section is a sum of three terms, viz., static, polarization, and interference terms.

The static channel is associated with the emission of a photon by the incident particle as a result of its acceleration in the static potential of the target. The polarization channel can be interpreted as scattering of the self-field of the incident particle (virtual photon) into a real photon on the electron shell of the target. It occurs even when the acceleration of the incident particle is negelected. This process is similar to the scattering of radiation on an atom (an ion with an electron core), but, unlike the latter, always (except in some exotic cases) has a static ''addition,'' which facilitates interchannel interference. The interference term becomes especially significant for the target ion,[1] producing several interesting features in the spectral-amplitude characteristics of the bremsstrahlung: asymmetry of the shape of the spectral line and dips on it (for a near-resonant process), as well as a minimum in the spectral dependence of the recombination rate.[1,4,5] In fact, in the case of a target in the form of a neutral atom the regions of existence of the static and polarization channels differ sharply with respect to the incident particle scattering angle:[7] the polarization channel dominates for small angles, and the static channel dominates for large angles. Thus, the interference term contributes little to the total bremsstrahlung cross section on a neutral atom for all incident particle scattering angles, although for the angular differential cross section this interference (at certain frequencies) can cause the total spectral bremsstrahlung cross section to vanish for some incident particle scattering angles.[8]

We note here that the stripping of an atom in bremsstrahlung as a consequence of consideration of the polarization channel, which was first discovered in Ref. 9, is not an interference effect, but describes the addition to the ranges of incident particle scattering angle that are important for the static and polarization channels.

The situation for bremsstrahlung (and recombination) on ions with an electron core is qualitatively different. In this case the ranges of significant incident particle scattering angles for the two channels overlap, making the interference term in the total bremsstrahlung cross section important for all incident particle scattering angles. The role of this interference for the near-resonant stimulated bremsstrahlung of quasiclassical electrons on highly charged ions was investigated in Refs. 1, 4, and 5, and its role for recombination was examined in Ref. 10.

Stimulated bremsstrahlung in a laser field naturally also takes place along two channels, a static channel and a polarization channel. In the latter case the process can be described as stimulated scattering of the self-field of the incident particle into a laser mode.

We shall consider the situation in which the presence of a laser field weakly influences the motion of the incident particle, i.e, the approximation in which the current can be viewed as given (see the basis for it in Ref. 1), so that first-

order perturbation theory with respect to the interaction between the incident particle and the laser radiation can be used to calculate the amplitude of the static channel. At the same time, since a laser frequency close to one of the eigenfrequencies of the ion core is being considered, the interaction of the laser radiation with a near-resonant transition in the core can be strong. It can be taken into account within the rotating-wave approximation (for further details, see Ref. 1).

For our further discussion it is important to stress that the treatment in Refs. 1, 4, 5, and 10 was performed in the dipole approximation with respect to the interaction of the incident particle with the target. This approximation was justified for bremsstrahlung that is near-resonant to electron transitions in the ion core with $\Delta n=0,1$, since in the spirit of the rotational approximation[11] in the Kramers limit ($\omega \gg \omega_{\rm Coul}=v^3/Z_i$, where v is the velocity of the incident particle, Z_i is the charge of the ion, and atomic units are used everywhere) the radius of the portion of the incident particle trajectory responsible for emission of the specific (near-resonant) frequencies under consideration exceeds the radius of the outer shell of the ion, which corresponds to the upper level of the resonant transition in the ion (for further details, see below).

However, a more detailed treatment (not based on the rotational approximation) showed that the influence of the nondipole nature of the incident-particle–target interaction on the total cross section of the process for all incident particle scattering angles is very significant: it smears out the dips on the amplitude-frequency characteristics of the stimulated bremsstrahlung and eliminates the asymmetry of the spectral line shape for $\Delta n=0$ and, in addition, significantly reduces the bremsstrahlung cross section for $\Delta n=1$ over the entire near-resonant portion of the spectrum.[12]

The nondipole effects in stimulated bremsstrahlung and inverse bremsstrahlung (SBIB) were investigated in Ref. 12 for the case of an isotropic (with respect to the initial velocities) distribution of the incident particles, where the dependence of the process on the polarization of the external radiation vanishes.

The purpose of the present work is to take into account just this dependence. Therefore, below we calculate the SBIB cross section without averaging over the direction of the initial incident particle velocity for various angles between the initial velocity vector of the incident particle and the electric field intensity vector of the linearly polarized laser radiation.

2. BASIC FORMULAS

In Ref. 1 we justified the specified-current approximation for calculating the bremsstrahlung cross section on highly charged ions with a core for arbitrary values of the Born parameter $\eta=Z_i/v$. For sufficiently slow incident particles, which satisfy the relation $\eta\geqslant 1$, good results are provided by the so-called semiclassical approximation, in which the classical theory of bremsstrahlung is used together with several quantum restrictions.[13,14] Here we shall also employ the semiclassical approximation to calculate the polarization channel of SBIB in the part concerning the incident particle; we shall, of course, use a systematic quantum-mechanical approach[12] to calculate the characteristics of the bound core electrons.

Within this picture the spectral amplitude of the SBIB for the incident particle scattering angle ϑ [or the eccentricity of the orbit $\varepsilon=1/\sin(\vartheta/2)$] is the sum of two terms, one of which is proportional to the time Fourier transform of the dipole moment of the scattered incident particle [$\mathbf{D}^{\rm st}(\omega,\varepsilon)$ — the static channel], while the other is proportional to the Fourier transform of the dipole moment induced by the incident particle in the ion core [$\mathbf{D}^{\rm pol}(\omega,\varepsilon)$ — the polarization channel].

Next, let us consider SBIB at a frequency ω of the external radiation that is close to the eigenfrequency ω_0 of a transition in the electronic ion core of the form $(n_i,s)\rightarrow(n,p)$, assuming that the frequency mismatch $\Delta=\omega-\omega_0$ exceeds the linewidth of the transition, so that the real excitation of the ion core can be neglected.

The corresponding formulas have the form

$$d\sigma^{\rm tot}(\omega,\varepsilon,\varphi)=\frac{1}{4}\frac{Z_i^2}{v^4}|\mathbf{E}_0\mathbf{D}^{\rm tot}(\omega,\varepsilon)|^2\varepsilon d\varepsilon\, d\varphi, \tag{1}$$

$$\mathbf{D}^{\rm tot}(\omega,\varepsilon)=\mathbf{D}^{\rm st}(\omega,\varepsilon)+\mathbf{D}^{\rm pol}(\omega,\varepsilon), \tag{2}$$

$$D_j^{\rm pol}(\omega,\varepsilon)=\frac{1}{6\pi}\frac{\langle n_i,l=0||d||n,l=1\rangle}{\Omega}\int_{-\infty}^{\infty}dt\, e^{i\omega t}\frac{R_j(t,\varepsilon)}{R(t,\varepsilon)} \times\langle n,l=1||a_1(r_{\rm core},R(t,\varepsilon))||n_i,l=0\rangle, \tag{3}$$

$$a_1(r,R)=\theta(R-r)\frac{r}{R^2}+\theta(r-R)\frac{R}{r^2}.$$

Here $R_j(t,\varepsilon)$ is the jth projection of the radius vector of the incident particle for a given orbital eccentricity as a function of time; $\theta(x)$ is a Heaviside step function; $\Omega=(\Delta^2+\Omega_0^2)^{1/2}$ is the generalized Rabi frequency; $\Omega_0=d_0E_0$ is the resonant Rabi frequency; d_0 is the dipole moment matrix element of the resonant transition; and E_0 is the amplitude of the electric field in the external radiation.

The expression (3) was obtained in first-order perturbation theory with respect to the interaction of an incident particle with an electronic transition in an ion core + near-resonant laser field system. Its systematic derivation within the dressed-state model was given in Ref. 12. This approach permits allowance for the influence of the near-resonant laser field on the radiative transition in the ion core within the rotating-wave approximation and, consequently, to obtain the dependence of the SBIB cross section on the electric field intensity in the external radiation.

We note that consideration of the influence of the laser field on the near-resonant transition formally reduces to replacement of the frequency mismatch from resonance ($\Delta=\omega-\omega_0$) by the generalized Rabi frequency (Ω) in formula (3). The expression for the total cross section for all incident particle scattering angles has the form

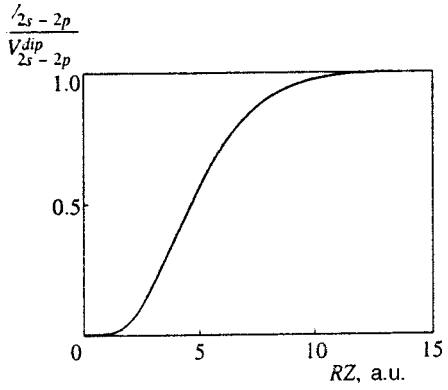

FIG. 1. Deviation of the interaction potential of an incident particle with the $2s-2p$ transition of a bound electron in a hydrogenic ion from its dipole approximation (Z is the charge of the ion nucleus).

$$d\sigma^{\rm tot}(\omega,\alpha)=\frac{\pi}{2}\frac{Z_i^2}{v^4}\mathbf{E}_0^2\int_1^\infty[(|D_x^{\rm tot}(\omega,\varepsilon)|^2)f_x(\alpha,\varepsilon)+(|D_y^{\rm tot}(\omega,\varepsilon)|^2)f_y(\alpha,\varepsilon)]\varepsilon^{-1}\,d\varepsilon, \tag{4}$$

$$f_x(\alpha,\varepsilon)=\cos^2\alpha+\frac{\varepsilon^2-1}{2}\sin^2\alpha,$$

$$f_y(\alpha,\varepsilon)=(\varepsilon^2-1)\cos^2\alpha+\frac{1}{2}\sin^2\alpha.$$

In these equations x and y are the axes of the focal coordinate system assigned by the initial and final incident particle velocity vectors,[14] and α is the angle between the initial incident particle velocity vector and the electric field intensity vector of the external radiation. The derivation of (4) from (1) involved integration over the azimuthal scattering angle φ.

Formulas (1)–(4) are the basic equations for our investigation of the polarization-interference effects in the SBIB of quasiclassical electrons on ions with a core.

3. INFLUENCE OF THE NONDIPOLE BEHAVIOR OF THE PROJECTILE-TARGET INTERACTION ON THE RADIATING DIPOLE MOMENT

Polarization-interference effects stem from the nondipole nature of the interaction between the incident particle and the near-resonant transition in the ion core. In the dipole approximation the interference term in the bremsstrahlung cross section does not depend on the polarization of the external radiation.[1,4,5] Thus, the difference between the incident-particle–ion-core interaction potential and the dipole analog (the incident-particle–point-ion interaction potential) is essential for the appearance of this dependence.

Figure 1 shows how the ratio of the exact interaction potential to the dipole potential varies with the parameter $x=ZR$ (where R is the distance between the incident particle and the nucleus) for an incident particle interacting with the $2s-2p$ transition in a hydrogenic ion.

A calculation in the model-potential approximation reveals that a similar dependence holds for transitions of the outer electron in a lithium-like ion.

It follows from Fig. 1 that a deviation from thedipole appproximation in the interaction of the incident particle with the near-resonant transition appears at $x_0\approx 12$ (for $\Delta n=0$), while the value of x corresponding to the radius of the $2p$ state of a hydrogenic ion is of order 4. Thus, nondipole behavior begins to show up at distances between the incident particle and the nucleus as small as three times the characteristic radius of the upper resonant state. As we know (see, for example, Ref. 15), in the Kramers limit the portion of the incident particle trajectory responsible for the emission of a photon of frequency ω is located at a distance of the order of $R_\omega(Z_i)\approx(Z_i/\omega^2)^{1/3}$ (for a Coulomb field) from the center of the field (this estimate corresponds to the rotational approximation[11,15]). In the case of lithium-like ions, the eigenfrequency of the $2s-2p$ transition can be approximated by the expression $\omega_{2s-2p}(Z)=0.0707Z-0.120$ a.u. (Z is the charge of the ion nucleus). The expressions presented can be used to estimate the nondipolarity parameter

$$\chi_{\rm nd}(x_\omega)=\frac{V_{2s-2p}(x_\omega)}{V_{2s-2p}^{\rm dip}(x_\omega)}$$

[here $x_\omega(Z_i)=Z_iR_\omega(Z_i)$] as a function of the charge of the lithium-like ion Z_i for a near-resonant $2s-2p$ transition. Simple calculations give

$$\chi_{\rm nd}(1,2,3,4,5)=0.25,0.8,0.96,0.99 \quad \text{for} \quad \Delta n=0.$$

Thus, this estimate, which is based on the use of the rotational approximation, shows that for transitions with no change in the principal quantum number in lithium-like ions the nondipolarity parameter is close to unity for fairly large ion charges ($Z_i\geqslant 4$). A similar estimate for a transition with $\Delta n=1$ shows that the corresponding nondipolarity parameter $x_\omega(Z_i)$ is significantly smaller than unity for all Z_i.

It should, however, be borne in mind that the rotational approximation corresponds to the replacement of summation of the contributions of different incident particle trajectories over the impact parameter by emission at a certain effective distance $R_\omega(Z_i)$. It is not clear *a priori* that such a replacement is correct for taking into account the interference of the static and polarization channels occurring for each fixed impact parameter, especially in the region of destructive interference of these channels, where the total cross section is small. As was mentioned above, the calculations performed in Ref. 12 in the approximation of a specified incident particle Coulomb current show that the rotational approximation is, generally speaking, inadequate for describing fine interference effects in the region of the spectral-amplitude dips, but remains approximately correct in the region of the constructive interference of the static and polarization channels for transitions with no change in the principal quantum number. Therefore, for a correct description of this interference over the entire range of values of the parameters we must start out from the general unsimplified formula (4).

Bearing in mind the calculation of SBIB for lithium-like ions, we shall henceforth use the Coulomb approximation for the incident particle current, in which the incident particle trajectory is assigned by the familiar classical expressions[14] for the motion of a charged particle in an attractive Coulomb

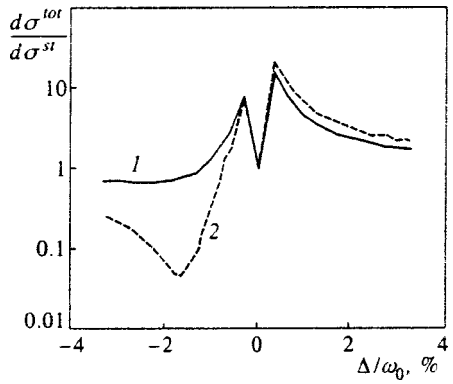

FIG. 2. Spectral dependence ($\Delta=\omega-\omega_0$) of the SBIB cross section of a quasiclassical electron on an N^{4+} ion normalized to the static values for two values of α ($\alpha=0$ — curve *1*, $\alpha=\pi/2$ — curve *2*) between the initial velocity vector of the incident particle ($v=0.6$ a.u.) and the electric field intensity vector of the external radiation ($E_0=10^{-3}$ a.u.) for a frequency of the external field near the eigenfrequency of a near-resonant transition in the ion core without alteration of the principal quantum number ($2s-2p$).

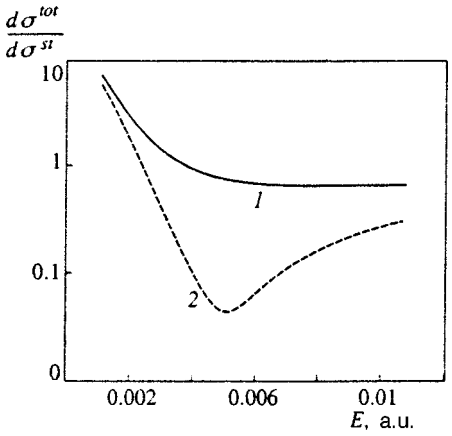

FIG. 3. Dependence of the SBIB cross section normalized to the static value on the electric field intensity of the external radiation for two values of α ($\alpha=0$ — curve *1*, $\alpha=\pi/2$ — curve *2*), a negative detuning from resonance $(\omega-\omega_0)/\omega_0=-0.3\%$ near the eigenfrequency of the $2s-2p$ transition in the N^{4+} ion, and an incident particle velocity $v=0.6$ a.u.

field with an effective charge equal to $Z_{\text{eff}}=Z_i+1/2$. A calculation of the dependence of the focal components of the induced dipole moment $D_{x,y}(\varepsilon)$ of the ion core on the eccentricity of the incident particle orbit shows that for transitions of bound electrons of the ion without a change in the principal quantum number the y projection of the dipole moment is close to its dipole analog, while the x projection of the dipole moment for eccentricities fairly close to unity differs strongly from the value calculated in the dipole approximation. In particular it vanishes and changes sign near $\varepsilon=1$ for a certain eccentricity of the incident particle orbit.

The difference just indicated underlies the dependence studied herein of the polarization-induced effects in SBIB on the angle between the initial incident particle velocity vector and the electric field intensity vector of the electromagnetic wave.

Physically, these features in the behavior of $D_x(\varepsilon)$ and $D_y(\varepsilon)$ stem from the different effects of penetration of the incident particle into the ion core on the projections of the Fourier transform of the induced dipole moment in the focal coordinate system. Penetration into the ion core has a stronger effect on the x component of the dipole moment than on the y component, since the x component of the radius vector of the incident particle changes sign during its motion from the point of closest approach to the center of the field to infinity, while the y component has a single sign on the portion of the trajectory under consideration.

4. DISCUSSION OF RESULTS

The results of the calculations of the SBIB cross section within the approach under consideration are presented in Figs. 2–4 and for spontaneous bremsstrahlung in Fig. 5. The calculations were performed for the $2s-2p$ and $2s-3p$ transitions in the lithium-like N^{4+} ion. The functions of the model-potential method were employed as the wave functions of the valence electron.

The total SBIB cross sections for all incident particle scattering angles are presented in all the figures.

Figure 2 shows the spectral dependence of the SBIB cross section normalized to the static value near the eigenfrequency of the transition with no change in the principal quantum number in the ion core for two values of α (the angle between the initial velocity vector of the incident particle and the electric field intensity vector of the external radiation) for an initial incident particle velocity $v=0.6$ a.u. It follows from this figure that the interference effects are displayed most vividly for $\alpha=\pi/2$, while they are reduced significantly for $\alpha=0$. This difference is greatest in the region where the static and polarization channels experience destructive interference ($\omega<\omega_0$), where the calculations in the dipole approximation with respect to the interaction of the incident particle with the core[4] exhibit a deep minimum, which is caused by mutual compensation of the polarization and static terms in the amplitude of the process. In the region of constructive interference ($\omega>\omega_0$) the difference in the

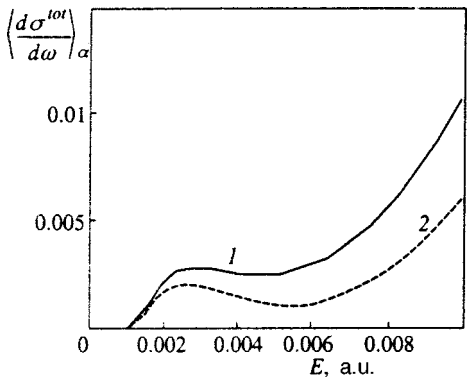

FIG. 4. Dependence of the SBIB cross section averaged over the angle α [for an angular distribution of the form $D_p(\alpha)=\cos^p\alpha/(p+1)$ with a distribution axis perpendicular to the polarization of the external radiation] on the electric field strength of the external radiation for two values of the angular distribution parameter ($p=2$ — curve *1*, $p=12$ — curve *2*), scattering of electrons with $v=0.6$ a.u. on N^{4+} ions, and detuning of the frequency of the external field from the eigenfrequency of the $2s-2p$ transition in the ion core equal to $(\omega-\omega_0)/\omega_0=-0.3\%$.

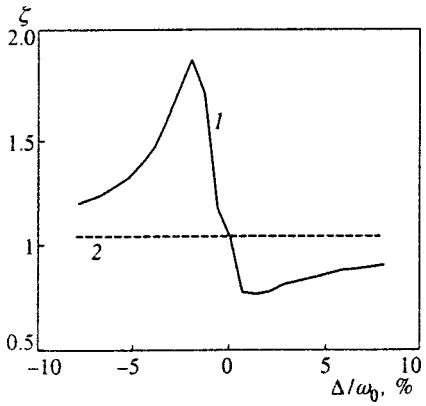

FIG. 5. Polarization of the spontaneous bremsstrahlung $\zeta=2\sigma(\alpha=0)-\sigma(\alpha=\pi/2)/\sigma(\alpha=0)+\sigma(\alpha=\pi/2)$ as a function of the relative frequency mismatch from the eigenfrequency of the $2s-2p$ transition in the N^{4+} ion with (curve *1*) and without (curve *2*) consideration of the polarization channel ($Z_{\text{eff}}=4.5$, $v=0.9$ a.u.), $\zeta_{\min}=0.77$, $\zeta_{\max}=1.88$.

spectral dependence of the cross sections for the two values of α is small.

The polarization features in the spectral SBIB cross section under consideration are attributed to the difference discussed in the preceding section between the effects of the nondipole nature of the interaction of the incident particle with the near-resonant transition on the focal components of the radiating dipole moment induced in the ion core: a stronger effect on its x component and a weaker influence (for the transition with $\Delta n=0$) on the y component. In the case of $\alpha=0$ the main contribution to the radiation (for sufficiently small incident particle velocities) is made by D_x due to the form of the function $f_{x,y}(\alpha,\varepsilon)$ [Eq. (5)]. In fact, for $\alpha=0$ we have $f_x=1$ and $f_y(0,\varepsilon)=\varepsilon^2-1$, and if it is now taken into account that for sufficiently small values of the incident particle velocity (in the Kramers limit[15]) the main contribution to the total cross section of the process for all eccentricities of the incident particle orbit is made by $\varepsilon\approx 1$ (close collisions), it follows from the preceding relations that $f_x(0,\varepsilon)\gg f_y(0,\varepsilon)$. Similar arguments show that the reverse inequality $f_x(\pi/2,\varepsilon)\ll f_y(\pi/2,\varepsilon)$ holds for $\alpha=\pi/2$ and that, accordingly, the main contribution to the process is made by the y component of the radiating dipole moment.

Thus, the angle α controls the relative contribution of the various focal components of the radiating dipole moment. If the main contribution to the radiation is made by D_x, the strong influence of the nondipole nature of the interaction of the incident particle with the ion core causes the mutual compensation of the amplitudes of the static and polarization channels (where they destructively interfere, $\omega<\omega_0$) to occur only for fairly large eccentricities of the incident particle orbit. This compensation no longer occurs for fairly small eccentricities. Moreover, at a certain value of ε_0 the function $D_x(\varepsilon)$ changes sign, and the interference between the static and polarization channels takes on a constructive character. As a result, the interference dip in the total SBIB cross section for all eccentricities of the incident particle orbit is smeared out, as is clearly seen in Fig. 2.

On the high-frequency wing of the spectral line at $\omega>\omega_0$ the situation is reversed: the stronger manifestation of the interference of the channels leads to a larger value of the SBIB cross section for $\alpha=\pi/2$ than for $\alpha=0$. However, in this case the relative role of the interference term is small, since the total cross section of the process is large. This circumstance is associated with the closeness of the nondipolarity parameter χ_{nd} to unity for $\Delta n=0$ and $Z_i\geqslant 2$ mentioned in the preceding section.

On the low-frequency wing of the spectral line, where the total cross section is small, on the other hand, the role of the interference is very significant, and the dependence of the SBIB cross section on the polarization of the external radiation under consideration is most pronounced.

A similar spectral dependence was also obtained for the transition in which the principal quantum number changes, $2s-3p$. In this situation the difference between the cross sections in the region of destructive interference is not so great as for the transition with $\Delta n=0$. This is attributed to the greater influence of the nondipole behavior on the magnitude of the y component of the dipole moment induced in the ion core due to the larger radius of the $3p$ state compared to the $2p$ state.

Figures 3 and 4 present the amplitude dependence of the total SBIB cross section for all incident particle scattering angles on the electric field intensity in the external radiation, for which polarization-induced interference effects of the interaction between the static and polarization channels also exist.

Figure 3 presents the amplitude dependence of the ratio of the total (including the polarization term in the amplitude) SBIB cross section to its static analog for two values of the angle α ($\alpha=0$ — curve *1*, $\alpha=\pi/2$ — curve *2*) in the region of destructive interference between the channels ($\omega<\omega_0$). As is seen from the figure, in the former case the interference effects are very weak: they are manifested by the fact that as the electric field intensity is increased, the cross section of the process tends to a value somewhat smaller than the static value as a consequence of destructive interference, which is strongly suppressed by the nondipole behavior of the interaction of the incident particle with the ion core. There is direct evidence of a strong interference effect for $\alpha=\pi/2$: the amplitude dependence of the SBIB cross section has a broad dip caused by the mutual compensation of the state and polarization terms in the cross section of the process, which is characteristic of treatments within the dipole approximation with respect to the interaction of the incident particle with the ion core.[1,4] The reason for preservation of the dipole features in the amplitude dependence is the same as in the spectral dependence: for $\alpha=\pi/2$ the main contribution to the radiation is made by the y component of the induced dipole moment, which is weakly subject to the influence of the nondipole nature for transitions without a change in the principal quantum number.

Figure 4 presents plots of the amplitude dependence of the total SBIB cross section averaged over an angular distribution of the incident particles of the form

$$D_p(\alpha)=\cos^p\alpha/(p+1)$$

for two values of the angular distribution parameter ($p=2$

and $p=12$) and an angle between the distribution axis and the electric field intensity vector of the electromagnetic wave equal to $\pi/2$. The form of the angular distribution of the electrons chosen is characteristic of the ablation of a solid-state target under the action of high-power laser radiation, and the angle $\alpha=\pi/2$ corresponds to normal incidence of the radiation onto the surface of the solid.

It can be seen from the figure that the interference effect is displayed most strongly for the more pointed angular distribution ($p=12$), as would be expected, since the mean value of α is closer to $\pi/2$ in this case.

Figure 5 presents the spectral dependence of the polarization of the spontaneous near-resonant bremsstrahlung calculated with (solid curve) and without (dashed curve) consideration of the polarization channel. The low-frequency wing of the line exhibits a strong upward deviation of the polarization of the radiation parallel to the initial velocity vector of the incident particle from the perpendicular value $[\sigma(\alpha=0)/\sigma(\alpha=\pi/2)=32]$ due to the interference-induced suppression of the perpendicular polarization, which is less subject to the effect of the nondipole behavior of the interaction of the incident particle with the ion core. This upward deviation is an order of magnitude greater than the corresponding value obtained without consideration of the polarization term in the bremsstrahlung amplitude.

As can be seen from Fig. 5, the maximum value of the polarization of the spontaneous near-resonant bremsstrahlung ($\zeta_{\max}=1.88$) corresponds to the relative mismatch $\Delta/\omega_0=-2\%$, which is significantly greater than the Doppler width of the transition.

5. CONCLUSION

The polarization effects in the near-resonant SBIB of quasiclassical electrons on ions with a core due to the nondipole character of the interaction of the incident particle with a radiative transition in the ion upon formation of a polarization channel for the process have been analyzed within the specified-current approximation.

It has been shown that consideration of this nondipole behavior is essential for a correct description of the SBIB cross section in the region of destructive interference of the static and polarization channels, where the cross section is relatively small.

It is significant that a dependence of the spectral-amplitude characteristics of SBIB on such parameters of the process as the polarization of the external radiation, the absolute value of the velocity, and the angular distribution of the incident particles, which is not observed in the dipole treatment, appears in the more exact approach described here. In particular, a dependence of the interference dips in the total SBIB cross section (with consideration of the polarization channel) on the angle between the initial velocity vector of the incident particle and the electric field vector of the external radiation has been discovered. This dependence is observed for both the spectral and amplitude SBIB cross section and is most strongly manifested for near-resonant transitions in the ion core with $\Delta n=0$.

In the case of spontaneous bremsstrahlung this effect is observed in the spectral dependence of the polarization of the radiation, which differs fundamentally from the static analog.

The SBIB features discovered can be significant for analyzing the possibility of utilizing this phenomenon in the context of the general problem of generating short-wavelength radiation.

This research was carried out with partial financial support from the Russian Fund for Fundamental Research (Project No. 98-02-16763).

*)E-mail: astval@hotmail.com

[1] V. A. Astapenko and A. B. Kukushkin, Zh. Éksp. Teor. Fiz. **111**, 419 (1997) [JETP **84**, 229 (1997)].
[2] A. V. Korol and A. V. Solov'yov, J. Phys. B **30**, 1105 (1997).
[3] L. A. Bureyeva and V. S. Lisitsa, J. Phys. B **31**, 1477 (1998).
[4] V. A. Astapenko and A. B. Kukushkin, Laser Phys. **8**, 552 (1998).
[5] V. A. Astapenko, Laser Phys. **8**(5) (1998).
[6] A. V. Korol', A. G. Lyalin, O. I. Obolenskiĭ, and A. V. Solov'ev, Zh. Éksp. Teor. Fiz. **114**, 458 (1998) [JETP **87**, 251 (1998)].
[7] *Polarization Bremsstrahlung*, V. N. Tsytovich and I. M. Oiringel (Eds.), Plenum Press, New York (1992) [Russ. original, Nauka, Moscow (1987)].
[8] L. I. Trakhtenberg, Opt. Spektrosk. **44**, 863 (1978) [Opt. Spectrosc. **44**, 510 (1978)].
[9] V. M. Buĭmistrov and L. I. Trakhtenberg, Zh. Éksp. Teor. Fiz. **73**, 850 (1977) [Sov. Phys. JETP **46**, 447 (1977)].
[10] V. A. Astapenko, L. A. Bureyeva, and V. S. Lisitsa, in *Proceedings of the 16th ICAP, Abstracts of Contributed Papers*, Windsor, Ontario, Canada (1998), p. 183.
[11] V. I. Kogan and A. B. Kukushkin, Zh. Éksp. Teor. Fiz. **87**, 1164 (1984) [Sov. Phys. JETP **60**, 665 (1984)].
[12] V. A. Astapenko, Laser Phys. **8**(6) (1998).
[13] V. I. Gervids and V. I. Kogan, Pis'ma Zh. Éksp. Teor. Fiz. **22**, 308 (1975) JETP Lett. **22**, 142 (1975).
[14] I. Ya. Berson, Zh. Éksp. Teor. Fiz. **80**, 1727 (1981) [Sov. Phys. JETP **53**, 891 (1981)].
[15] V. I. Kogan, A. B. Kukushkin, and V. S. Lisitsa, Phys. Rep. **213**, 1 (1992).

Translated by P. Shelnitz

Gain coefficient of a high harmonic emitted during the above-threshold ionization of atoms

D. F. Zaretskiĭ*)

Kurchatov Institute Russian Scientific Center, 123182 Moscow, Russia

É. A. Nersesov

Moscow State Engineering-Physics Institute, 115409 Moscow, Russia

(Submitted 13 November 1998)
Zh. Éksp. Teor. Fiz. **115**, 1630–1641 (May 1999)

The problem of the amplification of high harmonics generated during the above-threshold ionization of atoms in a high-power laser wave field is examined for the first time. An estimate of the gain coefficient as a function of the parameters of the atom beam and the pump wave is given. © *1999 American Institute of Physics.* [S1063-7761(99)00805-7]

1. INTRODUCTION

The generation of high harmonics provides a source of coherent radiation in the hard-ultraviolet range. The significant progress made in the area of creating high-power (with a peak intensity in a pulse $I=10^{13}-10^{18}\,\mathrm{W\,cm^{-2}}$) ultrashort pulses ($\tau=600\,\mathrm{fs}-1$ ps) has permitted advances into the range of harmonic wavelengths $\lambda_s=10-100$ nm.[1–4]

In the last few years some new experimental results have been obtained in studies of the detailed properties of pulses of high harmonics. In particular, data on the angular distribution of the radiation,[5–7] its spectral profile[4] and temporal picture,[8] and the conditions for phase synchronism between the high-harmonic wave and the pump wave have been obtained. High-resolution cross-correlation experiments have led to a conclusion regarding the possible temporal duration of the generation of high harmonics.[9,10]

Advancements in laser technology have made it possible to attain a large number of high-harmonic photons during a pulse ($N_s\approx 10^6-10^{10}$) at comparatively high gas densities (with a pressure in the beam $P=10-100$ Torr). These achievements raised the question of the possibility of observing the amplification of high harmonics as they pass through the interaction region of the pump wave and the atom beam.

This paper examines the problem of the amplification of high harmonics generated in the above-threshold ionization of atoms in a high-power laser wave field for the first time. An estimate of the gain coefficient as a function of the parameters of the atom beam and the pump wave is given.

2. STATEMENT OF PROBLEM. BASIC EQUATIONS

We consider the amplification of a pulse of a high harmonic, where a weak probe wave of intensity I_2 with a frequency $\Omega\approx s\omega$ (s is the number of the harmonic, and ω is the frequency of the pump wave) is sent into the interaction region of the atom beam with the strong pump wave in the direction of that wave.

We assume that the specified-field approximation, under which the increment of the number of high-harmonic photons over the length of the interaction region is less than the number of photons in the amplified wave (the condition that the gain coefficient of the probe wave over the interaction length of the beam with the probe wave be small), holds.

The amplification effect is determined by the competition between the forward and reverse processes (with the emission and absorption of high-harmonic photons, respectively). The plots in Fig. 1 describe the processes taking place when both waves are simultaneously present in the interaction volume. Figure 1a corresponds to the process with the absorption of s quanta of the laser wave by an atom, a transition of an electron into a virtual state $\Psi_{\mathbf{p}}$ in the continuous spectrum [the familiar Volkov solution in a strong wave field (the Keldysh–Faisal–Reiss model[11])], and subsequent stimulated recombination to the ground state of the atom with the emission of one Ω photon. Figure 1b corresponds to the reverse process with the absorption of an Ω photon. The initial (Ψ_i) and final ($\Psi_{fe,a}$) functions of the electron–ion system belong to the ground state of the atom, but the functions Ψ_{fe} and Ψ_{fa} differ in the energy and momentum of the atom after it interacts with the electromagnetic fields.

The gain coefficient of the probe wave is determined by the total probability of the process:

$$G\propto\Delta w=w_e-w_a\,, \tag{1}$$

where $w_{e,a}$ are, respectively the probabilities of the stimulated emission and absorption of an Ω photon by the system of atoms in the interaction volume.

Let us focus our attention on one special feature of the problem under consideration. Since the atom is in the ground state before and after the interaction with the waves, the matrix elements for processes a and b (see Fig. 1) written without allowance for the motion of the center of mass of the electron–ion system are equal, and, accordingly, $\Delta w=0$. A nonzero result can appear in (1) only in a basis set of wave functions that include the combined motion of the electron and the ion. In this case we must take into account the recoil effect, which appears when this system emits or absorbs quanta of a field. When the problem under consideration is

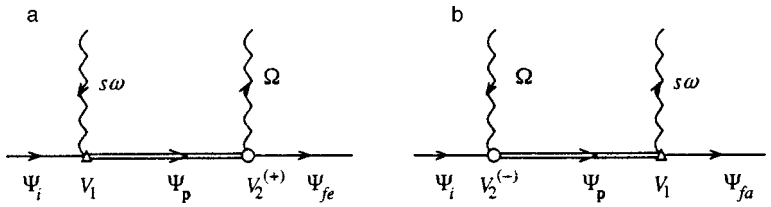

FIG. 1. The vertex in the form of a triangle with a wavy line corresponds to the interaction of an electron with the pump wave field, which leads to direct multiphoton ionization of the atom from the ground state to a virtual state in the continuous spectrum; the vertex depicted by a wavy line with a filled circle describes the interaction of the photoelectron in the continuum with the probe wave field, which leads to recombination to the ground state of the atom with the emission of a photon of a high harmonic of frequency Ω (case a); the double line with an arrow depicts the state of the photoelectron in the continuous spectrum appearing in the strong wave field (the Keldysh method). Case b corresponds to the reverse process, and the meanings of the notation are the same as in case a.

so formulated, it becomes similar to the familiar problem of the stimulated Compton emission of an electron (see, for example, Ref. 12). We note a formal feature shared by these problems. Since processes in specified wave fields are considered in them, the final result for the gain coefficient G must be averaged over the distribution functions of both the atom beam and the probe wave (this question is discussed in greater detail in Sec. 3).

Let us consider the question of the basis set of wave functions used to calculate the composite matrix elements of the diagrams in Fig. 1. Because the parameter associated with recoil of the atom is small, the amplification of a high harmonic can be appreciable only if the condition of phase synchronism of the emitters in the propagation direction of the waves is satisfied in the medium of atoms.[13] The phase of the atoms is determined by the phase of the coherent pump wave, and the calculation of the gain coefficient with allowance for phase synchronism requires consideration of the motion of the center of mass of the electron–ion system and assignment of the absolute coordinate of the electron in the laboratory reference frame.

The adiabatic approximation, within which the electron and the residual ion are treated as fast and slow subsystems respectively, is assumed to hold. This approximation makes it possible to separate the spatial variables of the particles (the terms in the Schrödinger equation that are discarded in the separation procedure are small, being $\sim m_e\varepsilon_{\mathbf{p}}/ME$, where M and m_e are the masses of the ion and the electron, respectively, $\varepsilon_{\mathbf{p}}$ is the energy of a photoelectron with the momentum $\mathbf{p}$, and E is the kinetic energy of the atom).

The psi function of the initial state of the system with consideration of the combined motion of the ion and the electron is written in the form ($\hbar=c=1$)

$$\Psi_i=\exp[-i(Et-\mathbf{P}\cdot\mathbf{R})]\exp(-iE_0t)\psi_0(\mathbf{r}) \tag{2}$$

and describes the free motion of an atom with the energy $E=P^2/2M$ and the momentum $\mathbf{P}$ in the ground state $\psi_0(\mathbf{r})$ with the energy E_0. Here $\mathbf{R}$ is the radius vector of the atom (ion), $\mathbf{r}$ is the relative radius vector of the electron, and the normalization plane-wave volume is set equal to unity. In accordance with the Keldysh model[14] it is assumed that the ground state of the atom is weakly perturbed by the laser wave field.

The wave function of the final state can be written in a similar manner:

$$\Psi_{fe,a}=\exp[-i(E_{e,a}t-\mathbf{P}_{e,a}\cdot\mathbf{R})]\exp(-iE_0t)\psi_0(\mathbf{r}), \tag{3}$$

where E_e, $\mathbf{P}_e$ and E_a, $\mathbf{P}_a$ are the energy and momentum values which the atom acquires as a result of the emission (e) and absorption (a) of a photon with the wave vector $\mathbf{K}$ and the frequency Ω.

The intermediate state $\Psi_{\mathbf{p}}$ in the figure corresponds to the ion and electron in the continuous spectrum and is written as follows:

$$\begin{aligned}\Psi_{\mathbf{p}}=&\exp[-i(E't-\mathbf{P}'\cdot\mathbf{R})]\\&\times\exp[-i(\tilde{\varepsilon}_{\mathbf{p}}t-\tilde{\mathbf{p}}\cdot(\mathbf{R}_0+\mathbf{R}+\mathbf{r}))]\\&\times\exp(i\alpha_1\cos\varphi_1)\exp(-i\alpha_2\sin 2\varphi_2).\end{aligned} \tag{4}$$

In (4) E' and P' are the energy and momentum of the ion in the intermediate state; $\tilde{\varepsilon}_{\mathbf{p}}=\varepsilon_{\mathbf{p}}+U_{\mathbf{p}}$ is the energy of the photoelectron with consideration of the ponderomotive energy $U_{\mathbf{p}}=(eE_0\lambda\!\!\!^{-}_0)^2/4m_e$ (E_0 is the amplitude value of the electric field intensity in the pump wave, and $\lambda\!\!\!^{-}_0=1/\omega$); $\tilde{\mathbf{p}}=\mathbf{p}+(U_{\mathbf{p}}/\omega)\mathbf{k}$ is the momentum of the photoelectron; $\mathbf{R}_0+\mathbf{R}+\mathbf{r}$ is the absolute radius vector of the electron ($\mathbf{R}_0$ is the fixed radius vector of the ion without consideration of its motion); the dimensionless parameters

$$\alpha_1=\frac{e\mathbf{E}_0\cdot\mathbf{p}}{m_e\omega^2},\quad \alpha_2=\frac{(eE_0\lambda\!\!\!^{-}_0)^2}{8m_e\omega}=\frac{U_{\mathbf{p}}}{2\omega} \tag{5}$$

are determined by the strength of the interaction of the electron with the laser wave field; and $\varphi_1=\omega t-\mathbf{k}(\mathbf{R}_0+\mathbf{R}+\mathbf{r})$ is the phase of the wave at the time t at the site of the electron.

We use the expression

$$\hat{V}_1=e\mathbf{A}_1(t)\cdot\hat{\mathbf{p}}/m_e+[eA_1(t)]^2/2m_e; \tag{6}$$

to define the interaction operator of the electron with the pump wave field. The vector potential of the wave is assigned by the classical expression $\mathbf{A}_1(t)=\mathbf{A}_{01}\sin\varphi_1$, where $\mathbf{A}_{01}$ is the amplitude of the vector potential. The pump wave field is used within the Keldysh method.[14]

The probe wave field is treated in first-order perturbation theory. The perturbation operator associated with this wave has the form

$$\hat{V}_2=e\mathbf{A}_2(t)\cdot\hat{\mathbf{p}}/m_e+e^2\mathbf{A}_1(t)\cdot\mathbf{A}_2(t)/m_e, \tag{7}$$

where $\mathbf{A}_2(t)=\mathbf{A}_{02}\sin[\Omega t-\mathbf{K}\cdot(\mathbf{R}_0+\mathbf{R}+\mathbf{r})]$ is the vector potential of the probe wave, and $\mathbf{K}$ and Ω are wave vector and frequency.

The probability amplitude of the transition of an electron from the ground state of the jth atom to the continuous spectrum with the absorption of several quanta of the pump wave followed by stimulated recombination to the ground state with the emission of a high-harmonic $\mathbf{K},\Omega$ photon by the time t (Fig. 1a) is given by the expression

$$A_e^j(t)=(-i)^2\int\frac{d\mathbf{P}'}{(2\pi)^3}\int\frac{d\mathbf{p}}{(2\pi)^3}\times\int_{-\infty}^{t}dt_1\int_{-\infty}^{t_1}dt_2\,\langle\Psi_{fe}^{j*}|V_2^{(+)}(t_1)|\Psi_{\mathbf{p}}^{j}\rangle\times\langle\Psi_{\mathbf{p}}^{j*}|V_1(t_2)|\Psi_i^j\rangle, \tag{8}$$

where $\mathbf{P}'$ and $\mathbf{p}$ are, respectively, the momenta of the ion and the photoelectron in the virtual intermediate state, and the $V_2^{(+)}$ terms, which are responsible for emission, have been left in the operator $V_2(t)$ (7). The analogous probability amplitude of the reverse process $A_a^j(t)$ (Fig. 1b) is a trivial problem.

The total probability amplitude of the stimulated emission (absorption) of a $\mathbf{K},\Omega$ photon by the entire system of atoms in the interaction volume of the beam with the waves appears as a result of summation of the amplitudes (8) over all the numbers j of the atoms in that volume and is distinguished from (8) by the multiplier[15]

$$\sum_j\exp[i(s\mathbf{k}-\mathbf{K})]\cdot\mathbf{R}_{0j}. \tag{9}$$

The sum (9) can easily be calculated in the continuous-medium approximation and gives rise to a diffractive multiplier in the expressions for the probabilities of the processes:[15]

$$\left|\sum_j\exp[i(s\mathbf{k}-\mathbf{K})]\cdot\mathbf{R}_{0j}\right|^2=\left(\frac{\pi}{6}\right)^2(V_{\rm int}n_a)^2\frac{\sin^2[(\Omega-n_\omega s\omega)l/2]}{[(\Omega-n_\omega s\omega)l/2]^2}, \tag{10}$$

where $V_{\rm int}=\pi\rho_0^2 l$ is the interaction volume of the atom beam with the waves; ρ_0 is the radius of the focus of the pump wave at its center; l is the longitudinal dimension of the interaction region in the direction of the wave (as a rule, in experiments on harmonic generation we have $l=d$, where d is the diameter of the beam of atoms sent in the direction transverse to the waves); $n_\omega=1-\omega_p^2/2\omega^2$ is the refractive index of the ionized medium for the pump wave (here $\omega_p=\sqrt{4\pi n_e e^2/m_e}$ is the plasma frequency); and n_a is the concentration of atoms in the beam.

From (8) with consideration of the diffractive multiplier (10) we obtain the following expression for the transition probability of the system of atoms to the partial final state per unit time:

$$w_{e,a}^{(s)}=\frac{d|A_{e,a}|^2}{dt}=8\pi\alpha^2\frac{\lambda\!\!\!^{-}_0^4}{s^2}\frac{I_1I_2}{\mathrm{Ry}^3}\left(\frac{\pi}{6}\right)^2(V_{\rm int}n_a)^2\frac{\sin^2 u}{u^2}\times\left|\sum_n E_{s,n}\right|^2\delta[E_{e,a}-E\pm(\Omega-s\omega)]. \tag{11}$$

In (11) I_1 and I_2 are the intensities of the pump wave and the probe wave, respectively; $u=(\Omega-n_\omega s\omega)l/2$; $E_{s,n}$ is an expression which appears when the integrals in the composite matrix element of the amplitude (8) are evaluated [its form is not shown in this paper, since the final result is expressed in terms of the probability of spontaneous emission, which contains this expression (see below)]; the summation index n runs through the sequence of numbers of the photoelectron peaks in the above-threshold ionization spectrum of the atoms; $\alpha=e^2/\hbar c$; $\mathrm{Ry}=m_e e^4/2\hbar^2$ is the Rydberg constant; the plus and minus signs in the argument of the δ function, which assigns the energy conservation law for the processes under consideration, refer, respectively, to the stimulated emission (e) and absorption (a) of a photon of frequency $\Omega\approx s\omega$; E is the initial kinetic energy of the atom; and $E_{e,a}$ is the kinetic energy of the atom in the final state, to which it passes as a result of its interaction with the waves with the emission (absorption) of a high-harmonic photon.

We note that the values of $E_{e,a}$ in the stimulated processes are determined from the momentum conservation law $\mathbf{P}_{e,a}-\mathbf{P}\pm(\mathbf{K}-s\mathbf{k})=0$ and equal

$$E_{e,a}=E\mp\frac{\mathbf{P}\cdot(\mathbf{K}-s\mathbf{k})}{M}+\frac{(\mathbf{K}-s\mathbf{k})^2}{2M} \tag{12}$$

(the minus and plus signs correspond to emission and absorption processes).

The probabilities of the simulated processes (11) can be related fairly simply to the probability $w_{\rm sp}^{(s)}$ of the spontaneous emission of a photon of frequency $\Omega\approx s\omega$. This relation has the form

$$w_{e,a}^{(s)}=4\rho_0^2\lambda_0 I_2\frac{\sin^2 u/u^2}{\sin^2 u_0/u_0^2}\frac{1}{s}w_{\rm sp}^{(s)}\delta\times[E_{e,a}-E\pm(\Omega-s\omega)], \tag{13}$$

where $u_0=s\omega(1-n_\omega)l/2$ is the argument of the diffractive multiplier u for the exact equality $\Omega=s\omega$.

The delta functions in (13), which assign the energy conservation law for the transition of the system to a partial final state, should be replaced in the ensuing calculations by integration over the distributions in the interacting objects.

3. GAIN COEFFICIENT

The dimensionless gain coefficient per transit of a probe wave of frequency $\Omega\approx s\omega$ is given by

$$G^{(s)}=\frac{\langle w^{(s)}\rangle\Omega}{\pi\rho_0^2 I_2}, \tag{14}$$

where $\langle\Delta w^{(s)}\rangle=\langle w_e^{(s)}-w_a^{(s)}\rangle$ is the total probability of the stimulated processes averaged over the distributions. The substitution of (13) into (14) leads to the result

$$G^{(s)}=\frac{8w_{\rm sp}^{(s)}}{I_2(\sin^2 u_0/u_0^2)}J, \tag{15}$$

where J denotes the integral

$$J=\int\int\frac{\sin^2 u}{u^2}\{\delta[E_e-E+(\Omega-s\omega)] - \delta[E_a-E-(\Omega-s\omega)]\}f(\mathbf{P})\,d\mathbf{P}\frac{dI_2}{d\Omega}\frac{dI_1}{d\omega}d\Omega\,d\omega, \tag{16}$$

$f(\mathbf{P})$ is the distribution function of the initial atom beam normalized to unity, and the functions $dI_1/d\omega$ and $dI_2/d\Omega$ describe the spectral distributions in the pump wave (probe wave) normalized to the total intensity I_1 (and I_2, respectively), whose widths are related to the pulse durations.

The integration procedure in (16) is described in Appendix 1. The final result for $G^{(s)}$ and the character of the integrations in (16) depend on the formulation of the problem. When an answer is needed in the form of the gain coefficient in a mode, i.e., at a specified frequency of the probe wave, it is sufficient to confine ourselves to integration over the distribution of the initial atom beam in (16). In this case, in accordance with Eq. (1.8) from Appendix 1 for the gain coefficient in a mode, we obtain

$$G^{(s)}(\Omega)=8w_{\rm sp}^{(s)}\frac{\sin^2 u/u^2}{\sin^2 u_0^2/u_0^2}\left(-\frac{16\sqrt{\pi}\tilde{M}^2}{\Delta P}\frac{\Omega-s\omega}{|\Omega-n_\omega s\omega|}\right)\times\exp\left[-\tilde{M}^2\left(\frac{\Omega-s\omega}{\Omega-n_\omega s\omega}\right)^2\right]. \tag{17}$$

As follows from (17), the gain coefficient vanishes when $\Omega=s\omega$ holds exactly and is positive for waves with frequencies $\Omega<s\omega$. An increase in the frequency mismatch $|\Omega-s\omega|$ is accompanied by a sharp decrease in the coefficient $G^{(s)}(\Omega)$ in view of the large value of $\tilde{M}$.

A situation in which averaging over the spectral distribution in the probe wave is also required in the integral J should be considered more characteristic of the conditions of experiments on high-harmonic generation. In this case the final result for J has the form [Eq. (1.12) in Appendix 1]

$$J=-\frac{16\sqrt{\pi}\,s\omega(1-n_\omega)}{\Delta\Omega M}\frac{\sin^2 u_0}{u_0^2}I_2\exp\left[-\left(\frac{\Omega_0-s\omega}{\Delta\Omega}\right)^2\right]\times\left[u_0\cot u_0+\frac{(\Omega_0-s\omega)s\omega(1-n_\omega)}{(\Delta\Omega)^2}\right]. \tag{18}$$

It follows from (15) and (18) that a positive value of the gain coefficient can be obtained in two limiting cases. If the density of the atoms in the beam is small, so that $u_0<1$ holds (we recall that u_0 depends on the extent of ionization and the concentration of atoms in the medium), amplification of the probe wave can be achieved only when the frequency Ω_0, which corresponds to the maximum of its spectral distribution, is smaller than $s\omega$. The other limiting case is realized for dense beams, in which u_0 lies in the range $1<u_0<\pi$. Then the spontaneous emission of the sth harmonic ($\Omega_0=s\omega$), which can be induced ahead of time under conditions similar to the conditions of the amplification process, can serve as the source of the probe wave. This radiation must then be directed into the region where the second beam interacts with the pump wave and be synchronized with the arrival of the pulse of this wave (a two-beam experiment scheme).

In the ensuing calculations we shall confine ourselves to this case and omit all the multipliers containing the difference $\Omega_0-s\omega$ in (18). The parameters n_ω and u_0 depend how much of the medium is ionized and, therefore, on the local characteristics of the field intensity in the focus of the pump wave. Therefore, after substituting (18) into (15), we must average the expression obtained over the coordinate ρ, which is transverse to the focal axis, to derive the final formula for the gain coefficient:

$$\langle G^{(s)}\rangle=-\frac{(16)^2 s\omega}{\sqrt{\pi}\Delta\Omega M}\frac{J_\rho}{\rho_0^2}w_{\rm sp}^{(s)}, \tag{19}$$

where

$$J_\rho=\int_0^{\rho_0}[1-n_\omega(\rho)]u_0(\rho)\cot u_0(\rho)\rho\,d\rho. \tag{20}$$

The details of the calculations of the integral (20) are described in Appendix 2. Using the result (2.3) obtained therein, we present the expression for the averaged gain coefficient (in ordinary units)

$$\langle G^{(s)}\rangle=-\frac{(16)^2\hbar c N_s}{\sqrt{\pi}lMc^2\Delta\Omega\tau_s}\left(\frac{\tilde{\rho}_0}{\rho_0}\right)^2\int_0^{u_{0m}}x\cot x\,dx, \tag{21}$$

where N_s is the number of photons of the sth harmonic emitted during a pump-wave pulse, which can be estimated from the relation $N_s\approx w_{\rm sp}^{(s)}\tau_s$; τ_s is the duration of the probe-wave pulse; $\tilde{\rho}_0$ is the characteristic constant of the Gaussian distribution of the electron density in the focus; and u_{0m} is the value of the argument on the focal axis (see Appendix 2).

At fairly large values of $u_0\sim 1$ (when $u_0\leqslant\pi$) we can use the expansion[16]

$$\int x\cot x\,dx=\sum_{k=0}^{\infty}(-1)^k\frac{2^{2k}B_{2k}}{(2k+1)!}x^{2k+1}, \tag{22}$$

where B_{2k} is Bernoulli's number.

4. ESTIMATES; ANALYSIS OF RESULTS

Let us estimate the gain coefficient for the case in which a Ti:sapphire laser (with a wavelength $\lambda\approx 800$ nm) serves as the source of the pump waves. Let us also assume that the probe wave is the spontaneous radiation appearing at the frequency of a high harmonic when the pump wave interacts with the atom beam. It is further assumed that this radiation crosses the second atom beam simultaneously with the pump-wave pulse (a two-beam experiment).

Let us consider the conditions needed for amplification of the probe wave. We first note that in the case of an atom beam with a comparatively low density ($u_0<1$), where $\cot u_0\approx 1/u_0$, the gain coefficient has a negative value [see

(18)], which corresponds to absorption of the probe wave in the interaction volume. However, the situation changes when u_0 is increased. In particular, when u_0 approaches π, the integral (21), which specifies the value of the averaged gain coefficient, is represented in the form of the series (22). For example, for $u_0=3$ the integral (21) is equal to -2.6, i.e., the gain coefficient is positive. Substituting this value into (21) and assuming that photoionization saturation is achieved over the high-power laser wave field in essentially the entire volume of the focus ($\tilde{\rho}_0 \approx \rho_0$), we obtain the estimate $\langle G^{(s)} \rangle \approx 16\%$ for $N_s = 10^7$. This number of spontaneous photons is observed in the plateau region in the case of a beam of Xe atoms for harmonics with the numbers $s=40$–60 (the corresponding frequencies are $\Omega = 60$–90 eV).

In conclusion, we note that the spectral composition of the probe wave can differ from the spectrum of the high-harmonic spontaneous emission. In this case Ω_0 can differ from $s\omega$. Then the value of the gain coefficient depends on the magnitude and sign of the detuning $\Delta\Omega = \Omega_0 - s\omega$ and can be positive even when $u_0 < 1$.

We thank M. V. Fedorov for discussing the statement of the problem, as well as P. Agostini for discussing the possibility of experimental verification of the results obtained.

APPENDIX 1:

Let us discuss the evaluation of the integral (16). The order of integration in it depends on the relative sharpness of the distribution functions of the quantities over which the integration is performed. The real conditions of an experiment on the generation of high harmonics in atom beams satisfy the inequalities $\Delta P/P \ll \Delta\Omega/\Omega \ll \Delta\omega/\omega$, where ΔP is determined by the temperature of the atoms in the accompanying reference frame, and $\Delta\Omega \sim 1/\tau_s$ and $\Delta\omega \sim 1/\tau_p$ are the widths of the spectral distributions of the waves, which are related to the duration of the corresponding pulses.

We use I to denote the following integral

$$I = \int \{\delta[E_e - E + (\Omega - s\omega)] - \delta[E_a - E - (\Omega - s\omega)]\} f(\mathbf{P})\, d\mathbf{P}. \tag{1.1}$$

The momentum distribution function of the beam atoms is assigned by the model dependence normalized to unity

$$f(\mathbf{P}) = \frac{4}{\sqrt{\pi}(\Delta P)^3} \exp\left[-\frac{(\mathbf{P}-\mathbf{P}_0)^2}{(\Delta P)^2} \right], \tag{1.2}$$

where $\mathbf{P}_0$ is the momentum of the atom beam as a whole and ΔP is the width of the momentum distribution.

The integration over the momenta of the atoms in (1.1) is preceded by a series of transformations. Using (12) we represent the δ function in the form

$$\delta_{\pm}\left[\frac{\mathbf{P}(\mathbf{K}-s\mathbf{k})}{M} - (\Omega - s\omega) \mp \frac{(\mathbf{K}-s\mathbf{k})^2}{2M} \right], \tag{1.3}$$

where the minus sign in the argument of the function refers to the emission of a $\mathbf{K},\Omega$ photon and the plus sign refers to the absorption process.

If we assume that the momentum P_0 is directed perpendicularly to the vectors $\mathbf{K}$ and $\mathbf{k}$, it is convenient to go over to the relative momentum of the atoms $\mathbf{a} = \mathbf{P} - \mathbf{P}_0$ in all the expressions. We note that sending a beam at an arbitrary angle ϑ to the propagation direction of the waves gives rise to insignificant corrections $\sim P_{0\|}/M$ in the expression for J ($P_{0\|} = P_0 \cos\vartheta$ is the component of $\mathbf{P}_0$ along $\mathbf{K}$ and $\mathbf{k}$).

Using the known properties of δ functions, we can represent the expressions (1.3) in the form

$$\delta_{\pm}\left[\frac{\mathbf{P}(\mathbf{K}-s\mathbf{k})}{M} - (\Omega - s\omega) \mp \frac{(\mathbf{K}-s\mathbf{k})^2}{2M} \right] = \frac{M}{|\Omega - n_\omega s\omega||\cos\theta|} \delta(a - a_{\pm}), \tag{1.4}$$

where

$$a_{\pm} = \frac{M}{\cos\theta} \frac{\Omega - s\omega}{|\Omega - n_\omega s\omega|} \left[1 \pm \frac{(\Omega - n_\omega s\omega)^2}{2M(\Omega - s\omega)} \right] \tag{1.5}$$

are the values of the initial relative momentum of an atom participating in the emission (a_+) or absorption (a_-) of a $\mathbf{K},\Omega$ photon, and θ is the angle between the vector $\mathbf{a}$ and the wave vectors of the waves.

When the frequency of the probe wave coincides with $s\omega$ ($\Omega = s\omega$), the pulses (1.5) are identical and equal

$$a_+ = a_- = \frac{|\Omega - n_\omega s\omega|}{2|\cos\theta|}.$$

In this case only atoms traveling in the propagation direction of the waves, i.e., with $|\theta| < \pi/2$, participate in the emission of a high-harmonic photon. On the other hand, only atoms traveling in the direction opposite to the waves, with $|\theta| > \pi/2$, participate in absorption processes. Because of the symmetry of the directions of the momenta $\mathbf{a}$ of the atoms in the beam relative to the vectors $\mathbf{K}$ and $\mathbf{k}$ and the equality $a_+ = a_-$ in the case of $\Omega = s\omega$, the gain coefficient of the probe wave vanishes at this frequency.

In the general case $\Omega \neq s\omega$ the momentum difference

$$a_+ - a_- = \frac{|\Omega - n_\omega s\omega|}{\cos\theta} \tag{1.6}$$

is related to the recoil experienced by the atom upon its interaction with the waves and determines the value of the gain coefficient.

Integration over the modulus of the momentum $\mathbf{a}$ in (1.1) leads, with consideration of (1.4) and (1.5), to the following result:

$$I = \frac{16\sqrt{\pi}M^2}{(\Delta P)^3} \frac{\Omega - s\omega}{|\Omega - n_\omega s\omega|} \left\{ \int_0^\pi \frac{\sin\theta\, d\theta}{|\cos\theta| \cos^2\theta} \times \exp\left[-\frac{\tilde{M}^2}{\cos^2\theta} \left(\frac{\Omega - s\omega}{\Omega - n_\omega s\omega} \right)^2 \right] - \tilde{M}^2 \left(\frac{\Omega - s\omega}{\Omega - n_\omega s\omega} \right)^2 \int_0^\pi \frac{\sin\theta\, d\theta}{|\cos\theta| \cos^4\theta} \right.$$

$$\times\exp\left[-\frac{\tilde{M}^2}{\cos^2\theta}\left(\frac{\Omega-s\omega}{\Omega-n_\omega s\omega}\right)^2\right]\Bigg\}, \tag{1.7}$$

where the value of the dimensionless parameter is $\tilde{M}\equiv M/\Delta P$.

The integrals in (1.7) are reduced by the replacement of a variable $x=\cos\theta$ to known integrals,[16] and, as a result, from (1.7) we obtain

$$I=-\frac{16\sqrt{\pi}\tilde{M}^2}{\Delta P}\frac{\Omega-s\omega}{|\Omega-n_\omega s\omega|}\exp\left[-\tilde{M}^2\left(\frac{\Omega-s\omega}{\Omega-n_\omega s\omega}\right)^2\right]. \tag{1.8}$$

The substitution of (1.8) into the definition of J (16) yields the result

$$J=-\frac{16\tilde{M}^2 I_2}{\Delta P\Delta\Omega}\int_{-\infty}^{\infty}\frac{\sin^2 u}{u^2}\exp\left[-\left(\frac{\Omega-\Omega_0}{\Delta\Omega}\right)^2\right]\times\frac{\Omega-s\omega}{|\Omega-n_\omega s\omega|}\exp\left[-\tilde{M}^2\left(\frac{\Omega-s\omega}{\Omega-n_\omega s\omega}\right)^2\right]d\Omega. \tag{1.9}$$

Equation (1.9) was obtained using a simple model dependence for the spectral distribution of the probe wave:

$$\frac{dI_2}{d\Omega}=\frac{I_2}{\sqrt{\pi}\Delta\Omega}\exp\left[-\left(\frac{\Omega-\Omega_0}{\Delta\Omega}\right)^2\right],$$

which is normalized to the total intensity I_2. By virtue of the inequality $\Delta\Omega/\Omega\ll\Delta\omega/\omega$ there is no need for subsequent integration over the frequency ω.

The sharpest function of Ω in the integrand in (1.9) is $\exp\{-\tilde{M}^2[(\Omega-s\omega)/(\Omega-n_\omega s\omega)]^2\}$, and the contributions to the value of J are a result of integration in a small vicinity of the point $\Omega=s\omega$ (an estimate of the characteristic width of the integration range $\delta\Omega$ is given below).

To calculate (1.9) we perform the replacement of a variable $x=(\Omega-s\omega)/(\Omega-n_\omega s\omega)$. As a result, (1.9) yields

$$J=-\frac{16M^2 I_2 s\omega(1-n_\omega)}{(\Delta P)^3\Delta\Omega}J_x, \tag{1.10}$$

where

$$J_x=\int_{-\infty}^{\infty}\frac{\sin^2 u(x)}{u^2(x)}\frac{x}{(1-x)^2}\exp\left[-\left(\frac{\Omega(x)-\Omega_0}{\Delta\Omega}\right)^2\right]\times\exp(-\tilde{M}^2x^2)\,dx, \tag{1.11}$$

where $u(x)=u_0/(1-x)$ (we recall that u_0 is the argument of the diffractive multiplier for the exact equality $\Omega=s\omega$).

As follows from (15) and Eqs. (1.10) and (1.11), J_x must be negative to obtain a positive value of G. For this reason, it is important to take into account the diffractive multiplier $\sin^2 u(x)/u^2(x)$ and the distribution function $\exp\{-[(\Omega-\Omega_0)/\Delta\Omega]^2\}$ in the lower orders with respect to x.

The effective integration range in (1.11) is determined by the width of the function $\exp(-\tilde{M}^2x^2)$ and amounts to $\delta x\sim 1/\tilde{M}=\Delta P/M$. The frequency interval $\delta\Omega\sim s\omega(1-n_\omega)\delta x$, which is significantly smaller than the distance between the point $s\omega$ and the special point $n_\omega s\omega$ on the Ω axis.

The result of the subsequent calculations depends on the value of u_0, the sharpness of the function $\exp\{-[(\Omega(x)-\Omega_0)/\Delta\Omega]^2\}$, and the possible detuning $|\Omega-s\omega|$.

In the general case, where the condition $\Omega_0\neq s\omega$ holds, as a result of expansion of all the multipliers in front of $\exp(-\tilde{M}^2x^2)$ in series in the small parameter x followed by integration we obtain

$$J=-\frac{16\sqrt{\pi}\,s\omega(1-n_\omega)I_2}{\Delta\Omega M}\frac{\sin^2 u_0}{u_0^2}\exp\left[-\left(\frac{\Omega_0-s\omega}{\Delta\Omega}\right)^2\right]\times\left[u_0\cot u_0+\frac{(\Omega_0-s\omega)s\omega(1-n_\omega)}{(\Delta\Omega)^2}\right]. \tag{1.12}$$

The result is valid if the following condition holds:

$$\frac{|\Omega_0-s\omega|s\omega(1-n_\omega)}{\tilde{M}(\Delta\Omega)^2}\ll 1.$$

APPENDIX 2:

Let us describe the dependence of the intensity of the pump wave on the transverse coordinate ρ by the simple Gaussian law

$$I_1(\rho)=I_{01}\exp[-\rho^2/\rho_0^2],$$

where I_{01} is the intensity of the wave on the focal axis. In accordance with this dependence we specify the electron density in the interaction volume by the expression

$$n_e(\rho)=n_{0e}\exp[-(\rho^2/\tilde{\rho}_0^2)]$$

(in view of the nonlinear dependence of the ionization probability of an atom on the intensity of the ionizing wave $\tilde{\rho}_0\leqslant\rho_0$). It follows from the definition of the argument of the diffraction multiplier u_0 that $1-n_\omega=2u_0(\rho)/s\omega l$. This makes it possible to represent the value of u_0 in the form of the dependence

$$u_0(\rho)=u_{0m}\exp[-(\rho/\tilde{\rho}_0)^2],$$

where u_{0m} is the value of the argument on the focal axis. The substitution of this dependence into the integral J_p (20) brings it into the form

$$J_\rho=\frac{u_{0m}^2\tilde{\rho}_0^2}{s\omega l}J_t, \tag{2.1}$$

where

$$J_t=\int_0^{(\rho_0/\tilde{\rho}_0)^2}e^{-2t}\cot(u_{0m}e^{-t})\,dt. \tag{2.2}$$

Finally, the substitution $x=u_{0m}e^{-t}$ reduces (2.2) to the familiar integral[16]

$$J_t\approx\frac{1}{u_{0m}^2}\int_0^{u_{0m}}x\cot x\,dx. \tag{2.3}$$

The lower integration limit $u_{0m}\exp[-(\rho_0/\tilde{\rho}_0)^2]$ is set equal to zero, since $\exp[-(\rho_0/\tilde{\rho}_0)^2]<1$, and the contributions to the value of the integral come mainly from the vicinity of the upper limit.

*)E-mail: zaretsky@imp.kiae.ru

[1] K. Miyazaki and H. Sakai, J. Phys. B **25**, L83 (1992).
[2] K. Kondo, H. Sarukura, K. Sajiki, and S. Watanabe, Phys. Rev. A **47**, R2480 (1993).
[3] J. J. Macklin, J. D. Kmetec, and C. L. Gordon III, Phys. Rev. Lett. **70**, 776 (1993).
[4] C.-G. Wahlstrom, J. Larsson, A. Persson *et al.*, Phys. Rev. A **48**, 4709 (1993).
[5] J. W. G. Tisch, R. A. Smith, J. E. Muffett *et al.*, Phys. Rev. A **49**, R28 (1994).
[6] P. Salieres, T. Ditmire, K. S. Budil *et al.*, J. Phys. B **27**, L217 (1994).
[7] P. Peatross and D. D. Meyerhofer, Phys. Rev. A **51**, R906 (1995).
[8] M. E. Faldon and M. H. R. Hutchinson, J. Opt. Soc. Am. B **9**, 2094 (1992).
[9] T. E. Glover, R. W. Schoenlein, A. H. Chin, and C. V. Shank, Phys. Rev. Lett. **76**, 2468 (1996).
[10] A. Bouhal, R. Evans, G. Grillon *et al.*, J. Opt. Soc. Am. B **14**, 950 (1997).
[11] H. Reiss, J. Opt. Soc. Am. B **4**, 726 (1987).
[12] M. V. Fedorov, *Electrons in Strong Light Fields* [in Russian], Nauka, Moscow (1991), p. 31.
[13] D. F. Zaretskiĭ and É. A. Nersesov, Zh. Éksp. Teor. Fiz. **107**, 79 (1995) [JETP **80**, 41 (1995)].
[14] L. V. Keldysh, Zh. Éksp. Teor. Fiz. **47**, 1945 (1964) [Sov. Phys. JETP **20**, 1307 (1965)].
[15] D. F. Zaretskiĭ and É. A. Nersesov, Zh. Éksp. Teor. Fiz. **109**, 1994 (1996) [JETP **82**, 1073 (1996)].
[16] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, New York (1965) [Russ. original, 4th ed., Fizmatgiz, Moscow (1963), p. 1].

Translated by P. Shelnitz

Lorentz ionization of atoms in a strong magnetic field

V. S. Popov

Institute of Theoretical and Experimental Physics, 117218 Moscow, Russia

B. M. Karnakov,*) V. D. Mur

Moscow State Engineering Physics Institute, 115409 Moscow, Russia

(Submitted 17 November 1998)
Zh. Éksp. Teor. Fiz. **115**, 1642–1663 (May 1999)

Lorentz ionization emerges due to the motion of atoms or ions in a strong magnetic field. We use the semiclassical approximation to calculate the probability w_L of Lorentz ionization. We also find the stabilization factor S, which takes into account the reduction by the magnetic field of the probability of ionization decay of the bound s state. We estimate the probabilities w_L in magnetic-cumulation experiments and in astrophysics. We also qualitatively examine the dynamics of the magnetic cumulation process with allowance for the conductivity of the shell. Finally, we discuss a paradox related to the use of the quasistationary solution at the shell expansion stage. © *1999 American Institute of Physics.* [S1063-7761(99)00905-1]

The semiclassical theory of ionization of atoms and ions by constant electric and magnetic fields has been developed in Refs. 1 and 2. The special case of calculating the probability w_L of Lorentz ionization (i.e., ionization of atoms and ions as a result of their motion in a constant or quasistationary magnetic field) is studied in the present paper.[1)] We list the estimates of w_L made through magnetic cumulation experiments (compression of an axial magnetic field by an explosion), in which record-breaking values of the magnetic field strength were obtained in laboratory conditions,[4–6] and from astrophysical data (magnetic white dwarfs). We also give a qualitative description of the dynamics of the magnetic cumulation process.

Throughout the paper we use the atomic system of units: $\hbar=e=m_e=1$ ($\mathscr{E}_a=5.14\times10^9\ \mathrm{V\,cm^{-1}}$, $\mathscr{H}_a=137\mathscr{E}_a=2.35\times10^9$ G, and $me^4/\hbar^3=4.13\times10^{16}\ \mathrm{s^{-1}}$ are the atomic units of the electric and magnetic field strengths and of the probability w_L) and the same notation as in Refs. 2 and 3.

1. THEORY OF LORENTZ IONIZATION

If an atom or ion enters a region of space where there is a constant magnetic field $\mathscr{H}$, the electric field $\mathscr{E}_0$ that appears as a result of the Lorentz transformation in the reference frame K_0 in which the atom or ion is at rest causes ionization, which has become known as Lorentz ionization. Here $\mathscr{E}_0\perp\mathscr{H}_0$,

$$\frac{\mathscr{E}_0}{\mathscr{H}}=\sqrt{\Gamma^2-1}\ \sin\varphi=\sqrt{a},$$

$$\frac{\mathscr{H}_0}{\mathscr{H}}=\sqrt{\Gamma^2\sin^2\varphi+\cos^2\varphi}=\sqrt{1+a}, \tag{1}$$

with $\Gamma=1/\sqrt{1-v^2/c^2}$ the Lorentz factor, v the velocity of the atom, φ the angle between $\mathbf{v}$ and $\mathscr{H}$, and

$$\rho\equiv\frac{\mathscr{E}_0}{\mathscr{H}_0}=\sqrt{\frac{a}{1+a}},\quad a=(\Gamma^2-1)\sin^2\varphi. \tag{2}$$

The semiclassical theory of Lorentz ionization follows directly from the results obtained in Refs. 1 and 2. The parameter $\gamma=\omega_c/\omega_t$ introduced in those papers ($\omega_c=e\mathscr{H}_0/m_ec$ is the cyclotron frequency and $\omega_t=\mathscr{E}/\kappa$ is the tunneling frequency), which determines the nature of subbarrier electron motion, is given by the expression

$$\gamma_L=\frac{\kappa\mathscr{H}_0}{c\mathscr{E}_0}=\frac{\kappa}{v_\perp}\sqrt{1-(1-\Gamma^{-2})\cos^2\varphi}, \tag{3}$$

where $v_\perp=v\sin\varphi$ is the transverse (with respect to the field $\mathscr{H}$) component of the atom's velocity (here v is expressed in atomic units, $v_a=e^2/\hbar$), $\kappa=\sqrt{2E_i}$ in terms of the ionization potential E_i of the atom (or ion), and $c=\alpha^{-1}=137$.

For nonrelativistic particles we have $\rho\ll1$, and γ_L may take any values:

$$\rho=\frac{v_\perp}{c},\quad \gamma_L=\frac{\kappa}{v_\perp},\quad \mathscr{E}_0=\frac{v_\perp}{c}\left(1+\frac{v^2}{2c^2}\right)\mathscr{H}. \tag{4}$$

On the other hand, for ultrarelativistic particles ($\Gamma\gg1$) we have

$$\rho=\begin{cases}1-\dfrac{1}{2\Gamma^2\sin^2\varphi}, & \varphi\gg\Gamma^{-1},\\[2ex] \dfrac{\varphi}{\sqrt{\Gamma^{-2}+\varphi^2}}, & 0<\varphi\ll1.\end{cases} \tag{5}$$

Hence crossed fields are generated in the reference frame K_0, i.e., the fields $\mathscr{E}_0$ and $\mathscr{H}_0$ are mutually perpendicular and equal in magnitude (except for particles entering the region within a narrow cone, $\varphi\lesssim\Gamma^{-1}$, around the direction of the magnetic field). Here $\mathscr{E}_0$ may be several times greater than the initial magnetic field $\mathscr{H}$, and

TABLE I.

γ	τ_0	$g(\gamma)$	$P(\gamma)$	$Q(\gamma)$	$f(\gamma)$
0.5	0.5071	1.0084	0.9583	1.058	2.81(−3)
0.8	0.8296	1.0219	0.8934	1.160	1.17(−2)
0.9	0.9425	1.0280	0.8651	1.209	1.68(−2)
1.0	1.0590	1.0348	0.8337	1.268	2.32(−2)
1.1	1.1795	1.0425	0.7990	1.339	3.12(−2)
1.2	1.3046	1.0511	0.7614	1.422	4.09(−2)
1.3	1.4349	1.0606	0.7208	1.523	5.25(−2)
1.4	1.5711	1.0710	0.6776	1.643	6.63(−2)
1.5	1.7137	1.0825	0.6319	1.787	8.25(−2)
1.6	1.8636	1.0950	0.5843	1.962	0.101
1.7	2.0214	1.1087	0.5351	2.174	0.123
1.8	2.1878	1.1234	0.4850	2.432	0.148
1.9	2.3635	1.1393	0.4346	2.750	0.176
2.0	2.5490	1.1564	0.3847	3.143	0.209
2.1	2.7450	1.1748	0.3362	3.630	0.245
2.2	2.9517	1.1943	0.2899	4.239	0.285
2.3	3.1694	1.2150	0.2465	5.001	0.330
2.4	3.3983	1.2368	0.2067	5.959	0.379
2.5	3.3683	1.2597	0.1709	7.168	0.433
2.6	3.3894	1.2837	0.1394	8.699	0.492
2.7	4.1515	1.3087	0.1121	1.064(1)	0.556
2.8	4.4244	1.3346	8.905(−2)	1.311(1)	0.625
2.9	4.7078	1.3613	6.981(−2)	1.625(1)	0.699
3.0	5.0018	1.3888	5.406(−2)	2.027(1)	0.778
3.1	5.3061	1.4170	4.136(−2)	2.540(1)	0.862
3.2	5.6207	1.4458	3.127(−2)	3.199(1)	0.951
3.3	5.9454	1.4752	2.338(−2)	4.044(1)	1.045
3.4	6.2802	1.5051	1.728(−2)	5.131(1)	1.145
3.5	6.6251	1.5355	1.263(−2)	6.532(1)	1.250
3.6	6.9801	1.5664	9.126(−3)	8.340(1)	1.359
3.8	7.7200	1.6292	4.613(−3)	1.370(2)	1.594
4.0	8.5000	1.6934	2.233(−3)	2.273(2)	1.849
4.5	1.0625(1)	1.8583	3.019(−4)	8.318(2)	2.575
5.0	1.300(1)	2.0280	3.135(−5)	3.156(3)	3.427
6	1.850(1)	2.3767	1.546(−7)	4.899(4)	5.507
7	2.500(1)	2.7332	2.722(−10)	8.159(5)	8.088
8	3.250(1)	3.0945	1.725(−13)	1.426(7)	1.117(1)
9	4.100(1)	3.4588	3.954(−17)	2.584(8)	1.475(1)
10	5.050(1)	3.8254	3.292(−21)	4.813(9)	1.884(1)
15	1.130(2)	5.6751	3.560(−48)	1.398(16)	4.675(1)

Note: In all tables $x(y)$ stands for $x\times 10^y$.

$$\gamma_L=\frac{\kappa}{137}\left[1+\frac{1}{2\Gamma^2\sin^2\varphi}\right]\ll 1.$$

In particular, if $\mathbf{v}\perp\mathscr{H}$, then

$$\mathscr{E}_0=\sqrt{\Gamma^2-1}\,\mathscr{H},\quad \mathscr{H}_0=\Gamma\mathscr{H},$$

$$\rho=\sqrt{1-\Gamma^{-2}},\quad \gamma_L=\frac{\kappa}{v}=\frac{\kappa}{137}\frac{1}{\sqrt{1-\Gamma^{-2}}}. \tag{6}$$

The probability w_L of Lorentz ionization of the atomic s level is (in the lab frame)

$$w_L=\Gamma^{-1}\kappa^2\times 2^{2\eta}C_\kappa^2\epsilon^{1-2\eta}P(\gamma_L)[Q(\gamma_L)]^\eta \times\exp\left\{-\frac{2}{3\epsilon}g(\gamma_L)\right\}. \tag{7}$$

Here

$$\epsilon=\mathscr{E}_0/\kappa^3\mathscr{E}_a=\Gamma\kappa^{-1}v_\perp h,\quad h=\mathscr{H}_0/\kappa^2\mathscr{H}_a; \tag{7'}$$

ϵ and h are the ''reduced'' values of the fields in the rest frame K_0 of the atom, $\eta=Z/\kappa$ is the Sommerfeld parameter for the discrete spectrum (or the effective quantum number n^*), Z is the charge of the atomic core, C_κ is the asymptotic coefficient (at infinity) of the wave function of the free atom,[2)] and $P(\gamma)$, $Q(\gamma)$, and $g(\gamma)$ are functions introduced in Ref. 2 and referring to the angle $\theta=\pi/2$ between the fields (see Table I). Note that the factor Γ^{-1} in (7) allows for the effect of relativistic retardation as one goes from the rest frame K_0 to the lab frame K.

It is convenient to write the probability of Lorentz ionization in the separable form

$$w_L=\Gamma^{-1}Sw(\mathscr{E}_0), \tag{8}$$

where $w(\mathscr{E}_0)$ is the probability of ionization by the electric field $\mathscr{E}_0$ alone, say, in the semiclassical approximation,

$$w_{\rm cl}(\mathscr{E}_0)=\kappa^2 2^{2\eta}|C_\kappa|^2\epsilon^{1-2\eta}\exp\{-2/3\epsilon\}. \tag{9}$$

The other quantity in (8) that need defining, S, is the ''stabilization factor,'' which determines the suppression of the decay of a bound state initiated by the magnetic field:

$$S=P(\gamma_L)[Q(\gamma_L)]^\eta\exp\left\{-\frac{1}{h}f(\gamma_L)\right\}. \tag{10}$$

Here

$$f(\gamma)=\frac{2}{3}\gamma[g(\gamma)-1] =\begin{cases}\frac{1}{45}\gamma^3\left(1+\frac{11}{252}\gamma^2+\cdots\right), & \gamma\ll 1,\\ \frac{1}{4}\gamma^2\left(1-\frac{8}{3\gamma}+\cdots\right), & \gamma\gg 1,\end{cases} \tag{11}$$

$$P(\gamma)[Q(\gamma)]^\eta =\begin{cases}1-\frac{1}{6}\left(1-\frac{4}{3}\eta\right)\gamma^2+\cdots, & \gamma\ll 1,\\ c_0c_1^\eta\gamma^{1-2\eta}\exp\left\{-\left(\frac{1}{2}\gamma^2-\pi\eta\gamma\right)\right\}, & \gamma\gg 1,\end{cases} \tag{12}$$

where c_0 and c_1 are numerical factors: $c_0=1.716$ and $c_1=0.0106$.

The Coulomb factor $Q(\gamma_L)$ substantially increases the probability of Lorentz ionization if $\gamma_L\gtrsim 1$ and $\eta>0$. On the whole, however, the pre-exponential factor in (7) decreases rapidly with increasing γ_L [see Eq. (12)]. We also note that although the functions $P(\gamma)$ and $Q(\gamma)$ change more rapidly than $g(\gamma)$ and $f(\gamma)$, the probability w_L is most sensitive to variations in g and f, since these functions enter into the exponents in (7) and (10) with large coefficients ($2/3\epsilon$ and $1/h$, respectively).

TABLE II. Stabilization factor S for the hydrogen atom (ground state $\kappa=1$).

$\mathscr{H}$	v 0.8	0.9	1.0	1.1	1.2	1.5	2.0	5	10
1		1.6(−33)	2.2(−24)	2.3(−18)	3.1(−14)	1.5(−7)	1.38(−3)	0.659	0.950
2	2.2(−24)	4.2(−17)	1.5(−12)	1.5(−9)	1.8(−7)	3.9(−4)	3.74(−2)	0.813	0.975
5	3.7(−10)	2.9(−7)	1.9(−5)	2.1(−4)	2.05(−3)	4.37(−2)	0.271	0.922	0.990
10	2.0(−5)	5.60(−4)	4.53(−3)	1.80(−2)	4.62(−2)	0.212	0.524	0.961	0.995
15	7.6(−4)	6.96(−3)	2.79(−2)	6.96(−2)	0.130	0.358	0.653	0.975	0.997
25	1.39(−2)	5.22(−2)	0.119	0.206	0.299	0.546	0.779	0.986	0.998
50	0.123	0.236	0.355	0.464	0.558	0.748	0.889	0.994	0.999
100	0.367	0.503	0.613	0.697	0.762	0.876	0.949	0.998	1000
350	0.799	0.863	0.905	0.933	0.952	0.981	0.995	1.000	1000

Note: The magnetic field strength is measured in megagauss, and the atomic velocity v is measured in atomic units $v_a=e^2/\hbar=2.19\times10^8$ cm s^{-1}. The fact that a cell is empty means that $S<10^{-40}$.

For slow particles the stabilization factor is exponentially small:

$$S\approx\exp\left\{-\frac{1}{4h}\left(\frac{\kappa}{v_\perp}\right)^2\right\},\quad v\ll\kappa, \tag{10′}$$

with $\gamma_L\gg1$. This factor, however, rapidly increases with atomic velocity and approaches unity for $v\gg0.3\kappa h^{-1/3}$, or $\gamma_L\ll3.5h^{1/3}$:

$$S=1+\frac{2}{9}\left(\eta-\frac{3}{4}\right)\gamma_L^2-\frac{1}{45h}\gamma_L^3+\cdots,\quad \gamma_L=\frac{\kappa}{v_\perp}. \tag{10″}$$

Note that for negative ions ($\eta=0$) the stabilization factor is always smaller than unity:

$$S(\gamma)=\begin{cases}1-\dfrac{1}{6}\gamma^2-\dfrac{1}{45h}\gamma^3-\dfrac{1}{3240}\gamma^4-\cdots, & \gamma\to0,\\ c_0\gamma\exp\left\{-\dfrac{1}{h}\left(\dfrac{1}{4}\gamma^2-\dfrac{2}{3}\gamma+\dfrac{1}{2}\right)\right\}, & \gamma\to\infty.\end{cases} \tag{13}$$

For neutral atoms and positive ions it can also be larger than unity, but this happens only within the range of parameters where the reduced field ϵ is of order unity, in view of which the semiclassical approximation breaks down.

Numerical calculations yield the curves for S shown in Fig. 3 of Ref. 3. This figure shows that $S\ll1$ for weak magnetic fields and for $\gamma_L\gtrsim1$. The effect of the Coulomb interaction on S becomes appreciable if $\gamma_L>1.5$, and allowing for this interaction increases the numerical value of w_L.

Static magnetic fields generated in laboratory conditions do not exceed one megagauss.[8] The method of magnetic cumulation (i.e., explosive compression of a magnetic field surrounded by a well-conducting shell), proposed by Sakharov in 1951, made it possible to reach record-breaking values $\mathscr{H}=25$ MG in the USSR (Refs. 4–6) and $\mathscr{H}=15$ MG in the USA (Ref. 8). Progress in this field of research could lead to field strengths of 3×10^7–10^8 G (Ref. 9). Other possibilities of generating ultrastrong magnetic fields are also discussed, e.g., the compression of a metallic shell by the pressure of light from a high-power laser.[10]

Bearing all these aspects in mind, we calculated the stabilization factor S for the hydrogen atom. Table II shows that in the region of values of $\mathscr{H}$ and v considered here the exponential suppression (10′) of the ionization probability is rapidly superseded by (10″), where the effect of the magnetic field can be neglected. For fast particles (i.e., for $v\gtrsim10\kappa$ and the more so for $\Gamma\gg1$) $S\approx1$, i.e., atomic-level ionization proceeds at the same rate as in the case of an electric field $\mathscr{E}_0$ acting alone in the rest frame of the atom. This distinguishes Lorentz ionization with $\Gamma\gg1$ from the well-known problem of pair production in vacuum, whose probability for crossed fields vanishes identically.[11,12] Physically, the reason for this difference is obvious: in the case of ionization of an atom there is a special reference frame K_0, whereas the vacuum is Lorentz-invariant and one can always select a reference frame in which the field strengths $\mathscr{E}$ and $\mathscr{H}$ are as small as desired (in this case, obviously, pairs are not produced).

The results of calculations for the H^- ion (the electron affinity energy $E_i=0.7542$ eV, and $\kappa=0.2355$) are listed in Table III. As for other negative ions with a small binding energy,[3)] the dependence of w_L on $\mathscr{H}$ and v is similar the one previously studied, but the region where $S\approx1$ begins at much smaller values of $\mathscr{H}$ and v.

2. NUMERICAL ESTIMATES

For $\mathscr{H}$ is less than one megagauss, the atom is actually stable, since the electric field $\mathscr{E}_0$ is too weak ($\mathscr{E}_0<0.01$ if

TABLE III. Stabilization factor S for a negative atomic hydrogen ion ($\kappa=0.236$).

$\mathscr{H}$	v 0.15	0.2	0.3	0.4	0.6	0.8	1.0	2.0
0.01	–	–	–	4.3(−27)	1.8(−8)	5.5(−4)	2.15(−2)	0.619
0.1	–	7.1(−23)	4.3(−7)	2.19(−3)	0.164	0.466	0.675	0.951
1.0	–	4.81(−3)	0.209	0.514	0.815	0.914	0.954	0.993
10	9.4(−11)	0.463	0.775	0.887	0.957	0.978	0.987	0.997
25	7.2(−5)	0.628	0.846	0.919	0.967	0.983	0.989	0.998

TABLE IV. Lorentz ionization probability for hydrogen atoms.

	$\mathscr{H}=25$ MG				$\mathscr{H}=350$ MG		
v	$\mathscr{E}_0$	S	w_L	v	$\mathscr{E}_0$	S	w_L
0.5	5.32(−3)	3.69(−9)	4.1(−44)	0.167	2.48(−2)	6.6(−19)	7.6(−12)
1.0	1.06(−2)	0.119	1.03(−9)	0.20	2.98(−2)	1.01(−11)	8.05(−3)
1.25	1.33(−2)	0.345	6.41(−4)	0.22	3.28(−2)	7.79(−9)	50.4
1.67	1.77(−2)	0.645	2.40	0.25	3.72(−2)	2.06(−6)	1.06(5)
2.0	2.13(−2)	0.779	1.55(5)	0.3	4.47(−2)	6.73(−4)	5.22(8)
2.5	2.66(−2)	0.882	5.56(7)	0.4	5.96(−2)	6.70(−2)	1.33(12)
3.33	3.55(−2)	0.950	2.14(10)	0.5	7.45(−2)	0.298	3.54(13)
5.0	5.32(−2)	0.984	6.24(12)	1.0	0.149	0.905	2.19(15)
10	0.107	0.998	7.73(14)				

Note: The values of v and $\mathscr{E}_0$ are measured in atomic units, while the value of the ionization probability w_L is measured in s^{-1}.

$\Gamma<25$). If the atom velocity is very low, Lorentz ionization can be observed in the region $\mathscr{H}\gtrsim 10$ MG. For instance, at $\mathscr{H}=25$ MG ($h=0.0106$), we obtain the values of w_L listed in Table IV; the values of the atomic velocity v (the case where $\mathbf{v}\perp\mathscr{H}$) and the electric field $\mathscr{E}_0$ acting in the rest frame of the atom are also listed Table IV. In the range of velocities in question, the situation of the atom changes from almost total stability to ionization in a time interval of roughly the atomic time.

The stabilization factor S is especially large when $\gamma_L>1$. For the values of $\mathscr{E}_0$ (and, accordingly, of the probability w_L) not to be too large, κ must be much less than unity; then $v=\kappa/\gamma_L\ll 1$ holds, i.e., we are dealing with the nonrelativistic case. Anomalously small values of κ are attained for Rydberg atomic states (e.g., in the hydrogen atom $\kappa=1/n$ for states with the principal quantum number n) and in solid state physics (for Wannier–Mott excitons[14] in semiconductors). For instance, since in the germanium crystal the effective electron mass is $m_*\approx 0.2m_e$ and the dielectric constant is $\varepsilon\approx 16$, we have $\kappa=m_*/\varepsilon m_e\approx 1/80$, with the result that the characteristic field strengths at which ionization is essentially instantaneous ($\epsilon\sim 1$) are of order $\kappa^3\mathscr{E}_a\sim 10\ \mathrm{kV\,cm^{-1}}$. Such fields can easily be generated in experiments.

Ultrastrong magnetic fields can also be encountered in astrophysics. In this connection we should mention magnetic white dwarf stars, which are being actively investigated at present.[15–18] The fields at the surface of such an object are huge (see, e.g., the table on p. 35 in Ref. 18, which lists about 50 such objects with fields ranging from 2 MG to roughly 1000 MG). For instance, the magnetic field in the star Grw+70°08247 corresponds to a magnetic dipole and varies from the maximum value $\mathscr{H}_m=350$ MG at the poles of the star to $0.5\mathscr{H}_m$ at the equator[17] (according to Ref. 18, in this case $\mathscr{H}_m\approx 240$ MG). What makes white dwarfs so special (compared, say, to neutron stars, where magnetic fields are much stronger) is that it is possible to study their optical spectra.[16,17] This allows us to study the effect of $\mathscr{E}$ and $\mathscr{H}$ on the atomic levels, primarily for the atoms of hydrogen and helium. Lorentz ionization of atoms may take place when such a star passes through a cloud of neutral hydrogen. The data listed in Table IV suggest that the probability of Lorentz ionization of hydrogen atoms in a field $\mathscr{H}=350$ MG ($h=0.149$) becomes appreciable for stellar velocities $v\gtrsim 400\ \mathrm{km\,s^{-1}}$, with w_L depending very strongly on v.

However, even if the stellar velocity with respect to the cloud of interstellar gas is relatively low, Lorentz ionization of the atoms of this gas still takes place because[4)] the atoms fall to the surface of the star with a velocity $v=\sqrt{2GM/R}$ (since the gas density is low, we can neglect the collisions between the atoms). Assuming that a white dwarf has a mass $M\sim M_\odot=2\times 10^{33}$ g and a radius $R\sim 10^4$ km, we get $v\simeq(4-5)\times 10^3\ \mathrm{km\,s^{-1}}$. Since $\mathscr{H}(r)\propto\mathscr{H}_m(r/R)^{-3}$ holds outside the star, Lorentz ionization of hydrogen atoms takes place at distances of the order of several R.

A remark concerning the calculations of the values of w_L listed in Table IV is in order. Being asymptotically exact in the limit of a weak field ($\mathscr{E}\to 0$), the semiclassical formula (7) ceases to be valid relatively soon. The complex-valued energies $E(\mathscr{E})=E_r-i\Gamma/2$ of the Stark resonances in the hydrogen atom have been calculated by many researchers; see, e.g., Refs. 19–27 (the ionization probability is $w(\mathscr{E})\equiv\Gamma(\mathscr{E})$ if the level width Γ is small). They used various numerical methods, including summation of divergent perturbation-theory series[26–30] by the use of Borel–Padé[20] and Padé–Hermite[21] approximants. We write the probability of ionization of the s level as follows:

$$w(\mathscr{E})=q(\mathscr{E})w_{\mathrm{cl}}(\mathscr{E}), \tag{14}$$

where w_{cl} is given by formula (9), and w is the result of numerical calculations. Figure 1 depicts the function $q(\mathscr{E})$

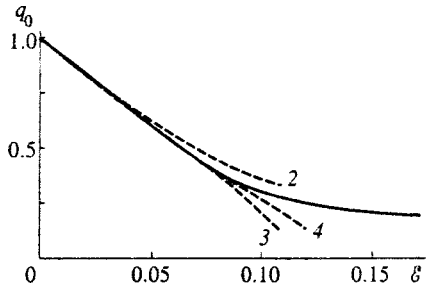

FIG. 1. The function $q_0(\mathscr{E})=w/w_{\mathrm{cl}}$ for the ground state of the hydrogen atom. The dashed curves *2*, *3*, and *4* correspond to allowance in (14′) of terms up to $\mathscr{E}^2$, and $\mathscr{E}^3$, and $\mathscr{E}^4$ inclusive.

for the ground state of the hydrogen ($q=q_0$); here we have used the results of Refs. 18–23. In the case at hand,[31,25–29]

$$w_{\rm cl}(\mathscr{E})=4\mathscr{E}^{-1}\exp\left\{-\frac{2}{3\mathscr{E}}\right\},\quad q_0(\mathscr{E})=\sum_{k=0}^{\infty}c_k\mathscr{E}^k, \tag{14'}$$

where $c_0=1$, $c_1=-107/12$, $c_2=7363/288$, $c_3=-158.75\pm0.07$, $c_4=469.0\pm1.0$, $c_5=10\,250\pm150$, ... (this expansion can be used for $\mathscr{E}<0.1$; see the dashed curves in Fig. 1). Note that the semiclassical asymptotic value of probability, $w_{\rm cl}(\mathscr{E})$, overestimates[5)] the exact value of the ionization probability; for $\mathscr{E}\sim0.2\mathscr{E}_a\sim10^9$ V cm the numerical values of w are five to six times smaller than $w_{\rm cl}$.

Within the range of velocities v where the probability w_L increases dramatically, the correction factor $q_0(\mathscr{E})$ is important and was taken into account in calculating the values of w_L listed in Table IV. Note that formula (10) for the stabilization factor S has a wider range of applicability, up to $\mathscr{H}\sim0.1$. This follows from a comparison of our results with those of numerical calculations done by Johnson *et al.*[32] for the case of parallel $\mathscr{E}$ and $\mathscr{H}$ fields and with the results of recent calculations done by Vaĭnberg and Gani[33] for the case of crossed fields.

3. IONIZATION IN THE PROCESS OF MAGNETIC CUMULATION

Up to this point it was assumed that the magnetic field does not vary in time. Now we turn to the case of a variable field $\mathscr{H}(t)$ generated by magnetic cumulation.

Below we follow the ideas of Sakharov.[5,6] The metallic cylinder squeezing the magnetic field (the inner radius of the cylinder is $R(t)$, the thickness of the shell is $d\ll R$, and its conductivity is σ) in the first approximation can be assumed an ideal conductor. Then

$$\mathscr{H}(t)=\mathscr{H}_0(R_0/R)^2,\quad W(t)=W_0(R_0/R)^2, \tag{15}$$

where $W=\mathscr{H}^2R^2/8$ is the magnetic-field energy (per unit length of cylinder). In terms of the dimensionless variables $\xi=R/R_0$ and $\tau=v_0t/R_0$, the energy conservation law yields

$$\dot\xi^2=1+\frac{1}{K}\left(1-\frac{1}{\xi^2}\right),\quad \xi(0)=-\dot\xi(0)=1, \tag{16}$$

where $K=Mv_0^2/2W_0=8\pi\rho_0v_0^2d_0/\mathscr{H}_0^2R_0$ is the ratio of the shell kinetic energy to the magnetic energy within the shell at the initial moment $t=0$ (real values of K are much larger than unity), ρ_0 is the density of the shell material, and the dot stands for a derivative with respect to τ. The solution of Eq. (16) is

$$\xi^{(0)}(t)=\sqrt{\xi_m^2+\frac{K+1}{K}(\tau-\tau_m)^2},$$
$$\mathscr{H}^{(0)}(t)=\mathscr{H}_m\left\{1+\frac{(t-t_m)^2}{(\Delta t)^2}\right\}^{-1}, \tag{17}$$

where $\mathscr{H}_m/\mathscr{H}_0=W_m/W_0=K+1$, the quantities $\xi_m=(K+1)^{-1/2}$ and $\tau_m=K/(K+1)$ refer to the momemt of maximum magnetic field, $2\Delta t=[\sqrt{K}/(K+1)]T_0$ is the time during which the field is close to its maximum value, and $T_0=2R_0/v_0$ is the characteristic of the entire process ($\Delta t/T_0\approx K^{-1/2}\ll1$). On the other hand, the electron tunneling time is $T_t\sim\kappa^2/2\mathscr{E}_0\sim(\kappa\epsilon)^{-1}\times10^{-17}c\approx(vh)^{-1}\times10^{-15}\,{\rm s}^{-1}\ll\Delta t$, with the result that the total Lorentz ionization probability can be calculated in the adiabatic approximation:[6)]

$$w_t=\int_0^\infty w_L(\mathscr{H}(t))\,dt=\mu w_L(\mathscr{H}_m)\Delta t, \tag{18}$$

$$\mu=\sqrt{\frac{\pi h_m}{f(\gamma_L)}}=\kappa^{-1}\sqrt{\frac{\mathscr{H}_m}{\mathscr{H}_a}}\begin{cases}11.9\gamma_L^{-3/2}, & \gamma_L\ll1,\\ 3.5\gamma_L^{-1}, & \gamma_L\gg1.\end{cases} \tag{18'}$$

Here we have assumed that the particle stays in the region with the magnetic field for a time $\gtrsim\Delta t$. This is true for ions, for which the Larmor radius is usually smaller than the minimum shell radius $R_m\equiv R(t_m)=K^{-1/2}R_0$. A magnetic field does not bend the path of neutral atoms, and so for atoms $\mu=Lv_0/R_mv$, where v_0 is the initial shell velocity, v is the atomic velocity, and L is the length of the path of the atom in the magnetic field ($L\sim R_m/\sin\varphi$). In this case $\mu\sim10^{-3}\ll1$, with the result that the probability of the atoms being ionized (in conditions of an explosion) is suppressed as compared to the probability for negative ions. Using these formulas, we can easily calculate w_t if we know the parameters of the magnetic cumulation process.

Note that the quantity

$$\mu\Delta t=\kappa^{-1}\sqrt{\frac{\pi}{f(\gamma_L)}}\frac{\mathscr{H}_0}{\mathscr{H}_a}\frac{R_0}{v_0} \tag{18''}$$

in (18) is expressed in terms of the parameters R_0, v_0, and $\mathscr{H}_0$, which refer to the initial moment ($t=0$).

In the next approximation we must include the ohmic losses due to the finite conductivity of the shell material. We have

$$\frac{d}{dt}\left\{\frac{1}{2}M\left(\frac{dR}{dt}\right)^2+\frac{1}{8}\mathscr{H}^2R^2\right\}=-J=-\frac{1}{2}cR\mathscr{E}\mathscr{H}, \tag{19}$$

where J is the electromagnetic energy flux through a wall of radius $R(t)$ in the wall rest frame. The Maxwell equations for the quasistationary field in a highly conducting medium,[34]

$$\frac{\partial\mathscr{H}}{\partial t}=D\Delta\mathscr{H},\quad \mathscr{E}=\frac{c}{4\pi\sigma}\,{\rm curl}\,\mathscr{H},\quad D=\frac{c^2}{4\pi\sigma}, \tag{20}$$

has the solution (cf. Ref. 5)

$$\mathscr{H}(x,t)=\mathscr{H}_0\exp\left(-\frac{x}{\delta}+\int_0^t\lambda(t')\,dt'\right),$$
$$\mathscr{E}=\sqrt{\frac{\lambda}{4\pi\sigma}}\mathscr{H}, \tag{21}$$

where $x=r-R(t)$, $\delta=c/\sqrt{4\pi\sigma\lambda}$ is the skin depth, and

$$\lambda=\frac{1}{\mathscr{H}}\frac{d\mathscr{H}}{dt}\approx-\frac{2}{R}\frac{dR}{dt}=-4T_0^{-1}\frac{\dot\xi}{\xi}. \tag{22}$$

This solution is valid if $\dot\delta/\delta\ll\dot{\mathscr{H}}/\mathscr{H}$, or

TABLE V. Results of numerical calculations.

	$K=100$			$K=1000$			$K=10^4$		
α	β	ξ_m	τ_m	β	ξ_m	τ_m	β	ξ_m	τ_m
0	0	0.100	0.990	0	0.0316	0.999	0	0.0100	1.000
0.05	0.158	0.114	0.971	0.281	0.0392	0.990	0.5	0.0138	0.996
		0.115	0.968		0.0400	0.988		0.0144	0.995
0.1	0.316	0.126	0.960	0.562	0.0445	0.986	1.0	0.0162	0.994
		0.129	0.953		0.0469	0.982		0.0178	0.992
0.15	0.474	0.135	0.951	0.844	0.0490	0.983	1.5	0.0181	0.993
		0.141	0.941		0.0531	0.977		0.0207	0.990
0.2	0.632	0.144	0.945	1.125	0.0528	0.980	2.0	0.0198	0.992
		0.153	0.931		0.0587	0.973		0.0234	0.988
0.3	0.947	0.158	0.933	1.687	0.0593	0.976	3.0	0.0225	0.991
		0.173	0.913		0.0688	0.965		0.0282	0.984

Note: Here ξ_m and τ_m refer to the stopping point, and $\beta=K^{1/4}\alpha$ is the inelasticity parameter. For given $\alpha>0$ and K the first and second rows correspond to planar and cylindrical geometries (Eqs. (24) and (31), respectively).

$$\frac{1}{2}\left|\frac{d}{dt}\frac{1}{\lambda(t)}\right|=\frac{1}{2}\left|1-\frac{\mathscr{H}\ddot{\mathscr{H}}}{\dot{\mathscr{H}}^2}\right|\ll 1, \tag{23}$$

which is similar to the condition of applicability of the WKB method in quantum mechanics. We also note that the approximation (21) is physically meaningful only in the case $\lambda(t)>0$ (compression of the shell) and cannot be used when the shell expands (the reason is explained in Sec. 5).

Equations (20)–(22) yield

$$J=\frac{1}{2}\lambda\mathscr{H}^2R\delta=\frac{4cW_0}{(\pi\sigma T_0)^{1/2}R_0}(-\dot{\xi})^{1/2}\xi^{-7/2},$$

where $W_0=\mathscr{H}_0^2R_0^2/8$, so that we finally arrive at the equation

$$\ddot{\xi}=\frac{1}{K\xi^3}\left(1+\frac{2\alpha}{\sqrt{-\xi\dot{\xi}}}\right),\quad 0<\tau<\tau_m, \tag{24}$$

with the same initial conditions as in Eq. (16). Here τ_m is the stopping point, where $\dot{\xi}(\tau_m)=0$, and α is the dimensionless loss coefficient:[5]

$$\alpha=\frac{c}{\sqrt{2\pi\sigma R_0 v_0}}\approx\frac{40}{\sqrt{\sigma R_0 v_0}} \tag{25}$$

(in this expression the conductivity σ is measured in $\Omega^{-1}\cdot\mathrm{cm}^{-1}$, R_0 in centimeters, and v_0 in kilometers per second).

As a guide we provide some numerical values. For $R_0=1$ cm, $v_0=1\ \mathrm{km\,s^{-1}}$, and $K=100$ (or 10^3) we have $\alpha=0.05$ (0.016) for $\sigma=6\times10^5\,\Omega^{-1}\cdot\mathrm{cm}^{-1}$ (which corresponds to the conductivity of copper at room temperature[13]), while for $\sigma=4\times10^4\,\Omega^{-1}\cdot\mathrm{cm}^{-1}$ ($T=1500\,^\circ$C) we have $\alpha=0.2$ (0.063). In the latter case copper is already in a liquid state ($T_m=1085\,^\circ$C). For $R_0=3$ cm, $v_0=10\ \mathrm{km\,s^{-1}}$, and $K=1000$ we have $\alpha=0.01$ and 0.037 for the above values of σ, which implies $\alpha\ll1$. The results gained from solving Eq. (24) are listed in Table V; several specific features of these solutions and the limits of this equation are discussed in the Appendix.

In deriving Eqs. (16) and (24) we assumed that the magnetic flux $\Phi=\pi R^2\mathscr{H}$ inside the conducting shell is conserved. Let us now estimate the flux losses in the process of cumulation. The solenoidal electric field generated in the shell is

$$\mathscr{E}=-\frac{1}{2\pi cR}\frac{\partial\Phi}{\partial t}=\sqrt{\frac{\lambda}{4\pi\sigma}}\mathscr{H},$$

from which it follows that

$$\frac{d\Phi}{dt}=-\frac{c}{R}\sqrt{\frac{\lambda}{\pi\sigma}}\Phi=-\frac{4\alpha}{T_0}\sqrt{-\frac{\dot{\xi}}{\xi^3}}R^2\mathscr{H}.$$

Passing to dimensionless time $\tau=2t/T_0$, we obtain

$$\frac{\dot{H}}{H}=-2\left(\frac{\dot{\xi}}{\xi}+\alpha\sqrt{-\frac{\dot{\xi}}{\xi^3}}\right), \tag{26}$$

where $\dot{H}=dH/d\tau$, etc. This equation can be used if the total flux variation is small. Bearing this in mind, we arrive at the following estimate (when the shell is at maximum compression):

$$\frac{\mathscr{H}_m}{\mathscr{H}_0}=\xi_m^{-2}\exp\left\{-2\alpha\int_0^{\tau_m}\sqrt{-\frac{\dot{\xi}}{\xi^3}}d\tau\right\}$$
$$\approx(K+1)\left\{1+2\alpha\int_0^{\tau_m}\left(1+\frac{2}{K\xi^2}\right)\sqrt{-\frac{\dot{\xi}}{\xi^3}}d\tau\right\}^{-1}. \tag{27}$$

The ratio $\rho_m=\xi_m/\xi_m^{(0)}=\sqrt{K+1}\,\xi_m$ indicates how much the minimum radius of the shell changes due to losses. In Sec. 4 we show that ρ_m depends not on K or α separately but on the parameter $\beta=K^{1/4}\alpha$. This statement is confirmed by numerical calculations (Fig. 2). For a given value of K, the values of the loss coefficient (increasing in the order of β in Fig. 2) are $\alpha=0.05$, 0.1, 0.2, and 0.3 (the last value is true only at $K=10$ and 100).

Allowing for losses in the cumulation process changes the value of τ_m. However, for the values of K and α being

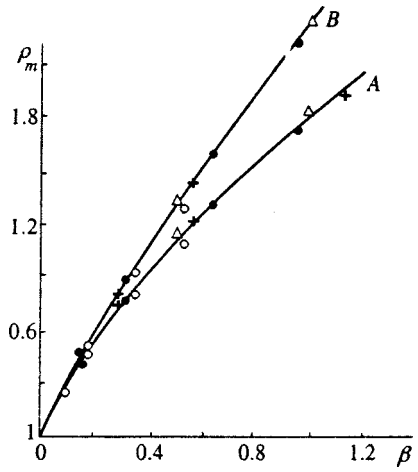

FIG. 2. The ratio $\xi_m/\xi_m^{(0)}$ as a function of $\beta=K^{1/4}\alpha$; the points correspond to the results of numerical calculations by (24) at $K=10$ (○), 100 (●), 1000 (+), and 10 000 (△). For a given value of K these points correspond to the following values of the loss coefficient α (increasing in the order of β): 0.05, 0.10, 0.20, and 0.30. The curves *A* and *B* correspond to planar and cylindrical geometries, respectively.

discussed this change is at most a few percent (see Table V) and rapidly decreases with increasing K. The drop in the maximum attainable magnetic field $\mathscr{H}_m^{(0)}=(K+1)\mathscr{H}_0$ (the ideal case: no losses) is shown in Fig. 3, where curve *A* corresponds to the case in which the magnetic flux is conserved ($\Phi(t)\approx$const and $\mathscr{H}_m\propto\xi_m^{-2}$), while curves *B*1, *B*2, and *B*3 have been calculated by (27), i.e., correspond to cases in which the decrease in $\Phi(t)$ is taken into account, for $K=100$, 10^3, and 10^8, respectively. The last value of K corresponds to the use of a low-yield underground atomic explosion to compress the shell rather than conventional explosives (see p. 85 in Ref. 6). When K changes by a factor of 10^6, the curves in Fig. 3 change only slightly, which confirms the assumption that the parameter of the problem is $\beta=K^{1/4}\alpha$.

Thus, at $K\sim1000$ and $\alpha\sim0.05$, the magnetic flux decreases by factor of 1.5–2 and the peak magnetic field decreases by a factor of 2–3 in comparison to the ideal case ($\alpha=0$). This estimate agrees with the brief remarks made in Ref. 5.

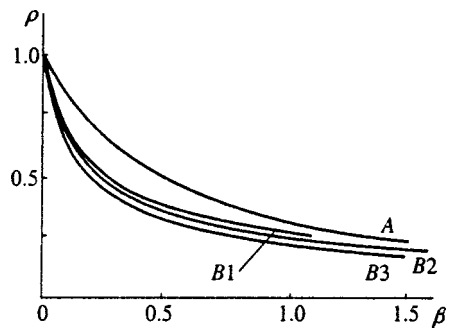

FIG. 3. Magnetic-field losses in the cumulation process; $\beta=\mathscr{H}_m/\mathscr{H}_m^{(0)}$. Curve *A* corresponds to the case in which the magnetic flux Φ is conserved, while curves *B*1, *B*2, and *B*3 correspond to cases in which the decrease in Φ is taken into account.

4. EFFECT OF THE GEOMETRY

Thus far we have neglected the curvature of the shell and simply multiplied the results obtain for the flat (one-dimensional) problem by $2\pi R$. However, real experiments correspond to cylindrical geometry rather than to planar, and this changes the formulas somewhat.

Assuming that $\mathscr{H}\propto\exp\{\lambda t\}$ with λ positive, we arrive at the following solution of Eqs. (20):

$$\mathscr{H}_z=\mathscr{H}(R)\frac{K_0(r/\delta)}{K_0(R/\delta)},\quad \mathscr{E}_\varphi=\sqrt{\frac{\lambda}{4\pi\sigma}}\mathscr{H}(R)\frac{K_1(r/\delta)}{K_0(R/\delta)}, \tag{28}$$

where $r\geqslant R$, $K_\nu(z)$ are modified Bessel functions of the second kind, $\delta=c/\sqrt{4\pi\sigma\lambda}$, and the other components of $\mathscr{H}$ and $\mathscr{E}$ vanish. The electromagnetic energy flux J flowing into the shell and the Joule heat liberated per unit time inside the shell, Q, are

$$J=2\pi R\frac{c}{4\pi}\mathscr{E}_\varphi\mathscr{H}_z|_{r=R}=J_0f_1(z),$$

$$Q=\int_R^\infty\sigma\mathscr{E}_\varphi^2\times2\pi r\,dr=\frac{1}{2}J_0f_2(z), \tag{29}$$

where $z=R/\delta$,

$$f_1(z)=\frac{K_1(z)}{K_0(z)},\quad f_2(z)=z\left\{\frac{K_2(z)}{K_0(z)}-\left[\frac{K_1(z)}{K_0(z)}\right]^2\right\}, \tag{29$'$}$$

and J_0 corresponds to the case of planar geometry:

$$J_0=\frac{1}{2}\lambda\mathscr{H}_0^2R\delta=\frac{1}{2}c\sqrt{\frac{\lambda}{\pi\sigma}}\mathscr{H}_0^2R.$$

The asymptotic expansion

$$\frac{K_\mu(z)}{K_\nu(z)}=1+\frac{\mu^2-\nu^2}{2z}+\frac{(\mu^2-\nu^2)(\mu^2-\nu^2-2)}{8z^2}+\cdots$$

yields[7)]

$$f_1(z)=1+\frac{1}{2z}+\cdots,\quad f_2(z)=1+\frac{1}{z}+\cdots,\quad z\to\infty. \tag{30}$$

Allowing for the correction $\propto1/z=\delta/R$ in the expression for J, we obtain the equation of shell motion in the form

$$\ddot{\xi}=\frac{1}{K\xi^3}\left(1+\frac{\alpha}{\sqrt{-\dot{\xi}\xi}}\right)^2, \tag{31}$$

which differs from (24) only in the α^2 order. It would seem that we can neglect this difference. We will show, however, that the characteristic parameter of the problem is $\beta=K^{1/4}\alpha$, which may be of order unity (see Table V).

Multiplying both sides of Eq. (31) by $\dot{\xi}$, integrating from zero to the stopping point τ_m, and allowing for the boundary condition, we arrive at the relationship

$$\xi_m^{-2}=\left(\frac{R_0}{R_m}\right)^2=K+1-2\int_0^{\tau_m}(\alpha^2+2\alpha\sqrt{-\dot{\xi}\xi})\xi^{-4}\,d\tau. \tag{32}$$

Combining this with (A6) and (A7), we see that $\xi_m=(K+1)^{-1/2}(1+c_1\beta+c_2\beta^2+\cdots)$. The coefficients c_1, c_2, etc.,

TABLE VI. Parameters of the magnetic cumulation process.

				A	
v	$f(\gamma_L)$	μ	$\alpha=0.05$	0.1	0.2
0.5	0.209	0.400	0.852	0.887	0.995
0.667	8.25(−2)	0.636	0.922	0.977	1.114
0.833	4.09(−2)	0.904	0.973	1.046	1.208
1.0	2.32(−2)	1.20	1.012	1.101	1.283
1.250	1.17(−2)	1.69	1.057	1.164	1.373
1.667	4.88(−3)	2.62	1.11	1.24	1.48
2.0	1.81(−3)	3.45	1.14	1.28	1.54
2.5	1.43(−3)	4.83	1.17	1.33	1.61
5.0	1.79(−4)	13.67	1.22	1.42	1.75

Note: The values of the factor A (Eq. (33)) are listed for $K=1000$ and $\mathscr{H}_m=25$ MG ($\kappa=1$ and $\gamma_L=1/v$).

can easily be calculated by (A7). What is more important, however, is that the ratio $\rho_m=\xi_m/\xi_m^{(0)}$, which determines the effect of the inelasticity of the process, seems to vary with β. This assumption is confirmed by the results of numerical calculations—see Fig. 2, in which the curves A and B refer to the cases of planar and cylindrical geometries (Eqs. (24) and (31), respectively). Comparison of curves A and B shows that although the effect of the geometrical factor on the magnetic cumulation process is significant, it does not change the results dramatically.

Now we can go back to the question of the probability of ionization of atoms in explosions (with allowance for losses). For $\alpha=0$ we have formula (18) for the total probability w_t. In the approximation in which $\Phi(t)=\text{const}$, which has already been used in deriving Eq. (31), we have $\mathscr{H}(t)=\mathscr{H}_m(\xi/\xi_m)^{-2}$. Using the definition (10) and the adiabatic approximation, we arrive at (18) with an additional factor A in the pre-exponential factor:

$$A=\frac{2(K+1)}{\mu\sqrt{K}}\int_0^{\tau_m}\exp\left\{-\frac{\pi}{\mu^2}\left[\left(\frac{\xi(\tau)}{\xi_m}\right)^2-1\right]\right\}d\tau, \tag{33}$$

where μ has been defined in (18′), and $\xi(\tau)$ is the solution of Eq. (31). At $\alpha=0$ we have $(\xi^{(0)}/\xi_m)^2=1+K^{-1}(K+1)^2(\tau-\tau_m)^2$ and $A\equiv 1$. If α is positive, the integral in (33) can be evaluated numerically. This calculation (for $K=10^3$; see Table V) shows that the factor A increases with the velocity v of the atom and can reach values ranging from 1.5 to 2, i.e., if we allow for losses the probability w increases somewhat (on the other hand, if v is less than 2×10^8 cm s^{-1} the factor A is usually smaller than unity; for more details see Table VI).

The data of Tables III and V suggest that $w_t=Aw_L\mu\Delta t\geqslant 1$ holds for $v\gtrsim 4\times10^8$ cm s^{-1}. This means that in the present case the ionization of hydrogen atoms is almost total.

Here is a numerical example. Taking $\mathscr{H}=25$ MG, $K=10^3$, $\alpha=0.1$, and $\Delta t=K^{-1/2}R_0/v_0=10^{-7}$ s, we obtain estimates for the total Lorentz ionization of hydrogen atoms in the magnetic cumulation process, $w_t=A\mu\Delta t w_L(\mathscr{H}_m)$: $w_t=8\times10^{-7}$, 9.8×10^{-3}, 0.067, 0.37, 0.64, and 1.00 at $v=5/3$, 1.9, 2.0, 2.1, 2.2, and 2.5, respectively. We see that w_t exhibits threshold behavior as a function of the atomic velocity v. On the other hand, at a fixed value of $\mathscr{H}_m$ the dependence of w_t on the loss coefficient α is very weak.

5. CONDITIONS OF APPLICABILITY OF THE SEMICLASSICAL SOLUTION

As noted earlier, for negative λ (i.e., when the shell expands), the approximation (21) has no physical meaning, since according to (21) as we move deeper into the metal the field $\mathscr{H}(x,t)$ does not decrease but oscillates as $x\to\infty$. Here δ becomes imaginary, and the calculation of Q yields an infinite result.

To resolve this paradox, we examine the time-dependent problem for the equation of heat conduction (diffusion) along the semiaxis $x>0$:

$$u_t=Du_{xx},\quad u(x,t_0)=0,\quad u(0,t)=f(t)\quad \text{for } t>t_0, \tag{34}$$

whose solution[8)]

$$u(x,t)=\frac{x}{\sqrt{4\pi D}}\int_{t_0}^{t}f(t')\exp\left\{-\frac{x^2}{4D(t-t')}\right\}\frac{dt'}{(t-t')^{3/2}} \tag{35}$$

determines the effect of a variable temperature (or magnetic field), $f(t)$, specified at the boundary of the medium, $x=0$. For $f(t)=e^{\lambda t}$ and $\lambda>0$ in (36) we can pass to the limit $t_0\to-\infty$ (adiabatic switch-on of field). Allowing for the values of the integral[36]

$$\int_0^\infty\exp\left\{-\left(\frac{a}{\tau}+b\tau\right)\right\}\tau^{-3/2}d\tau=\sqrt{\frac{\pi}{a}}\exp(-2\sqrt{ab})$$

$(a,b>0)$, we find that

$$u(x,t)=\exp(\lambda t-x/\delta),\quad \rho=x/\delta,\quad \delta=\sqrt{D/\lambda}, \tag{36}$$

which agrees perfectly with (21) at $\lambda=\text{const}$. But if λ is negative, we cannot put $t_0=-\infty$ in (36), and the formal solution

$$u(x,t)=\exp(\lambda t)\cos\rho,\quad \rho=x\sqrt{|\lambda|/D}, \tag{37}$$

corresponds to nonphysical initial ($t=0$) and boundary ($x\to\infty$) conditions and cannot describe the magnetic field when the shell is expanding.

Here are the results obtained by solving the boundary-value problem (34) numerically. Figures 4a and 4b, which correspond to $\lambda>0$ and $\lambda<0$ (the initial time is $t_0=0$), depict the function

$$a(x,t)=u(x,t)/u(0,t), \tag{38}$$

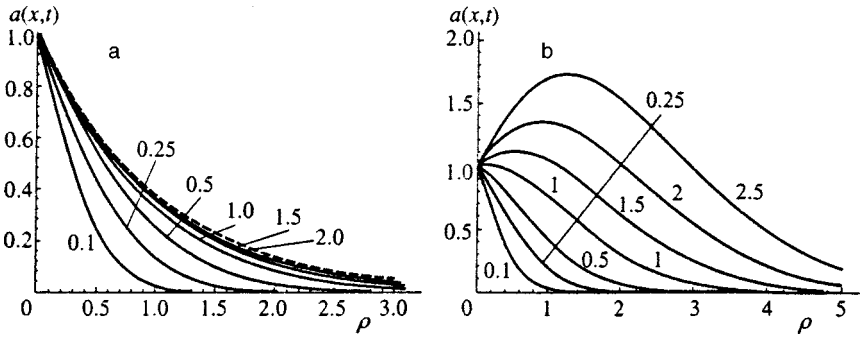

FIG. 4. Thermal wave profile for $\lambda>0$ (a) and $\lambda<0$ (b); the numbers next to the curves indicate the values of $|\lambda|t$. The dashed curve corresponds to the stable regime (37).

which determines the profile of the thermal wave in the medium. Figure 4a shows that when λ is positive, the thermal wave travels from the boundary $x=0$ into the shell (as t increases), where the distribution (36) holds (for $x \lesssim \sqrt{4D\lambda\, t}$). It can be shown that

$$a(x,t)=1-a_1\rho+\frac{1}{2}\rho^2+\cdots, \quad \rho\to 0. \tag{39}$$

For small times the coefficient $a_1=a_1(t)$ increases irrespective of the sign of λ:

$$a_1=(\pi|\lambda|t)^{-1/2}(1+\lambda t+\cdots), \quad |\lambda|t\ll 1, \tag{40}$$

with the result that the function $a(x,t)$ decreases with increasing ρ much faster than (36). In the limit $\lambda t\gg 1$ we have

$$a_1=1+O(\exp\{-\lambda t\}\,(\lambda t)^{-3/2}), \quad \lambda>0, \tag{40'}$$

and $a(x,t)$ is exponentially close to $\exp\{-\rho\}$. Thus, the semiclassical solution (26), (36) is an asymptotic solution, which obtains in the limit $\lambda t\gg 1$ and $x<x_*=\sqrt{4D\lambda\, t}$.

If λ is negative, a_1 vanishes at $\lambda t=-0.854$ and then changes sign:

$$a_1(t)\approx -\exp(|\lambda|t)/2\sqrt{\pi|\lambda|t}, \quad t\to\infty \tag{40''}$$

(see Fig. 4b). Actually, the profile function $a(x,t)$ loses all physical meaning in this case: the field $u(x,t)$ inside the conducting medium is determined not by the boundary value of $f(t)$ bit by the values $f(t')$ at the preceding moments, $0<t'\lesssim|\lambda|^{-1}$. When $x\ll x_*$ and $|\lambda|t\gg 1$, we find that

$$u(x,t)\approx\frac{\xi}{\sqrt{\pi|\lambda|t}}\exp\{-\xi^2\}, \quad \xi=\frac{x}{\sqrt{4Dt}} \tag{41}$$

(the maximum of u is at $\xi=\xi_0=2^{-1/2}$, or $x=x_0=\sqrt{2Dt}\ll x_*$). With the passage of time the distribution spreads in proportion to $\sqrt{t}$ (which is common in diffusion), and the Joule heat released (per unit time) in the process, Q, decreases in proportion to $t^{-3/2}$.

Similar results can be obtained for

$$f(t)=\frac{1}{1+\tau^2}, \quad \tau=\frac{t-t_m}{\Delta t}, \tag{42}$$

which corresponds to the variation of the magnetic field $\mathscr{H}(0,t)$ in the cumulation process (without allowance for losses). In this case the profile function (38) is

$$a=\frac{2}{\sqrt{\pi}}\int_0^\infty\frac{\exp\{-y^2\}}{1+Ay^{-2}+By^{-4}}\,dy,$$
$$A=-\frac{2\rho^2\tau}{1+\tau^2}, \quad B=\frac{\rho^4}{1+\tau^2}, \tag{43}$$

where, in particular, $a(0,t)\equiv 1$. Near the inner surface of the shell, i.e., when $\rho=x/\sqrt{4D\Delta t}\to 0$, for $a(x,t)$ we have an expansion of the form (39) in which

$$a_1=\left[\frac{\pi}{2}(\sqrt{1+\tau^2}+\tau)\right]^{1/2}\left(1-\frac{2\tau}{\sqrt{1+\tau^2}}\right), \tag{44}$$

with

$$a_1(\tau)=\begin{cases}1.5\sqrt{-\pi/\tau}, & \tau\to-\infty,\\ \sqrt{\pi/2}(1-\frac{3}{2}\tau+\cdots), & \tau\to 0,\\ 0, & \tau=3^{-1/2}=0.577,\\ -\sqrt{\pi\tau}, & \tau\to+\infty.\end{cases} \tag{44'}$$

Thus, there is a marked difference between the case of an increasing magnetic field at the boundary and the case of a decreasing magnetic field at the boundary: in the first case the calculation of the magnetic field inside the conducting shell and the ohmic losses can be done in the quasistationary approximation (21), while in the second case we must find the exact solution of the magnetic-field diffusion equation. The equation of motion of the shell becomes much more complicated in the latter case, but there is really no need to use this approach if we are interested solely in the maximum attainable magnetic field $\mathscr{H}_m$ and not in the shell expansion stage.

6. CONCLUSION

The following remarks are in order.

(1) We have found that the probability of Lorentz ionization of neutral atoms reaches a substantial value when the magnetic field strength is of order of tens of megagauss if the atomic velocity is $v\sim 10^8\,\mathrm{cm\,s^{-1}}$. Here w_L changes rapidly within a narrow range of the parameters $\mathscr{H}$ and v (which is characteristic of tunneling processes; see, e.g., Ref. 37). For negative ions (such as H^- and Li^-, where $\kappa\ll 1$), the threshold values of $\mathscr{H}$ and v are much smaller. Note that values of the magnetic field strength for which almost total stability of the atom is replaced by ionization in atomic times are en-

countered in outer space and can be generated in laboratory conditions (magnetic cumulation and excitons in semiconductors).

(2) We have assumed all along that the current density j and the electric field $\mathscr{E}$ are related locally, $j=\sigma\mathscr{E}$, with a static value of the conductivity σ. This is also true in the case of a variable field if[34]

$$\omega \ll \left(\sigma, \frac{1}{\tau}\right), \quad \delta \gg l, \tag{45}$$

where l and τ and the mean free path and time of conduction electrons. For a pulsed field (42), the Fourier spectrum is cut off at frequencies $\omega \sim 1/\Delta t \sim K^{1/2} v_0/R_0 \sim 10^7 - -10^8\ \mathrm{s}^{-1}$. Since for metals we have $\sigma \sim 10^{17}\ \mathrm{s}^{-1}$, $1/\tau = ne^2/m\sigma$, and $l = v_F\tau$ (here $v_F \sim 10^8\ \mathrm{cm\,s}^{-1}$ is the velocity corresponding to the Fermi boundary), we obtain the estimates $1/\tau \sim 10^{13}-10^{14}\ \mathrm{s}^{-1}$, $l \sim 10^{-6}-10^{-5}$ cm, and $\delta \sim c\sqrt{\Delta t/4\pi\sigma} \sim 0.05$ cm. We see that the conditions (45) are met by a large margin.

(3) In Eqs. (24) and (31) it is assumed that $\dot{\xi}(\tau)$ is negative, which means these equations can be used only for the compression stage. Calculations of the losses at the expansion stage can probably not be done in the quasistationary approximation.

(4) The energy flux J in (19) consists of two parts: $J = Q + J_H$, where Q is the Joule heat released (per unit time) in the process and J_H is the rate of variation of the energy of the magnetic field contained in the conducting shell:

$$Q = 2\pi R \int_0^\infty \sigma \mathscr{E}^2(x)\, dx = \frac{1}{4}\lambda \mathscr{H}^2 R\delta,$$

$$J_H = 2\pi R \frac{d}{dt}\left(\frac{1}{16\pi}\mathscr{H}^2\delta\right). \tag{46}$$

Interestingly, in the quasistationary approximation, $\delta \approx$ const, $Q = J_H = J/2$ irrespective of the value of σ and the time constant λ. With allowance for the curvature of the shell,

$$\frac{Q}{J} = \frac{z}{2}\left\{\frac{K_2(z)}{K_1(z)} - \frac{K_1(z)}{K_0(z)}\right\} = \frac{1}{2} + \frac{\delta}{4R} - \frac{11}{64}\left(\frac{\delta}{R}\right)^2 + \cdots. \tag{47}$$

Thus, the relation $J = 2Q$ is a specific feature of planar geometry.

Since the electric field rapidly decays as it gets deeper into the shell, the total magnetic flux Φ_{tot} is conserved. However, the approximation $\Phi_{\mathrm{tot}} = (\pi R^2 + 2\pi R\delta)\mathscr{H}$ breaks down when $\delta \sim R$, and to calculate the variation of the flux $\Phi(t) = \pi R^2 \mathscr{H}$ during the entire cumulation process requires the more precise expression (29), in which it is not assumed that $\delta \ll R$.

The authors are grateful to S. I. Blinnikov, A. S. Chernov, V. S. Ishmennik, and P. V. Sasorov for interesting discussions, to V. P. Kraĭnov for discussions concerning the results of the work, and to V. L. Morgunov, S. G. Pozdnyakov, A. V. Sergeev for the help provided in doing the numerical calculations. This work was made possible by a grant from the Russian Fund for Fundamental Research (Grant No. 98-02-17007).

APPENDIX

We discuss the properties of the solutions of Eqs. (24) and (31) and the procedure by which certain integrals can be evaluated.

(a) From Eq. (24) it immediately follows that

$$\xi(\tau) = \begin{cases} 1 - \tau + \dfrac{c_2}{2K}\tau^2 + \cdots, & \tau \to 0, \\ \xi_m + a(\tau_m - \tau)^{5/3} + b(\tau_m - \tau)^2 + \cdots, & \tau \to \tau_m, \end{cases} \tag{A1}$$

where τ_m is the stopping point, at which $\dot{\xi}(\tau_m) = 0$. Here

$$c_2 = 1 + 2\alpha, \quad a = a_1 \alpha^{2/3} K^{-2/3} \xi_m^{-7/3}, \quad b = b_1 K^{-1} \xi_m^{-3}, \tag{A2}$$

where $a_1 = 3^{5/3} \times 5^{-1} = 1.248$, and $b_1 = 3/8$ if α is positive and $b_1 = 1/2$ if α vanishes. Thus, the solution $\xi(\tau)$ has a weak singularity at point $\tau = \tau_m$, with $a/b \sim \alpha^{2/3} \ll 1$. The discontinuity of the coefficient b_1 at $a = 0$ is due to the nonlinearity of Eq. (24) and suggests that the dependence of $\xi(\tau)$ on the parameter α is significant even when $\alpha \ll 1$.

Interestingly, Eqs. (24) and (31) ''limit'' themselves at the stopping point: the solution $\xi(\tau)$ is not continued to the region where $\tau > \tau_m$ holds, since this would yield complex values of the radius $R(t)$. Note that the skin depth satisfies $\delta \propto (-\dot{\xi})^{-1/2} \to \infty$ as $\tau \to \tau_m$, and the condition $\delta \ll d$ is not met. Here we come to a region where the energy dissipation[34] varies as $Q \sim \lambda^2 \to 0$ (and not $Q \sim \lambda^{1/2}$, as earlier), which, however, has an effect on the behavior of $\xi(\tau)$ only for values of τ that are very close to τ_m. Indeed, the conditions $\delta \ll R$ and $\delta \lesssim d$ imply that, respectively,

$$-\dot{\xi}\xi \gg \frac{1}{4}\alpha^2 \quad \text{and} \quad -\frac{\dot{\xi}}{\xi^3} \gtrsim \frac{1}{4}\alpha^2\left(\frac{R_0}{d_0}\right)^2, \tag{A3}$$

where R and d are the radius of the conducting shell and the shell thickness, and α is the loss coefficient (25). In deriving these inequalities we used the relationships

$$\delta = \frac{c}{\sqrt{4\pi\sigma\lambda}} = \frac{1}{2}\alpha R_0 \sqrt{-\frac{\xi}{\dot{\xi}}}, \quad \frac{\delta}{R} = \frac{\alpha}{2}\frac{1}{-\dot{\xi}\xi}, \quad d = \frac{d_0}{\xi} \tag{A4}$$

($Rd = R_0 d_0$, i.e., the shell material is assumed incompressible). Plugging into (A3) in the zeroth approximation the expression $\xi^2(\tau) = K^{-1} + (\tau_m - \tau)^2$, we find that these conditions area violated (for $K \gg 1$) only near the stopping point. The condition for the applicability of Eq. (24) has the form

$$\tau_m - \tau \gg \max\left\{\frac{1}{4}\alpha^2, \frac{\alpha^2}{4K^2}\left(\frac{R_0}{d_0}\right)^2\right\}. \tag{A5}$$

If we now take $\sigma = 5 \times 10^{17}\ \mathrm{s}^{-1}$, which is the conductivity of copper, and assume $d_0/R_0 \sim 0.1$, $K \sim 10^2 - 10^3$, and $\alpha = 0.05 - 0.1$, we find $\tau_m - \tau \gg 10^{-4} - 10^{-3}$. Since $\tau_m \approx 1$ [for $K \gg 1$; see Eq. (17)], the equation of motion (24) can be used in almost all stages of shell compression.

If we replace (24) with (31), the coefficient c_2 in Eq. (A1) changes very little, $c_2 = (1+\alpha)^2$. We illustrate the parameter variations at the stopping point by a typical example:

at $K=10^3$ and $\alpha=0.1$ Eq. (24) yields $\xi_m=0.0445$ and $\tau_m=0.986$, while Eq. (31) yields $\xi_m=0.0469$ and $\tau_m=0.982$.

(b) Consider the integral

$$J_{\mu\nu}=\int_0^{\tau_m}\xi^{-\mu}(-\dot{\xi})^{\nu}\,d\tau, \quad \mu>1, \quad \nu>-1. \tag{A6}$$

Since $\alpha\ll 1$ holds, in (A6) we substitute $\xi(\tau)\approx\xi^{(0)}(\tau)=\sqrt{\xi_m^2+(\tau-\tau_m)^2}$, where $\xi_m\approx K^{-1/2}$ and $\tau_m\approx 1$. In this approximation we have $\dot{\xi}\xi=-\sqrt{\xi^2-\xi_m^2}$. Replacing integration with respect to τ in (A6) by integration with respect to ξ, we find that (for $K\gg 1$)

$$J_{\mu\nu}=c_{\mu\nu}K^{(\mu-1)/2}, \quad c_{\mu\nu}=\frac{\Gamma\left(\frac{\mu-1}{2}\right)\Gamma\left(\frac{\nu+1}{2}\right)}{2\Gamma\left(\frac{\mu+\nu}{2}\right)} \tag{A7}$$

(however, for $K=10^3$ and $\alpha\gtrsim 0.1$ this asymptotic expression is not very exact). The value of $J_{\mu\nu}$ decreases with increasing loss coefficient α, which is due to the corresponding increase of ξ_m. For instance, at $K=1000$ numerical calculations yield $J_{3/2,1/2}=8.87$, 7.96, 7.32, and 6.45 at $\alpha=0.05$, 0.1, 0.15, and 0.25, respectively (for the case of cylindrical geometry).

*)E-mail: karnak@theor.mephi.msk.su

1) Some of the results discussed in this paper have been announced in Ref. 3.

2) The values of C_κ are known from numerical calculations by the Hartree–Fock method and from experiments. An estimate of the values of C_κ for neutral atoms can be made if one uses a simple analytic formula[7] that follows from an effective range expansion. Note that we have $Z=1$ when a (neutral) atom is ionized, and $Z=\eta=0$ when a negative ion is ionized.

3) For example, Ca^- ($E_i=0.0184$ eV and $\kappa=0.037$), Fe^- ($E_i=0.151$ eV and $\kappa=0.105$), and Sr^- ($E_i=0.11$ eV and $\kappa=0.090$).[13]

4) We are thankful to S. I. Blinnikov who pointed out this fact to us.

5) The same is true of excited states, $n\geqslant 2$.

6) This expression is valid as long as $w_t\ll 1$. In the opposite case it must be replaced by $1-\exp\{-w_t\}$. See also formula (33).

7) This approximation has a good accuracy up to $z\sim 1$. For instance, from tables of Bessel functions we find that $f_1(1)=1.4296$, $f_1(3)=1.1560$, and $f_1(10)=1.0489$, which differ from (30) by 4.9%, 0.9%, and 0.1%, respectively.

8) See, e.g., Ref. 35. The Duhamel integral (35) can be transformed into an expression more convenient for numerical calculations.

[1] V. S. Popov and A. V. Sergee, JETP Lett. **63**, 417 (1996).
[2] V. S. Popov, V. M. Karnakov, and V. D. Mur, Phys. Lett. A **229**, 306 (1997); Zh. Éksp. Teor. Fiz. **113**, 1579 (1998) [JETP **86**, 860 (1998)].
[3] B. M. Karnakov, V. D. Mur, V. S. Popov, JETP Lett. **65**, 405 (1997).
[4] A. D. Sakharov, R. Z. Lyudaev, E. N. Smirnov *et al.*, Dokl. Akad. Nauk SSSR **165**, 65 (1965). [Sov. Phys. Dokl. **10**, 1045 (1966).
[5] A. D. Sakharov, Usp. Fiz. Nauk **88**, 725 (1966). [Sov. Phys. Usp. **9**, 294 (1966)].
[6] *Collected Papers of A. D. Sakharov* [in Russian], Tsentrkom, Moscow (1995).
[7] V. D. Mur, V. S. Popov, and B. M. Karnakov, Zh. Éksp. Teor. Fiz. **115**, 521 (1999) [JETP **88**, 286 (1999)].
[8] F. Bitter, Sci. Am. July 1965, p. 64.
[9] A. I. Pavlovskiĭ, in *Collected Papers of A. D. Sakharov* [in Russian], Tsentrkom, Moscow (1995), p. 85.
[10] G. A. Askar'yan, in *He Lived Among Us: Memories of A. D. Sakharov* [in Russian], Praktika, Moscow (1996), p. 125.
[11] J. Schwinger, Phys. Rev. **82**, 664 (1951).
[12] E. Brezin and C. Itzykson, Phys. Rev. D **2**, 1191 (1970).
[13] *Handbook of Chemistry and Physics*, D. R. Lide (Ed.), CRC Press, London (1994).
[14] A. S. Davydov, *Solid-State Theory* [in Russian], Nauka, Moscow (1976).
[15] D. T. Wickramasinghe and L. Ferrario, Astrophys. J. **327**, 222 (1988).
[16] I. Seipp, K. T. Taylor, and W. Schweizer, J. Phys. B **29**, 1 (1996).
[17] I. Seipp and W. Schweizer, Astron. Astrophys. **318**, 990 (1997).
[18] S. N. Fabrika and G. G. Balyavin, Preprint 129, Special Astrophysical Observatory of the Russian Academy of Sciences, Nizhniĭ Arkhyz (1998).
[19] M. Hehenberger, H. V. McIntosh, and E. Brändas, Phys. Rev. A **10**, 1494 (1974).
[20] V. Franceschini, V. Grecchi, and H. J. Silverstone, Phys. Rev. A **32**, 1338 (1985).
[21] V. S. Popov, V. D. Mur, A. V. Sergeev, and V. M. Vaĭnberg, Phys. Lett. A **124**, 77 (1987); **149**, 418 (1990).
[22] V. S. Popov, Phys. Lett. A **173**, 63 (1993).
[23] V. D. Mur and V. S. Popov, Laser Phys. **3**, 462 (1993).
[24] F. M. Fernández, Phys. Rev. A **54**, 1206 (1996).
[25] R. J. Damburg and V. V. Kolosov, J. Phys. B **11**, 1921 (1978).
[26] I. W. Herbst and B. Simon, Phys. Rev. Lett. **41**, 67 (1978). **41**, 1759 (1978).
[27] L. Benassi, V. Grecchi, E. Harrell, and B. Simon, Phys. Rev. Lett. **42**, 704 (1979); **42**, 1430 (1979).
[28] H. J. Silverstone, B. G. Adams, J. Cizek, and P. Otto, Phys. Rev. Lett. **43**, 1498 (1979).
[29] A. V. Eletskiĭ and V. S. Popov, Dokl. Akad. Nauk SSSR **250**, 74 (1980) [Sov. Phys. Dokl. **25**, 27 (1980).
[30] S. P. Alliluev, V. L. Eletsky, and V. S. Popov, Phys. Lett. A **73**, 103 (1979); **78**, 43 (1980).
[31] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-relativistic Theory*, 3rd ed., Pergamon Press, Oxford (1977).
[32] B. R. Johnson, K. F. Scheibner, and D. Farrely, Phys. Rev. Lett. **51**, 2280 (1963).
[33] V. M. Vaĭnberg and V. A. Gani, in *Abstract of Reports Presented at the 16th Conference on Fundamental Atomic Spectroscopy* [in Russian], Spectroscopy Council of the Russian Academy of Sciences, Moscow (1998), p. 130.
[34] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, Pergamon Press, Oxford (1984).
[35] H. S. Carslaw, *Introduction to the Mathematical Theory of the Conduction of Heat in Solids*, Dover, New York (1945).
[36] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Sums, Series and Products*, Academic Press, New York (1980).
[37] N. B. Delone and V. P. Kraĭnov, Usp. Fiz. Nauk **168**, 531 (1998) [Phys. Usp. **41**, 469 (1998)].

Translated by Eugene Yankovsky

Anomalous light-induced drift of linear molecules

A. I. Parkhomenko*)

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

(Submitted 18 November 1998)
Zh. Éksp. Teor. Fiz. **115**, 1664–1679 (May 1999)

The sudden approximation in energy is used to derive analytic formulas that describe the anomalous light-induced drift (LID) of linear molecules absorbing radiation in the rovibrational transition nJ_i–mJ_f (n and m are the ground and excited vibrational states, and J_α is the rotational quantum number in the vibrational state $\alpha=m,n$). It is shown that for all linear molecules with moderate values $B\lesssim 1\ \mathrm{cm}^{-1}$ of the rotational constant, anomalous LID can always by observed under the proper experimental conditions; temperature T, rotational quantum number J_i, and type of transition (P or R). The parameter $\gamma=B[J_i(J_i+1)-J_f(J_f+1)]\nu_n/2k_BT(\nu_m-\nu_n)$ is used to derive a condition for observing anomalous LID: $\gamma\sim 1$ (k_B is the Boltzmann constant and ν_α is the transport rate of collisions of molecules in the vibrational state α and buffer particles at moderate molecular velocities $v\ll\overline{v}_b$, where $\overline{v}_b$ is the most probable velocity of the buffer particles). For $\nu_m>\nu_n$ anomalous LID can be observed only in P-transitions, while for $\nu_m<\nu_n$ it can be observed only in R-transitions. It is shown that anomalous LID is possible for all ratios $\beta=M_b/M$ of the masses of the buffer particles (M_b) and of the resonant particles (M) and any absorption-line broadening (Doppler or homogeneous). The optimum conditions for observing anomalous LID are realized when the absorption line is Doppler-broadened in an atmosphere of medium-weight ($\beta\sim 1$) and heavy ($\beta\gg 1$) buffer particles. In this case, anomalous LID can be observed in the same transition within a broad temperature interval $\Delta T\sim T$. If the buffer particles are light ($\beta\ll 1$) or if the broadening of the absorption line is homogeneous, anomalous LID in the same transition can be observed only within a narrow temperature range $\Delta T\ll T$. © *1999 American Institute of Physics.* [S1063-7761(99)01005-7]

1. INTRODUCTION

Light-induced drift (LID), predicted by Gel'mukhanov and Shalagin[1] and observed for the first time in experiments by Antsygin *et al.*,[2] is being actively investigated both theoretically and experimentally (see, e.g., the article in Ref. 3, the monograph in Ref. 4, the reviews in Refs. 5–8, and the literature cited therein). The effect consists in the formation of a macroscopic flux of absorbing particles that interact with a traveling light wave and collide with buffer gas particles. The LID effect is due to velocity-selective excitation of the particles absorbing the light (the Doppler effect) and to a change in the values of the transport characteristics of the absorbing particles. The magnitude of the LID effect is proportional to the relative difference of the transport collision rates of the collision of resonant particles in the ground and excited states and buffer particles.

Until fairly recently, all data on LID experiments could be described satisfactorily by a LID theory in which the transport collision rates were velocity independent.[4–10] This theory yielded a dispersion-curve-like (the frequency derivative of the absorption lineshape) frequency dependence of the drift velocity $u_L(\Omega)$ with one zero where the detuning Ω of the radiation frequency vanishes (see curves *5* in Fig. 1 below). In 1992, while studying LID in C_2H_4 molecules with Kr acting as a buffer gas, van der Meer *et al.*[11] detected an unexpectedly large deviation of the frequency dependence of the drift velocity $u_L(\Omega)$ from a dispersionlike curve: an anomalous spectral profile of LID velocity with three zeros instead of the one zero predicted by the then existing theory. Today we have a vast body of data, both experimental[11,13–18] and theoretical,[12,14,18–24] gathered in anomalous-LID studies. It has been found that anomalous LID is entirely due to the dependence of transport collision rates on the velocity v of the resonant particles, and the anomaly can arise only if the difference $\Delta\nu(v)$ of the transport collision rates on the combining (i.e., affected by the radiation) levels changes its sign as a function of v.

In view of such specific requirements for the behavior of $\Delta\nu(v)$ for anomalous LID to exist, the following question arises: is the anomalous LID observed in some molecules an exotic feature of the behavior of the potential representing the interaction of molecules in combining states and buffer particles, or is there a pattern in the manifestation of anomalous LID that has yet to be discovered? The theoretical solution of this problem would be of interest not only to the physics of the LID phenomenon but also to the physics in intermolecular interactions.

The present paper discusses the problem of anomalous LID for the case of linear molecules in the energy sudden approximation, which is effective when applied to molecules

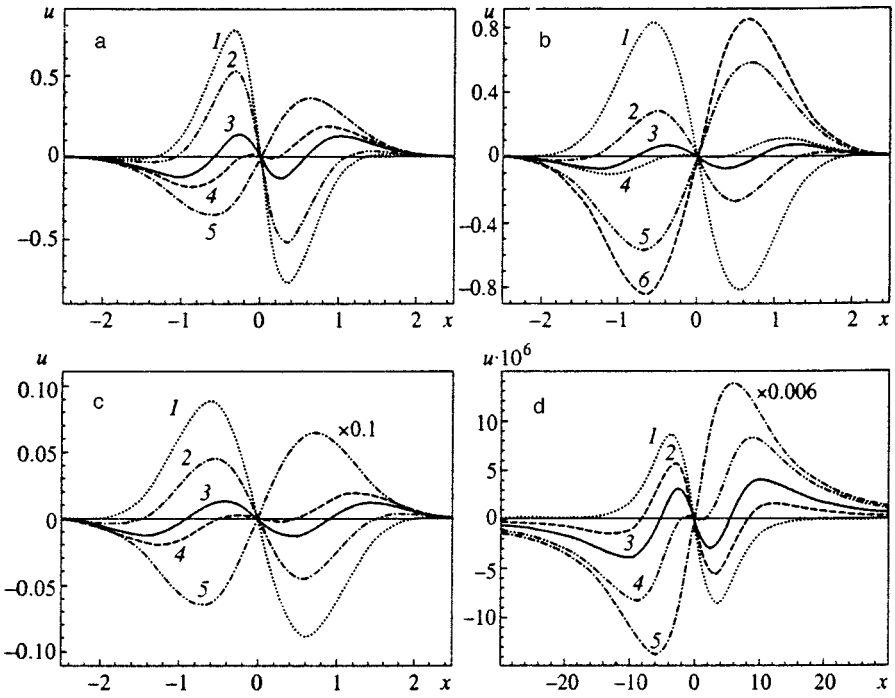

FIG. 1. Dimensionless drift velocity $u(x)$ as a function of the dimensionless frequency detuning $x=\Omega/k\bar{v}$ of the radiation, at $\xi=12$ in the cases: (a) $\beta=10$, $y=0.01$, and $\gamma=1.7$ (curve *1*), 1.3 (curve *2*), 0.65 (curve *3*), 0.4 (curve *4*), 0 (curve *5*), -0.5 (curve *6*); (b) $\beta=1$, $y=0.01$, and $\gamma=2.5$ (curve *1*), 1.5 (curve *2*), 1.1 (curve *3*), 0.95 (curve *4*), 0 (curve *5*); -0.5 (6) (c) $\beta=0.1$, $y=0.01$, and $\gamma=1.5$ (curve *1*), 1.41 (curve *2*), 1.34 (curve *3*), 1.31 (curve *4*), 0 (curve *5*); and (d) $\beta=1$, $y=10$, and $\gamma=1.29$ (curve *1*), 1.288 (curve *2*), 1.286 (curve *3*), 1.283 (curve *4*), 0 (curve *5*).

with moderate rotational constants. It was found that under properly selected experimental conditions anomalous LID can always be observed.

2. GENERAL RELATIONSHIPS

We consider the LID effect in the field of a traveling monochromatic light wave. The interaction between the light and the molecules in the buffer gas is described by the following transport equations:

$$\frac{d}{dt}\rho_m(J_m,\mathbf{v})=S_m(J_m,\mathbf{v})+NP(\mathbf{v})\delta_{J_mJ_f},$$
$$\frac{d}{dt}\rho_n(J_n,\mathbf{v})=S_n(J_n,\mathbf{v})-NP(\mathbf{v})\delta_{J_nJ_i}. \tag{1}$$

Here $\rho_\alpha(J_\alpha,\mathbf{v})$ is the population distribution over velocity $\mathbf{v}$ and the rotational levels J_α for the absorbing molecules in the vibrational state α ($\alpha=n$ is the ground vibrational state, $\alpha=m$ is a vibrational excited state, and J_α denotes the set of rotational quantum numbers characterizing the rotational state), $S_\alpha(J_\alpha,\mathbf{v})$ is the collision integral reflecting the collisions of buffer particles and molecules in the vibrational state α and the rotational state J_α, and $N=N_m+N_n$ is the concentration of the absorbing molecules, with

$$N_\alpha=\sum_{J_\alpha}\int\rho_\alpha(J_\alpha,\mathbf{v})d\mathbf{v}.$$

The probability $P(\mathbf{v})$ (per unit time) for a molecule with fixed velocity $\mathbf{v}$ to absorb light is given by the formula

$$NP(\mathbf{v})=\frac{BI}{\pi}Y(\mathbf{v})[\rho_n(J_i,\mathbf{v})-\rho_m(J_f,\mathbf{v})], \tag{2}$$

where

$$Y(\mathbf{v})=\frac{\Gamma(v)}{\Gamma^2(v)+(\Omega-\mathbf{k}\cdot\mathbf{v})^2},$$
$$B=\frac{\lambda^2A}{4\hbar\omega}\frac{2J_f+1}{2J_i+1},\quad \Omega=\omega-\omega_{mn}, \tag{3}$$

I is the light intensity; λ, ω, and $\mathbf{k}$ are the wavelength, frequency, and wave vector of the light, which we assume to be in resonance with the rovibrational transition nJ_i–mJ_f; ω_{mn} is the frequency of the nJ_i–mJ_f transition; A is the probability of spontaneous emission of light in the nJ_i–mJ_f transition; and $\Gamma(v)$ is the homogeneous halfwidth of the absorption line on the nJ_i–mJ_f transition. In Eqs. (1) we have neglected radiative relaxation, since it affects rovibrational transitions only at very low pressures.

For the diagonal collision integral we will use the velocity-isotropic model of ''arrival'':

$$S_\alpha(J_\alpha,\mathbf{v})=-\nu_\alpha(v,J_\alpha)\rho_\alpha(J_\alpha,\mathbf{v})+S_\alpha^{(2)}(J_\alpha,v), \tag{4}$$

where the arrival term $S_\alpha^{(2)}(J_\alpha,v)$ is a function of the absolute value of velocity, $v=|\mathbf{v}|$, and $\nu_\alpha(v,J_\alpha)$ has the dimensions and meaning of collision rate. The collisional model (4) allows for the dependence of the collision rate on velocity and at the same time makes it possible to obtain an analytic solution of the problem.

We will now establish the dependence of the collision rate $\nu_\alpha(v,J_\alpha)$ in Eq. (4) on the velocity and the rotational quantum number. To this end we will examine the internal friction force $\mathbf{F}_\alpha$ due to collisions of molecules in a vibrational state α and particles of the buffer gas. On the one hand, this force is given by the well-known expression in terms of the collision integral,

$$\mathbf{F}_\alpha=M\sum_{J_\alpha}\int\mathbf{v}S_\alpha(J_\alpha,\mathbf{v})d\mathbf{v}, \tag{5}$$

which with allowance for (4) becomes

$$\mathbf{F}_\alpha = -M\sum_{J_\alpha}\int \mathbf{v}\rho_\alpha(J_\alpha,\mathbf{v})\nu_\alpha(v,J_\alpha)d\mathbf{v}, \tag{6}$$

where M is the mass of a molecule. Comparing this expression with the general expression for the friction force in terms of the transport collision rate (see Refs. 4 and 25), we conclude that $\nu_\alpha(v,J_\alpha)$ in (4) has the meaning of a transport collision rate, and its dependence on the characteristics of an elementary scattering act is given by the formula (see Refs. 4 and 25)

$$\nu_\alpha(v,J_\alpha)=\frac{q}{v^3}\int_0^\infty u^2\exp\left(-\frac{u^2+v^2}{\overline{v_b^2}}\right)F(uv)\sigma_\alpha^{\mathrm{tr}}(u,J_\alpha)du, \tag{7}$$

where

$$F(uv)=\frac{2uv}{\overline{v_b^2}}\cosh\frac{2uv}{\overline{v_b^2}}-\sinh\frac{2uv}{\overline{v_b^2}}, \tag{8}$$

$$q=\frac{\mu}{M}\frac{N_b\overline{v}_b}{\sqrt{\pi}},\quad \mu=\frac{MM_b}{M+M_b},\quad \overline{v}_b=\sqrt{\frac{2k_BT}{M_b}}.$$

Here N_b and M_b are the buffer particle concentration and mass, k_B is the Boltzmann constant, T is the temperature, and u is the relative velocity of the colliding particles before collision. The quantity $\sigma_\alpha^{\mathrm{tr}}(u,J_\alpha)$ has the meaning of the transport scattering cross section of molecules on the rotational level J_α in the vibrational state α and is given by the expression

$$\sigma_\alpha^{\mathrm{tr}}(u,J_\alpha)=\sum_{J_{\alpha1}}\int\left(1-\frac{u_{1J_\alpha J_{\alpha1}}}{u}\cos\theta\right)\times\sigma_\alpha(u,\theta,\phi;J_\alpha\to J_{\alpha1})d\Omega,$$

$$\sigma_\alpha(u,\theta,\phi;J_\alpha\to J_{\alpha1})=\frac{u_{1J_\alpha J_{\alpha1}}}{u}|f_\alpha(u,\theta,\phi;J_\alpha\to J_{\alpha1})|^2,$$

$$u_{1J_\alpha J_{\alpha1}}=\sqrt{u^2-\frac{2\Delta\varepsilon_\alpha(J_{\alpha1}J_\alpha)}{\mu}},\quad d\Omega=\sin\theta\,d\theta\,d\phi. \tag{9}$$

Here $\sigma_\alpha(u,\theta,\phi;J_\alpha\to J_{\alpha1})$ is the differential scattering cross section in the channel $J_\alpha\to J_{\alpha1}$, $f_\alpha(u,\theta,\phi;J_\alpha\to J_{\alpha1})$ is the amplitude for scattering through angles θ and ϕ in the channel $J_\alpha\to J_{\alpha1}$, where θ and ϕ are the polar and azimuthal scattering angles; $\Delta\varepsilon_\alpha(J_{\alpha1}J_\alpha)=\varepsilon_\alpha(J_{\alpha1})-\varepsilon_\alpha(J_\alpha)$ is the change in rotational energy of the molecules due to inelastic transitions $J_\alpha\to J_{\alpha1}$ in the vibrational state α, $\varepsilon_\alpha(J_\alpha)$ is the rotational energy of state J_α of vibrational level α, and $u_{J_\alpha J_{\alpha1}}$ is the value of the relative velocity of the colliding particles after collision. In deriving formula (7) we assumed that the buffer particles are structureless and their velocity distribution is Maxwellian.

In the particular case of elastic collisions [here we must formally set $\Delta\varepsilon_\alpha(J_{\alpha1}J_\alpha)$ to zero in (9)], formula (7) becomes the well-known formula for the transport collision rate in elastic scattering.[4]

The LID velocity of molecules is given by the formula

$$\mathbf{u}_L\equiv\frac{\mathbf{j}_m+\mathbf{j}_n}{N},\quad \mathbf{j}_\alpha=\sum_{J_\alpha}\int\mathbf{v}\rho_\alpha(J_\alpha,\mathbf{v})d\mathbf{v}, \tag{10}$$

where $\mathbf{j}_\alpha$ is the partial flux of molecules in the vibrational state α. In steady-state and spatially homogeneous conditions, Eqs. (1), (4), and (10) yield an expression for the LID velocity:

$$\mathbf{u}_L=\int\tau_{if}(v)\mathbf{v}P(\mathbf{v})d\mathbf{v},\quad \tau_{if}(v)=\frac{\nu_n(v,J_i)-\nu_m(v,J_f)}{\nu_n(v,J_i)\nu_m(v,J_f)}. \tag{11}$$

As noted in the introduction, anomalous LID can occur only if the difference of transport collision rates, $\Delta\nu(v)\equiv\nu_m(v,J_f)-\nu_n(v,J_i)$ on the combining levels (or equivalently, the factor $\tau_{if}(v)$) changes its sign as a function of v. Here molecules with both positive and negative $\tau_{if}(v)$ contribute to the drift velocity $\mathbf{u}_L$, which may cause the frequency dependence of the LID velocity $\mathbf{u}_L(\Omega)$ to differ greatly from a dispersion-curve-like dependence, including the occurrence of additional zeros at certain values of Ω.

3. THE ENERGY SUDDEN APPROXIMATION

Let us study how the transport collision rate $\nu_\alpha(v,J_\alpha)$ of the molecules depend on their velocity v and rotational state J_α. For subsequent analysis it is convenient to transform the integral for $\nu_\alpha(v,J_\alpha)$ in (7) and represent it as a sum of two terms,

$$\nu_\alpha(v,J_\alpha)=\nu_\alpha^t(v,J_\alpha)+\nu_\alpha^c(v,J_\alpha), \tag{12}$$

where

$$\nu_\alpha^t(v,J_\alpha)=\frac{q}{v^3}\int_0^\infty u^2\exp\left(-\frac{u^2+v^2}{\overline{v_b^2}}\right)\times F(uv)\sigma_\alpha^t(u,J_\alpha)du, \tag{13}$$

$$\nu_\alpha^c(v,J_\alpha)=\frac{q}{v^3}\sum_{J_{\alpha1}}\int_0^\infty u^2\exp\left(-\frac{u^2+v^2}{\overline{v_b^2}}\right)F(uv)\times\left[1-\frac{u_{1J_\alpha J_{\alpha1}}}{u}\right]\sigma_\alpha^c(u;J_\alpha\to J_{\alpha1})du. \tag{14}$$

Here we have introduced the notation

$$\sigma_\alpha^t(u,J_\alpha)=\int(1-\cos\theta)\left(\sum_{J_{\alpha1}}\sigma_\alpha(u,\theta,\phi;J_\alpha\to J_{\alpha1})\right)d\Omega, \tag{15}$$

$$\sigma_\alpha^c(u;J_\alpha\to J_{\alpha1})=\int\cos\theta\,\sigma_\alpha(u,\theta,\phi;J_\alpha\to J_{\alpha1})d\Omega.$$

For subsequent calculations we use the well-known formula that relates the differential cross sections of RT-transitions in linear molecules when they collide with atoms,[26–29] which is derived in the energy sudden approximation commonly used in the theory of inelastic molecular collisions:

$$\sigma_\alpha(u,\theta,\phi;J_\alpha\to J_{\alpha 1})=\left[1+\frac{2\varepsilon_\alpha(J_\alpha)}{\mu u^2}\right](2J_{\alpha 1}+1)\times\sum_{L=|J_\alpha-J_{\alpha 1}|}^{L=J_\alpha+J_{\alpha 1}}\begin{pmatrix}J_\alpha & J_{\alpha 1} & L\\ 0 & 0 & 0\end{pmatrix}^2\sigma_\alpha\times\left(u+\frac{2\varepsilon_\alpha(J_\alpha)}{\mu u},\theta,\phi;0\to L\right). \quad (16)$$

where $\begin{pmatrix}a & b & c\\ 0 & 0 & 0\end{pmatrix}$ is the Wigner $3j$-symbol.[30] This formula is valid if the atom–molecule collision time $\tau_{\rm col}$ is less than the molecular rotation period $\tau_{\rm rot}$ (see Ref. 31):

$$\frac{\tau_{\rm col}}{\tau_{\rm rot}}\ll 1. \quad (17)$$

For diatomic molecules this condition (17) becomes[32,33]

$$\sqrt{\frac{\mu\Delta\varepsilon}{M_r k_B T}}\ll 1, \quad (18)$$

where μ is the reduced mass of the colliding particles, M_r is the reduced mass of the atoms comprising the diatomic molecule, and $\Delta\varepsilon=|\Delta\varepsilon_\alpha(J_{\alpha 1}J_\alpha)|$.

Khare[28] derived formula (16) in the energy sudden approximation, i.e., by replacing the rotational energy operator of the molecule with a constant. This energy sudden approximation neglects the dependence of the cross section on the energy in the outgoing channel. The factor $1+2\varepsilon_\alpha(J_\alpha)/\mu u^2$ in (16) appears by virtue of detailed balance.[28,32] The sudden approximation is effective when the change in the rotational energy is small compared to the total energy.

Taking into account the orthogonality relation[30]

$$\sum_{J_{\alpha 1}}(2J_{\alpha 1}+1)\begin{pmatrix}J_\alpha & J_{\alpha 1} & L\\ 0 & 0 & 0\end{pmatrix}^2=1, \quad (19)$$

and combining (15) with (16), we obtain

$$\sigma^t_\alpha(u,J_\alpha)=\left[1+\frac{2\varepsilon_\alpha(J_\alpha)}{\mu u^2}\right]\sigma^t_{0\alpha}\left(u+\frac{2\varepsilon_\alpha(J_\alpha)}{\mu u}\right), \quad (20)$$

where

$$\sigma^t_{0\alpha}\left(u+\frac{2\varepsilon_\alpha(J_\alpha)}{\mu u}\right)=\int(1-\cos\theta)\left[\sum_L\sigma_\alpha\times\left(u+\frac{2\varepsilon_\alpha(J_\alpha)}{\mu u},\theta,\phi;0\to L\right)\right]d\Omega. \quad (21)$$

Since in the sudden approximation it is assumed that there is only a small change in rotational energy, Eq. (20) is valid if $\varepsilon_\alpha(J_\alpha)\ll\mu u^2$. To first order in the small parameter

$$\frac{\varepsilon_\alpha(J_\alpha)}{k_B T}\ll 1, \quad (22)$$

Eqs. (13) and (20) yield

$$\nu^t_\alpha(v,J_\alpha)=\nu^{\rm vib}_{0\alpha}(v)+\frac{\varepsilon_\alpha(J_\alpha)}{k_B T}\nu^{\rm vib}_{1\alpha}(v), \quad (23)$$

where

$$\nu^{\rm vib}_{0\alpha}(v)=\frac{q}{v^3}\int_0^\infty u^2\exp\left(-\frac{u^2+v^2}{\overline{v_b^2}}\right)F(uv)\sigma^t_{0\alpha}(u)du, \quad (24)$$

$$\nu^{\rm vib}_{1\alpha}(v)=\frac{q\overline{u^2}}{v^3}\int_0^\infty\exp\left(-\frac{u^2+v^2}{\overline{v_b^2}}\right)F(uv)\left[\sigma^t_{0\alpha}(u)+u\frac{d\sigma^t_{0\alpha}(u)}{du}\right]du,\quad \overline{u}=\sqrt{\frac{2k_BT}{\mu}}, \quad (25)$$

$$\sigma^t_{0\alpha}(u)=\int(1-\cos\theta)\left(\sum_L\sigma_\alpha(u,\theta,\phi;0\to L)\right)d\Omega. \quad (26)$$

The quantities $\nu^{\rm vib}_{0\alpha}(v)$ and $\nu^{\rm vib}_{1\alpha}(v)$ in (23) have the dimensions of collision rate and depend only on the velocity v and the vibrational state α. The entire dependence of $\nu^t_\alpha(v,J_\alpha)$ on the rotational state J_α is in the factor $\varepsilon_\alpha(J_\alpha)/k_BT$. The quantity $\sigma^t_{0\alpha}(u)$ in (26) is the total (i.e., the elastic and inelastic) transport scattering cross section of a molecule on vibrational level α and in rotational state $J_\alpha=0$.

We now simplify Eq. (14) for $\nu^c_\alpha(v,J_\alpha)$. Since the main contribution to the integral in (14) is provided by velocities $u\sim\overline{u}$, in view of condition (22) the term $u_{1J_\alpha J_{\alpha 1}}$ [defined in (9)] in the integrand can be expanded in powers of the small quantity $\Delta\varepsilon_\alpha(J_{\alpha 1}J_\alpha)/\mu u^2$ (velocities $u\to 0$ contributes essentially nothing to the integral). To first order in the small parameter

$$\frac{|\Delta\varepsilon_\alpha(J_{\alpha 1}J_\alpha)|}{k_BT}\ll 1, \quad (27)$$

Eq. (14) yields

$$\nu^c_\alpha(vJ_\alpha)=\frac{q}{\mu v^2}\sum_{J_{\alpha 1}}\Delta\varepsilon_\alpha(J_{\alpha 1}J_\alpha)\int_0^\infty\exp\left(-\frac{u^2+v^2}{\overline{v_b^2}}\right)\times F(uv)\sigma^c_\alpha(u;J_\alpha\to J_{\alpha 1})du. \quad (28)$$

If in the linear approximation we allow for the correction term $\varepsilon_\alpha(J_\alpha)$ in (16), substituting (16) in (28) gives excessive numerical precision. Hence in (28) we must substitute (16) with $\varepsilon_\alpha(J_\alpha)=0$. We also allow for the fact that for linear molecules the energy of the rotational level J_α is

$$\varepsilon_\alpha(J_\alpha)=B_\alpha J_\alpha(J_\alpha+1), \quad (29)$$

where B_α is the rotational constant for vibrational level α. Next, employing the fact that[34,35]

$$\sum_{J_{\alpha 1}}J_{\alpha 1}(J_{\alpha 1}+1)(2J_{\alpha 1}+1)\begin{pmatrix}J_\alpha & J_{\alpha 1} & L\\ 0 & 0 & 0\end{pmatrix}^2=J_\alpha(J_\alpha+1)+L(L+1) \quad (30)$$

and, taking (19) into account, from (28) we obtain

$$\nu_\alpha^c(v,J_\alpha)=\nu_\alpha^c(v)\equiv\frac{q}{\mu v^3}\sum_L \varepsilon_\alpha(L)\int_0^\infty \exp\left(-\frac{u^2+v^2}{v_b^2}\right)\times F(uv)\sigma_\alpha^c(u;0\to L)du. \quad (31)$$

It is clear from (31) that in the linear approximation, $\nu_\alpha^c(v,J_\alpha)$ is independent of the initial rotational state of the molecule, J_α.

Thus, if the conditions (18), (22), and (27) are met for the molecules, then to the first approximation the transport collision rate $\nu_\alpha(v,J_\alpha)$ in (11) is

$$\nu_\alpha(v,J_\alpha)=\nu_\alpha^{\rm vib}(v)+\frac{\varepsilon_\alpha(J_\alpha)}{k_BT}\nu_{1\alpha}^{\rm vib}(v),$$
$$\nu_\alpha^{\rm vib}(v)\equiv v_{0\alpha}^{\rm vib}(v)+\nu_\alpha^c(v), \quad (32)$$

where $\nu_\alpha^{\rm vib}(v)$ and $\nu_{1\alpha}^{\rm vib}(v)$ depend only on the velocity v and vibrational state α. The dependence of the transport rate $\nu_\alpha(v,J_\alpha)$ on the initial rotational state J_α is due only to the factor $\varepsilon_\alpha(J_\alpha)/k_BT$, whose value can be assumed known.

4. THE LIGHT-INDUCED DRIFT VELOCITY

If we allow for (32), the factor $\tau_{if}(v)$ in the expression (11) for the LID velocity can be written

$$\tau_{if}(v)=\frac{\nu_n^{\rm vib}(v)-\nu_m^{\rm vib}(v)}{\nu_n^{\rm vib}(v)\nu_m^{\rm vib}(v)}+\frac{\varepsilon_n(J_i)\nu_{1n}^{\rm vib}(v)-\varepsilon_m(J_f)\nu_{1m}^{\rm vib}(v)}{k_BT\nu_n^{\rm vib}(v)\nu_m^{\rm vib}(v)}. \quad (33)$$

Here, in view of condition (22), we have neglected the terms $\varepsilon_\alpha(J_\alpha)\nu_{1\alpha}^{\rm vib}(v)/k_BT$ in the denominator, which has no effect on the possible change of sign of $\tau_{if}(v)$ and, therefore, does not influence the description of anomalous LID by formula (11). In (33) we can also neglect the difference between the values of the rotational energy $\varepsilon_m(J)$ and $\varepsilon_n(J)$ with the same value of J in different vibrational states, since it is at most a few percent,[36] and allowing for it results in excessive accuracy. Since experiments have shown that the relative difference of the transport collision rates of molecules in the ground and excited vibrational states, $|\nu_m^{\rm vib}(v)-\nu_n^{\rm vib}(v)|/\nu_n^{\rm vib}(v)$, is usually $\lesssim 1\%$ (see Refs. 8,10 and 13–16), in (33) we also neglect the difference between $\nu_{1m}^{\rm vib}(v)$ and $\nu_{1n}^{\rm vib}$ because it is expected to be of the same order. In the denominator of (33) we neglect the difference between $\nu_m^{\rm vib}(v)$ and $\nu_n^{\rm vib}(v)$ as well, since, in view of what we have just said, allowing for it results in excessive accuracy. As a result, from (33) we obtain

$$\tau_{if}(v)=\tau_{\rm vib}(v)+\tau_{\rm rot}(v), \quad (34)$$

where

$$\tau_{\rm vib}(v)=\frac{\nu_n^{\rm vib}(v)-\nu_m^{\rm vib}(v)}{[\nu_n^{\rm vib}(v)]^2},$$
$$\tau_{\rm rot}(v)=\frac{\varepsilon(J_i)-\varepsilon(J_f)}{k_BT}\frac{\nu_{1n}^{\rm vib}(v)}{[\nu_n^{\rm vib}(v)]^2}. \quad (35)$$

Here $\varepsilon(J)=BJ(J+1)$, with B the rotational constant. Thus, the factor $\tau_{if}(v)$ is a sum of two independent terms, the vibrational term $\tau_{\rm vib}(v)$, which depends only on the vibrational numbers m and n, and the rotational term $\tau_{\rm rot}(v)$, which depends only on the rotational numbers J_i and J_f.

The behavior of the vibrational term $\tau_{\rm vib}(v)$ is due to the nature of the colliding particles and cannot be controlled by the experimenter. As for the rotational term $\tau_{\rm rot}(v)$, thanks to the factor $[\varepsilon(J_i)-\varepsilon(J_f)]/k_BT$, the sign and value of this term can easily be controlled by the experimenter by selecting the proper temperature T, the rotational quantum number J_i, and type of transition (P or R). In P-transitions $\tau_{\rm rot}(v)$ is positive and in R-transitions, negative. For instance, in $P(J_i)$- and $R(J_i-1)$-transitions (i.e., in $J_i\to J_f=J_i-1$ and $J_i-1\to J_f=J_i$ transitions, respectively) $\tau_{\rm rot}(v)$ has the same absolute value but opposite signs:

$$[\tau_{\rm rot}(v)]_{P(J_i)}=-[\tau_{\rm rot}(v)]_{R(J_i-1)}=\frac{2BJ_i}{k_BT}\frac{\nu_{1n}^{\rm vib}(v)}{[\nu_n^{\rm vib}(v)]^2}. \quad (36)$$

The value of $\tau_{\rm rot}(v)$ is proportional to the ratio J_i/T and can easily be controlled by the experimenter.

Equations (11) and (34) yield an important conclusion: if $\nu_n^{\rm vib}(v)\neq\nu_m^{\rm vib}(v)$, the experimenter can always observe anomalous LID by selecting the proper experimental conditions (temperature T, rotational number J_i, and type of transition). Indeed, as noted earlier, anomalous LID is possible when the sign of $\tau_{if}(v)$ changes. If $\nu_n^{\rm vib}(v)=\nu_m^{\rm vib}(v)$, the factor $\tau_{if}(v)$ is equal to $\tau_{\rm rot}(v)$ and its sign does not change (is independent of v), so that no anomalous LID can be observed. But if $\nu_n^{\rm vib}(v)\neq\nu_m^{\rm vib}(v)$, the possibility of changing the sign and value of the rotational term $\tau_{\rm rot}(v)$ at will ensures a controlled choice of the behavior of the factor $\tau_{if}(v)$ needed for observing anomalous LID. From general physical considerations it follows that the collision rates in different vibrational states are usually different, $\nu_n^{\rm vib}(v)\neq\nu_m^{\rm vib}(v)$. This conclusion is also confirmed by LID experiments involving molecules.[8,10,13–16] Hence by selecting the proper experimental conditions we can always observe anomalous LID of linear molecules.

Note that a relationship similar to (34) exists for the relative transport collision rates $\Delta\nu/\nu$ on combining (i.e., affected by radiation) levels:

$$\frac{\Delta\nu}{\nu}\equiv\frac{\nu_m(v,J_f)-\nu_n(v,J_i)}{\nu_n(v,J_i)}=\left(\frac{\Delta\nu}{\nu}\right)_{\rm vib}+\left(\frac{\Delta\nu}{\nu}\right)_{\rm rot}, \quad (37)$$

where

$$\left(\frac{\Delta\nu}{\nu}\right)_{\rm vib}=\tau_{\rm vib}(v)\nu_n^{\rm vib}(v),\quad \left(\frac{\Delta\nu}{\nu}\right)_{\rm rot}=\tau_{\rm rot}(v)\nu_n^{\rm vib}(v). \quad (38)$$

The representation (37) of the quantity factor $\Delta\nu/\nu$ (measured in LID experiments) as a sum of independent vibrational and rotational terms was suggested earlier on qualitative grounds by Chapovsky *et al.*[13] and has been used to process the data of LID experiments.[13,15,16,37]

We also note that, according to (36), the sum of the LID velocities for $P(J_i)$- and $R(J_i-1)$-transitions does not contain a rotational component,

$$(\mathbf{u}_L)_{P(J_i)}+(\mathbf{u}_L)_{R(J_i-1)}=2\int \tau_{\rm vib}(v)\mathbf{v}P(\mathbf{v})d\mathbf{v}, \tag{39}$$

while the difference of LID velocities does not contain a vibrational component,

$$(\mathbf{u}_L)_{P(J_i)}-(\mathbf{u}_L)_{R(J_i-1)}=2\int [\tau_{\rm rot}(v)]_{P(J_i)}\mathbf{v}P(\mathbf{v})d\mathbf{v}. \tag{40}$$

To do specific calculations, we will limit ourselves to the case of low-intensity radiation:

$$I\ll \frac{8\pi^2\hbar c(\Gamma+k\overline{v})}{w_{ni}\lambda^3}, \tag{41}$$

where Γ is the characteristic (average) value of $\Gamma(v)$, $\overline{v}=\sqrt{2k_BT/M}$ is the most probable velocity of the absorbed particles, and w_{ni} is the relative population of the level nJ_i in the absence of radiation. If condition (41) is met, in (2) we can neglect the excited-level population ($\rho_m(J_f,\mathbf{v})=0$), and the velocity distribution of the populations in the ground state can be assumed to be close to Maxwellian ($\rho_n(J_i,\mathbf{v})=w_{ni}NW(\mathbf{v})$, where $W(\mathbf{v})$ is the Maxwell distribution). Here from (2) we find that

$$P(\mathbf{v})=\frac{BIw_{ni}}{\pi}Y(\mathbf{v})W(\mathbf{v}). \tag{42}$$

Substituting (42) in (11) and integrating over the directions of $\mathbf{v}$, we arrive at

$$\mathbf{u}_L\equiv\frac{\mathbf{k}}{k}u_L,\quad u_L=\frac{2BIw_{ni}}{k\pi^{3/2}}\int_0^\infty t\tau_{if}(t)f(t)\exp\{-t^2\}dt, \tag{43}$$

where

$$f(t)=x\psi(t)+\frac{y}{2}\ln\frac{y^2+(t-x)^2}{y^2+(t+x)^2},$$

$$\psi(t)=\arctan\frac{t+x}{y}+\arctan\frac{t-x}{y}, \tag{44}$$

$$y=\frac{\Gamma(v)}{k\overline{v}},\quad x=\frac{\Omega}{k\overline{v}},\quad t=\frac{v}{\overline{v}},$$

and the factor $\tau_{if}(t)\equiv\tau_{if}(t\overline{v})\equiv\tau_{if}(v)$ has been defined in (34) and (35).

5. ANALYSIS OF RESULTS

To calculate the LID velocity by formula (43), we must know the dependence of $\tau_{if}(t)$ on the dimensionless velocity t. The collision rates in the expressions (34) and (35) for $\tau_{if}(t)$ can be calculated by assuming that the total transport scattering cross section of a molecule, $\sigma^t_{0\alpha}$, depends on the relative velocity u in (26) according to a power law:

$$\sigma^t_{0\alpha}(u)=\sigma_\alpha(\overline{v}_b)\left(\frac{u}{\overline{v}_b}\right)^{-4/\xi_\alpha},\quad \alpha=m,n. \tag{45}$$

When the scattering is elastic, this dependence of the cross section on the relative velocity corresponds to the power-law interaction potential $U\propto r^{-\xi_\alpha}$ (Ref. 38). Since under the conditions specified by (27) inelastic rotational transitions have a small effect on the path of the colliding particles, the total transport scattering cross section (45) also approximately corresponds the potential $U\propto r^{-\xi_\alpha}$.

Inserting (45) into (24) and (25) yields

$$\nu^{\rm vib}_{0\alpha}(t)=\nu_\alpha\Phi(a_\alpha-1;\tfrac{5}{2};-\beta t^2), \tag{46}$$

$$\nu^{\rm vib}_{1\alpha}(t)=\nu_{1\alpha}\Phi(a_\alpha;\tfrac{5}{2};-\beta t^2), \tag{47}$$

where

$$\nu_\alpha=\nu^{\rm vib}_{0\alpha}(0)=\frac{4}{3\sqrt{\pi}}\frac{\mu}{M}N_b\overline{v}_b\sigma_\alpha(\overline{v}_b)\Gamma\left(3-\frac{2}{\xi_\alpha}\right),$$

$$\nu_{1\alpha}=\nu^{\rm vib}_{1\alpha}(0)=\frac{\xi_\alpha-4}{2(\xi_\alpha-1)}(1+\beta)\nu_\alpha,$$

$$a_\alpha=\frac{2}{\xi_\alpha}+\frac{1}{2},\quad \beta\equiv\frac{M_b}{M}\quad(\xi_\alpha>1), \tag{48}$$

$\Phi(a;b;x)$ is Kummer's confluent hypergeometric function, and $\Gamma(x)$ is the gamma function. The quantity $\nu^{\rm vib}_{0\alpha}(t)$ decreases monotonically with increasing t for $\xi_\alpha<4$ and increases monotonically with t for $\xi_\alpha>4$. At $\xi_\alpha=4$, $\nu^{\rm vib}_{0\alpha}(t)$ is constant and independent of velocity. The quantity $\nu^{\rm vib}_{1\alpha}(t)$ monotonically decreases for all values of ξ_α.

To simplify matters, we will examine the case of equal parameters $\xi_m=\xi_n\equiv\xi$ in the power-law potentials describing the interaction of molecules in the ground and excited vibrational states and buffer particles. This is equivalent to assuming that the collision rates (46) and (47) are similar:

$$\frac{\nu^{\rm vib}_{0m}(t)}{\nu^{\rm vib}_{0n}(t)}=\text{const},\quad \frac{\nu^{\rm vib}_{1m}(t)}{\nu^{\rm vib}_{1n}(t)}=\text{const} \tag{49}$$

($\nu_m\neq\nu_n$ and $\nu_{1m}\neq\nu_{1n}$). Moreover, in view of the condition (22), we can neglect $\nu^c_\alpha(v)$ in Eq. (32) for $\nu^{\rm vib}_\alpha$, and in calculating $\tau_{if}(t)$ in Eq. (43) for the LID velocity we can assume $\nu^{\rm vib}_\alpha(v)\approx\nu^{\rm vib}_{0\alpha}(v)$. Here it is convenient to write the expression for the LID velocity u_L, obtained by (43), in the form

$$u_L=u_0u(x), \tag{50}$$

where we have introduced the parameter u_0 with dimensions of velocity,

$$u_0=\frac{\nu_n-\nu_m}{\nu_n\nu_m}\frac{2BIw_{ni}}{k\pi^{3/2}}, \tag{51}$$

and the dimensionless velocity $u(x)$ as a function of the dimensionless frequency detuning x of the radiation,

$$u(x)=\int_0^\infty t\tau(t)f(t)\,\exp\{-t^2\}dt. \tag{52}$$

Here

$$\tau(t)=\frac{1-\gamma K(\beta,t)}{\Phi(a-1;\frac{5}{2};-\beta t^2)},\quad \gamma=\frac{\varepsilon(J_i)-\varepsilon(J_f)}{2k_BT}\frac{\nu_n}{\nu_m-\nu_n}, \tag{53}$$

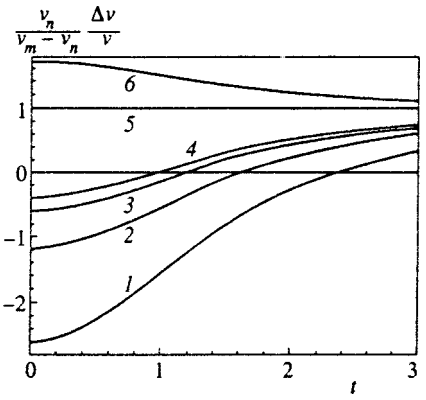

FIG. 2. Relative difference of transport collision rates, $\Delta\nu/\nu$, of molecules on combining (i.e., affected by radiation) levels as a function of the dimensionless velocity $t=v/\bar{v}$ of molecules at $\beta=1$, $\xi=12$, and different values of the parameter γ: curve *1*, $\gamma=2.5$; curve *2*, $\gamma=1.5$; curve *3*, $\gamma=1.1$; curve *4*, $\gamma=0.95$; curve *5*, $\gamma=0$; and curve *6*, $\gamma=-0.5$.

$$K(\beta,t)=\frac{(1+\beta)(\xi-4)}{\xi-1}\,\frac{\Phi(a;\frac{5}{2};-\beta t^2)}{\Phi(a-1;\frac{5}{2};-\beta t^2)},\quad a=\frac{2}{\xi}+\frac{1}{2}.$$

For $\xi\neq 4$ the function $K(\beta,t)$ is finite and decreases monotonically with increasing velocity t. In LID experiments, the sign and value of the parameter γ in (53) can be specified at will by selecting the type of transition (P or R), the value of the initial rotational number J_i, and the temperature T. For positive γ, the factor $\tau(t)$ can change its sign with increasing t, so that anomalous LID can be observed only if γ is positive. From this condition it follows that for $\nu_m>\nu_n$ anomalous LID can be observed only in P-transitions, while for $\nu_m<\nu_n$ it can be observed only in R-transitions.

Under the same conditions [$\xi_m=\xi_n$ and $\nu_\alpha^{\rm vib}(v)\approx\nu_{0\alpha}^{\rm vib}(v)$], the relative difference of the transport collision rates (37) is given by the formula

$$\frac{\Delta\nu}{\nu}=\frac{\nu_m-\nu_n}{\nu_n}[1-\gamma K(\beta,t)]. \tag{54}$$

Figures 1 and 2 illustrate the results of numerical calculations by formulas (52)–(54).

Figures 1a–1c depict the dependence of LID velocity on the frequency detuning of the radiation for different ratios β of the masses of resonant and buffer particles in the case of Doppler broadening of the absorption line of molecules ($y=0.01\ll 1$) and for equal values of the parameter ξ in the power-law interaction potentials. Curves *2*, *3*, and *4* in these figures correspond to anomalous LID, with curves *2* and *4* marking the approximate limits of plots of anomalous LID and curve *3* corresponding to maximum manifestation of anomalous LID (near detunings $x>0$ or $x<0$ the amplitudes of the minimum and maximum of the drift velocity $u(x)$ are approximately the same). Anomalous LID is possible for all ratios β of the masses of resonant and buffer particles. For heavy buffer particles ($\beta\gg 1$), the interval of values of γ at which anomalous LID is observed is $\Delta\gamma\approx 1$ (curves *2* and *4* in Fig. 1a). For medium-weight buffer particles ($\beta\sim 1$), $\Delta\gamma\approx 0.6$ (curves *2* and *4* in Fig. 1b). Finally, for light buffer particles ($\beta\ll 1$; see Fig. 1c), $\Delta\gamma\approx\beta\ll 1$. When anomalous LID is at its maximum (curves *3* in Figs. 1a–1c), the maximum (or minimum) value of the drift velocity $u(x)$ is almost completely independent of the parameter β for $\beta\gtrsim 1$ and decreases in proportion to β in the case of light buffer particles ($\beta\ll 1$).

Anomalous LID can also be observed when the broadening of the absorption line of the molecules is homogeneous ($y\gg 1$; see Fig. 1d). Figure 1d shows that at $y=10$ the interval of values of γ at which anomalous LID is observed is very narrow, $\Delta\gamma\approx 0.005$.

Numerical analysis shows (partially the results of such analysis are shown in Fig. 1) that anomalous LID can be observed only for $\gamma\sim 1$.

The foregoing implies that in Doppler broadening of an absorption line in the case of heavy or medium-weight buffer particles, anomalous LID in the same transition (at fixed values of the rotational numbers J_i and J_f) can be observed over a broad temperature interval

$$\Delta T\sim T\frac{\Delta\gamma}{\gamma}\sim T, \tag{55}$$

while for light buffer particles anomalous LID can be observed only in a narrow temperature interval

$$\Delta T\sim T\frac{\Delta\gamma}{\gamma}\sim\beta T\ll T. \tag{56}$$

When the absorption-line broadening is homogeneous, the temperature interval ΔT is also extremely narrow. For instance, at $y=10$ we have $\Delta T\sim T\Delta\gamma/\gamma\sim 0.005T$.

Figure 2 shows how the relative difference $\Delta\nu/\nu$ in the transport collision rates depends on the velocity of the molecules for the same values of the parameters ξ, β, and γ as in Fig. 1b. Figures 2 and 1b show that at $\beta=1$ anomalous LID (curves *2*, *3*, and *4* in Fig. 1b) can be observed only if the difference of the transport collision rates, $\Delta\nu(v)$, on combining levels changes its sign as a function of v near the most probable velocity $\bar{v}$ of the resonant particles.

The values of the parameter ξ in the power-law interaction potential that are much greater than unity are usually most suitable for describing realistic interaction potentials. The results of numerical calculations depicted in Figs. 1 and 2 we obtained at $\xi=12$. Numerical analysis shows that the conclusions of the present section are valid for any other value of ξ that is much larger than unity, i.e., are weakly sensitive to the details of the interaction potential.

We present an example in which our results will be used to make estimates. Suppose we have a linear molecule with a rotational constant $B=0.5\,{\rm cm}^{-1}$ and suppose that $(\nu_m-\nu_n)/\nu_n=10^{-2}$ (this parameter can be measured in LID experiments). What transition is involved in anomalous LID at room temperature? Since we have $\nu_m>\nu_n$, anomalous LID can be observed only in $P(J_i)$-transitions (i.e., in $J_i\to J_f=J_i-1$ transitions). Then formula (53) for the parameter γ yields $\varepsilon(J_i)-\varepsilon(J_f)=2BJ_i$, and hence $\gamma=100BJ_i/k_BT$. If we allow for the condition for observing anomalous LID, $\gamma\sim 1$, we get $J_i\sim k_BT/100B$. At $T=300\,{\rm K}$ and $B=0.5\,{\rm cm}^{-1}$ we obtain $J_i\sim 4$, i.e., anomalous LID is observed in the $P(4)$-transition.

6. CONCLUSION

On the basis of the well-known factorization relationship for the cross sections of *RT*-transitions for linear molecules in collisions with atoms (a relationship valid within the energy sudden approximation), we have derived formulas that describe anomalous LID of linear molecules. It should be expected that at room temperature these formulas are valid for linear molecules with moderate values of the rotational constant $B \lesssim 1\ \mathrm{cm}^{-1}$ and for moderate values of the initial rotational number J_i, since the energy sudden approximation assumes that the variations in rotational energy are small compared to the kinetic energy of relative motion of the colliding particles.

The theory that has been developed in this paper is based on the relationship (16) for factorization of cross sections and can be verified in LID experiments involving molecules with small values of the rotational constant. LID experiments are much less involved that experiments with molecular beams, and verifying the accuracy and establishing the limits of formula (16) in LID experiments would be of interest from the standpoint of the physics of intermolecular interactions.

To date, of all linear molecules only the HF molecule has been studied in LID experiments.[15,16] For this molecule, the condition (18) for the applicability of the energy sudden approximation is not met due to the large value of the rotation constant, $B \approx 21\ \mathrm{cm}^{-1}$, so that the formulas derived in the present paper cannot be used to quantitatively describe anomalous LID of HF molecules. Nevertheless, by their experiments, van Duijn *et al.*[15,16] have confirmed the prediction of the given theory that for $\nu_m > \nu_n$ anomalous LID can be observed only in *P*-transitions; according to Ref. 16, $(\nu_m - \nu_n)/\nu_n = 10^{-2}$ for the HF–Ar mixture, and anomalous LID (with $u(x)$ close in shape to curves *3* in Figs. 1a–1c) was observed by van Duijn *et al.*[15,16] in the $P(4)$-transition.

If we use the data of Ref. 16 [$(\nu_m - \nu_n)/\nu_n = 10^{-2}$ for the HF molecule in an argon atmosphere at room temperature] to calculate the parameter γ by (53) for the HF molecule in the $P(4)$-transition, we get $\gamma \approx 40$. But according to the theory, γ must be of order unity for anomalous LID to occur. Formally this means that for the HF molecule the dependence of the parameter γ on J_i is 40 times weaker than that given by formula (53). This result can be explained by the fact that collisional rotational transitions, which determine the J_i-dependence of γ, arise solely because of the anisotropy (nonsphericity) of the intermolecular interaction potential. When the condition (17) is not met, the angle through which the molecule rotates during the collision time τ_{col} is not small and the interaction potential is effectively averaged over the angle. As a result, the effective anisotropy of the interaction potential decreases, which in the case of an HF molecule ensures a J_i-dependence of γ that is weaker than the one that follows from (53).

ACKNOWLEDGMENTS

The author would like to thank A. M. Shalagin for useful remarks and discussions of the results. The present work was sponsored by the Russian Fund for Fundamental Research (Grant No. 98-02-17924) and the State Scientific and Technical Program (Laser Physics, Grant No. 7.41).

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

[1] F. Kh. Gel'mukhanov and A. M. Shalagin, JETP Lett. **29**, 711 (1979).
[2] V. D. Antsygin, S. N. Atutov, F. Kh. Gel'mukhanov, G. G. Telegin, and A. M. Shalagin, JETP Lett. **30**, 261 (1979).
[3] A. M. Shalagin, ''Light-induced drift,'' in *Physics Encyclopedia* [in Russian], Vol. 4, A. M. Prokhorov (Ed.), Bol'shaya Rosiĭskaya Entsiklopediya, Moscow (1994), p. 468.
[4] S. G. Rautian and A. M. Shalagin, *Kinetic Problems of Non-linear Spectroscopy*, North-Holland, Amsterdam (1991).
[5] G. Nienhuis, Phys. Rep. **138**, 151 (1986).
[6] H. G. C. Werij and J. P. Woerdman, Phys. Rep. **169**, 145 (1988).
[7] E. R. Eliel, Adv. At., Mol., Opt. Phys. **30**, 199 (1992).
[8] L. J. F. Hermans, Int. Rev. Phys. Chem. **11**, 289 (1992).
[9] V. P. Mironenko and A. M. Shalagin, Izv. Akad. Nauk SSSR, Ser. Fiz. **45**, 995 (1981).
[10] P. L. Chapovskiĭ, Izv. Akad. Nauk SSSR, Ser. Fiz. **53**, 1069 (1989).
[11] G. J. van der Meer, J. Smeets, S. P. Pod'yachev, and L. J. F. Hermans, Phys. Rev. A **45**, R1303 (1992).
[12] F. Kh. Gel'mukhanov and A. I. Parkhomenko, Phys. Lett. A **162**, 45 (1992).
[13] P. L. Chapovsky, G. J. van der Meer, J. Smeets, and L. J. F. Hermans, Phys. Rev. A **45**, 8011 (1992).
[14] G. J. van der Meer, J. Smeets, E. R. Eliel, P. L. Chapovsky, and L. J. F. Hermans, Phys. Rev. A **47**, 529 (1992).
[15] E. J. van Duijn, H. I. Bloemink, E. R. Eliel, and L. J. F. Hermans, Phys. Lett. A **184**, 93 (1993).
[16] E. J. van Duijn, R. Nokhai, and L. J. F. Hermans, J. Chem. Phys. **105**, 6375 (1996).
[17] F. Yahyaei-Moayyed and A. D. Streater, Phys. Rev. A **53**, 4331 (1996).
[18] B. Nagels, P. L. Chapovsky, L. J. F. Hermans, G. J. van der Meer, and A. M. Shalagin, Phys. Rev. A **53**, 4305 (1996).
[19] F. Kh. Gel'mukhanov and A. I. Parkhomenko, Zh. Éksp. Teor. Fiz. **102**, 424 (1992) [Sov. Phys. JETP **75**, 225 (1992)].
[20] F. Kh. Gel'mukhanov, G. V. Kharlamov, and S. G. Rautian, Opt. Commun. **94**, 521 (1992).
[21] I. Kuščer, L. J. F. Hermans, P. L. Chapovsky, J. J. M. Beenakker, and G. J. van der Meer, J. Phys. B **26**, 2837 (1993).
[22] F. Kh. Gel'mukhanov and A. I. Parkhomenko, J. Phys. B **28**, 33 (1995).
[23] A. I. Parkhomenko, Opt. Spektrosk. **80**, 604 (1996) [Opt. Spectrosc. **80**, 541 (1996)].
[24] F. Kh. Gel'mukhanov, A. I. Parkhomenko, T. I. Privalov, and A. M. Shalagin, J. Phys. B **30**, 1819 (1997).
[25] A. I. Parkhomenko and A. M. Shalagin, Zh. Éksp. Teor. Fiz. **113**, 1649 (1998) [JETP **86**, 897 (1998)].
[26] R. Goldflam, D. J. Kouri, and S. Green, J. Chem. Phys. **67**, 5661 (1977).
[27] G. A. Parker and R. T. Pack, J. Chem. Phys. **67**, 1585 (1977).
[28] V. Khare, J. Chem. Phys. **68**, 4631 (1978).
[29] M. Faubel, Adv. At. Mol. Phys. **30**, 199 (1992).
[30] D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskiĭ, *Quantum Theory of Angular Momentum*, World Scientific, Singapore (1987).
[31] A. S. Dickinson, Comput. Phys. Commun. **17**, 51 (1979).
[32] Z. H. Top and D. J. Kouri, Chem. Phys. **37**, 265 (1979).
[33] A. V. Bogdanov, G. V. Dubrovskiĭ, and A. I. Osipov, Khim. Fiz. **4**, 1155 (1985).
[34] M. L. Strekalov, Khim. Fiz. **7**, 1182 (1988).
[35] A. V. Storozhev and M. L. Strekalov, Khim. Fiz. **16**(2), 17 (1997).
[36] A. A. Radtsig and B. M. Smirnov, *Reference Data on Atoms, Molecules, and Ions*, Springer, Berlin (1985).
[37] H. I. Bloemink, J. M. Boon-Engering, P. L. Chapovsky, L. J. F. Hermans, and E. R. Eliel, J. Phys. B **27**, 4559 (1994).
[38] L. D. Landau and E. M. Lifshitz, *Mechanics*, 3rd ed., Butterworth–Heinemann, Boston (1982).

Translated by Eugene Yankovsky

Ultrahigh-spatial-resolution photoelectron projection microscopy using femtosecond lasers

S. K. Sekatskiĭ,*) S. V. Chekalin, A. L. Ivanov, V. O. Kompanets, Yu. A. Matveets, A. G. Stepanov, and V. S. Letokhov

Institute of Spectroscopy, Russian Academy of Sciences, 142092 Troitsk, Moscow Region, Russia
(Submitted 19 November 1998)
Zh. Éksp. Teor. Fiz. **115**, 1680–1688 (May 1999)

Ultrahigh spatial resolution of two-photon photoelectron images (as high as 3 nm, which is the best value that has been achieved to date in photoelectron microscopy with spatial resolution) is obtained when silicon nanotips are irradiated by the second harmonic of a pulsed femtosecond Ti : sapphire laser. In addition, the absolute value of the two-photon external photoeffect coefficient is measured. © *1999 American Institute of Physics.* [S1063-7761(99)01105-1]

1. INTRODUCTION

In the first experiments,[1,2] which were performed comparatively recently, an improvement in the possibilities of laser photoionization projection microscopy was demonstrated: a spatial resolution of $\simeq 30$ nm was achieved by irradiating the tips of needles (with a radius of curvature $r_c \simeq 0.5-1\ \mu$m) of single-crystal LiF containing F_2 color centers with continuous radiation from an argon laser, and single F_2 color centers were detected as light spots in the photoelectron image of the needle tip. Further progress in this area should be associated both with increasing the spatial resolution of the method and with the use of more powerful lasers for achieving the external photoeffect, since the intensity of continuous sources is clearly insufficient for obtaining effective photoemission from most samples.

An analysis of the characteristics of presently existing lasers revealed that the ideal source for observing resonant and nonresonant two-photon photoelectron images is a femtosecond laser with a high (in the megahertz or higher range) pulse repetition rate, particularly, the femtosecond Ti : sapphire laser, which has recently drawn wide notice. The femtosecond pulses generated by such a laser, even without amplification, have a mean power of 3–10 mW in the second harmonic at a wavelength of 410 nm, which corresponds to a photon energy $h\nu = 3.02$ eV. Such powers and energies of single pulses are low enough to avoid the problems encountered in strong electric fields, such as optical destruction and melting of the samples. At the same time, the parameters of our laser, viz., a pulse duration $\tau = 40$ fs, a pulse repetition rate $f = 82$ MHz, spot diameters equal to $0.01-0.1$ mm^2, and a corresponding power, provide an intensity $I = 10^6 - 3 \times 10^7$ W/cm^2, which is fully sufficient for obtaining bright photoelectron images of a broad range of samples. In addition, the use of femtosecond lasers can make it possible to perform experiments on the visualization of single molecules implanted in a matrix, as was previously discussed in Refs. 3 and 4, since such an approach makes it possible to overcome the difficulties associated with the fairly fast transfer of the excitation energy of a molecule to the surrounding medium. The potential possibility of using femtosecond lasers to achieve ultrahigh temporal resolution is also noteworthy, but a discussion of this subject would be far beyond the scope of this paper.

As for the spatial resolution of projection microscopy, it is determined by the mean kinetic energy E_0 of the transverse motion of the electron emitted and the radius of curvature r_c of the microtip (see below). The value of E_0 can be determined as half of the difference between the energy of one photon of laser radiation (or many photons in the case of a multiphoton photoeffect) and the external photoelectric work function of the sample being examined. This difference usually has a value of $\simeq 1$ eV and cannot be much lower. Thus, it can be presumed that investigating microtips with significantly smaller radii of curvature can be regarded as a systematic approach to improving the spatial resolution. Below we shall present several estimates, which show that for $r_c = 20$ nm the spatial resolution can reach values of $\simeq 3$ nm. Such values are far better than the spatial resolution that has been achieved with ''classical'' photoelectron microscopes [here we are referring to a microscope equipped with a system of electrostatic or magnetic lenses for forming magnified photoelectron images; their spatial resolution is about 40 nm (Ref. 5)] and approaches the theoretical limit of the spatial resolution of a photoelectron microscope.[6]

In this paper we report the first experimental implementation of this approach. We obtained photoelectron images of ultrasharp silicon nanotips ($r_c = 20$ nm) by irradiating them with the output of a femtosecond Ti : sapphire laser and achieved a spatial resolution of $\simeq 3$ nm. The possibility of using this microscope to perform quantitative measurements of the two-photon external photoeffect for the samples examined was also analyzed, and its quantum yield at a wavelength of 410 nm for silicon was determined.

2. LASER PHOTOELECTRON MICROSCOPE

A photoelectron projection microscope is shown in Fig. 1. The sample investigated is a sharpened needle with a radius of curvature r_c, which is tightly held in a special holder

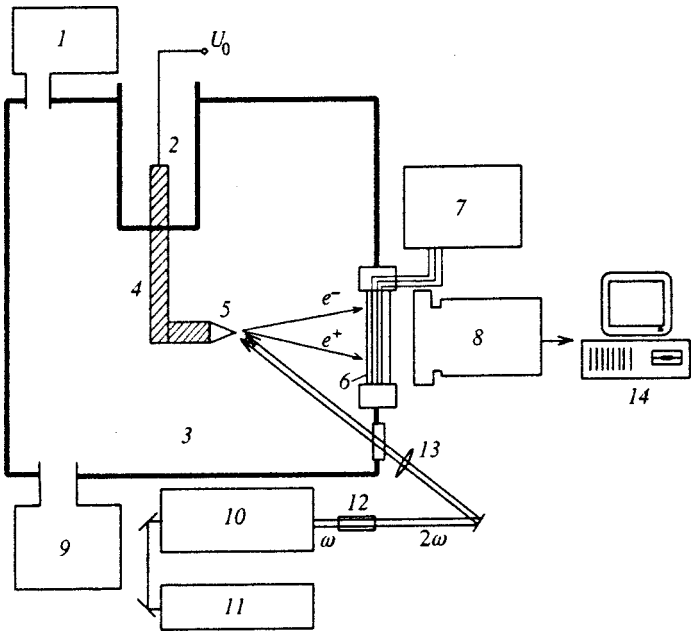

FIG. 1. Diagram of a laser photoelectron microscope equipped with a femtosecond Ti : sapphire laser: *1* — gas admission system; *2* — liquid-nitrogen cooling system; *3* — vacuum chamber; *4* — electrode; *5* — nanotip examined; *6* — microchannel plates and phosphorescent screen; *7* — power supply; *8* – chamber with a CCD matrix; *9* — pumping system; *10* — femtosecond Ti : sapphire laser; *11* — argon laser; *12* — KDP crystal; *13* — focusing lens; *14* — Argus-50 data-processing system.

at a distance $L=10$ cm from a detector consisting of a microchannel plate and a phosphorescent screen (Hamamatsu Photonics K. K., Japan). A voltage U_0 in the range 0–4 kV is applied to the sample (the entrance to the microchannel plate is grounded), and if this voltage is sufficiently high, effective field (tunneling) emission of electrons from the needle tip takes place. The radial electric field around the tip directs the electrons emitted to the detector, forming a magnified image of the needle on the screen. The magnification $K=L/\gamma r_c$, where γ is a numerical factor, which appears because of the deviation of the geometry of the system from a simple spherical capacitor and is equal to 1.5–2 (see any monograph on field electron–ion microscopy, for example, Refs. 7 and 8). When the photoelectron image of the needle tip, which does not appear upon field emission, is investigated, the potential of the needle is reduced to a level at which the tunneling emission of electrons from the tip is equal to zero, and electrons are emitted exclusively as a result of the external photoeffect in the needle material caused by the second-harmonic output of the Ti : sapphire laser (the parameters of the laser were indicated in the Introduction). As in the case of field emission, the electric field around the tip directs the photoelectrons to the detector, forming a photoelectron image rather than a field image of the tip with the same magnification.

A more detailed description of a laser photoelectron projection microscope can be found in Ref. 1.

The motion of the photoelectrons emitted in the radial electric field of a projection microscope can easily be calculated with consideration of the angular momentum conservation law:[7,8]

$$p_\theta = mr^2\theta = \text{const} = mr_c v_0,$$

where m is the mass of the electron, $v_0=\sqrt{2E_0/m}$ is the initial transverse velocity of the electron, E_0 is the initial energy of the electron, and θ is the angle between r and the symmetry axis of the system (Fig. 2). Omitting several simple steps, we can write the following expression for the diameter of the spot formed in the detector as the image of the emitting point source:

$$d=2L\theta=4\gamma r_c\sqrt{E_0/eU_0}. \quad (1)$$

The same formula can be used to estimate the spatial resolution of the microscope. In this case it must be understood that the mean, rather than the maximum, energy is be used in the calculation. For $E_0=0.75$ eV, i.e., half of the difference between the energy of two photons of the laser radiation ($2h\nu=6.04$ eV) and the work function of silicon [4.5 eV (Ref. 9)], $r=20$ nm, and $U_0=1$ kV, formula (1) gives $d\approx 3$ nm.

We note that this value is close to the theoretical limit for the resolution of a projection microscope determined by the Heisenberg uncertainty principle.[7,8] In fact, according to (1), an electron emitted from an area of diameter d_0 on the tip surface will have a transverse velocity $v_0\sim h/2md_0$ and will create an image in the form of a circle of diameter $\gamma r_c\sqrt{h/med_0^2U_0}$ in the detector. The total diameter of the emitting region will thus equal

$$d=\sqrt{(Kd_0)^2+2\gamma h^2/med_0^2U_0}. \quad (2)$$

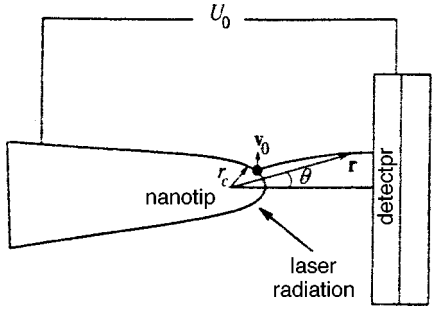

FIG. 2. Illustration of the motion of particles in the microscope (not to scale) and its spatial resolving power.

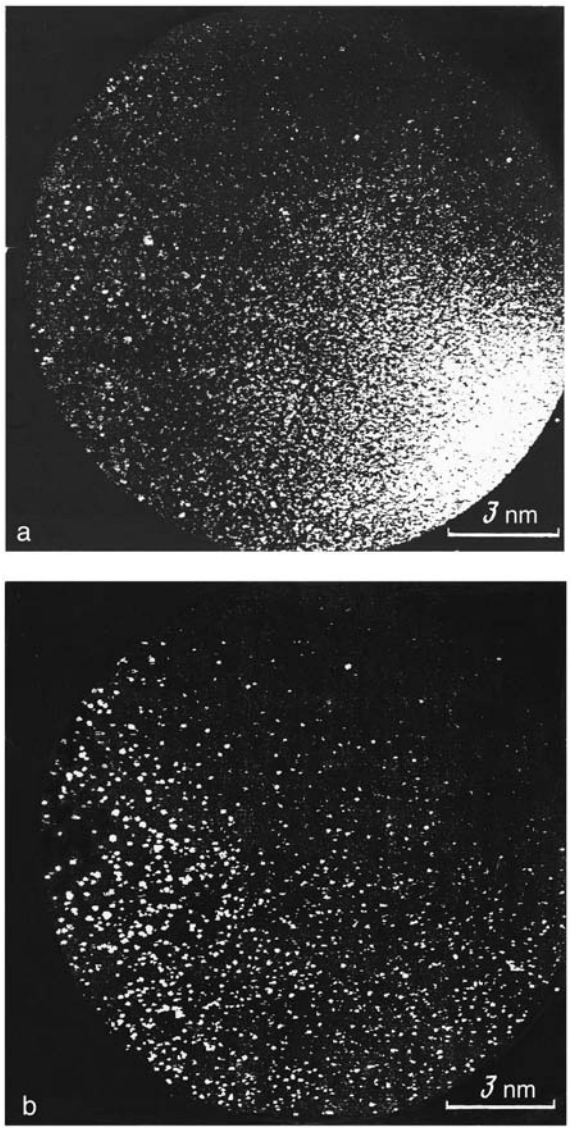

FIG. 3. Emission images of ultrasharp silicon nanotips with a radius of curvature equal to 20 nm: a — laser photoelectron image, $U_{tip}=1.0$ kV, $I \simeq 3\times 10^6$ W/cm^2, b — field-emission image, $U_{tip}=1.5$ kV.

After minimization of this diameter as a function of d_0, we obtain the following estimate of the best achievable spatial resolution:

$$d_{lim}=2(\gamma h^2/med_0^2U_0)^{1/4}. \quad (3)$$

Our experimental parameters imply $d_{lim}\approx 1$ nm, i.e., the fundamental resolution limit is, in fact, of the same order as the estimate of the spatial resolution (3) for the microscope.

3. EXPERIMENT WITH SILICON TIPS: PHOTOELECTRON IMAGES AND ANALYSIS

Highly conductive single-crystal silicon microtips were grown on a (111) surface of a silicon rod measuring $1\times 1\times 100$ mm^3 in the Institute of Crystallography of the Russian Academy of Sciences (Moscow). The growth surface was polished and etched in an HF-HNO_3 solution. Its dimensions after preparation for growth were about 0.5×0.5 mm^2. The microtips grown were first sharpened by wet etching, and

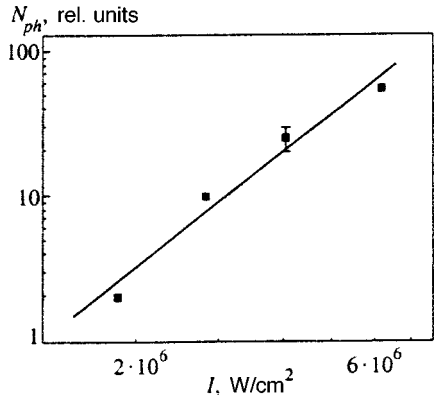

FIG. 4. Photoelectric current N_{ph} (integrated over the entire image) as a function of the intensity I of the second harmonic of the Ti : sapphire laser irradiating the silicon tip.

then they were subjected to repeated temperature-induced oxidation followed by removal of the oxide by HF. The tips with a height of 100 μm thus prepared had a radius of curvature less than 25 nm (the sharpness of the needles was verified using a high-resolution electron microscope, and in some cases the tips were so sharp that even single silicon atoms could be seen on the apex of the needle[10]). The growth techniques and the sharpening procedures are described in greater detail in Ref. 10.

The laser photoelectron image of an ultrasharp silicon tip is shown in Fig. 3a. The tip potential U_{tip} was insufficient for field emission from a nanotip, and thus the photocurrent scarcely depended on the tip potential and exhibited a quadratic dependence on the laser intensity (Fig. 4). No photocurrent was observed with irradiation by pulses of the first harmonic of the femtosecond Ti : sapphire laser (the wavelength was 820 nm, and the photon energy was 1.56 eV) up to an intensity of order 10^9 W/cm^2.

These experimental observations unequivocally indicate that the photoelectron images of the needles are produced by the two-photon photoemission of silicon subjected to femtosecond laser pulses with a photon energy of 3.02 eV, rather than, for example, by the laser-induced field emission of silicon, since in the latter case the dependence of the photocurrent on the tip potential would be much sharper than a quadratic dependence (such data were presented in, for example, Ref. 11). This conclusion is fully consistent with the known data on the work function of silicon $W=4.5$ eV (Ref. 9), as well as with the previous studies of the linear and nonlinear external photoeffect in silicon (see, for example, Refs. 12 and 13 and the references cited therein).

The image of the silicon tip itself obtained as a result of field emission at a higher tip potential without laser irradiation is shown in Fig. 3b. Neither image exhibits a well developed structure, and this is characteristic of all the silicon tips examined. We attribute this absence of a clear structure to the amorphous character of the tip surface. Moreover, in comparison to the case of a well developed surface crystal structure, the local variations in the work function are much smaller, and the field electron (field ion) images are not so clearly structured.[7,8] Such an assumption is consistent with

the absence of high-quality (high-resolution) images of these ultrasharp silicon needles in the field ion microscope (many attempts were made to observe them). Thorough thermal annealing and a procedure of field evaporation in an ultrahigh vacuum are needed to obtain silicon tips with a clearly expressed crystal structure.[8,14]

At the same time, we observed a distinct difference between the laser photoelectron image and the image obtained using field emission for one of the silicon tips examined (Fig. 3). The photoelectron image contains a bright spot, which is not observed in the field-emission image. Taking into account the magnification of the microscope in this case, we can conclude that the diameter of the object observed is less than 3 nm. We assume that this spot is caused by the presence of a strongly absorbing defect (or several such defects) near the surface of the silicon tip (the photoelectron escape depth for silicon has been estimated[12] as 1.2 nm). We are referring to impurities which can be photoionized by a one- or multiphoton mechanism under the action of radiation with a wavelength of 410 nm more efficiently than pure impurity-free silicon. The photoionization of such impurities can lead to a local (near an impurity) significant increase in the photocurrent from a silicon tip and thus to the appearance of an additional bright spot on the photoelectron images. The possibility of observing single light-absorbing centers using laser photoelectron projection microscopy has already been demonstrated for LiF:F_2 crystals, where single defects in the LiF lattice were observed on the photoelectron images of tips as bright spots having diameters specified by the spatial resolving power of the microscope.[1,2]

Utilizing the possibilities of our recording system (Argus-50), we also measured the local dependence of the photocurrent on the irradiation intensity for the bright photoinduced spot in Fig. 3a. A quadratic dependence of $N_{\rm ph}(I)$ was observed, and a similar dependence was observed for the total current (Fig. 4).

Various impurities on a tip surface can be manifested in the field-emission image in the form of bright spots, which are associated with the resonant tunneling of electrons through unoccupied impurity energy levels (see, for example, Refs. 15 and 16 and the references cited therein). This did not occur in our case, most probably because the impurity (defect) was located somewhere in deeper layers of the tip, rather than directly on the surface. In this case such a center is invisible to field emission, but is visible to the external photoelectric effect.

Many different types of impurities and defects are known for silicon (see, for example, Refs. 17 and 18 and the references cited therein), but at present we do not have sufficient information to link the bright spot observed to a specific type of defect.

4. DETERMINATION OF THE ABSOLUTE VALUE OF THE EXTERNAL TWO-PHOTON PHOTOEFFECT COEFFICIENT

In this section we would like to focus attention on the fact that the method under consideration also makes it possible to determine the absolute value of the two-photon photoeffect coefficient β_2 for the materials studied. We define β_2 as the ratio of the photoelectron flux associated with two-photon photoemission $N_{\rm ph}^{\rm pulsed}$ [photoelectrons/cm$^2\cdot$s] to the pulsed light flux P [photons/cm$^2\cdot$s]: $N_{\rm ph}^{\rm pulsed}=\beta_2 P^2$. In fact, the area S of the region from which the photoelectrons are collected is specified by the geometry of the microscope (the distance between the tip and the detector $L=10$ cm, and the working region of the microchannel plane has a diameter $a=32$ mm) and can easily be calculated:

$$S=(\xi\pi/4)(\gamma a r_c/L)^2\simeq 0.3 r_c^2.$$

This relation clearly reflects the fact that the diameter of this region is equal to the diameter a of the working region of the microchannel plate divided by K. The coefficient ξ is close to unity and takes into account that the tip surface is hemispherical, rather than flat (a simple calculation based on elementary geometry gives $\xi\simeq 1.03$ for our experimental conditions, so that it can be neglected), and the spatial variations of K can be neglected.[7,8] Although the dimensions of the emitting region are small, they are usually much greater than the typical values of the photoelectron escape depth $l_{\rm esc}$ for the materials studied [the latter usually lies between 0.1–10 nm (Ref. 6); for example, for silicon $l_{\rm esc}=1.2$ nm (Ref. 12)]. For this reason, our calculations can be compared with the results obtained by classical experimental methods based on the irradiation of flat surfaces by a focused laser beam (see, for example, Refs. 13,19, and 20 and the references cited therein).

The intensity of the laser radiation I can also easily be measured to a high accuracy, since strong focusing is not employed in the experiments described. In a photoelectron-counting mode with a microchannel plate having 100% efficiency (a coefficient smaller than unity can be introduced where necessary) the total number of photoelectrons per second can easily be measured:

$$N_{\rm ph}=\beta_2 P^2 S f\tau=\beta_2 I_0^2 S/(h\nu)^2 f\tau. \tag{4}$$

For simplicity, here we introduced the mean intensity of the laser radiation I_0 [W/cm^2], which can be measured directly by experimental means: $P=I_0/h\nu f\tau$. Thus, when $N_{\rm ph}$ is measured, we can easily calculate β_2. For our experimental conditions

$$\beta_2=N_{\rm ph}(h\nu)^2 f\tau/I_0^2 S=7.65\times 10^{-43} N_{\rm ph}/I_0^2 S. \tag{5}$$

Using formula (5) we determined the external two-photon photoeffect coefficient for silicon under the action of light with a wavelength of 410 nm (after averaging): $\beta_2=1.5\times 10^{-32}$ cm^2 s. This value corresponds to the well-known value of the analogous coefficient for a wavelength of 355 nm: $\beta_2=2.5\times 10^{-32}$ cm^2 s (Ref. 13).

Our measured value of β_2 lies in the same range as the other known values of the two-photon photoeffect coefficient for semiconductors and insulators.[13,19,20] This means that bright photoelectron images can also be recorded for these materials using a femtosecond Ti : sapphire laser and that a photoelectron microscope equipped with such a laser can become a universal tool suitable for studying virtually any metal, semiconductor, or insulator. Our approach has an additional advantage over the other classical methods for measuring the external photoeffect, since we measure not only

the averaged value of β_2, but also its local values (with a spatial resolution as high as 3 nm). Note that we have neglected the reflected light from the surface studied and the variation of the angle of the light sphere in the calculation of β_2, but such neglect is typical of all the methods used to measure the external two-photon photoeffect coefficient.

5. CONCLUSIONS

Thus, the data presented convincingly demonstrate the broad possibilities discovered in using femtosecond lasers with a megahertz pulse repetition rate in laser photoelectron projection microscopy: this makes it possible not only to obtain images with ultrahigh spatial resolution for an extensive list of samples, but also to perform exact quantitative measurements of the external photoeffect. In addition, the best spatial resolution of a photoelectron microscope to date (3 nm) has been achieved, which is close to the theoretical limit.

Such investigations will undoubtedly be useful for analyzing nanostructures in materials and for measuring the local values of the work function, defect densities, etc. We believe that two-photon femtosecond laser projection microscopy can also be used for the direct detection of biomolecules in specially prepared matrices (such as paraffins or their analogs), as was briefly mentioned in the Introduction.

This work was partially supported by the Russian Fund for Fundamental Research, the Defense Department of the U.S.A., and the U.S. Civilian Research and Development Foundation for the Independent States of the Former Soviet Union (CRDF) (Award RP2-154). We thank Hamamatsu Photonics K. K. (Japan) for supplying the necessary experimental equipment, as well as V. V. Zhirnov and E. I. Givargizov for preparing the silicon tips.

*)E-mail: sekats@lls.isan.troitsk.ru

[1] V. N. Konopsky, S. K. Sekatskii, and V. S. Letokhov, Opt. Commun. **132**, 251 (1996).
[2] S. K. Sekatskiĭ and V. S. Letokhov, JETP Lett. **65**, 491 (1997)].
[3] V. S. Letokhov, in *Laser Spectroscopy IX*, M. S. Feld, J. E. Thomas, and A. Mooradian (Eds.), Academic Press, New York (1989), p. 494.
[4] V. S. Letokhov and S. K. Sekatskii, Appl. Phys. B **55**, 177 (1992).
[5] Model PEEM-350, Staib Instrumente GmbH, Langenbach, Germany.
[6] O. H. Griffith and G. F. Rempfer, Adv. Opt. Electron. Microsc. **10**, 269 (1987).
[7] R. Gomer, *Field Emission Field Ionization*, Harvard University Press, Cambridge, MA (1961).
[8] T. T. Tsong, *Atom-Probe Field Ion Microscopy*, Cambridge University Press, Cambridge (1990).
[9] R. M. Broudy, Phys. Rev. B **1**, 3430 (1970).
[10] E. I. Givargizov, J. Vac. Sci. Technol. B **11**, 449 (1993).
[11] V. N. Konopsky, V. V. Zhirnov, N. S. Sokolov *et al.*, J. Phys. IV **6**, C5-129 (1996).
[12] C. Sebenne, D. Bolmont, G. Guichar, and M. Balkanski, Phys. Rev. B **12**, 3280 (1975).
[13] T. L. F. Leung and H. M. van Driel, Appl. Phys. Lett. **45**, 683 (1984).
[14] H. M. Liu and T. T. Tsong, J. Appl. Phys. **63**, 1532 (1987).
[15] J. W. Gadzuk, Phys. Rev. B **47**, 12 832 (1993).
[16] V. N. Konopsky, S. K. Sekatskii, and V. S. Letokhov, J. Phys. IV **6**, C5-125 (1996).
[17] *Deep Centers in Semiconductors*, S. T. Pantelides (Ed.), Gordon and Breach, New York (1986).
[18] J. Bourgoin and M. Lannoo, *Point Defects in Semiconductors*, Springer, Berlin (1983).
[19] E. F. Lazneva, *Laser Desorption* [in Russian], Izd. Leningrad. Univ., Leningrad (1990).
[20] E. M. Logothetis and P. L. Hartman, Phys. Rev. **187**, 460 (1969).

Translated by P. Shelnitz

Microwave generation using a superluminal source

Yu. N. Lazarev and P. V. Petrov

Russian Federal Nuclear Center, All-Russian Institute of Technical Physics, 456770, Snezhinsk, Chelyabinsk Region, Russia

(Submitted 30 June 1998)
Zh. Éksp. Teor. Fiz. **115**, 1689–1707 (May 1999)

The power produced by existing sources of microwave radiation falls off with decreasing wavelength. To solve this problem a new concept is proposed for generating microwave radiation, based on the use of a superluminal source formed when electrons are emitted into vacuum from a medium and the emission front propagates along the surface with a speed greater than that of light. Such generators are shown to have a number of completely unique properties: they radiate extremely short pulses (as short as picoseconds); their power exceeds that of existing sources by orders of magnitude; and unlike existing sources, it increases as the wavelength is reduced. © *1999 American Institute of Physics.* [S1063-7761(99)01205-6]

1. INTRODUCTION

The dropoff in power with decreasing wavelength (pulse duration) is an inherent aspect of existing microwave technology.[1–4] To be sure, the power output from microwave generators is ever-increasing,

but so is the level of difficulty of the problems that must be overcome in order to achieve the desired result. In the meantime, the wavelength dependence of the power remains unchanged.

To solve this problem we propose a new design concept for microwave generation, based on a type of superluminal source, namely, one created by electron emission from the interface between vacuum and a material medium when the emission front propagates along the surface faster than the speed of light, c (Refs. 5 and 6). In this case the vacuum propagation conditions combine beneficially with the high current densities emitted by the solid surface, and there is no problem of extracting the radiation.

Any superluminal source produces a coherent tightly collimated beam.[5–8] In addition to everything else, however, a superluminal source resulting from emission can also have a very short pulse length T_0, since in this case the duration of the radiation pulse is determined by the density n_e of the emitted electrons:[7,9,10]

$$T_0 \sim \frac{2\pi}{\omega_{pe}} \sim \frac{10^{-4}}{\sqrt{n_e}}, \quad \omega_{pe} = \sqrt{\frac{4\pi e^2 n_e}{m_e}}. \tag{1}$$

Here m_e is the electron mass and ω_{pe} is the electron plasma frequency

The electron density increases as a function of the radiation intensity and can be greater than or of order $10^{12}\,\mathrm{cm}^{-3}$. Consequently the microwave pulse length can be less than or of order $10^{-10}\,\mathrm{s}$.

The amplitude of the radiated electromagnetic field is proportional to the first or second time derivative of the dipole moment surface density. But as is well known, the latter depends only on the energy of the electrons driven out of the surface, $P_0 \propto \varepsilon$. It is therefore unsurprising that the intensity and total energy of the output radiation increase as functions of the energy of the emitted electrons. Calculations show that in order to obtain electromagnetic radiation with energies and intensities that are interesting, the emitted electrons must have energies of at least tens of keV. If we restrict ourselves to sources of electromagnetic radiation suitable for creating electron emission such as optical lasers, this problem can be solved by combining the process of electron production with the formation of a radiating dipole layer. The laser light is used only to drive off electrons with as small an energy as possible; the electrons making up the radiating layer attain a final energy of hundreds or thousands of keV from the acceleration they undergo in the external electric field.

It follows that an element of a device radiating in the microwave range typically has the form shown schematically in Fig. 1.

Using a superluminal source of electromagnetic radiation makes it possible to raise the radiation power output by orders of magnitude and shorten the pulse length to the picosecond range. The power output from a superluminal source rises as the wavelength decreases and can be increased by simply scaling up the device.[11–13] A full-scale device with a large radiated energy can be constructed out of small ''elementary'' sources in the same way a house is built out of bricks.

The purpose of the present work is to study the properties of such an elementary source, develop the theoretical groundwork, and outline a program for building a systematic theoretical and computational foundation for the new concept and its hardware implementation.

2. SIMPLE SUPERLUMINAL SOURCES

Generally speaking, sources are called superluminal if they travel faster than the phase velocity of light,

$$v > v_{\mathrm{ph}} = c/n,$$

where n is the index of refraction in the medium. It is a

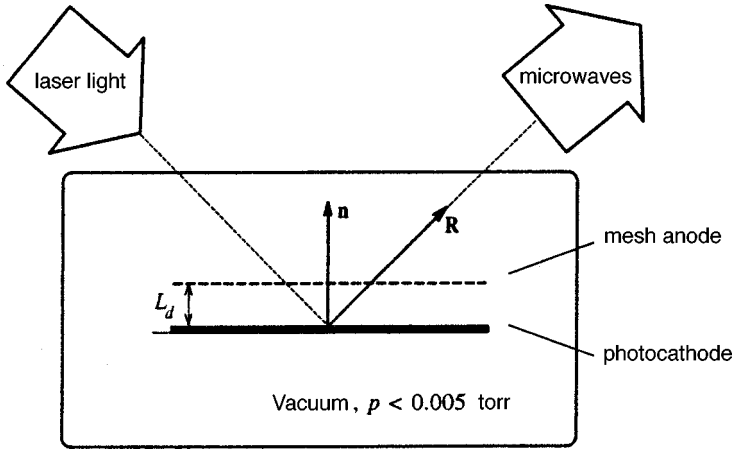

FIG. 1. Elementary microwave source.

familiar fact that such sources can exist in media with $n>1$. It is much less well known that they can also exist in a vacuum.[8]

Below we will consider two very simple superluminal sources for which solutions exist in closed form. Such solutions help in understanding the nature of the problem and the implications of various approximations.

Throughout what follows we assume $T\omega_{pe}>1$, where T is the pulse length of the ionizing radiation.

2.1. Superluminal source in free space

The simplest superluminal source is a current pulse propagating faster than light along a conducting surface. Such a current pulse arises, e.g., when an ionizing plane wave illuminates a conducting planar surface (Fig. 2).[6,7,9] The spatial current distribution of the emitted electrons creates an antenna near the surface excited in phase so that the radiated electromagnetic wave propagates in the direction corresponding to specular reflection of the ionizing radiation.

We identify the x-y plane with an infinite planar surface along which a superluminal current pulse

$$j_z=j_z(z,t-x/v), \quad v>c,$$

is propagating, composed of electrons emitted from the surface. The Maxwell equations for the nonzero components of the electromagnetic field take the form

$$\frac{1}{c}\frac{\partial E_x}{\partial t}=-\frac{\partial H_y}{\partial z},$$

$$\frac{1}{c}\frac{\partial E_z}{\partial t}=-\frac{4\pi}{c}j_z-\frac{\partial H_y}{\partial x}, \tag{2}$$

$$\frac{1}{c}\frac{\partial H_y}{\partial t}=\frac{\partial E_x}{\partial x}-\frac{\partial E_x}{\partial z},$$

together with the boundary conditions

$$E_x=E_z=H_y|_{t-x/v=0}=0, \quad E_x|_{z=0}=0, \tag{3}$$

We specify that in the limit $z\to\infty$ the wave is outgoing.

Since the current j_z (which is the only thing responsible for the presence of a field) and the boundary conditions depend on x and t only through the combination $t-x/v$, all the other quantities have the same dependence on x and t. This enables us to reduce the dimensionality of Eqs. (2) by going from the independent variables t, x, and z to $\tau=t-x/v$ and z. In terms of these variable we obtain the following equation for E_x:

$$\frac{\partial^2 E_x}{\partial z^2}-\frac{1}{c^2}\left(1-\frac{c^2}{v^2}\right)\frac{\partial^2 E_x}{\partial \tau^2}=\frac{4\pi}{v}\frac{\partial j_z}{\partial z}. \tag{4}$$

As may easily be verified by substitution, the solution of Eqs. (3) and (4) can be written in the form

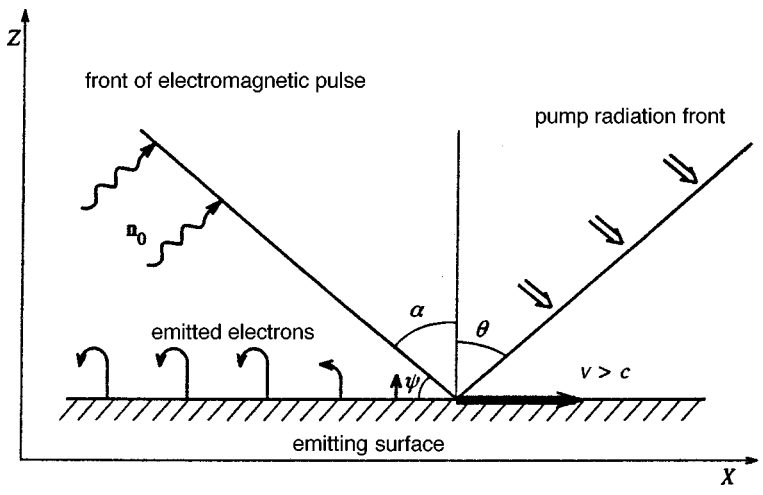

FIG. 2. Schematic of electromagnetic pulse generation by a superluminal current source.

$$E_x(\tau,z)=\frac{2\pi}{v}\left\{\int_0^z dz_1 j_z\left(\tau-\frac{z-z_1}{c}\sqrt{1-\frac{c^2}{v^2}},\ z_1\right)+\int_0^\infty dz_1 j_z\left(\tau-\frac{z+z_1}{c}\sqrt{1-\frac{c^2}{v^2}},\ z_1\right)-\int_z^\infty dz_1 j_z\left(\tau+\frac{z-z_1}{c}\sqrt{1-\frac{c^2}{v^2}},z_1\right)\right\}. \quad (5)$$

If the current source is localized in some interval Δz, then for $z>\Delta z$

$$E_x(\tau,z)=\frac{2\pi}{v}\left\{\int_0^\infty dz_1 j_z\left(\tau-\frac{z-z_1}{c}\sqrt{1-\frac{c^2}{v^2}},\ z_1\right)+\int_0^\infty dz_1 j_z\left(\tau-\frac{z+z_1}{c}\sqrt{1-\frac{c^2}{v^2}},\ z_1\right)\right\}. \quad (6)$$

If in addition we have

$$\frac{\Delta z}{c}\sqrt{1-\frac{c^2}{v^2}}<T_p, \quad (7)$$

where T_p is the time scale on which the current varies, then

$$E_x(\tau,z)\simeq\frac{4\pi}{v}\int_0^\infty dz_1 j_z\left(\tau-\frac{z}{c}\sqrt{1-\frac{c^2}{v^2}},\ z_1\right). \quad (8)$$

It is obvious that $\Delta z\sim\overline{v}T_p$, where $\overline{v}$ is the average velocity of the emitted electrons. Since we have $v<c$, the inequality (7) always holds.

The integral

$$-e\int_0^\infty dz_1 z_1 n_e(t,z_1)$$

is nothing other than the surface density of the dipole moment. Then

$$\dot{P}\equiv\frac{dP}{dt}=-e\int_0^\infty dz_1 z_1\frac{\partial n_e}{\partial t}=-\int_0^\infty dz_1 z_1\frac{\partial j_z}{\partial z_1}=\int_0^\infty dz_1 j_z(t,z_1). \quad (9)$$

Hence returning to (8) we have

$$E_x(\tau,z)\simeq\frac{4\pi}{v}\dot{P}\left(\tau-\frac{z}{c}\sqrt{1-\frac{c^2}{v^2}}\right)=\dot{P}\left(t-\frac{x}{v}-\frac{z}{c}\sqrt{1-\frac{c^2}{v^2}}\right). \quad (10)$$

Correspondingly,

$$H_y=\frac{4\pi}{\sqrt{v^2-c^2}}\dot{P}\left(t-\frac{x}{v}-\frac{z}{c}\sqrt{1-\frac{c^2}{v^2}}\right). \quad (11)$$

Thus, the radiated electromagnetic wave propagates in the direction corresponding to specular reflection of the incident radiation (Fig. 2):

$$\cos\alpha=\frac{c}{v},\quad \alpha=\frac{\pi}{2}-\psi,\quad \alpha=\theta. \quad (12)$$

The solution given above shows that at sufficiently large distances ($z>\Delta z$) the radiation field depends only on the surface density of the dipole moment, or rather, on its first derivative with respect to time. To find this quantity it is necessary to solve the problem of the formation of the electron boundary layer which arises in the course of electron emission from the surface. It turns out that at sufficiently large angles θ the radiated field E_w drops out of the problem.

It is quite clear that the radiated field E_w can be neglected when it is small compared with the space-charge field $E_{\text{s-c}}$,

$$E_w<E_{\text{s-c}}.$$

In order of magnitude,

$$E_w\sim\frac{\varepsilon}{\lambda\!\!\!^{-}\tan\theta},\quad E_{\text{s-c}}\sim\frac{\varepsilon}{\lambda_D}.$$

Here we have written $\varepsilon=m_e\overline{v^2}/2$ for the average energy of the emitted electrons and set $\lambda\!\!\!^{-}=cT_p$ and $\lambda=\overline{v}T_p$. Physically, $2\pi\lambda\!\!\!^{-}$ is the characteristic wavelength of the radiation and λ_D is the electron Debye radius. Consequently, if

$$\cos\theta<\frac{1}{\sqrt{1+\overline{v^2}/c^2}},$$

it follows that the radiation field can be disregarded in the calculation of the surface density of the dipole moment.

But this is by no means the only simplification. In the vast majority of cases of practical interest we can restrict ourselves to solving a one-dimensional problem in the variables z and τ when we calculate the electron dipole layer.

Thus, analysis of the above exact solution allows us to draw the following conclusions:

1. Only the surface distribution of the first time derivative of the dipole moment surface density is needed in order to calculate the radiative properties of a superluminal source, and the calculation may be done using an electrodynamic code with prescribed sources.

2. The time dependence of the formation of the electron boundary layer can be studied via a one-dimensional approach.

3. For sufficiently large θ the radiation field can be neglected and only one of the three Maxwell equations need be solved in order to get the polarization field E_z.

As already noted, the emitted electrons all radiate coherently. Consequently, at sufficiently large distances R (i.e., in the far field) the illuminated surface will radiate in the direction of propagation of the electromagnetic pulse like a point dipole with moment equal to the sum of the dipole moments due to the individual electrons. It follows immediately that

$$E_w\sim H_w\sim\frac{\ddot{P}S}{c^2R}\sim\frac{PS}{c^2T_p^2R}. \quad (13)$$

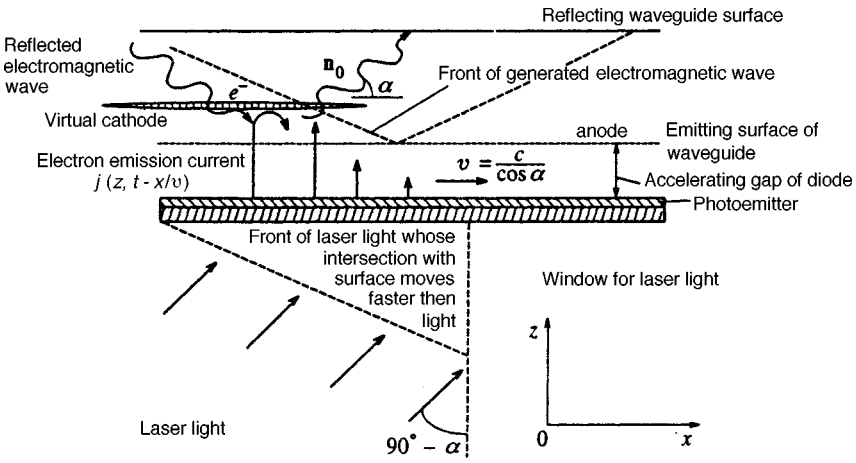

FIG. 3. Schematic of microwave generator based on superluminal source in a waveguide.

Here $\ddot{P} \sim P/T_p^2$ is the second time derivative of the dipole moment surface density, $P = \varepsilon/2\pi e$, S is the surface area, and e is the electron charge.

From (13) it is easy to estimate the radiated power Q and energy U:

$$Q \sim 2 \times 10^{12} (\gamma - 1)^2 \frac{S}{\lambda\!\!\!^{-2}} \text{ W}, \tag{14}$$

$$U \sim 70 (\gamma - 1)^2 \frac{S}{\lambda\!\!\!^{-}} \text{ J}, \tag{15}$$

where $2\pi\lambda\!\!\!^{-} = cT_0$, $\gamma = 1 + \varepsilon/m_e c^2$, m_e is the electron mass, $[S] = \text{m}^2$, and $[\lambda\!\!\!^{-}] = \text{cm}$. For $S = 0.1\,\text{m}^2$, $\lambda\!\!\!^{-} = 0.1\,\text{cm}$, and $\gamma - 1 = 0.7$, Eqs. (14) and (15) yields $Q \sim 10^{13}\,\text{W}$ and $U \sim 350\,\text{J}$.

2.2. Superluminal source in a waveguide

Another interesting and simple example of a superluminal source is an electron current pulse propagating superluminally along one of the surfaces of a planar waveguide[13] (Fig. 3).

As is well known, the amplitude of the radiation field adjacent to the radiating surface is proportional to the first derivative with respect to time of the dipole moment surface density (cf. Eqs. 10 and 11). For example, in the case of a slab superluminal source we have

$$H_w = \frac{4\pi}{c} \frac{\dot{P}}{\tan\alpha}, \quad E_{wz} = \frac{4\pi}{c} \frac{\cos^2\alpha}{\sin\alpha} \dot{P}. \tag{16}$$

If the dipole moment surface density oscillates at some frequency ω_0 equal to an integral multiple of an eigenfrequency of the waveguide,

$$\omega_0 = \frac{\pi c}{a \sin\alpha} k, \quad k = 1,2,\ldots, \tag{17}$$

then conditions obtain under which the radiated field can be enhanced by a large factor. It is readily seen that the achievable amplification length L is determined in principle by the duration T_{os} of the dipole moment oscillations:

$$L = cT_{\text{os}} \frac{\cos\alpha}{\sin^2\alpha}. \tag{18}$$

The conditions under which space charge (and hence the dipole moment) builds up undergo a change, however, since the field of the electromagnetic wave grows during the amplification process. Relation (18) is therefore valid, strictly speaking, only when the field of the electromagnetic wave is smaller than the field of the space charge.

When the z-component of the electric field E_{wz} of the wave becomes of order the space-charge field $E_{\text{s-c}}$,

$$E_{wz} \sim E_{\text{s-c}} \sim 4\pi e n_e \lambda_D \cong \frac{2\varepsilon}{e\lambda_D}, \tag{19}$$

its influence on the formation of the dipole layer can no longer be disregarded.

Deceleration of the emitted electrons in the wave field both reduces the dipole moment surface density and increases the oscillation frequency, and so violates condition (17). The linear stage of the amplification process is superseded by the nonlinear stage.

All these estimates are made within the restrictions of the linear approach. Moreover, it should be noted that only the case $E_{wz} > E_{wx}$ ($\cos\alpha \sim 1$) is being treated. This means that it is possible to disregard the presence of the x-component of the wave field and thus avoid unnecessary complications.

The condition $E_{wz} \leqslant E_{\text{s-c}}$ yields a bound on the growth length:

$$\frac{L}{\lambda} \leqslant \frac{4}{\sin\alpha} \frac{\gamma}{\sqrt{\gamma^2 - 1}}, \quad \lambda = 2\pi\lambda\!\!\!^{-}. \tag{20}$$

At the output end of a waveguide of length L the wave field is amplified by a factor N, where

$$N = \frac{L \tan\alpha}{2a}. \tag{21}$$

Accordingly, the energy flux density is amplified by a factor N^2:

$$J_x=\frac{4\pi}{c}\frac{\cos^3\alpha}{\sin^2\alpha}\dot{P}^2N^2. \tag{22}$$

The energy flux per unit length of the transverse cross section of the waveguide grows by the same factor:

$$\int dzJ_x\sim\frac{2\pi}{c}\frac{\cos^3\alpha}{\sin^2\alpha}\dot{P}^2N^2a. \tag{23}$$

Using these results we can show that if radiation is incident with power Q and energy U per unit length of the transverse cross section, then

$$Q\sim\frac{\pi}{4}\sin\alpha\cos\alpha\frac{m_e^2c^5}{e^2}\frac{(\gamma-1)^2}{\lambda}\left(\frac{L}{\lambda}\right)^2$$

$$\sim 2\times10^9\frac{(\gamma-1)^2}{\lambda}\left(\frac{L}{\lambda}\right)^2\ \mathrm{W\,cm^{-1}}, \tag{24}$$

$$U\sim\frac{\pi}{6}\sin\alpha\cos\alpha\frac{m_e^2c^4}{e^2}(\gamma-1)^2\left(\frac{L}{\lambda}\right)^2$$

$$\sim 4.8\times10^{-2}(\gamma-1)^2\left(\frac{L}{\lambda}\right)^3\ \mathrm{J\,cm^{-1}}.$$

For $\lambda=0.1$ cm, $\gamma-1=0.25$, and $L/\lambda=10$ we have $Q\sim10^{11}\,\mathrm{W\,cm^{-1}}$ and $U\sim3\,\mathrm{J\,cm^{-1}}$.

One can check these estimates by looking at an analytical solution of the problem. Suppose a current pulse with current density $j_z=j_z(\tau)$, $\tau=t-x/v$, is propagating in the x direction along an interior surface ($z=0$; cf. Fig. 3) of a planar metal waveguide of width a. The electromagnetic wave will be described by Eqs. (2) with the following initial and boundary conditions:

$$E_x=E_z=H_y|_{\tau=0}=0,\quad E_x|_{z=0}=E_x|_{z=a}=0. \tag{25}$$

This problem has an analytical solution, which can be written in the dipole approximation in the form

$$H_y(z,\tau)\to_{z>\lambda_d}\frac{4\pi}{\sqrt{v^2-c^2}}\sum_{n=0}^{\infty}\left(\dot{P}\left(\tau-\frac{2na+z}{c}\sin\alpha\right)\right.$$

$$\left.+\dot{P}\left(\tau-\frac{2(n+1)a-z}{c}\sin\alpha\right)\right), \tag{26}$$

$$E_z(z,\tau)\to_{z>\lambda_d}-\frac{c}{v}H_y.$$

From these expressions it follows that the electromagnetic wave will be amplified if the oscillation frequency of the dipole moment surface density is close to a waveguide eigenfrequency given by (17). If relations (17) and (20) hold for the length L of the waveguide, then on exiting from the waveguide the electromagnetic wave will have the following properties:

$$H_y\approx NH_0,$$

$$H_0=\frac{4\pi}{c\tan\alpha}\left(\dot{P}\left(\tau-\frac{z}{c}\sin\alpha\right)+\dot{P}\left(\tau-\frac{2a-z}{c}\sin\alpha\right)\right),$$

$$E_x\approx NE_0,$$

$$E_0=\frac{4\pi}{c}\sin\alpha\left(\dot{P}\left(\tau-\frac{z}{c}\sin\alpha\right)-\dot{P}\left(\tau-\frac{2a-z}{c}\sin\alpha\right)\right), \tag{27}$$

$$J_x=\frac{c}{4\pi}H_yE_z\approx J_0N^2,\quad J_0=\frac{c}{4\pi}H_0^2\sin\alpha.$$

Here $N=[(L/2a)\tan\alpha]$, where [] signifies that the integral part of the enclosed number is taken. Expressions (27) are in complete agreement with the estimates made earlier.

Thus, superluminal sources of radiation open new, totally unique prospects for generating high-power microwaves. The estimates (14), (15), and (24) show that superluminal sources can generate short high-power pulses of microwaves. The power and energy of these pulses grow as the characteristic wavelength falls off.

2.3. Estimate of the properties of a radiating element

A radiating element (see Fig. 1) consists of a diode with a cathode made from a photemissive material and and a mesh anode. The electrons expelled by the incident light are accelerated by the electric field between the electrodes. After passing through the mesh they form a radiating dipole layer in the space above the mesh. Generally speaking, the acceleration process alters the space–time distribution of the current density, so the time dependence of the current at the anode and at the cathode can differ in functional form. If the gap between the electrodes is much smaller than the size of the electrodes themselves, then when a superluminal pulse develops on the cathode the pulse of accelerated electrons at the anode will also be superluminal. Hence a synchronously radiating dipole layer forms above the mesh too. A diode with charge distributed in this manner is also a superluminal source of electromagnetic radiation and therefore will share all the properties of such sources. By varying the parameters of the radiating element (diode) and choosing different light sources and power inputs to the diode we can obtain a broad range of devices producing electromagnetic pulses in the microwave range.

Let us consider a planar diode with a gap L_d between electrodes charged to a potential φ_0, where $e\varphi_0/mc^2=\gamma-1$. The discharge of the diode is intiated by a planar flux of radiation obliquely incident on the cathode through the mesh anode with intensity

$$q=q_tt. \tag{28}$$

Photons of the incident radiation expel electrons from the surface of the cathode, forming a pulse with current density

$$j_e=j_tt,\quad j_t=\frac{eY}{\varepsilon_k}q_t, \tag{29}$$

where ε_k is the energy of a photon of the incident radiation and Y [electrons per photon] is the yield.

Let $\Delta\tau$ be the time taken by an electron to traverse the interelectrode gap,

$$\Delta\tau=\frac{L_d}{c}\sqrt{\frac{\gamma+1}{\gamma-1}}. \tag{30}$$

It is clear that if this time satisfies

$$\left|\frac{d\Delta\tau}{dt}\right| \ll 1, \tag{31}$$

then the current density is conserved and the current density at the anode must equal that at the cathode. In this case the current density in the space above the mesh is also given by Eq. (29), and we can estimate the characteristic time T_p for the formation of the dipole layer, which is obviously related to the electron plasma frequency by

$$\omega_{pe} T_p = 1, \tag{32}$$

and the characteristic wavelength of the broadband electromagnetic pulse radiated by the layer, $\lambda\!\!\!^{-} = cT_p$. From (32) it follows that

$$T_p^{-2} = \frac{4\pi e^2 n_e}{m_e \gamma} = \frac{4\pi e^2 j_t T_p}{m_e c\sqrt{\gamma^2-1}} \tag{33}$$

and hence

$$\lambda\!\!\!^{-} = \left(\frac{m_e c^4}{4\pi e}\,\frac{\sqrt{\gamma^2-1}}{j_t}\right)^{1/3}. \tag{34}$$

In the diode discharge process the potential across the electrodes varies, and the time of flight across the gap varies in consequence. If we take γ to be time-dependent according to $\gamma = e\varphi/m_e c^2 + 1$, where $\varphi = EL_d$ and E is the electric field in in the gap, then from (30) it follows that

$$\frac{d\Delta\tau}{dt} = -\frac{L_d}{c}\,\frac{1}{(\gamma-1)\sqrt{\gamma^2-1}}\,\frac{d\gamma}{dt}.$$

Since

$$\frac{dE}{dt} = -4\pi j_e,$$

we finally obtain

$$j_e \ll \frac{1}{4\pi}\,\frac{m_e c^3}{e}\,\frac{\gamma-1}{L_d^2}\,\sqrt{\gamma^2-1}, \tag{35}$$

and hence

$$j_e \ll \frac{1}{4\pi}\,\frac{m_e c^4}{e\lambda\!\!\!^{-}}\,\frac{\gamma-1}{L_d^2}\,\sqrt{\gamma^2-1}. \tag{36}$$

After substituting for $\lambda\!\!\!^{-}$ we get from (36)

$$j_e \ll \frac{1}{4\pi}\,\frac{m_e c^4}{e\lambda\!\!\!^{-}}\,\frac{(\gamma-1)^2}{L_d^2}\,\sqrt{\gamma+1}. \tag{37}$$

In what follows we will not consider the case $\gamma \gg 1$, since presumably the higher the accelerating voltage the greater the difficulties in producing and maintaining this potential, and *a fortiori* in building a workable design for the radiating element as a whole.

If we assume that the value of j_t is one-eighth of the upper limit imposed by the inequality (37), then the radiation produced in this case will have a characteristic wavelength

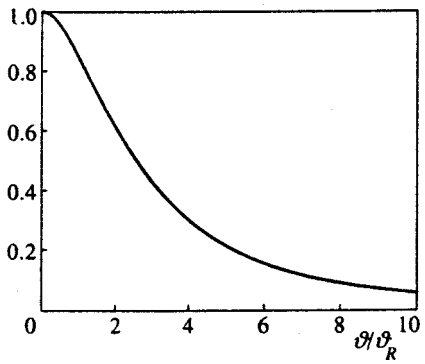

FIG. 4. Angular distribution of electromagnetic radiation from a square slab radiator.

$$\lambda\!\!\!^{-} \simeq \frac{2L_d}{\sqrt{\gamma-1}}. \tag{38}$$

Since the wavelength scales as $\lambda\!\!\!^{-} \propto j_t^{-1/3}$, letting the current density j_t assume its maximum value reduces $\lambda\!\!\!^{-}$ by a factor of 2:

$$\lambda\!\!\!^{-}_{\lim} = \frac{L_d}{\sqrt{\gamma-1}}. \tag{39}$$

We can estimate the electromagnetic field this radiator produces in the wave zone (the far field) by making use of the retarded potential. The easiest part is finding the magnetic field:

$$\mathbf{H} = \frac{1}{c^2 R}\int dv\ \mathbf{r}\times\frac{\partial \mathbf{j}}{\partial t}, \quad H \sim \frac{S\ddot{P}}{c^2 R}. \tag{40}$$

Here

$$\ddot{P} = \frac{m_e c^2}{2\pi e T_p^2}(\gamma-1)$$

is the second derivative with respect to time of the dipole moment surface density, S is the surface area of the radiator, and R is the distance from the radiator to the point of observation.

By using (40) we can obviously get the same values as before for the properties of the radiation, e.g., intensity of the microwaves:

$$J = \frac{c}{4\pi}H^2 \sim 1.8\times 10^{11}\,\frac{S^2}{R^2}\,\frac{(\gamma-1)^2}{\lambda\!\!\!^{-4}}\ \mathrm{W\,cm^{-2}}; \tag{41}$$

energy density:

$$V = JT_p \sim 5.9\,\frac{S^2}{R^2}\,\frac{(\gamma-1)^2}{\lambda\!\!\!^{-3}}\ \mathrm{J\,cm^{-2}}; \tag{42}$$

divergence (diffraction limit):

$$\vartheta_R = \frac{2\lambda\!\!\!^{-}}{D}, \tag{43}$$

where D is the length scale of the radiator.

Figure 4 displays a plot of the angular dependence of the

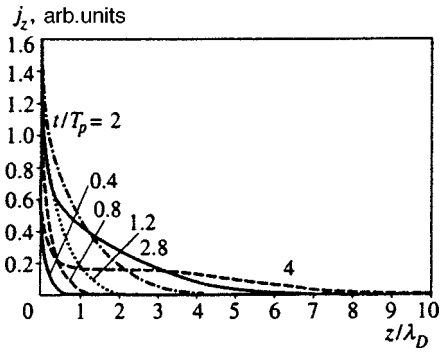

FIG. 5. Current density as a function of distance from the emission surface at various times.

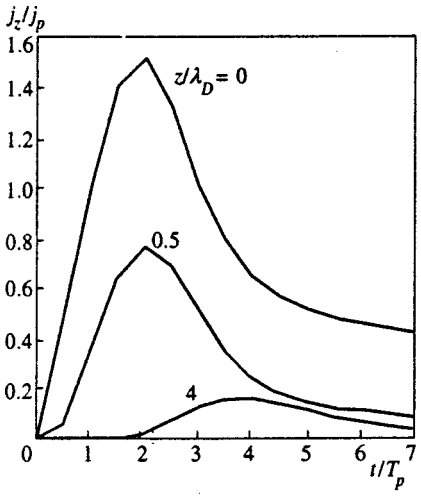

FIG. 6. Current density at various distances from the emission surface.

radiated power relative to the direction of specular reflection of the laser light for a square slab radiator with a dipole moment whose second time derivative has the functional form

$$\ddot{P}(\tau) = \frac{d}{d\tau}[\tau^2 \exp(-\tau)], \tag{44}$$

which is typical for a current with a linear time dependence. From the plot it is clear that essentially all the energy is concentrated within a cone with opening angle $3\vartheta_R$.

The energy stored in the gap between the electrodes is approximately equal to

$$W_a \approx 116 \frac{(\gamma-1)^2}{L_d} S \text{ J}. \tag{45}$$

In the formulas given above the units are $[R]=\text{m}$, $[S]=\text{m}^2$ and $[\text{L}]=[\lambda]=\text{cm}$.

The requirement on the intensity of the light source can be found by substituting the expression for the current density in Eqs. (28) and (29):

$$q(t=T_p) = \frac{1}{16\pi} \frac{m_e c^3}{e^2} \frac{\varepsilon_k}{Y} \frac{\gamma-1}{L_d^2} \sqrt{\gamma^2-1}, \tag{46}$$

from which it follows that

$$q \approx 3.4\times 10^2 \frac{\varepsilon_k}{Y} \frac{\gamma-1}{L_d^2} \sqrt{\gamma^2-1} \text{ W cm}^{-2}. \tag{47}$$

When j_t assumes its limiting value given by Eq. (37), q increases by a factor of 4:

$$q_{\text{lim}} \approx 4q. \tag{48}$$

If we take $\gamma=1.5$, $L=0.1$ cm, and $S=0.02\text{ m}^2$, then these formulas show that microwave radiation is generated with wavelength $\lambda \approx 0.28$ cm, divergence $\vartheta_R \sim 4\times 10^{-2}$ rad, total energy output ≈ 1.3 J, and total power 1.4×10^{11} W. The energy stored in the gap between the electrodes is 5.8 J, the light intensity is $\sim 2\times 10^5 \text{ W cm}^{-2}$ ($\varepsilon_k \approx 2$ eV) and the yield is $Y \approx 0.2$.

3. RESULTS OF NUMERICAL MODELING

It is evident that the problems that must be solved in connection with radiation from superluminal sources made up of radiating elements constructed as shown in Fig. 1 can be broken down into three categories:

1. Problems related to the study of the electromagnetic radiation produced by the superluminal source. In these problems the properties of the electron layer near the surface (above the mesh) are assumed to be given. These problems are purely electrodynamic, with sources prescribed.

2. Problems related to the properties of the electron layer near the surface (above the mesh).

3. ''Systems'' problems, i.e., problems in which the radiation processes are considered together with the development of the electron spatial and energy distributions.

For the vast majority of problems a two-dimensional code suffices. Hence the set of programs developed for the studying superluminal sources numerically includes one- and two- dimensional codes, both purely electrodynamic codes that only solve Maxwell's equations and unified codes that solve Maxwell equations for the electromagnetic fields together with the Vlasov equation for the electrons,

Numerical solutions were obtained for the following cases:

a) An infinite plane, using a one-dimensional code to solve the combined Maxwell–Vlasov system of equations (Figs. 5–7).

b) An infinite conducting strip of width $\Delta = 50$ cm, using a three-dimensional electrodynamic code with specified sources, and using a two-dimensional code for the Maxwell equations and a particle description for the electrons (Fig. 8).

The current of the emitted electrons was specified according to

$$j_z(t) \propto \eta\left(t - \frac{x}{c}\cos\alpha\right), \quad \alpha = \frac{\pi}{4}, \tag{49}$$

where η is the Heaviside unit step function.

The results of these calculations demonstrate that a directed electromagnetic field is produced and confirm the the-

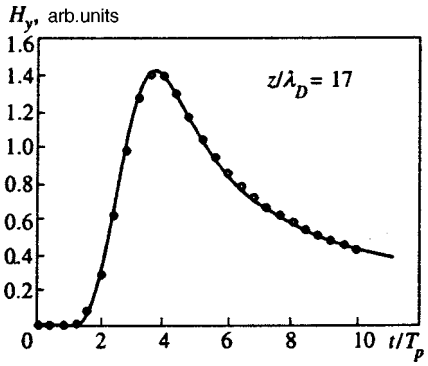

FIG. 7. Magnetic field strength versus time for $z/\lambda_D = 17$. The filled circles are taken from the one-dimensional numerical model and the solid curve is the analytical solution.

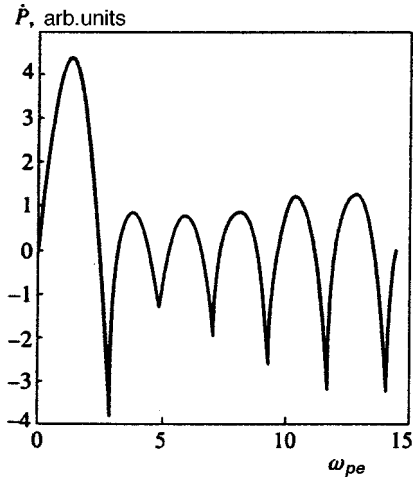

FIG. 9. First time derivative of the dipole moment of an electron cloud for a monoenergetic emission electron spectrum.

oretical estimates. The possibility of amplifying the electromagnetic field in a waveguide is related to the existence of stable oscillations of the dipole moment surface density.

The time dependence of the dipole moment surface density is determined by the time dependence of the pulse of emission current and the energy spectrum of the emitted electrons. In the numerical calculations a square-wave shape was asssumed for the emission current pulse:

$$j_z(t) = \text{const} \quad \text{for} \quad 0 < t < T \gg \omega_{pe}^{-1}. \tag{50}$$

Two forms of the electron spectrum were considered:

a) a monoenergetic spectrum with $\varepsilon = \overline{\varepsilon}$ and electrons emitted normal to the surface;

b) electrons distributed uniformly in energy within $\varepsilon \leqslant 2\overline{\varepsilon}$ and vanishing for $\varepsilon > 2\overline{\varepsilon}$, with the electrons again emitted normal to the surface.

Results of the calculations are shown in Figs. 9 and 10.

The simulations show that stable oscillations of the dipole moment surface density develop only in the case of monoenergetic electrons.[14] In particular, such an electron spectrum is found when emission electrons with a small initial energy traverse the accelerating gap at constant potential. The oscillations in the electron flow observed in the case of a monenergetic electron spectrum are analogous to the well known oscillations of a virtual cathode.[15] Thus, the electromagnetic field is amplified when conditions are satisfied for the formation of a virtual cathode.

The behavior of the electromagnetic field generated in a waveguide by a superluminal current pulse was studied

a) using a two-dimensional electrodynamic code for a current density distributed with a prescribed space–time dependence; the width of the waveguide was $a \simeq 4$ cm; the superluminal current pulse propagated in the x direction along the surface $z = 0$ with velocity $\sqrt{2}\,c$; and the electron current density was nonzero only within a thin layer of depth 0.1λ at the surface of the waveguide. The resulting time dependence of the current is shown in Fig. 9; the time dependence of the magnetic field at a distance from the beginning of the waveguide is that shown in Fig. 11.

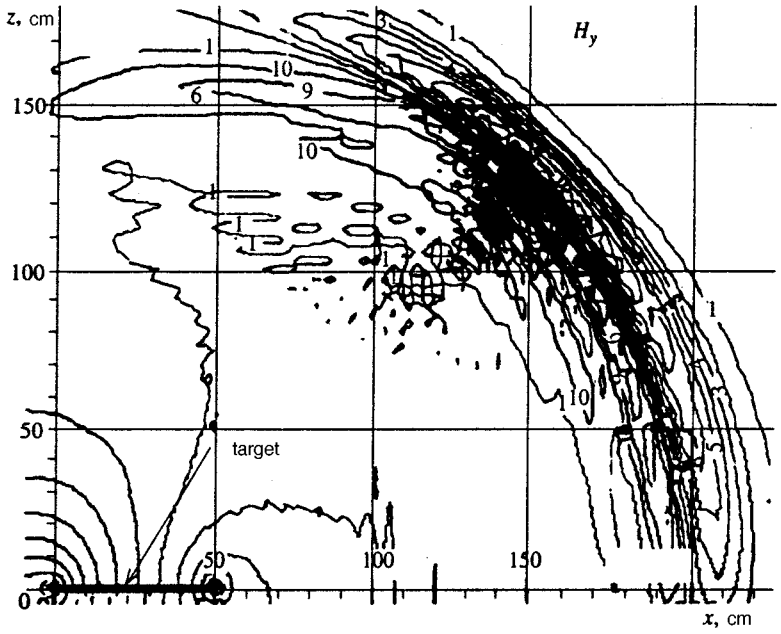

FIG. 8. Contours of constant magnetic field strength produced by an electron current pulse propagating with velocity $v \equiv v_x = \sqrt{2}\,c$.

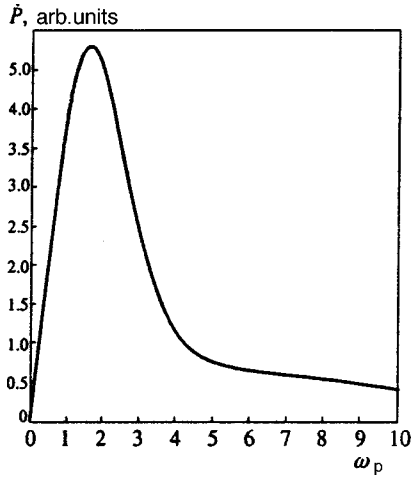

FIG. 10. First time derivative of the dipole moment of an electron cloud for a broad emission electron energy spectrum.

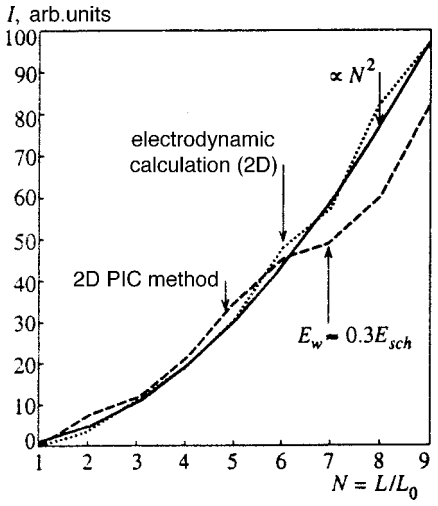

FIG. 12. Strength of the electromagnetic wave as a function of distance traveled along the waveguide.

b) using a two-dimensional particle-in-cell (PIC) code which solved the coupled Vlasov–Maxwell system of equations; the spectrum of the emitted electrons was prescribed at $z=0$: electrons with energy $\varepsilon=500$ keV were emitted by the surface over an extended period of time T and the emission front propagated along the surface with velocity $\sqrt{2}\,c$. Figure 12 shows the intensity of the electromagnetic wave as a function of the distance traveled by the wave.

The results of the numerical calculations agree with the theoretical estimates.

In the radiating element shown in Fig. 1 the energy of the electrons comprising the current pulse propagating with superluminal velocity along the surface of the cathode increases as a result of acceleration in an external field. During the acceleration process, however, the time dependence of the pulseform can change, so that the current pulse at the anode may differ from that at the cathode.

From the theoretical discussion it follows that for current densities below a maximum value—which of course is determined by the length of the accelerating gap and the potential across it—the current density is a conserved quantity, and the time dependence of the current pulse is therefore not altered. That is, the current density of the electrons dislodged from the surface of the anode will have the same time dependence as at the cathode, and if a superluminal current pulse propagates along the cathode then one will also propagate along the anode.

Calculations were performed for a two-dimensional planar diode. The length of the diode was 10 cm, the interelectrode gap was 0.1 cm, and the applied potential was 100 kV. Electrons were injected perpendicular to the surface of the cathode:

$$j_z(t,x,z=0)=\eta\left(t-\frac{x}{v}\right)j_0\left(t-\frac{x}{v}\right),$$

$$j_0=2.75\times10^{33}t \ \text{ electron cm}^{-2}\,\text{s}^{-1}, \quad v=\sqrt{2}\,c. \qquad (51)$$

The results of the calculations are shown in Figs. 13 and 14. According to the previous section, we should find $\chi\approx0.39$ cm, $T_p\approx1.3\times10^{-11}$ s, and $\dot{P}\sim4\times10^{12}$[cgs]. In the

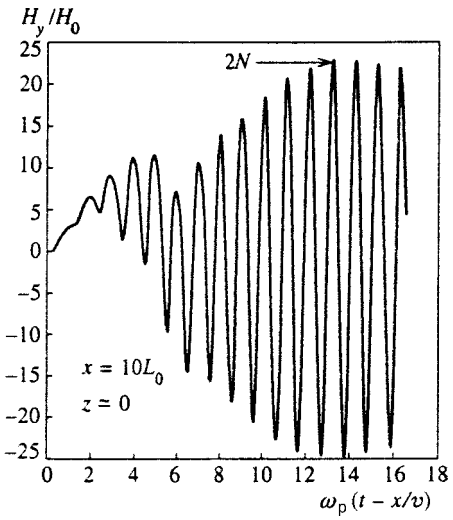

FIG. 11. Time dependence of magnetic field at a fixed point in the waveguide.

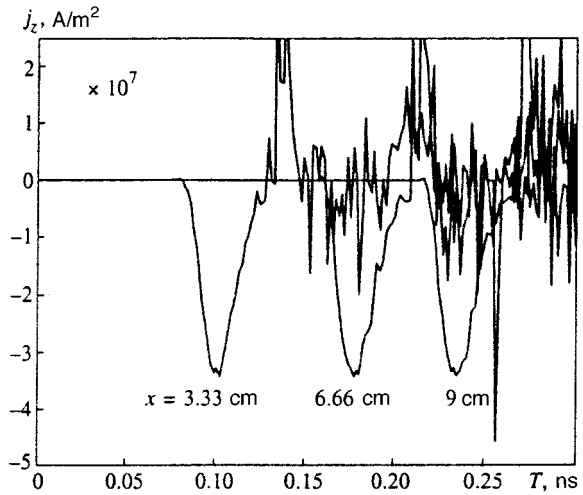

FIG. 13. Time dependence of the electron current above the anode surface.

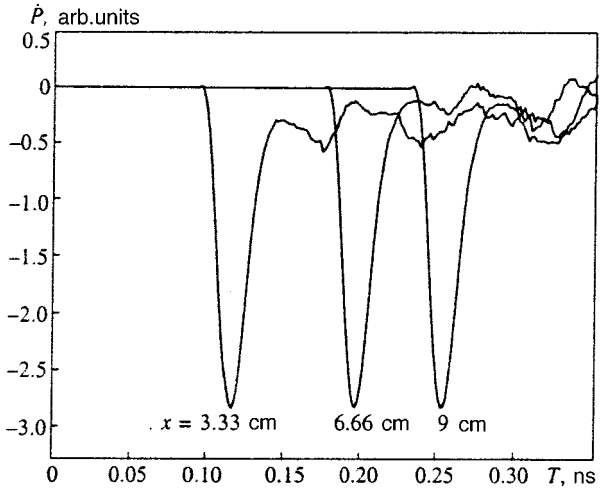

FIG. 14. Time dependence of the first time derivative of the dipole moment above the anode surface.

calculation the current density within the gap reached a value of $3.5\times10^3\,\mathrm{A\,cm^{-2}}$, which is only a factor of 2 less than the maximum value. The value of $\dot{P}$ estimated using the formulas of the preceding section is about 1.4 times larger than that obtained in the calculation. For $\varkappa$ and T_p the calculated and estimated values are practically the same. The current pulse travels along the surface of the anode at a velocity $\sqrt{2}\,c$. Note that the upper limit for the current density across the diode is about $7.4\times10^3\,\mathrm{A\,cm^{-2}}$, and $j_0(t=T_p)\approx5\times10^3\,\mathrm{A\,cm^{-2}}$.

Thus, the validity of the estimates is confirmed, and they are found to hold for current densities close to the limiting value.

4. CONCLUSION

Wideband generators of microwave radiation based on a superluminal emission source enjoy a number of absolutely unique properties. In principle such devices may find extensive practical applications under suitable circumstances because of these advantages:

1. The radiation is produced in extremely short pulses, down to picosecond time scales.

2. The shape of the radiating surface ensures highly directional radiation and beam formation.

3. The generator is compact and relatively light in weight.

4. In terms of power they surpass existing devices by orders of magnitude, and unlike other designs the power increases as the wavelength decreases.

The physics of such generators is extremely simple. The elementary concepts of classical electrodynamics suffice for a basic understanding and for obtaining estimates. A rigorous theory can be based on the Vlasov–Maxwell equations.

The theory presented here permits us both to predict the parameters of the radiation produced by an arbitrary emitting surface and to obtain specifications for the components of a generator, namely, the source of the ionizing radiation, the accelerating gap, and the power supply.

[1] K. A. Zheltkov, *Picosecond High-Current Electron Accelerators* [in Russian], Energoatomizdat, Moscow (1991).
[2] L. N. Kazanskiĭ and A. A. Rukhadze, Pis'ma Zh. Tekh. Fiz. **20**(3), 26 (1994) [Tech. Phys. Lett. **20**, 101 (1994)].
[3] A. N. Didenko, Dokl. Akad. Nauk **356**, 470 (1997) [Phys. Dokl. **42**, 528 (1998)].
[4] E. J. Nalos, Proc. IEEE **66**, 276 (1978).
[5] B. M. Bolotovskiĭ and V. P. Bykov, Usp. Fiz. Nauk **160**, 141 (1990) [Sov. Phys. Usp. **33**, 477 (1990)].
[6] Yu. N. Lazarev and P. B. Petrov, All-Union Technical Physics Research Institute, Preprint No. 9 (1991).
[7] Yu. N. Lazarev and P. V. Petrov, in *Proc. of Intern. School of Plasma Physics*, ed. by D. Akulina, E. Sindoni, and C. Warton, Piero Caldirola, Italy, Varenna (1991), ISSP-10, p. 565.
[8] V. L. Ginzburg, *Theoretical Physics and Astrophysics*, Pergamon, Oxford (1979) [Russian 2nd ed., Nauka, Moscow (1981)].
[9] N. J. Carron and C. L. Longmire, Proc. IEEE **NS-23**, 1897 (1976).
[10] N. J. Carron and C. L. Longmire, Proc. IEEE **NS-25**, 1329 (1978).
[11] Yu. N. Lazarev and P. V. Petrov, JETP Lett. **60**, 634 (1994).
[12] Yu. N. Lazarev and P. V. Petrov, in *Intense Microwave Pulses III*, ed. by H. E. Brand, Proc. SPIE **2557**, 159 (1995).
[13] Yu. N. Lazarev and P. V. Petrov, in *Intense Microwave Pulses IV*, ed. by H. E. Brand, Proc. SPIE **2843**, 197 (1996).
[14] R. Stettner, IEEE Trans. Nucl. Sci. **NS-24**, 2461 (1976).
[15] J. Benford and R. Swegle, *High Power Microwaves*, Artech House, Boston, MA (1991).

Translated by David L. Book

Supercritical convection associated with ultrafast MHD rotation

S. V. Starchenko*)

''Borok'' Geophysical Observatory, Branch of the Joint Geophysical Institute of the Russian Academy of Sciences, 152742, Borok, Yaroslavl Region, Russia

(Submitted 24 July 1998)
Zh. Éksp. Teor. Fiz. **115**, 1708–1720 (May 1999)

Highly nonlinear buoyant convection is investigated analytically under conditions typically encountered in the liquid cores of planets in the solar system. As a result of the supercritical behavior (enormous Rayleigh number) and ultrafast rotation (small Ekman number) typical of such flows, diffusion and viscosity act only in layers that are asymptotically thin in comparison with the radius of the core. These boundary layers control the buoyancy, the large-scale velocity, and the magnetic field observed at the planetary surface. The interchange of the internal layers determines the small-scale (unobservable) fields and the prevailing symmetry of the large-scale magnetic fields. It is proved for the first time that axisymmetric azimuthal flows dominate at large scales, while convection cells elongated parallel to the axis of rotation dominate at small scales. A system of equations is derived which is optimum for describing magnetoconvection of planetary cores on both large and small scales. It yields estimates in superb agreement with expensive numerical and experimental models of supercritical convection associated with rapid rotation. Such models will be capable of solving the MHD dynamo problem only when their algorithms are made consistent with the asymptotic limits presented here. © *1999 American Institute of Physics.* [S1063-7761(99)01305-0]

1. INTRODUCTION AND FORMULATION OF THE PROBLEM

The problem of the magnetohydrodynamic (MHD) dynamo of the earth and the other planets of the solar system was identified by Einstein as one of the five fundamental problems of physics, but an approach in principle to finding a solution has begun to appear only recently. For this reason, until now the dynamics of the most active part of the planet, its liquid core, has not been understood, although the core may be responsible for the evolution of all planetary interiors. The large-scale magnetic field is the sole evidence at the planetary surface for activity of the liquid core. Its evolution can be traced from times comparable with the age of the planet (paleomagnetic investigations) down to the present age (space investigations). The magnetic fields are a unique means of thoroughly evaluating the properties, dynamics, and structure of the deep-lying cores of the earth and other planets, since these regions are inaccessible to direct probes.

The simplest fully self-consistent system for describing an MHD dynamo in the Boussinesq approximation includes the Navier–Stokes equation for a divergence-free velocity $\mathbf{V}$ (in the coordinate frame rotating about the z axis with the angular velocity Ω of the mantle), the Maxwell equation for the divergence-free magnetic field $\mathbf{B}$, and the equation for the diffusion of the light constituent which describes the acceleration A due to Archimedes' Principle:

$$A\mathbf{r}/L=(2\mathbf{\Omega}\times+D/Dt-\nu\nabla^2)\mathbf{V}-(\nabla\times\mathbf{B})\times\mathbf{B}/\mu_0\rho+\nabla P, \tag{1}$$

$$\left(\frac{D}{Dt}-\frac{1}{\mu_0\sigma}\nabla^2\right)\mathbf{B}=(\mathbf{B}\cdot\nabla)\mathbf{V}, \quad \left(\frac{D}{Dt}-k\nabla^2\right)A=\beta\frac{V_r}{r^2}.$$

Here $D/Dt\equiv\partial/\partial t+\mathbf{V}\cdot\nabla$ and P is the total pressure. For the liquid core of the Earth, with an outer radius of $L=3.5\times10^6$ m, density of $\rho=10^4$ kg/m^3, and conductivity $\sigma=4\times10^5$ S m^{-1}, we have $\Omega=7.3\times10^{-6}$ s^{-1}, specific kinetic energy of convection β greater than or of order 10^2 J kg^{-1}, and thermal diffusivity k and viscosity ν less than 10^{-5} m^2 s^{-1}. At the boundary the velocity satisfies slip conditions, the magnetic field is continuous, and the conditions specifying the buoyancy field A, the basic source of the magnetoconvection, are applied.

The thickness of the viscous (Ekman) and that of the diffusive (Archimedean) boundary layers are directly proportional to the small quantitites δ and ε:

$$\delta^2\equiv\frac{\nu}{\Omega L^2}\ll1, \quad \frac{1}{\varepsilon^3}=R\equiv\frac{\beta}{k\Omega}\gg1,$$
$$\text{where } \beta\equiv\begin{cases} gG/4\pi k\rho w, \\ g\alpha Q/4\pi k\rho c_p. \end{cases} \tag{2}$$

Here Q is the thermal power and G is the gravitational power of the source of the convection, which is located at the inner boundary of the liquid core; g is the acceleration due to gravity at the outer boundary of the mantle, and α and c_p are the thermodynamic parameters of the liquid in the core. The purely gravitational source of convection is more effective than the thermal.[1] Accordingly, the specific gravitational energy w is considerably smaller than the specific thermal energy c_p/α in the liquid cores of planets. The power of the thermal source can be substantially greater than that of the gravitational source in giant planets,[2] however, so both effects must be retained in (2).

Direct numerical simulations of Eqs. (1) are extraordinarily difficult because of the strong convection of the physical fields A, $\mathbf{V}$, and $\mathbf{B}$ in the boundary and shear layers, whose dimensionless thickness is set by $\delta \leqslant 10^{-7}$ and $\varepsilon \leqslant 10^{-4}$ from (2). Recently, by dint of enormous expenditures of computer time, direct calculations have been carried out in which these quantities had values of 10^{-2} (Refs. 2–8). For the first time the behavior of the magnetic field has begun to resemble that of the geomagnetic field, now that these dimensionless quantities are small numbers. But simulations of the geodynamo via this approach using realistic values of the small parameters will not be feasible in the foreseeable future.

More promising is the asymptotic approach, in which these parameters are formally assumed to be vanishingly small, which makes it possible to resolve the structure of the boundary layers with high accuracy.

The self-consistent model of the geodynamo is supercritical, since from Eq. (2) its Rayleigh number R exceeds the critical value $R_c \approx \delta^{-2/3}$ for the onset of convection by many orders of magnitude. Supoercritical behavior is characteristic for dynamo systems, since magnetic fields can only be generated when the strength of the sources exceeds some threshold value. The main goal of the present work is to find an asymptotic form of Eqs. (1) in the supercritical limit, when the sources responsible for generating all the physical fields can formally be taken to be infinite. The strengths of the sources responsible for generating the velocity field $\mathbf{V}$ and the buoyancy A are characterized by huge numbers: the inverse Ekman number $1/E = \delta^{-2} \leqslant 10^{14}$ and the Rayleigh number $R = \varepsilon^{-3} \leqslant 10^{12}$. The accuracy of the supercritical approximation is determined by E and $1/R$, which appear as small coefficients multiplying the Laplacian Δ in Eqs. (9) below, which optimize (1). Thus, the supercritical approximation for the system (1) describing planetary cores should be extremely good, since its accuracy is $\sim \max(E, 1/R) \leqslant 10^{-12}$, which is far more accurate than the familiar direct models run for realistic lengths of time.[2–8]

Supercritical convection of the form described by (1) and (2) has been modeled in laboratories[2] and in space;[3] the critical Rayleigh number R_c was exceeded by one to two orders of magnitude, but these models have not been able to reproduce planetary hydrodynamics. Subsequent expensive numerical experiments[4] revealed that highly nonlinear ($R \geqslant 10^2 R_c$) supercritical convection differs in principle from supercritical. The methods known prior to the present work required colossal amounts of computer time in order to simulate rapidly rotating planetary cores with $R \geqslant 10^3 R_c \approx 10^6$ and $\delta^2 \leqslant 10^{-6}$. The most impressive model is that of Glatzmaier and Roberts,[5] which until now was the only model of supercritical MHD convection, since the approaches developed before now entailed amazing expense. Glatzmaier and Roberts[5] used 2000 hours of time on a Cray C-90 in order to reach $\delta^2 = 1.7 \times 10^{-6}$ and $R = 9.1 \times 10^6$, but were nevertheless unable to reproduce even the order of magnitude of the geomagnetic field and its symmetry. In subsequent work[6] Glatzmaier and Roberts, at the cost of even longer running times, managed to get their model to generate magnetic fields approximately equal to the actual geomagnetic field by varying the parameters of the system. The question of how these parameters determine the magnitude and structure of the magnetic, velocity, and buoyancy fields, however, remains unanswered.

In the present work A_*, V_*, and B_*, the scale values of the buoyancy, velocity, and magnetic field, respectively, are expressed analytically in terms of the basic parameters of the system. We have also made estimates of the characteristic large-scale, small-scale, boundary-layer, and internal magnetoconvection structures analytically. This supplies a simple explanation for the failure of Ref. 5 and the success of Ref. 6, and also points out an approach in principle to correctly reproducing the magnetic fields of the planets which arise as a result of supercritical turbulence for typically gigantic quantities $R \geqslant 10^{12}$ and extraordinarily small quantities $\delta^2 \leqslant 10^{-14}$. Obviously, both direct numerical simulation and experiment are necessarily totally unable to attain such extreme values. Here we propose an asymptotic approach that must be included in the numerical model if it is to satisfactorily solve the problem of supercritical magnetoturbulence for ultrafast rotation.

For all planets under consideration we are entitled to adopt the boundary conditions of impenetrability, continuous magnetic field, and constancy of the buoyancy and its flux:

$$V_r = 0, \quad \mathbf{B} = \mathbf{B}_1 = \mathbf{B}_c, \quad A(r=1) = 0, \quad \partial A(r=cL)/\partial r = 0, \tag{3}$$

which are imposed on the system (1) at the interface between the (possibly incompletely) solid inner core, labeled ''c,'' at $r = cL$ and at the mantle–core boundary (labeled ''1'') $r = L$. The boundary conditions (3) are satisfied by the no-slip condition $\mathbf{V} = \mathbf{V}_c$ at the boundary with the solid core (if there is one). In terrestrial planets a similar no-slip condition $\mathbf{V} = \mathbf{V}_1$ holds at the outer boundary as well. In the giant planets Uranus and Neptune the latter (and possibly the former) is partly or wholly replaced by free-surface conditions.

The solid core rotates with velocity

$$\mathbf{V}_c = (0, V_{c\theta}, V_{c\varphi}),$$

and the magnetic field inside it (for $r \leqslant cL$) satisfies the equation

$$\mu_0 \sigma_c D\mathbf{B}_c/Dt = \nabla^2 \mathbf{B}_c$$

with boundary conditions

$$\mathbf{B}_c = \mathbf{B}(r = cL).$$

The core–mantle boundary rotates with velocity

$$\mathbf{V}_1(r=L) = (0, V_{1\theta}, V_{1\varphi}),$$

and the magnetic field in the conducting part of the mantle (for $L \leqslant r \leqslant (1+d_1)L$) is given by the equation

$$\partial \mathbf{B}_1/\partial t = \nabla \times (\mathbf{V}_1 \times \mathbf{B} - \nabla \times \mathbf{B}_1/\mu_0 \sigma_1)$$

with boundary conditions

$$\mathbf{B} = \mathbf{B}_1(r = L).$$

In the upper (poorly conducting) part of the mantle the field becomes irrotational, $\mathbf{B}_1 = -\nabla U$, and is derivable from a potential satisfying the Laplace equation $\nabla^2 U = 0$ for

$r \geqslant (1+d_1)L$. It is this magnetic field that is observed with spacecraft and on the planetary surface and serves as the main source of information about the dynamics of the liquid core, which is inaccessible to observation.

Ordinary differential equations describing the time dependence of the mantle and the solid-core dynamics complete the formulation of the problem. The essential part of the description of the dynamics of the solid core is the balance between the viscous and electromagnetic forces, while effects not treated here (singular layers, the gravitational–precessional effect, etc.) can make an equally large contribution to the mantle dynamics. To lowest order the massive mantle does not move in our coordinate system, which rotates at angular velocity Ω, while the considerably lighter solid core rotates with angular velocity $\sim V_*/L \leqslant 10^{-3}\Omega$.

2. BOUNDARY LAYERS AND TURBULENCE

The balance between the Archimedean and Coriolis forces gives a relation between the buoyancy and the velocity:

$$A_* = \Omega V_*,$$

which follows from the first equation of the set (1) in the center of a layer. The variables appearing in (1) change abruptly as a function of radius at the boundaries, and a small radial component V_r arises there due to the solid-wall condition. Using the dominance of the radial derivatives $\partial/\partial r \sim 1/l_*$ and taking $V_r \approx (l_*/r)V_*$ in the Archimedean boundary layer $l_* \ll L$ we therefore obtain

$$\frac{\partial A}{\partial t} + \mathbf{V}\cdot\nabla A - k\frac{\partial^2 A}{\partial r^2} - \beta\frac{V_r}{r^2} = 0, \quad l_* = r\left(\frac{\Omega k L}{\beta r}\right)^{1/3}. \quad (4)$$

Here l_* is the thickness of the Archimedean layer at the outer boundary $r=L$ (where $l_* = LR^{-1/3}$) or at the inner boundary $r=cL$. Comparing the second term in (4) with the last one, we find typical magnitudes for the buoyancy and velocity:

$$A_* = \Omega V_* \simeq \beta\frac{l_*}{r^2}, \quad V_* = \frac{k^{1/3}\beta^{2/3}}{\Omega^{2/3}L} = R^{2/3}\frac{k}{L}. \quad (5)$$

Using the parameters of Glatzmaier and Roberts,[5] we find from (4) a thickness $l_* = 17$ km and from (5) a velocity $V_* = 0.4\ \mathrm{cm\,s^{-1}}$ and typical temperature $T_* = 3\times 10^{-3}$ K ($T_* = A_*/g\alpha$, $A_* = 3\times 10^{-7}\ \mathrm{m\,s^{-2}}$), in splendid agreement with the results of Ref. 5. Four modifications of Ref. 5 treated in Ref. 6 also give excellent agreement with our asymptotic estimates. The results shown in Eqs. (4) and (5) are in good agreement with unmagnetized hydrodynamics modeling done in the laboratory,[2] in space,[3], and numerically.[4] Our results (4) and (5) even agree satisfactorily with treatments[7,8] in which only weakly supercritical parameters were used, which confirms the high accuracy and reliability of this asymptotic approach.

Thus, the scale value of the velocity depends weakly on the magnetic fields, which nevertheless have a strong influence on the nature of the MHD flow. Consequently, the absence of an intrinsic magnetic field in a planet such as Venus is not surprising, since even strong flows in the core can not satisfy the kinematic conditions for generation.[9]

In the first two equations of the system (1) the terms $\sim D/Dt$ are negligible and $V_r = O(d_*/L)$ holds inside hydromagnetic boundary layers of thickness $d_* \ll L$, where the MHD equations (1) simplify to

$$2i\Omega\cos(\theta)\mathscr{V} - \nu\frac{\partial^2\mathscr{V}}{\partial r^2} = \frac{B_r}{\mu_0\rho}\frac{\partial\mathscr{B}}{\partial r}, \quad \frac{-1}{\mu_0\sigma}\frac{\partial\mathscr{B}}{\partial r} = B_r\mathscr{V} + \mathscr{C}. \quad (6)$$

Here the complex velocity $\mathscr{V} = V_\varphi - iV_\theta$ and magnetic field $\mathscr{B} = B_\varphi - iB_\theta$ vary strongly, while the integration constant $\mathscr{C}$ and radial magnetic field B_r experience essentially no change across the boundary layer, $\partial(\mathscr{C}, B_r)/\partial r \simeq 0$. The constants of integration $\mathscr{V}_0, \mathscr{V}_c, \mathscr{B}_0$ in the general solution of Eqs. (6) also do not vary at the inner boundary where $\Delta r = r - cL \sim \delta L \ll L$:

$$\begin{aligned}
\mathscr{B} &= \mathscr{B}_0 - \mu_0\sigma B_r\left[\int_{r=cL}^{r=r}\mathscr{V}dr + \mathscr{C}\Delta r\right],\\
\mathscr{V} &= \mathscr{V}_0 + \mathscr{V}_c\exp\left[-\left(2i\frac{\Omega}{\nu}\cos\theta + \frac{B_r^2\sigma}{\nu\rho}\right)^{1/2}\Delta r\right]\\
&\quad + \frac{-\sigma B_r\mathscr{C}}{2i\rho\Omega\cos\theta + \sigma B_r^2}.
\end{aligned} \quad (7)$$

The solution at the outer boundary looks similar, with different constants of integration and $\Delta r = L - r \sim \delta L \ll L$.

By using (6) and (7) to impose force balance and applying the boundary conditions (3) we find the thickness d_* of the boundary layer, the characteristic magnetic field, and the ratio $M^2B_*^2/\rho\mu_0V_*^2$ of the magnetic energy to the kinetic energy:

$$\frac{d_*}{L} = \begin{cases}\delta\\ \delta/S,\end{cases} \quad \frac{B_*}{\sqrt{\Omega\rho/\sigma}} = \begin{cases}\sqrt{S}\\ S\end{cases}, \quad \frac{M^2B_*^2}{\rho\mu_0V_*^2} = M^2\frac{\delta}{\epsilon}\begin{cases}1,\\ S,\end{cases} \quad (8)$$

where

$$S \equiv R^{2/3}q\,\delta = \mu_0\sigma V_* L\ \delta = \frac{\mu_0\sigma\nu^{1/2}k^{1/3}\beta^{2/3}}{\Omega^{7/6}L} = \begin{cases}\leqslant 1,\\ \geqslant 1.\end{cases}$$

Here $\epsilon \equiv V_*/L\Omega \ll 1$ is the very small Rossby number, $q \equiv k\mu_0\sigma < 1$ is the small Roberts number, and we have $M \geqslant 2$ because the peak magnetic field is shifted away from the boundary (where magnetic flux generation is strongest) by several d_* (Refs. 9 and 10).

The radial component of the magnetic field B_r is generated in the main volume, outside the region of peak flux generation at the edges of the layer where the components $B_\varphi, B_\theta \approx MB_*$ are produced, which are perpendicular to the radius r. Accordingly, the typical magnitude $B_r \approx B_*$ observed at the boundary of the core is usually much less than the (unobserved) fields in the interior, MB_*. Consequently, from (8) we find using the results of Ref. 11 that $S \leqslant 1$ holds for the magnetic fields of Earth, Jupiter, Saturn, Uranus, and Neptune, while for the other planets and satellites we have $S < 1$.

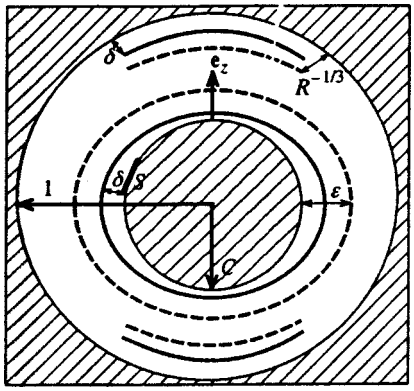

FIG. 1. The coordinate system rotates about the unit vector $\mathbf{e}_z$ which defines the typical rotation of the central sphere *c*. The outer sphere *1* is almost motionless. Shown are the hydromagnetic boundary layers: the Archimedean ($\varepsilon = R^{-1/3}$), the Ekman (δ) and the Hartmann (δ/S) layers.

The parameters of Ref. 5 yield $S=6>1$, so smaller values are reasonable in (8). Hence we find the exceedingly large value $MB_* \geqslant 2\times 10^2$ G and energy ratio $M^2 B_*^2/\rho\mu_0 V_*^2 \geqslant 2\times 10^3$. This result is in excellent agreement with Ref. 5. We therefore conclude that this model can reproduce the geomagnetic field, but only for $S \leqslant 1$. Glatzmaier and Roberts[5] were able to reproduce the geomagnetic field satisfactorily in their subsequent work[6] by varying the thermal flux at the mantle–core boundary, which is equivalent to requiring that S decrease when β does. However, as can be seen from Eq. (8), S can also be decreased by substantially reducing ν and k, which in Refs. 5 and 6 are considerably larger than quantities typical of the Earth's core.

If we measure distances in units of L, velocity in $V_* = R^{2/3}k/L$, time in L/V_*, buoyancy in $A_* = \Omega V_*$, and magnetic field in $B_* = \sqrt{\mu_0\rho\Omega V_* L\delta}$, Eqs. (1) go over to the optimum system, in which all variables are of order unity (cf. Fig. 1):

$$(\epsilon D/Dt + 2\mathbf{e}_z\times - \delta^2\nabla^2)\mathbf{V} + \nabla P - \delta(\nabla\times\mathbf{B})\times\mathbf{B} = A\mathbf{r},$$

$$\frac{D\mathbf{B}}{Dt} - \frac{\delta}{S}\nabla^2\mathbf{B} = (\mathbf{B}\cdot\nabla)\mathbf{V}, \quad \varepsilon\frac{DA}{Dt} - \varepsilon^3\nabla^2 A = \frac{V_r}{r^2}, \qquad (9)$$

where

$$\varepsilon \equiv R^{-1/3} \ll R_c^{-1/3} \approx (\delta^{-2/3})^{-1/3} \ll 10^{-1}, \quad S \gg \delta.$$

Here the dimensionless measure of the diffusion, ε^3, is small for supercritical convection, the smallness of δ^2 corresponds to ultrafast rotation, and the condition on S derived from (8) is necessary for exciting a magnetic field.

Since all variables in (9) are of order unity, we can readily estimate the turbulent diffusion that results from averaging (indicated by writing a bar above the symbol) the nonlinear term $\mathbf{V}\cdot\nabla A$:

$$\overline{\mathbf{V}\cdot\nabla A} \approx \overline{\mathbf{V}\cdot\nabla\int(\mathbf{V}\cdot\nabla A)dt} \approx \overline{\int(\mathbf{V}\otimes\mathbf{V})dt}\,\nabla^2\overline{A} \approx \nabla^2\overline{A}.$$

Similarly, we estimate the turbulent viscosity by taking into account the extreme smallness of $\epsilon \ll 10^{-3}$ in the MHD equations (9):

$$\overline{(\nabla\times\mathbf{B})\times\mathbf{B}} \approx \overline{\left(\nabla\times\int\nabla\times(\mathbf{V}\times\mathbf{B})dt\right)\times\mathbf{B}} \approx \nabla\times(\nabla\times\overline{\mathbf{V}}).$$

Comparing the averaged nonlinear terms with the linear terms in the averaged Eq. (9), for $S\approx 1$ we find the turbulent quantities (marked with a bar)

$$\overline{\delta^2} \approx \delta, \quad \overline{\varepsilon} \approx \sqrt[3]{\varepsilon} = R^{-1/9}, \quad \overline{q} = \overline{\varepsilon^2}/\overline{\delta}. \qquad (10)$$

In the Earth's core the Rayleigh number is large enough, $\overline{R} = (\overline{\varepsilon})^{-3} \approx 10^6$, and the Ekman number is small enough, $\overline{\delta^2} \approx 10^{-6}$, for the above results to be used in the boundary layers. The corresponding value of the average scale velocity $\overline{V}_* \simeq 10^{-3}$ m s^{-1} from (5) is in excellent agreement with the results of the most recent seismographic studies of the Earth's core.[12] The average of the typical value of the magnetic field obtained from (8), $\overline{B}_* \simeq 1$ mT$=10$ G, also agrees well with the results of extended observations of the magnetic field at the core–mantle boundary.[13] Our theory also does a good job reproducing the magnetic fields of the planets (for specifics see the Conclusion).

The details of the dynamics are determined by the averaged system derived below from (9), the solution of which can be compared directly with the observed fields of the planets. Equations (9) comprise a fourteenth-order system, since they satisfy fourteen boundary conditions. By virtue of the smallness of $\overline{\delta} \simeq 10^{-2}$–$10^{-4}$ and $\overline{\varepsilon} \simeq 10^{-1}$–$10^{-3}$, the order is reduced to four in the main part of the layer, where the equations are easily solved analytically. The original order is recovered in the boundary layers: there the analytical solution (7) holds for **V** and **B** and the simplified equation (4) for *A*.

Thus, the difficult problem of solving the three-dimensional set (1) of seven equations is replaced by the simple one of solving a single almost two-dimensional equation (4). Our theory also allows turbulent solutions of (9) to be obtained directly for $\delta \ll 10^{-3}$ and $\varepsilon \ll 10^{-2}$. An additional difficulty in studying supercritical magnetoturbulence directly is that of resolving the structure of the hydromagnetic layers, described in Ref. 14.

3. FIELDS ON LARGE AND SMALL SCALES

The large-scale fields vary over distances considerably greater than the thickness of the boundary and internal (see Sec. 4) layers. Thus, in Ref. 5 the dimensionless thickness of these layers is of order $\delta = 1.3\times 10^{-3}$ and $\varepsilon = 5\times 10^{-3}$. Even the most powerful existing computers are not capable of resolving the structure of such thin layers, and so Glatzmaier and Roberts[5,6] used hyperviscosity to avoid having to resolve small scales, as is typically done in all present-day numerical simulations.[6–8] This hyperviscosity is in outstanding agreement with our estimate (10) for the magnitude of the turbulent coefficients. For values of the parameters taken from Ref. 5 we find from (10) $\overline{\delta} \approx \delta^{1/2} = 1/28$ and $\overline{\varepsilon} \approx \varepsilon^{1/3} = 1/6$, in agreement with the thickness of the boundary layers in Fig. 3 of Ref. 5. However, the estimate (10), like the use of hyperviscosity, is a crude and rather artificial approximation. The required correction to the turbulent coefficients or the mag-

nitude of the hyperviscosity should be based on the influence of small-scale effects on large-scale effects and vice versa, described below.

To be specific, let us consider the magnetostrophic approximation $\epsilon=0$ and the case $S\leqslant 1$, which is typically valid for planetary cores. Substituting for the variables **V**, **B**, P, A the quantities $\mathbf{V}+\mathbf{v}$, $\mathbf{B}+\mathbf{b}$, $P+p$, and $A+a$ and performing averages over large scales (denoted by a bar over the symbol) we obtain a system of equations for the large-scale fields **V**, **B**, P, and A:

$$A\mathbf{r}=2\mathbf{e}_z\times\mathbf{V}-\delta^2\nabla^2\mathbf{V}+\nabla P+\delta(\mathbf{B}\times(\nabla\times\mathbf{B})+\mathscr{L}),$$

$$\frac{\partial\mathbf{B}}{\partial t}-\frac{\delta}{S}\nabla^2\mathbf{B}=\nabla\times(\mathbf{V}\times\mathbf{B}+\mathscr{E}),$$

$$\frac{V_r}{r^2}=\varepsilon\frac{\partial A}{\partial t}+\varepsilon(\mathbf{V}\cdot\nabla A+\mathscr{A})-\varepsilon^3\nabla^2 A, \tag{11}$$

$$\mathscr{L}=\overline{\mathbf{b}\times(\nabla\times\mathbf{b})},\quad \mathscr{E}=\overline{\mathbf{v}\times\mathbf{b}},\quad \mathscr{A}=\overline{\mathbf{v}\cdot\nabla a}.$$

Here the influence of the small scales on the large ones is taken into account through the averaged (turbulent) Lorentz force $\mathscr{L}$, electromotive force $\mathscr{E}$, and buoyant force $\mathscr{A}$. It should be noted that of these three, only the electromotive force $\mathscr{E}=(\alpha+\beta\nabla\times)\mathbf{B}$ has received special attention in previous work.[8–10]

Neglecting the small quantities δ and $\delta\mathscr{L}$, we find **V** from the incompressibility condition $\nabla\cdot\mathbf{V}=0$ and the angular momentum equation:

$$V_r=\frac{z}{r}V_z-\frac{1}{2r}\frac{\partial P}{\partial\varphi},\quad V_\theta=-\frac{z}{2rs}\frac{\partial P}{\partial\varphi}-\frac{s}{r}V_z,$$

$$V_\varphi=\frac{1}{2}\frac{\partial P}{\partial s}-\frac{s}{2}A. \tag{12}$$

Here

$$P=\int^z Az\,dz,\quad 2V_z=\int^z(\partial A/\partial\varphi)\,dz,$$

and the lower limits of integration include unspecified functions of (t,s,φ), which are needed in order to satisfy the boundary conditions.

Neglecting ε and $\varepsilon\mathscr{A}$ we find $V_r=0$, which implies axisymmetry, $\partial(A,P)/\partial\varphi=0$, in accordance with (12). Thus, the large-scale velocity is axisymmetric, has only an azimuthal component $\mathbf{V}=V(t,r,\theta)\hat{\mathbf{e}}_\varphi$, and is uniquely determined by the buoyancy A. The axisymmetric velocity V, the buoyancy A, and in the general case the nonaxisymmetric magnetic field **B** obey a simple set of large-scale equations:

$$V=\int^z\frac{\partial A}{\partial s}\frac{z\,dz}{2}-\frac{s}{2}A,\quad \frac{\partial A}{\partial t}=-\mathscr{A},$$

$$\frac{\partial\mathbf{B}}{\partial t}=\nabla\times(\mathbf{V}\times\mathbf{B}+\mathscr{E}). \tag{13}$$

Clearly, the dynamics of the large-scale fields is significantly affected by the small-scale fields **v**, **b**, p, and a, which obey a set of equations that follows from (9) and (11):

$$a\mathbf{r}=2\mathbf{e}_z\times\mathbf{v}-\delta^2\nabla^2\mathbf{v}+\nabla p+\delta[\mathbf{b}\times\nabla\times\mathbf{B}+\mathbf{B}\times\nabla\times\mathbf{b}+(\mathbf{b}\times\nabla\times\mathbf{b}-\mathscr{L})],$$

$$\frac{\partial\mathbf{b}}{\partial t}-\frac{\delta}{S}\nabla^2\mathbf{b}=\nabla\times[\mathbf{V}\times\mathbf{b}+\mathbf{v}\times\mathbf{B}+(\mathbf{v}\times\mathbf{b}-\mathscr{E})], \tag{14}$$

$$\frac{v_r}{r^2}=\varepsilon\frac{\partial a}{\partial t}+\varepsilon[\mathbf{V}\nabla a+\mathbf{v}\nabla A+(\mathbf{v}\nabla a-\mathscr{A})]-\varepsilon^3\nabla^2 a.$$

The complete set of equations (13) and (14) is closed by the boundary conditions (3) and the conditions that follow directly from (3) if we replace **V**, **B**, P, and A by $\mathbf{V}+\mathbf{v}$, $\mathbf{B}+\mathbf{b}$, $P+p$, and $A+a$. Additionally, the small-scale fields must satisfy solid-wall boundary conditions (see next Section), and the small-scale fields in the boundary layer should fall off sharply away from the boundary.

In the linearized layer the small-scale Ekman–Hartmann boundary field at a solid wall is completely given in terms of the large-scale and interior small-scale fields:

$$v_\varphi-iv_\theta=\mathscr{V}+(\mathscr{V}_{1,c}-\mathscr{V})\exp\left(\pm\sqrt{2i\cos\theta+SB_r^2}\,\frac{\Delta r}{\delta}\right),$$

$$b_\varphi-ib_\theta=\mathscr{B}_{1,c}\pm SB_r(\mathscr{V}_{1,c}-\mathscr{V})\times\frac{1-\exp(\pm\sqrt{2i\cos\theta+SB_r^2}\,\Delta r/\delta)}{\sqrt{2i\cos\theta+SB_r^2}}, \tag{15}$$

$$\{v,b\}_r=-\frac{1}{rs}\int_{\Delta r=0}^{r}\left[\frac{\partial(\sin\theta\{v,b\}_\theta)}{\partial\theta}+\frac{\partial\{v,b\}_\varphi}{\partial\varphi}\right]r\,dr.$$

Here we have written $\mathscr{V}=V+v_\varphi-iv_\theta$ at the upper boundary of the layer, where $v_{\varphi,\theta}$ is the (exclusively) internal field; the complex velocity at the other boundary is $\mathscr{V}_{1,c}\equiv V_{\varphi 1,c}-iV_{\theta 1,c}$; the upper sign $(+)$ and $\Delta r=r-1$ apply to the boundary at the mantle, which rotates at velocity $\mathbf{V}_1$; the lower sign $(-)$ and $\Delta r=r-c$ apply to the boundary at the inner core, which rotates with velocity $\mathbf{V}_c$ [see the expressions for $\mathbf{V}_1$ and $\mathbf{V}_c$ following Eq. (3)]. The radial velocity v_r and the magnetic field b_r were obtained by integrating the divergence-free condition.

The small-scale buoyancy a in the Archimedean boundary layer is found from the equation

$$\frac{v_r}{\varepsilon r^2}+\frac{\partial^2 a}{\partial x^2}-\frac{\partial a}{\partial t}=\frac{v_r}{\varepsilon}\frac{\partial a}{\partial x}+\frac{v_\theta}{r}\frac{\partial(A+a)}{\partial\theta}+\frac{V+v_\varphi}{s}\frac{\partial a}{\partial\varphi}-\mathscr{A}, \tag{16}$$

where we have used the same values of v_r v_θ, and v_φ as in Eq. (15), and the ''stretched'' variable is $x=(r-1)/\varepsilon$ at $r=1$ and $x=(r-c)/\varepsilon$ at $r=c$.

The system of equations (13) corresponding to large-scale effects, the boundary solution (15), and Eq. (16) together with the equations given below for the internal small-scale fields completely determine the solution of the problem. In obtaining it we use only the boundary conditions (3) without the solid-wall condition, since the the other boundary conditions are already taken into account in (15).

4. INTERNAL AXIAL CONVECTION

In treating the internal small-scale fields we will assume that the convection cells are elongated parallel to the axis of rotation and

$$\partial s \sim s_* \ll \partial \varphi \sim \varphi_* \ll \partial z \sim 1.$$

Here the asterisks denote characteristic magnitudes for the internal convection relative to the directions in which the cylindrical variables (z,s,φ) increase. We also assume the following scaling, which comes from condition (14) for the vanishing of the divergence:

$$V \sim 1 \geqslant v_z \sim v_* \gg v_\varphi \sim \varphi_* v_* \gg v_s \sim s_* v_* .$$

And finally, we neglect the magnetic terms in the first force equation (14). Later, when we have derived our final result, we will show that these initial assumptions are justified.

Neglecting the corresponding terms of (14), far from the boundaries we find

$$\frac{v_z \cos\theta}{\varepsilon r^2} + \varepsilon^2 \frac{\partial^2 a}{\partial s^2} = \frac{V}{s}\frac{\partial a}{\partial \varphi} + \frac{\partial (a-A)}{\partial t}, \quad za = -\delta^2 \frac{\partial^2 v_z}{\partial s^2}. \tag{17}$$

Setting the magnitudes of the terms of (17) equal, we find the relations

$$v_*/\varepsilon = \varepsilon^2 a_*/s_*^2 = a_*/\varphi_*, \quad a_* = \delta^2 v_*/s_*^2,$$

which yield the characteristic axial velocity $v_* \sim v_z$ and buoyancy $a_* \sim a$, as well as the scaling with respect to the cylindrical radius s_* and azimuthal angle φ_*:

$$\chi \equiv v_*/a_* = \varepsilon^{3/2}/\delta, \quad s_* = \varepsilon^{3/4}\delta^{1/2}, \quad \varphi_* = \delta/\varepsilon^{1/2}. \tag{18}$$

Balancing terms similarly in the equation for the electromagnetic induction

$$-(\delta/S)\partial^2 b_z/\partial s^2 \simeq B_s \partial v_z/\partial s$$

yields $b_* = S v_* s_*/\delta$ for typical B_s of order unity.

In the Ekman–Hartmann boundary layers (15) we have $\partial z \sim \delta$, and far from the equator $(s \neq c,1)$ balancing the buoyancy equation

$$\varepsilon^2 \partial^2 a/\partial z^2 \simeq v_z \partial a/\partial z$$

yields $\varepsilon^2/\delta = v_*$. Finally, we find estimates of the characteristic internal velocity v_*, magnetic field b_*, and electromotive force $\alpha_* = v_* b_*$, and the fundamental restriction on the initial choice of parameters:

$$v_* = \varepsilon^2/\delta, \quad b_* = S\varepsilon^{11/4}/\delta^{3/2},$$
$$\alpha_* = S\varepsilon^{19/4}/\delta^{5/2}, \quad S\varepsilon^{3/2} \ll \delta. \tag{19}$$

This estimate, together with (18), is used in the Conclusion to describe the magnetohydrodynamic interiors of planets. Our original assumptions are valid if $\varepsilon^2 \leqslant \delta$ and $S\varepsilon^{3/2} \ll \delta$ hold, as they do in most planetary interiors. The second condition usually overlaps the first; it is therefore the main limitation on the use of turbulent [Eq. (10)] or hyperviscosity values $\bar{\varepsilon}$ and $\bar{\delta}$ instead of the smaller ε and δ. In the work of Glatzmaier and Roberts[5] one of the conditions in (19) is not met, so the failure of their results to reflect the symmetry of the geomagnetic field imposed by the internal convection should come as no surprise (see Conclusion).

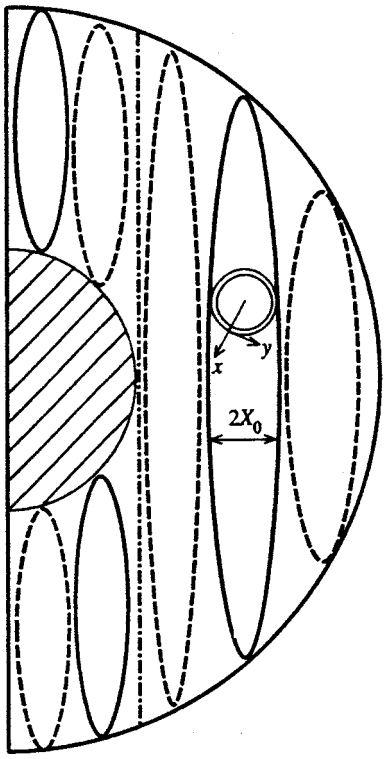

FIG. 2. Structure of convection cells in the $\varphi = \text{const}$ cross section. Cells with current flowing clockwise and counterclockwise are represented by solid and broken traces, respectively. The local x coordinate is transverse to the cell and the local y coordinate is along the circumference. The dot-dash curve indicates the generating tangent cylinder. A transition zone aligned with the cylinder (not shown) and the inner sphere separate the various convection zones.

We will describe the spatial structure of an individual convection cell, which far from the boundaries is elongated parallel to the axis of of rotation z (see Fig. 2). In the cross section $z = \text{const}$ this cell is bounded by the curve $X(s,\varphi) = \text{const}$, which is adjacent to neighboring curves, somewhat as in a honeycomb. From (18), the variable which is ''stretched'' along the cell is $x = X(s,\varphi)/s_*$. We complete our local coordinate system by introducing $y = y(s,\varphi)$, which encircles the cell and is orthogonal to x and z.

We use the fact that the components b_x and v_x are of order $s_* \equiv \varepsilon^{3/4}\delta^{1/2}$, which follows from the vanishing of the divergence, and satisfy $\partial(v_x, b_x)/\partial x \simeq 0$. Using the smallness of p, b, and v_x in the z and y components of the force law, we find from (14) and (17)

$$-\frac{1}{\chi}\frac{\partial^2 v_z}{\partial x^2} = r\cos\theta a, \quad -\frac{1}{\chi}\frac{\partial^2 v_y}{\partial x^2} = y_r r a,$$
$$-\frac{\partial^2 a}{\partial x^2} = \frac{v_z \cos\theta + y_r v_y}{\chi r^2},$$
$$\{v_z, v_y\} = -\chi r\{\cos\theta, y_r\}[C(e^{kx} - \cos(kx)) + D\sin(kx)], \tag{20}$$
$$a = k^2[C(e^{kx} + \cos(kx)) - D\sin(kx)].$$

Here $x \leqslant 0$, $|x| \sim 1$, $y_r \equiv \hat{\mathbf{e}}_y \cdot \hat{\mathbf{e}}_r$; C and $D(t,z,y)$ are free functions that are determined in the boundary layers; and $k=[\chi r(y_r^2+\cos^2\theta)]^{1/4}$ is a root of the characteristic equation. The boundary condition is the condition of intermittency: $v_z, v_y = 0$ at $x=0$.

The radius of the cell is determined by the distance

$$X_0 = x_0 s_* = \frac{\pi}{k} s_* = \frac{\pi s_*}{[\chi r(y_r^2+\cos^2\theta)]^{1/4}}$$

between zeros of the sine function in (20). Hence in the $\varphi = \text{const}$ cross section through the cells we have

$$N_0 \simeq \frac{1}{2X_0} \simeq \frac{1}{10(\varepsilon\delta^2)^{3/8}},$$

which is typically large. Thus, we have $N_0 \simeq 10^2$ for the values $\varepsilon \simeq 10^{-2}$ and $\delta \simeq 10^{-3}$ of Ref. 5. With so many cells their stucture can not possibly be resolved numerically. This is why Glatzmaier and Roberts[5] used hyperviscosity, which increases the value of δ and ϵ by an order of magnitude at small scales. Consequently there were only seven cells, which agrees perfectly with Fig. 3 of Ref. 5.

5. CONCLUSION

Let us compare our results with the observed planetary magnetic fields[11] in order to check our theory and describe convection in planetary cores. We direct our attention primarily at the magnitudes of the intrinsic magnetic fields of Earth, Saturn, Uranus, and Neptune, which are very close even though their interior structures are fundamentally different.[2,15–17] This happens because the flows in the cores of these planets are highly turbulent, since only then can the basic parameters $\overline{S} \leqslant 1$, $\overline{\delta} \approx \sqrt{\delta} \approx 10^{-3}$, $\overline{\varepsilon} \approx \sqrt[3]{\varepsilon} \approx 10^{-2}$ be close to one another; they differ little even when the molecular constituents vary greatly. The stronger field of Jupiter is explained by noting that its liquid metal core is relatively much larger than in the others.[17] Consequently, the observed field of Jupiter differs little from the field (8) at the core–mantle boundary.

The symmetry of the observed large-scale field results from the relationship between the internal small-scale turbulence and large-scale azimuthal flow, which from Eqs. (8) and (19) (cf. Refs. 8, 9, and 15) are characterized respectively by the following Reynolds numbers:

$$R_\alpha \simeq \overline{\varepsilon}^{19/4} R_\Omega / \overline{\delta}^{5/2} (\approx 10) \quad \text{and} \quad R_\Omega \simeq \overline{S}/\overline{\delta} (\approx 10^3). \qquad (21)$$

If the rough estimates given in parentheses are satisfied, then R_α^2 is a factor of ten smaller than R_Ω for these planets. Under this condition axisymmetric field strengths exceed nonaxisymmetric ones by almost an order of magnitude. This is seen on Earth and Jupiter from the small angle (about ten degrees) between the angle of rotation and the axis of the magnetic dipole. The somewhat smaller size of Saturn's core compared with that of Jupiter makes $\overline{\delta}$ almost one and a half times larger (according to Eq. (2) in Ref. 17). Consequently, other factors being equal, R_Ω in Saturn's core exceeds R_α^2 by almost two orders of magnitude. This marked superiority of the axisymmetric dynamo mechanism over the asymmetric mechanism is responsible for the very small observed inclination of Saturn's magnetic dipole.

The cores of Uranus and Neptune should be somewhat less active than that of Jupiter. This raises the value of $\overline{\varepsilon}$ and therefore, from (2), R_α^2 in Eq. (21) becomes comparable with R_Ω. As a result the asymmetric and axisymmetric fields are comparable, which is reflected in an angle of inclination of the magnetic dipole of nearly 50 degrees for both planets.[15]

One should expect even less activity in the cores of Mercury and Mars,[17] where from Eq. (8) $\overline{S} \ll 1$. As a result, the dynamo number $R_\alpha R_\Omega$ is small; probably it is barely bigger than the critical value for generating large-scale magnetic fields in the planetary core. This means that only very small irregular large-scale fields appear.

The activity of the core of Venus is of order $1/\varepsilon^3$ and its effective viscosity is of order δ^2, according to (2). This is two orders of magnitude higher than in Earth's core; the reason is that Venus rotates 243 times slower. This high activity and high viscosity, according to (19) and (21) almost totally suppress internal small-scale turbulence. As a result, the dynamo number in the core of Venus is far less than the critical value, and no large-scale field is generated there. Thus, raising the activity of a planet can lower its observed field. This is easy to see if one compares the fields of Uranus and Neptune and their activities.

We have shown that our theory is able to relate the observed magnetic fields of the planets to well known properties of their structure and dynamics. The more detailed behavior of the field is determined by the optimum set of equations (9), which simplifies to (13)–(17) if we distinguish between large- and small-scale fields. Numerical solution of these sets of equations together with (10) and (19) will be far cheaper and more realistic than using the widely accepted direct approaches employing hyperviscosity.

In conclusion we summarize our main results.

1) The typical magnitudes of quantities determining the nature and principal structures of supercritical MHD convection associated with ultrafast rotation have been estimated for the first time.

2) For the first time an approach has been developed that in principle permits a solution of the planetary MHD dynamo problem and the associated problem of supercritical convection.

3) A simple set of equations has been derived which describes in detail buoyant magnetoconvection in planetary interiors on both large and small scales. For the first time it has been shown that axisymmetric azimuthal flows dominate at large scales.

This work was performed with financial support from the Russian Fund for Fundamental Research (Grant Nos. 970564402 and 960564048).

*)E-mail: starche@borok.adm.yar.ru

[1] S. I. Braginskiĭ, Dokl. Akad. Nauk **149**, 8 (1963).
[2] F. H. Busse, Chaos **4** (2), 123 (1994).
[3] J. E. Hart, G. A. Glatzmaier, and J. Toomre, J. Fluid Mech. **173**, 519 (1986).
[4] G. A. Glatzmaier and P. Olson, Geophys. Astrophys. Fluid Dyn. **70**, 113 (1993).
[5] G. A. Glatzmaier and P. H. Roberts, Phys. Earth Planet. Inter. **91**, 63 (1995).
[6] G. A. Glatzmaier and P. H. Roberts, Contemp. Phys. **38**, No. 4, 269 (1997).
[7] A. Tilgner and F. H. Busse, J. Fluid Mech. **332**, 359 (1997).
[8] R. Hollerbach and C. A. Jones, Nature (London) **365**, 541 (1993).
[9] S. V. Starchenko, Geophys. Astrophys. Fluid Dyn. **77**, 55 (1994).
[10] S. V. Starchenko and M. Kono, Geophys. Astrophys. Fluid Dyn. **82**, 93 (1996).
[11] L. V. Ksanfomaliti, Astronom. Vestnik **32**, 37 (1998).
[12] X. Song and P. G. Richards, Nature (London) **382**, 221 (1996).
[13] J. Bloxham and A. Jackson, J. Geophys. Res. **97**, 19537 (1992).
[14] S. V. Starchenko, Zh. Éksp. Teor. Fiz. **112**, 2056 (1997) [JETP **85**, 1125 (1997)].
[15] A. A. Ruzmaikin and S. V. Starchenko, Icarus **93**, 82 (1991).
[16] D. E. Loper, Geophys. J. R. Astron. Soc. **54**, 389 (1978).
[17] V. N. Zharkov and V. P. Trubitsyn, *Physics of Planetary Interiors*, Pachart Publ. House, Tucson (1978) [Russian original, Nauka, Moscow (1980)].

Translated by David L. Book

Thermal diffusion in disperse systems

K. I. Morozov*)

Institute of Mechanics of Continuous Media, Ural Branch of the Russian Academy of Sciences, 614013 Perm, Russia

(Submitted 12 November 1998)
Zh. Éksp. Teor. Fiz. **115**, 1721–1726 (May 1999)

This paper discusses a theory for a new effect, the migration of solid dispersed particles initiated by a nonuniform temperature field. The reason for the motion is the inhomogeneity of the properties of a thin protective layer around a particle. The example of ionic dispersion shows that the sign of the coefficient of thermodiffusion depends on the magnitude of the electrostatic potential at the particle surface and the thickness of the Debye layer and that the coefficientis larger than the values known for molecular systems by a factor of 100 to 10000. In contrast to molecular systems, in disperse systems thermodiffusion should play a much more important role.
 [S1063-7761(99)01405-5]

The phenomenon of separation of substances in a binary mixture initiated by a nonuniform temperature field has bee known for more than a hundred years and is called thermodiffusion, or the Soret effect.[1] Thermal diffusion is observed in an enormous number of molecular systems, such as mixtures of gases, liquids, and saline solutions. The Soret coefficient S, which is a parameter characterizing the separation of the substances in the mixture (see below), is very small for these systems, which makes the use of thermodiffusion in applications highly problematical.[2] The Soret effect in disperse systems, such as colloidal solutions and suspensions, has significantly more possibilities for practical applications. Such an effect was detected in experiments recently conducted by Blums *et al.*[3] and Lenglet *et al.*,[4] who observed strong migration of solid colloidal particles in a nonuniform temperature field. The Soret effect of such thermodiffusion exceeds the record-breaking values of S for molecular systems by a factor of 100 to 1000. However, the nature of this phenomenon is unknown. In the present paper a theory of this novel phenomenon is studied.

The main difference between disperse and molecular systems is that the solid particles in disperse systems are about 100 Å in size (for colloids), which is much larger than the molecules of a liquid solvent. This simplifies the theoretical investigation of the problem, since the liquid can be regarded as a continuous medium. Note also that it is common to divide disperse systems into two large categories, ionic and surfactant, which differ in stability against particle coagulation.[5] The first category is characterized by electrostatic stabilization, which is achieved by imparting an electric charge to the particles. The second category incorporates disperse systems whose particles are covered by an additional layer of surface-active substances. In the present paper we will limit ourselves to ionic systems.

Suppose that a positively charged spherical particle of radius R is in an electrolyte solution. For simplicity we assume the electrolyte to be symmetric. Let the cation and anion charges be e and $-e$, respectively and the cation and anion concentration far from the particle be n. At a constant temperature T_0, the ions obey a Boltzmann distribution about the seed:

$$n_+^0 = n\exp(-\Psi_0), \quad n_-^0 = n\exp(\Psi_0), \tag{1}$$

where $\Psi_0 = e\varphi_0/kT_0$, with φ_0 the potential of the electrostatic field and k the Boltzmann constant. The dimensionless potential Ψ_0 satisfies Poisson's equation

$$\Delta\Psi_0 = \sinh\Psi_0/D^2, \tag{2}$$

where $D=(\varepsilon kT_0/8\pi ne^2)^{1/2}$ is the Debye length, with ε the dielectric constant of the liquid. The potential at the particle surface, $\zeta=\Psi_0(R)$, is assumed to be a parameter of the problem.

Note that the label ''0'' in Eqs. (1) and (2) designates a state in thermodynamic equilibrium, where the temperature of the system is constant. In this case the electric field $\mathbf{E}_0 = -\mathrm{grad}\varphi_0$ around the seed is spherically symmetric, so that the net force acting on the particle is zero. Similarly, the electric bulk force on the electrolyte is balanced by the pressure gradient in the liquid.

Let us now create a nonuniform temperature field in the liquid, with $\mathbf{A}$ the given temperature gradient far from the particle. The characteristic time of propagation of thermal perturbations, $\tau_T \sim R^2/\chi$, with χ the thermal diffusivity of the liquid, is much shorter than the time of rotational Brownian diffusion of the seed, $\tau_B \sim \eta R^3/kT_0$, where η is the viscosity of the liquid.[6] Indeed, using the data for an aqueous solution, $\eta \sim 1$ cP and $\chi \approx 0.0015\,\mathrm{cm^2 s^{-1}}$, we find $\tau_T/\tau_B \sim 10^{-3}$. Hence we may assume that the temperature distribution $T(\mathbf{r})$ near the particle is time-independent. Its form is well known:[6]

$$T(\mathbf{r}) = T_0 + T_1 = T_0 + [1-\kappa(R/r)^3]\mathbf{A}\cdot\mathbf{r}, \tag{3}$$

where $\kappa=(\kappa_1-\kappa_2)/(\kappa_1+\kappa_2)$, with κ_1 and κ_2 the thermal conductivity coefficients of the particle and electrolyte, respectively.

In a nonuniform temperature field, the ion concentrations $n_+(\mathbf{r})$ and $n_-(\mathbf{r})$ and potential $\varphi(\mathbf{r})$ are no longer equal to the equilibrium values n_+^0, n_-^0, and φ_0. As a result, the

electric potentials at the opposite sides of the particle do not balance each other and the particle begins to move.

To determine the seed velocity **U**, we must first find the ion distributions $n_+(\mathbf{r})$ and $n_-(\mathbf{r})$ in the temperature field. We do this using the example of cations. The diffusion flux of cations in an electrostatic field is

$$\mathbf{j}_+ = -D_+\left(\operatorname{grad} n_+ + \frac{n_+ e \operatorname{grad}\varphi}{kT}\right), \tag{4}$$

where D_+ is the cation diffusion coefficient. The distribution $n_+(\mathbf{r})$ can be found by solving the equation of steady-state diffusion in the approximation of small Péclet numbers with the additional condition that the particle surface is impermeable:[6]

$$\operatorname{div}\mathbf{j}_+ = 0, \quad j_{+n} = 0. \tag{5}$$

The anion concentration was calculated by using formulas similar to (4) and (5) with $\mathbf{j}_+$ replaced by $\mathbf{j}_-$, D_+ by D_-, and e by $-e$.

The next stage in simplifying the equations of ion diffusion is the linearization of these equations in the small values of the temperature gradient, $AR/T_0 \ll 1$. Here, however, we will not perform the necessary calculations, since they are rather cumbersome. One result of such linearization is that it is unnecessary to solve the separate problem of finding the electric field $\varphi(\mathbf{r})$ around the charged seed in a nonuniform temperature field: for $\varphi(\mathbf{r})$ we can take the result $\varphi_0(r)$ from the isothermal problem (2).

We will now investigate the hydrodynamic part of the problem. Suppose that the frame of reference is tied to the body of the seed, so that the particle in this frame is at rest. The liquid velocity **v** far from the seed determines the desired value of the particle velocity **U** with respect to the laboratory reference frame according to the obvious relationship $\mathbf{U} = -\mathbf{v}(r \to \infty)$. In the approximation of small Reynolds numbers the time-independent velocity field $\mathbf{v}(\mathbf{r})$, in turn, satisfies the Stokes equation[6]

$$\eta\Delta\mathbf{v} - \operatorname{grad} p + \mathbf{f} = 0, \quad \operatorname{div}\mathbf{v} = 0, \tag{6}$$

where p and η are the pressure in the liquid and the viscosity of the liquid. The bulk force **f** of electrical origin is given by the formula[7]

$$\mathbf{f} = -e(n_+ - n_-)\operatorname{grad}\varphi - \frac{(\operatorname{grad}\varphi)^2}{8\pi}\frac{\partial\varepsilon}{\partial T}\operatorname{grad} T. \tag{7}$$

Clearly, the value of the bulk force is determined by two terms: the Coulomb term, related to the presence of free charges (ions) in the liquid, and the dielectrophoretic term, reflecting the temperature dependence of the dielectric constant ε of the liquid. Since $\partial\varepsilon/\partial T$ is negative, the dielectrophoretic term forces the liquid to move along the temperature gradient **A** [basically, $\operatorname{grad} T$ is directed along **A**; see Eq. (3)]. This is equivalent to the particle moving in the opposite direction, i.e., into colder layers of the liquid. In contrast to the second term, the projection of the Coulomb term on the direction of **A** has no definite sign, i.e., it can be either positive or negative. As a result the particle can migrate either up or down the temperature gradient (see below).

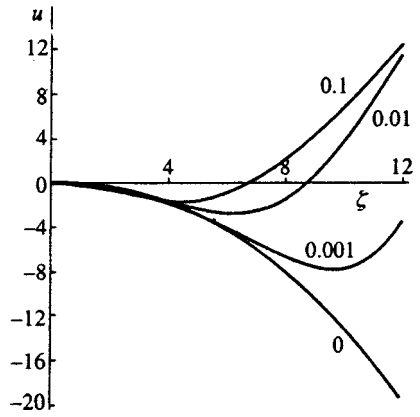

FIG. 1. Particle migration velocity in units of $AnkD^2/\eta$ as a function of the electric potential at the particle surface for several values of λ.

Equation (7) must be linearized in the value of the temperature gradient [just as we linearized the diffusion equation (6)]. However, to preserve clarity we will not clutter up the problem with details.

Equations (6) with the additional conditions of adhesion at the particle surface ($\mathbf{v}(r=R)=0$) and of balance of the electric and viscous forces ($\mathbf{F}=0$) acting on the particle form, together with Eqs. (2) and (4), constitute a complete set of equations for finding the migration velocity of a dispersed particle.

Generally speaking, the solution of this system of equations depends on four dimensionless parameters: the potential ζ at the particle surface, the ratio $\lambda = D/R$ of the Debye length to the particle radius, κ [see the commentary to Eq. (3)], and $\alpha = d\ln\varepsilon/d\ln T$, the logarithmic derivative of the dielectric constant with respect to temperature. We write the velocity of particle motion in the temperature field as $\mathbf{U} = \mathbf{A}nkD^2u/\eta$, specifying the dimensionless migration velocity u explicitly. For an infinitely thin double layer ($\lambda \to 0$), the value of u can be found analytically:

$$u = -\frac{2}{3}(1-\kappa)\left[\zeta^2 - 8(3+\alpha)\ln\cosh\frac{\zeta}{4}\right]. \tag{8}$$

For arbitrary values of λ only a numerical solution of the system of equations is possible. Hence we will fix the values of the parameters $\kappa = 0.75$ and $\alpha = -1.5$. Both values correspond to the experiment conducted by Lenglet *et al.*[4] and are characteristic of aqueous disperse systems, for which the thermal conductivity coefficient of the dispersed phase exceeds that of the solvent by a factor of 10.

Figure 1 depicts the results of numerical calculations of the dimensionless particle velocity as a function of ζ, the surface potential, for several values of λ (the respective values are placed at the curves). For an infinitely thin double layer [see the lower curve in Fig. 1 and Eq. (8)], the particle migration velocity is always directed opposite the temperature gradient. At finite values of λ this is the case only for small values of electric potential. As ζ increases, all the curves with $\lambda \neq 0$ first pass through a minimum and then, at $\zeta = \zeta_0$, through zero. The latter means that the direction of particle motion reverses.

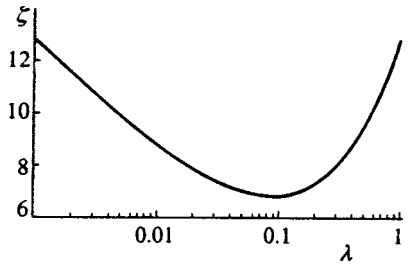

FIG. 2. ''Phase'' diagram of thermodiffusion in terms of dimensionless surface potential and Debye length.

Figure 2 depicts the ''neutral'' curve $\zeta_0=\zeta_0(\lambda)$ corresponding to the disappearance of thermodiffusion. The reason the seed ceases to move on this curve becomes clear if we analyze Fig. 3, which depicts the streamlines and the lines of constant anion concentration near the particle. The dielectrophoretic bulk force [see Eq. (7)] forms the flow chiefly along the temperature gradient **A** and forces the seed to move in the opposite direction (against **A**). However, the Coulomb force opposes this motion, so that near the ''hot'' pole of the seed an excess of negatively charged ions forms. For the values of λ and ζ depicted in Fig. 3, the tendencies are balanced. If $\zeta<\zeta_0$ holds, the first tendency is predominant, i.e., the particle migrates into the low-temperature region; if $\zeta>\zeta_0$ holds, the Coulomb term is predominant and the seed moves into the hot layers of the liquid (see Fig. 2).

Since the velocity of thermodiffusion is known, we can calculate the Soret dispersion coefficient S. Following Refs. 1 and 2, we define this coefficient by the equation

$$\operatorname{grad}\Phi+S\Phi(1-\Phi)\mathbf{A}=0, \tag{9}$$

where $\operatorname{grad}\Phi$ is the time-independent gradient of the bulk concentration of the disperse particles, a gradient thatdevelops in a temperature field with a gradient **A**. Now, setting equal to zero the sum of the thermodiffusion flux of the seeds $\mathbf{j}_T=n_p\mathbf{U}$ (here n_p is the particle concentration) and the ordinary diffusion flux $\mathbf{j}_d=-(kT_0/6\pi\eta R_H)\operatorname{grad}n_p$ (here R_H is the hydrodynamic radius of the particle), we arrive at an expression for the Soret coefficient:

$$S=-\frac{3}{4}\frac{R_H}{L}\frac{u}{(1-\Phi)T_0}, \tag{10}$$

where $L=e^2/\varepsilon kT_0$ is a characteristic length, interpreted as the distance at which the energy of ion interaction is equal to the thermal energy. For aqueous electrolytes, $L\approx 7$ Å.

Equation (10) and Fig. 1 suggest that the Soret coefficient may be either positive or negative, depending on the magnitudes of the electrostatic surface potential ζ and the dimensionless Debye radius λ. Let us estimate S using the experimental data of Lenglet *et al.*,[4] who studied thermodiffusion in an aqueous colloidal solution. If we substitute the experimental values $R_H=200$ Å, $T_0=323$ K, $\Phi=0.06$, $\lambda=0.1$ (Ref. 4), and $\zeta=8.5$ in (10), we get $S=-0.5\,\mathrm{K}^{-1}$. This value is in good agreement with the results of Ref. 4 and exceeds the coefficient of thermodiffusion of molecular systems[2] by a factor of 100. The Soret coefficient of suspensions, whose particles are roughly a hundred times larger than colloidal particles, must amount [in accordance with Eq. (10)] to $S\sim 100\,\mathrm{K}^{-1}$. Such large values of the Soret coefficients of colloids and suspensions make practical applications of thermodiffusion as a source of enrichment of disperse systems highly promising.

The author is grateful to A. F. Pshenichnikov and B. L. Smorodin for useful remarks and for discussions concerning this problem. The work was supported by a grant from the Russian Fund for Fundamental Research (Grant No. 97-03-32119).

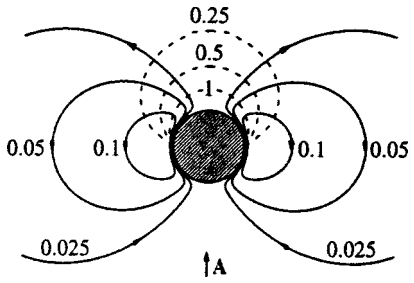

FIG. 3. Stream functions (solid curves) in units of $AnkD^2R^2/\eta$ and contours of constant excess in the concentration of oppositely diffusing ions (dashed curves) in units of AnR/T_0 in the absence of particle migration ($\lambda=0.1$ and $\zeta_0=6.82$).

*)E-mail: mrk@icmm.ru

[1] S. R. deGroot and P. Mazur, *Non-equilibrium Thermodynamics*, North-Holland, Amsterdam (1972).
[2] G. D. Rabinovich, R. Ya. Gurevich, and G. I. Bobrova, *Thermal-Diffusion Separation of Liquid Mixtures* [in Russian], Nauka i Tekhnika, Minsk (1971).
[3] E. Blums, A. Mezulis, M. Maiorov, and G. Kronkalns, J. Magn. Magn. Mater. **167**, 220 (1997).
[4] J. Lenglet, N. M. Heegaard, A. Bourdon *et al.*, submitted to Phys. Rev. E (1998).
[5] B. V. Derjaguin [Deryagin] , *Theory of Stability of Colloids and Thin Films*, Consultants Bureau, New York (1989) [Russian original, Nauka, Moscow (1986)].
[6] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, 2nd English ed., Pergamon, Oxford (1987) [Russian original, Nauka, Moscow (1986)].
[7] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, Pergamon Press, Oxford (1984).

Translated by Eugene Yankovsky

Giant magnetoacoustic effect in $KMnF_3$ due to nuclear spin waves

Kh. G. Bogdanova,*) V. A. Golenishchev-Kutuzov, V. E. Leont'ev, M. R. Nazipov, and M. M. Shakirzyanov

E. K. Zavoĭskiĭ Kazan' Physicotechnical Institute, Russian Academy of Sciences, 420029 Kazan', Russia

M. I. Kurkin

Institute of Metal Physics, Ural Branch of the Russian Academy of Sciences, 620219 Ekaterinburg, Russia

S. V. Petrov

P. L. Kapitsa Institute of Physics Problems, Russian Academy of Sciences, 117334 Moscow, Russia
(Submitted 16 October 1998)
Zh. Éksp. Teor. Fiz. **115**, 1727–1739 (May 1999)

An explanation is proposed for the gigantic magnetoacoustic effect that we observed in $KMnF_3$ in previous work {Kh. G. Bogdanova, V. A. Golenishchev-Kutuzov, M. I. Kurkin *et al.*, Zh. Éksp. Teor. Fiz. **112**, 1830 (1997) [JETP **85**, 1001 (1997)]}. The effect entails a tenfold amplitude reduction of an acoustic pulse in a magnetic field that varies over the range 0–8 kOe. It is shown that this effect is due to the interference of two nuclear magnetoelastic waves propagating in the sample under magnetoacoustic resonance conditions, if this resonance occurs in the region of strong spatial dispersion of nuclear spin waves. The effect is said to be gigantic because it exceeds in magnitude the magnetoacoustic effects observed previously in magnetically ordered materials even though it is due to nuclear magnetism, which is 10^5 times weaker than electronic magnetism. We observe a concomitant anomalous dependence of the dispersion of the velocity of sound on the external magnetic field. © *1999 American Institute of Physics.* [S1063-7761(99)01505-X]

1. INTRODUCTION

Magnetoacoustic effects are ordinarily understood to refer to the dependence of the acoustic parameters (amplitude u, frequency ω, propagation velocity V, and polarization of an acoustic wave) on the magnetic field H . Most investigations are concerned with the effect of the field on the velocity and polarization (rotation of the polarization plane and ellipticity) of sound.[1,2] This effect is ordinarily small even in magnetically ordered materials. The variations reach 100% only in exceptional cases.[3] The change produced in the amplitude u of an acoustic wave by a field H is even less appreciable, because the magnetoelastic interactions are small compared with phonon anharmonicity, which is the principal determinant of acoustic damping in solids. Ordinarily, the behavior of $u(H)$ is studied using acoustic excitation of ESR and NMR in paramagnets.[4]

Hence it is understandable how unexpected it was for us to observe a dependence $u(H)$ corresponding to a tenfold decrease in the amplitude u of an acoustic pulse in $KMnF_3$ single crystals.[5] When it became clear that such a strong magnetoacoustic effect is due to nuclear magnetism, which is 10^5 times weaker than electronic magnetism, we could not resist the temptation to call this effect gigantic. In our previous work[5] it was shown that the observed effect agrees well with the crystalline and magnetic symmetry conditions of the compound $KMnF_3$, so it cannot be due to random factors (contamination, measurement errors, or other factors). In addition, the effect was observed in the acoustic frequency range 630–670 MHz, which falls in the frequency band of nuclear spin waves in $KMnF_3$.[6,7] Since this band $\delta\omega_n$ lies below the AFMR frequencies, the observed dependence $u(H)$ is due to nuclear and not electronic magnetism.

It remained to determine how a weak magnetoelastic interaction such as the coupling of elastic deformations to vibrations of the magnetic moments of nuclei can lead to such large effects in magnetoacoustics. The work reported here attempts to resolve this issue.

2. NATURE OF THE ATTENUATION OF ACOUSTIC TRANSMITTANCE IN $KMnF_3$

The character of the observed dependence $u(H)$ in the form of minima (see Figs. 3 and 4 in Ref. 5) suggests that the observed effectis due to a magnetoelastic resonance of acoustic vibrations with the vibrations of nuclear spins. However, it cannot be related to nuclear magnetic relaxation, since its rate is much lower than the decay rate of sound due to phonon anharmonicity. In this case the contribution of phonon damping to nuclear magnetic relaxation can be large, but not vice versa.

The decisive factor in the search for nondissipative mechanisms of attenuation of the amplitude u of an acoustic pulse passing through a $KMnF_3$ sample was that we were not able to observe a dependence $u(H)$ in another weak ferromagnet $FeBO_3$. It was well known that the character of the vibrations of nuclear spins in $KMnF_3$ is substantially different from that in $FeBO_3$. In $KMnF_3$ at liquid helium tempera-

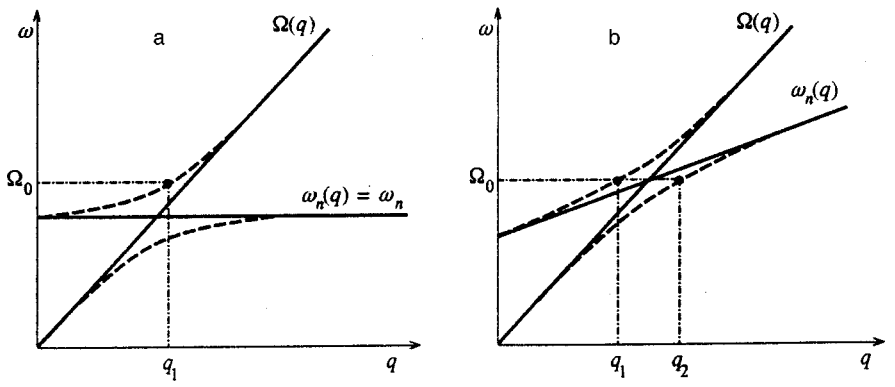

FIG. 1. Spectra of nuclear magnetoelastic waves near nuclear magnetoacoustic resonance: a) no dispersion in the nuclear spin wave spectrum; b) nuclear magnetoacoustic resonance falls in the region of strong dispersion in the nuclear spin wave spectrum.

tures there is reasonably strong interaction between nuclear spins at different sites of the crystal lattice; this gives rise to the formation of nuclear spin waves,[8] analogous to electronic spin waves in magnetically ordered materials. Significantly, in $KMnF_3$ there are two branches of nuclear spin waves with frequencies $\omega_{n1}(q,H)$ and $\omega_{n2}(q,H)$ that depend differently on H. The effect of this is that the nuclear magnetoacoustic resonance conditions, i.e., equality of the frequencies and wave numbers of the acoustic waves $[\Omega(q)]$ and nuclear spin waves $[\omega_n(q)]$,

$$\Omega(q)=\omega_{n1}(q,H_1), \quad \Omega(q)=\omega_{n2}(q,H_2), \tag{1}$$

are satisfied at different values of the field H. The values of H_1 and H_2 correspond to two minima observed in the curves $u(H)$ (see Figs. 3 and 4 in Ref. 5).

According to Ref. 7, the frequencies $\omega_{n1}(q,H)$ and $\omega_{n2}(q,H)$ in $KMnF_3$ lie in a frequency range $\delta\omega_n$ of approximately 100 MHz. The dependence of ω_{n1} and ω_{n2} on q (spatial dispersion) is therefore strong in all processes where nuclear spin waves participate, including nuclear magnetoacoustic resonance.

Estimates obtained for the range $\delta\omega_n$ in $FeBO_3$ using the well-known formulas of Ref. 8 yield a value 10^5 times smaller than in $KMnF_3$, i.e., about 1 kHz. This is due to major differences between the magnetic moments and natural abundances of ^{57}Fe and ^{55}Mn, and between the frequencies ω_n for $FeBO_3$ and $KMnF_3$. For such small $\delta\omega_n$ in $FeBO_3$ (1 kHz is much less than the width of the NMR line), spatial dispersion in the nuclear spin wave spectrum should not lead to any observable effects.

The existence of nuclear spin waves is the reason for the strong dependence $u(H)$ in $KMnF_3$. This can be seen in Figs. 1a and 1b, which show the dispersion curves for acoustic and nuclear spin waves near a nuclear magnetoacoustic resonance in $FeBO_3$ (Fig. 1a), where $\omega_n(q)$ does not depend on q, and for $KMnF_3$, where the branches $\Omega(q)$ and $\omega_n(q)$ cross in the region of strong spatial dispersion of the nuclear spin waves (Fig. 1b). The dashed lines show the dispersion curves for magnetoelastic waves containing elastic and magnetic components. It is evident in this figure that in the absence of nuclear spin waves (Fig. 1a) only one magnetoacoustic wave, possessing a wave vector $\mathbf{q}_1$, propagates in the sample at the given frequency Ω_0. However, if the magnetoacoustic resonance falls in the region of strong q-dependence of ω_n (Fig. 1b), then two magnetoelastic waves with the same frequency Ω_0 and different wave vectors $\mathbf{q}_1$ and $\mathbf{q}_2$ can propagate in the sample. As these waves propagate through the sample, a phase difference $\delta\varphi$ will accumulate between their acoustic components u_1 and u_2. This phase difference can be described in the coordinate system $\mathbf{y} \parallel \mathbf{q}_1 \parallel \mathbf{q}_2$ by the expression

$$\delta\varphi(y)=(q_1-q_2)y. \tag{2}$$

The existence of $\delta\varphi$ has a strong effect on the amplitude u_0 of an acoustic pulse propagating into a nonmagnetic medium ($LiNbO_3$) through the surface $y=\mathscr{L}$ (Fig. 2). The point here is that u depends not only on the amplitudes u_1 and u_2 but also on the phase difference $\delta\varphi(\mathscr{L})$ with which they approach the surface $y=\mathscr{L}$. If u_1 and u_2 are in antiphase, i.e.,

$$\delta\varphi(\mathscr{L})=(q_1-q_2)\mathscr{L}=2(n+1)\pi, \tag{3}$$

(where n are integers), then the amplitude u will be determined by the difference $|u|=|u_1-u_2|$, so that for $u_1=u_2$ the surface $y=\mathscr{L}$ becomes opaque to such magnetoelastic waves.

As will be shown below, $u_1=u_2$ at exact nuclear magnetoacoustic resonance, i.e., a strong decrease in u compared with the amplitude u_0 of the incident acoustic wave is possible only when (1) and (3) hold simultaneously. Although such a coincidence is a random event, it is nonetheless possible because the quantities in Eqs. (1) and (3) depend on the magnitude H and orientation of the magnetic field and on the frequency Ω_0 of the incident acoustic pulse. Variation of just these three parameters makes it possible for the two conditions to be satisfied simultaneously to high precision. However, the search for the required parameter values requires a large number of measurements. For example, the curves of the function

$$K(H)=u(H)/u_0, \tag{4}$$

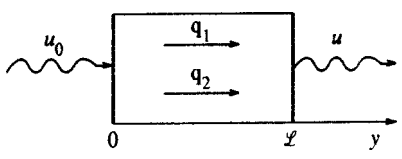

FIG. 2. Transmission of an acoustic pulse through the $KMnF_3$ sample under nuclear magnetoacoustic resonance conditions.

shown in Ref. 5, corresponding to a tenfold decrease in $u(H)$ [at the minimum of $K(H)$], amount to only several percent of the total number of curves measured. In what follows we derive equations that make it possible to obtain quantitative estimates of the parameters of the curves $K(H)$.

3. EQUATIONS FOR MAGNETOELASTIC WAVES NEAR A NUCLEAR MAGNETOACOUSTIC RESONANCE ON NUCLEAR SPIN WAVES

In this section we present numerical results for the magnitude of the indirect interaction of elastic waves with nuclear spin waves via vibrations of the electronic magnetic moments $\mathbf{M}(r_j)$ and the vibrational amplitudes of magnetoelastic waves. To describe the elastic waves we employed the standard equation of elasticity theory[9]

$$\rho \frac{\partial^2 u_\alpha}{\partial t^2} = \frac{\partial \sigma_{\alpha\beta}}{\partial r_\beta}, \tag{5}$$

where u_α $(\alpha = x,y,z)$ are the components of the displacement of points of the elastic medium, ρ is the density of the medium, t is the time, r_α are spatial coordinates, and $\sigma_{\alpha\beta}$ is the stress tensor, which is a variational derivative

$$\sigma_{\alpha\beta} = -\delta F / \delta U_{\alpha\beta} \tag{6}$$

of the free energy F with respect to the deformations $U_{\alpha\beta} = \partial u_\alpha / \partial r_\beta$. For a magnetically ordered elastic medium with a hyperfine interaction, the expression for F can be written in the form[8,10]

$$F = F_E + F_M + F_{ME} + F_{HF}, \tag{7}$$

where F_E is the elastic deformation energy, which determines the velocity of the acoustic waves, F_M is the energy responsible for the properties of the ordered magnetic moments $\mathbf{M}(r_j)$ (it includes the exchange interaction, the magnetic anisotropy, the Dzyaloshinskiĭ interaction responsible for weak ferromagnetism, and interaction with the magnetic field H), F_{ME} is the magnetoelastic interaction energy of the vectors $\mathbf{M}(r_j)$ with the elastic deformations $U_{\alpha\beta}$, and F_{HF} is the hyperfine interaction energy of the nuclear and electronic magnetic moments $\mathbf{m}(r_j)$ and $\mathbf{M}(r_j)$, respectively.

For longitudinally polarized elastic waves with wave vector $\mathbf{q} \parallel \mathbf{y}$ (Fig. 3), the expression for F_{ME}, taking account of the crystal symmetry and magnetic structure[11,12] of $KMnF_3$, can be written in the form[5]

$$F_{ME} = -\int dr \{B_1 L_y^2(r) + B_2[L_x^2(r) + L_z^2(r)] + B_3[M_x(r)L_z(r) + M_z(r)L_x(r)]\} U_{yy}(r), \tag{8}$$

where $U_{yy} = \partial u_y(r)/\partial y$ is the only nonvanishing component of the deformation tensor for the elastic waves under consideration;

$$\mathbf{L} = \mathbf{M}_1 - \mathbf{M}_2, \quad \mathbf{M} = \mathbf{M}_1 + \mathbf{M}_2 \tag{9}$$

are antiferromagnetism and ferromagnetism vectors for a two-sublattice antiferromagnet; $\mathbf{M}_1$ and $\mathbf{M}_2$ are the magnetizations of the sublattices; and, B_1, B_2, and B_3 are the magnetoelastic interaction constants. We note that in Eq. (8) only the first two terms of the expansion in powers of M are retained.

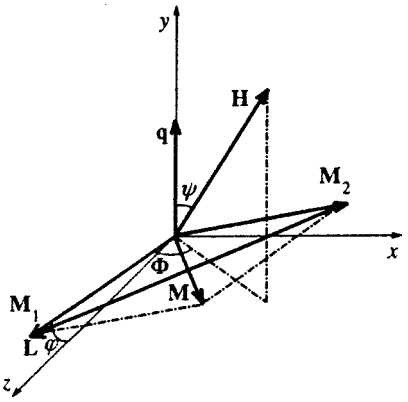

FIG. 3. Orientation of the wave vector **q** of the incident elastic wave, the magnetic field **H**, the sublattice magnetizations $\mathbf{M}_1$ and $\mathbf{M}_2$, and the antiferromagnetism **L** and ferromagnetism **M** vectors relative to the axes of the $KMnF_3$ crystal.

Since in $KMnF_3$ the vector **L** (Fig. 3) lies in the xz plane (the (001) plane), it is sufficient to retain in Eq. (8) only the terms containing B_3, since

$$\frac{d}{dy}(L_x^2 + L_z^2) = \frac{d}{dy} L^2 = 0$$

(up to terms quadratic in M). In this approximation Eq. (5) can bewritten

$$[\omega^2 - \Omega^2(q)] u_y(q,\omega) = iq\rho^{-1} B_3 [M_x(q,\omega)L_z(q,\omega) + M_z(q,\omega)L_x(q,\omega)], \tag{10}$$

where

$$u_y(q,\omega) = \int dy \int dt \exp(i\omega t + iqy) u_y(y,t) \tag{11}$$

is the Fourier transform of the elastic displacements $u(y,t)$ (the quantities $M_{x,z}(q,\omega)$ and $L_{x,z}(q,\omega)$ are defined similarly), and $\Omega(q) = Vq$ and V are the frequency and velocity of longitudinal elastic waves.

The relation (10) can be further simplified by expanding the expressions for M_x and L_x in terms of small vibrations due to their interaction with the elastic displacements $u_y(y,t)$ and vibrations of the nuclear magnetizations $m_{1,2}(y,t)$ of the sublattices. To do so, it is necessary to transform from **L** and **M** (9) to the sublattice magnetizations $\mathbf{M}_1$ and $\mathbf{M}_2$, writing them in coordinate systems (x_1,y_1,z_1) and (x_2,y_2,z_2) associated with the equilibrium orientations of the sublattices $\mathbf{M}_{10}$ and $\mathbf{M}_{20}$ ($\mathbf{z}_1 \parallel \mathbf{M}_{10}$, $\mathbf{z}_2 \parallel \mathbf{M}_{20}$). Then, to a first approximation it is sufficient to take account of only the components M_{x_j} and M_{y_j}, where j is the sublattice index ($j = 1,2$). As a result, Eq. (10) becomes

$$[\omega^2-\Omega^2(q)]u_y(q,\omega)$$
$$=iq\rho^{-1}B_3M_0\ \sin 2\varphi\{\cos\ \psi[M_{y_1}(q,\omega)-M_{y_2}(q,\omega)]$$
$$-2\sin\ \psi[M_{x_1}(q,\omega)+M_{x_2}(q,\omega)]\}, \quad (12)$$

where

$$M_{\alpha_j}(q,\omega)=\int dy\int dt\exp(i\omega t+iqy)M_{\alpha_j}(y,t), \quad (13)$$

$\alpha=x,y$; M_0 is the equilibrium value of **M** (9); ψ is the polar angle of the field **H** (Fig. 3); and φ is the azimuthal angle of the vector **L** in the xz plane. The quantity φ can be found by solving the fairly complicated trigonometric equation (10); it depends on the intensity of the field **H** and its orientation relative to the crystal axes. In the present paper we do not analyze these dependences, since they are immaterial to a description of the gigantic decrease in the acoustic transmittance in $KMnF_3$.

The components M_{x_j} and M_{y_j} in Eq. (12) were found by minimizing the magnetic part of Eq. (7) in the linear approximation in F_{ME} (8), with

$$F_{HF}=-\int dr\,A\sum_{j=1}^{2}M_j(r)m_j(r). \quad (14)$$

The components $M_{\alpha_j}(q,\omega)$ (13), $u_y(q,\omega)$ (11), and

$$m_{\alpha_j}(q,\omega)=\int dy\int dt\exp(i\omega t+iqy)m_{\alpha_j}(y,t) \quad (15)$$

satisfy the relations

$$M_{x_1}(q,\omega)=M_{x_2}(q,\omega)$$
$$=\chi_1(q,H)\{A[m_{x_1}(q,\omega)+m_{x_2}(q,\omega)]-2B_3M_0$$
$$\times\sin 2\varphi\ \sin\ \psi u_y(q,\omega)\}, \quad (16)$$
$$M_{y_1}(q,\omega)=-M_{y_2}(q,\omega)$$
$$=\chi_2(q,H)\{A[m_{y_1}(q,\omega)+m_{y_2}(q,\omega)]+B_3M_0$$
$$\times\sin 2\varphi\ \cos\ \psi u_y(q,\omega)\}, \quad (17)$$

where $\chi_1(q,H)$ and $\chi_2(q,H)$ are the components of the sublattice susceptibilities, which depend on the same quantities as φ.[5] Just as for φ, these dependences were not analyzed in the present work. Nonetheless we singled out the q and H dependences, since for what follows it is important that such dependences exist.

The components m_{α_j} were found from the Bloch equations[10]

$$\partial m_j/\partial t=\gamma_n\mathbf{m}_j\times\mathbf{H}_j, \quad (18)$$

where $H_j=\delta F_{HF}/\delta m_j$ are the effective fields acting on $\mathbf{m}_1$ and $\mathbf{m}_2$ due to $\mathbf{M}_1$ and $\mathbf{M}_2$. In the linear approximation in m_{α_j} these equations are

$$\partial m_{x_j}/\partial t=\gamma_nAM_0m_{y_j}-\gamma_nAm_0M_{y_j},$$
$$\partial m_{y_j}/\partial t=-\gamma_nAM_0m_{x_j}+\gamma_nAm_0M_{x_j}, \quad (19)$$

where M_0 and m_0 are the equilibrium values of $\mathbf{M}_j$ and $\mathbf{m}_j$ and γ_n is the nuclear gyromagnetic ratio. Substituting (16) and (17) into Eqs. (12) and (19) yields the system of equations

$$[\omega^2-\Omega^2(q,H)]u_y(q,\omega)-i\rho^{-1}q\alpha_2(q,H)l(q,\omega)$$
$$-i\rho^{-1}q\alpha_1(q,H)m(q,\omega)=0, \quad (20)$$
$$[\omega^2-\omega_{n1}^2(q,H)]m(q,\omega)+2iq\gamma_nm_0\alpha_1(q,H)u_y(q,\omega)=0, \quad (21)$$
$$[\omega^2-\omega_{n2}^2(q,H)]l(q,\omega)+2iq\gamma_nm_0\alpha_2(q,H)u_y(q,\omega)=0, \quad (22)$$
$$l(q,\omega)=m_{y_1}(q,\omega)-m_{y_2}(q,\omega),$$
$$m(q,\omega)=m_{x_1}(q,\omega)+m_{x_2}(q,\omega), \quad (23)$$

where

$$\Omega(q,H)=q\{V+\rho^{-1}B_3^2M_0^2\ \sin^2 2\varphi[\chi_2(q,H)\ \cos^2\varphi$$
$$+4\chi_1(q,H)\ \sin^2\psi]\}^{1/2} \quad (24)$$

is the frequency of the longitudinal acoustic wave with allowance for the magnetoelastic interaction F_{ME} (8), and

$$\omega_{n1}(q,H)=\gamma_n\{AM_0[AM_0-2A^2m_0\chi_1(q,H)]\}^{1/2}, \quad (25)$$
$$\omega_{n2}(q,H)=\gamma_n\{AM_0[AM_0-2A^2m_0\chi_2(q,H)]\}^{1/2} \quad (26)$$

are the frequencies of the two branches of the nuclear spin wave. The second terms in Eqs. (25) and (26) describe the so-called dynamic shift of the NMR frequency. On account of the factors χ_1 and χ_2 they depend on the magnetic field H. This makes it possible to tune to the nuclear magnetoacoustic resonance frequency (1) by varying the field intensity H. The parameters

$$\alpha_1(q,H)=2AB_3M_0\chi_1(q,H)\ \sin 2\varphi\ \sin\ \psi, \quad (27)$$
$$\alpha_2(q,H)=AB_3M_0\chi_2(q,H)\ \sin 2\varphi\ \cos\ \psi \quad (28)$$

determine the magnitude of the indirect interaction of the longitudinal elastic waves with the first and second branches of the nuclear spin wave.

4. DISCUSSION

Some properties of the function $K(H)$ (4) can be obtained even without solving Eqs. (20)–(22).

1) The nuclear magnetoacoustic resonance condition (1) can be satisfied by varying H while holding the frequency Ω_0 of the incident acoustic pulse constant. This is possible because of the strong field dependence of the frequencies $\omega_{n1}(q,H)$ (25) and $\omega_{n2}(q,H)$ (26), which were investigated experimentally in Ref. 6.

2) As follows from Eq. (28), for $\psi=90°$ the magnetoelastic interaction of sound with the second branch of the nuclear spin wave, which is described by the variable $l(q,\omega)$ (23), vanishes. The satisfaction of this requirement was specially checked in Ref. 5 (see Figs. 3 and 4 in Ref. 5).

3) According to Eqs. (27) and (28) the magnetoelastic coupling $u_y(q,\omega)$ with both branches of the nuclear spin wave should vanish at $\sin 2\varphi=0$. Since the angle φ depends on the azimuthal angle Φ of the field H, the curves $u(\Phi)/u_0$

should contain points where $u(\Phi)/u_0=1$. This conclusion also agrees with the experimental data of Ref. 5 (see Figs. 5 and 6 in Ref. 5).

4) The magnitude of the indirect interaction of the elastic oscillations $u_y(q,\omega)$ with the nuclear spin wave, just like the dynamic shifts of the NMR frequencies (25) and (26), is proportional to m_0 and $\chi_{1,2}(q,H)$. This means that the magnetoelastic effects due to nuclear spins should decrease together with the dynamic shift. The absence of appreciable magneto acoustic effects in $FeBO_3$, where the width of the band of nuclear spin waves (which coincides with the magnitude of the dynamic frequency shift[4]) is 10^5 times smaller than in $KMnF_3$, agrees with this result.

To analyze the other properties of $K(H)$ (4) it is necessary to know the explicit form of the solutions of Eqs. (20)–(22). These will be obtained for the case in which the frequencies of the nuclear spin wave $\omega_{n1}(q,H)$ (25) and $\omega_{n2}(q,H)$ (26) differ substantially in magnitude (nondegenerate case). Then the condition for nuclear magnetoacoustic resonance (1) with each branch of the nuclear spin wave will be satisfied for different values of H, so that Eqs. (20)–(22) can be solved, setting

$$\omega_{n1}^2(q,H)=\Omega_0^2(q,H), \quad l(q,\omega)=0 \tag{29}$$

in the case of nuclear magnetoacoustic resonance with the first branch of the nuclear spin wave, and

$$\omega_{n2}^2(q,H)=\Omega_0^2(q,H), \quad m(q,\omega)=0 \tag{30}$$

for resonance with the second branch.

The characteristic equation for the system (20)–(22) determines the wave vectors $\mathbf{q}_1$ and $\mathbf{q}_2$ for the two branches of magnetoelastic waves in each of the indicated variants (29) and (30) of nuclear magnetoacoustic resonance. The corresponding equations with the conditions (29) are

$$[\Omega_0^2-\Omega^2(q,H)][\Omega_0^2-\omega_{n1}^2(q,H)]-q^2\alpha_1^2(q,H) \times(2\rho^{-1}\gamma_n m_0)=0, \tag{31}$$

$$[\Omega_0^2-\Omega^2(q,H)][\Omega_0^2-\omega_{n2}^2(q,H)]-q^2\alpha_2^2(q,H) \times(2\rho^{-1}\gamma_n m_0)=0. \tag{32}$$

Each of these equations possesses two roots: q_1 and q_2. The analysis of these roots falls outside the scope of the present work, since we did not investigate the properties of the parameters $\alpha_{1,2}(q,H)$ (27), (28). Only one property of these roots will be important below:

$$\Delta q/q_0=|q_1-q_2|/q_0\ll 1, \tag{33}$$

where q_0 is the value of q satisfying the conditions (29) or (30). The inequality (33) follows from the requirement that the second terms in Eqs. (31) and (32) be small compared with $\Omega^2(q,H)$.

Using q_1 and q_2, the expression for the elastic component of the magnetoelastic waves in the sample can be written

$$u_y(y,\omega)=u_1(\omega)\ \exp(iq_1y)+u_2(\omega)\ \exp(iq_2y), \tag{34}$$

where the coefficients u_1 and u_2 must be found from the boundary conditions for Eqs. (5) and (19). For our experiments these conditions must be prescribed in a form corresponding to a Cauchy problem for the sample surface $y=0$ on which the acoustic pulse is incident. Strictly speaking, the initial conditions must also be taken into account. But for the sake of simplicity, we decided to neglect transient processes and to work with the Fourier transforms with respect to time t. For the latter, the boundary conditions at $y=0$ are

$$m(0,\omega)=l(0,\omega)=0,$$

$$u_1(\omega)+u_2(\omega)=u_y(0,\omega)\exp(i\varphi_0), \tag{35}$$

where $m(0,\omega)$ and $l(0,\omega)$ are the values of $m(y,\omega)$ and $l(y,\omega)$ at $y=0$,

$$u_y(0,\omega)=\int dt\, e^{i\omega t}u_y(0,t) \tag{36}$$

is the Fourier transform of the incident pulse, from which the reflected pulse is subtracted, and φ_0 is the initial phase of the wave that has entered the sample. The center frequency $u(0,\omega)$ of the spectrum equals the frequency Ω_0 of the acoustic generator, and the width $\delta\omega$ of this spectrum is determined by the pulse duration τ

$$\delta\omega=2\pi/\tau. \tag{37}$$

Under nuclear magnetoacoustic resonance conditions (29) or (30) we obtain from Eqs. (20)–(22) and (35)

$$u_1(\omega)=u_2(\omega)=\frac{1}{2}u_y(0,\omega)\exp(i\varphi_0). \tag{38}$$

Substituting the expression (38) into Eq. (34) we obtain for $u_y(y,\omega)$ at the second end surface of the sample ($y=\mathscr{L}$)

$$u_y(\mathscr{L},\omega)=u_y(0,\omega)\exp(i\varphi_0)[\exp(iq_1\mathscr{L})+\exp(iq_2\mathscr{L})]/2. \tag{39}$$

The detected acoustic signal is determined by the function $u_y(\mathscr{L},t)$, which is related to $u_y(\mathscr{L},\omega)$ (once again with the exception of the reflected wave) by an inverse Fourier transform

$$u_y(\mathscr{L},t)=\frac{1}{2\pi}\int d\omega\, e^{-i\omega t}u_y(\mathscr{L},\omega). \tag{40}$$

The result of the integration in Eq. (40) depends on how strong the frequency dispersion of the magnetoelastic wave velocities

$$V_1(\omega)=\omega/q_1(\omega), \quad V_2(\omega)=\omega/q_2(\omega) \tag{41}$$

is within the frequency spectrum $\delta\omega$ (37) of the incident pulse. If the ω dependence of V_1 and V_2 can be neglected, then Eq. (40) can be written

$$u_y(\mathscr{L},t)=\frac{1}{2}\{u_y[0,(t-\mathscr{L}/V_1)]+u_y[0,(t-\mathscr{L}/V_2)]\}, \tag{42}$$

where $u(0,t)$ is the amplitude of the elastic displacements excited at the sample surface $y=0$ by the incident pulse [see Eq. (36)], and

$$t_1=\mathscr{L}/V_1, \quad t_2=\mathscr{L}/V_2 \tag{43}$$

are the propagation times of the first and second magnetoelastic waves through the sample.

It follows from Eq. (42) that the incident pulse excites in the sample two magnetoelastic pulses with equal amplitudes and with the same shape as the incident pulse. This fact was already used above (see Sec. 2). However, such an equality occurs only under conditions of nuclear magnetoacoustic resonance (29) and (30) [see also Eq. (1)], where the relation (38) is satisfied. As the distance from resonance increases, the two magnetoelastic waves in question become less and less alike, so that they even have different names: elastic-like and spin-like waves. This inequivalence is manifested, specifically, in the fact that the amplitude of the elastic-like wave excited by the incident acoustic pulse is greater than that of the spin-like wave, and this difference increases with distance from resonance.

We do not present here the corresponding expressions for $u_y(\mathscr{L},\omega)$ for two reasons. In the first place, they are much more complicated than Eq. (42). In the second place, they can be used only if the explicit form of the functions $\chi_1(q,H)$ and $\chi_2(q,H)$ is known [see Eqs. (16) and (17)]. We have yet to do such an analysis, but for the time being we confined ourselves to analyzing only Eq. (42) for two cases: $|t_1-t_2|\ll\tau$ and $|t_1-t_2|>\tau$.

a) In the case

$$|t_1-t_2|\ll\tau \tag{44}$$

the pulses of the magnetoelastic waves excited by the incident acoustic pulse emerge at the sample surface ($y=\mathscr{L}$) virtually simultaneously, so that the difference between the velocities V_1 and V_2 will be manifested only in the phase difference between these oscillations. In this approximation Eq. (42) can be written in the form

$$u_y(\mathscr{L},t)=u_y[0,(t-\mathscr{L}/V_0)]\cos(\Delta q\mathscr{L}/2), \tag{45}$$

where $\Delta q=q_1-q_2$ and $V_0=\omega/q_0=(V_1+V_2)/2$ is the average propagation velocity of magnetoelastic waves. It follows from Eq. (45) that at time $t_0=\mathscr{L}/V_0$ after the exciting pulse reaches the opposite surface of the sample ($y=\mathscr{L}$), an acoustic pulse with the same shape $u_y(0,t)$ as the incident pulse but with a different amplitude because of the factor $\cos(\Delta q\mathscr{L}/2)$ is formed. This factor describes the influence of the interference effects that were discussed in Sec. 2 [its argument $\Delta q\mathscr{L}$ is identical to the expression for the phase difference $\delta\varphi(\mathscr{L})$ (3)].

Of course, Eq. (45) cannot be used to describe the experimental dependences of the ratio $K(H)=u(H)/u_0$ which are presented in Ref. 5. This is because it is valid only at the magnetoelastic resonance frequency (1). Since its position depends on H, Eq. (45) can describe only one point on the curve $K(H)$. But this point can fall in the region where $\Delta q\mathscr{L}\approx\pi$ and therefore $\cos(\Delta q\mathscr{L}/2)\approx 0$. Thus Eq. (45) demonstrates the possibility of achieving a large magnetoacoustic effect by the interference mechanism.

b) The case

$$|t_1-t_2|>\tau \tag{46}$$

is interesting primarily because under this condition the magnetoacoustic pulses excited at the $y=0$ surface of the sample separate completely when they emerge at the surface $y=\mathscr{L}$. After such separation they no longer interfere with one another, and therefore they can be studied separately. To assess the possibility of satisfying the inequality (46) under our experimental conditions we used Eqs. (41) and (43) and the parameters $\tau\approx 10^{-6}$ s, $V\approx 10^5$ cm/s, and $\mathscr{L}\approx 10$ mm. This made it possible to write the inequality (46) as

$$\Delta q/q_0>0.1, \tag{47}$$

where q_0 is determined just as in Eq. (33). Comparing Eqs. (47) and (33) shows that the inequality (47) is consistent with Eq. (33) over a narrow range of Δq. Moreover, the values of Δq satisfying the inequality (47) must be large enough that the ω dependence of V_1 and V_2 (41) starts to come into play. As is well known, such a dependence leads to distortions of the pulse shape and size. We assume that these effects made it impossible for us to consistently observe the splitting of the acoustic pulse transmitted through the $KMnF_3$ sample in the experiments discussed in the next section.

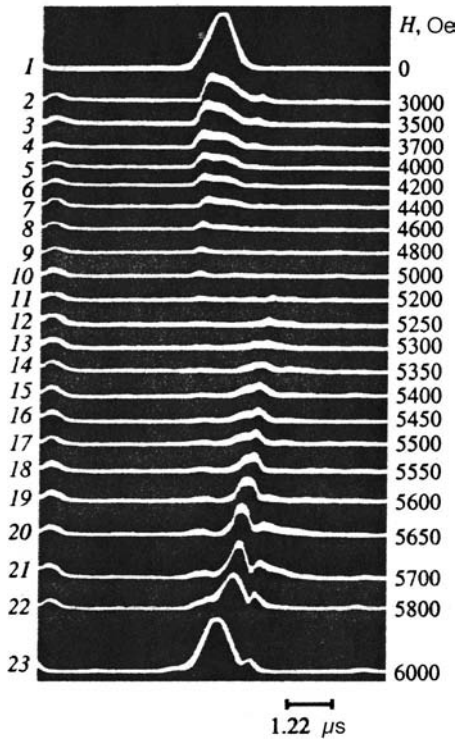

FIG. 4. Oscillograms characterizing the transmission dynamics of an ultrasonic pulse through a sample as a function of H.

5. MEASUREMENTS OF THE VELOCITY OF MAGNETOELASTIC WAVES NEAR A NUCLEAR MAGNETOACOUSTIC RESONANCE

Single-crystal $KMnF_3$ samples in the form of $4\times4\times7$ mm^3 parallelepipeds were used in the experiments. The crystal structure and the geometry of the experiment are described in Ref. 5.

The time interval t between the radio pulse of the transmitter and the first ultrasonic pulse transmitted through a compound resonator[13] was measured to determine the speed of sound V. The single-passage time of an ultrasonic pulse through the sample was determined as $t_{\text{sample}}=t-2t_{\text{trans}}$ and

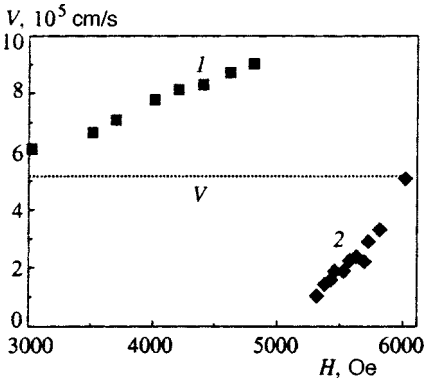

FIG. 5. Velocities V_1 (*1*) and V_2 (*2*) versus H (V is the speed of sound with no magnetic field).

$V=\mathscr{L}/t_{\text{sample}}$, where t_{trans} is the transmission time of a pulse through piezoelectric transducers of the same length.

The measurement error in the speed of sound did not exceed 20%, and depended on parameters such as the thickness of the bonding layer between the sample and the piezoelectric transducer, the matching of their acoustic impedances, and the pulse duration and shape.

Figure 4 displays a collection of oscillograms demonstrating the evolution of the amplitude and velocity of the transmitted acoustic pulse as a function of the magnetic field H. The numbers of the oscillograms are given on the left side and the field strength corresponding to each oscillogram is given on the right side. Time is plotted horizontally.

The oscillogram *1* corresponds to the first acoustic pulse transmitted through the sample with transmission time t_{sample}, velocity V, and amplitude u_0 at $H=0$. The large decrease in the amplitude of this pulse becomes appreciable in fields $H\geqslant 3000$ Oe (oscillograms *2–9*). At $H\approx 5200$ Oe (oscillograms *10*, *11*) the amplitude drops to the noise level. In fields $H\geqslant 5200$ Oe (oscillograms *12–22*) the reverse process of restoration of the pulse amplitude is observed.

The change in the transmission time of the signal through the sample, which is associated with the change in the speed of sound V, is clearly seen in the oscillograms. Thus, an increase of the speed of sound is observed in the oscillograms *2–11* and a decrease is observed in the oscillograms *11–13*.

The oscillograms *14–22* show the restoration of the transmission time. The dependence $V(H)$ is displayed in Fig. 5. The broadening of the transmitted acoustic pulse (oscillograms *2–22*) and the change in the pulse shape attest to a substantial frequency dispersion of the speed of sound within the frequency spectrum of the pulse. If this dispersion were weaker, so that the transmitted pulse had an appreciable amplitude at magnetoelastic resonance, it might have been possible to observe the splitting of the acoustic pulse, as mentioned in the preceding section.

We thank V. I. Grebennikov, V. V. Men'shenin, V. V. Nikolaev, and I. A. Garifullin for discussions and a number of valuable remarks.

This work was supported by the Russian Fund for Fundamental Research (Project 96-02-16489).

*)E-mail: bogdanova@dionis.kfti.kcn.ru

[1] V. I. Ozhogin and V. L. Preobrazhenskiĭ, Zh. Éksp. Teor. Fiz. **73**, 988 (1977) [Sov. Phys. JETP **46**, 523 (1977)].
[2] V. G. Bar'yakhtar, B. A. Ivanov, and A. L. Sukstanskiĭ, Zh. Éksp. Teor. Fiz. **78**, 1509 (1980) [Sov. Phys. JETP **51**, 757 (1980)].
[3] V. I. Ozhogin and V. L. Preobrazhenskiĭ, Usp. Fiz.Nauk **155**, 593 (1988) [Sov. Phys. Usp. **31**, 713 (1988)].
[4] Kh. G. Bogdanova, V. A. Golenishchev-Kutuzov, M. I. Kurkin *et al.*, Zh. Éksp. Teor. Fiz. **103**, 163 (1993) [JETP **76**, 89 (1993)].
[5] Kh. G. Bogdanova, V. A. Golenishchev-Kutuzov, M. I. Kurkin *et al.*, Zh. Éksp. Teor. Fiz. **112**, 1830 (1997) [JETP **85**, 1001 (1997)].
[6] A. M. Portis, G. L. Witt, and A. G. Heeger, J. Appl. Phys. **34**, 1052 (1963).
[7] W. J. Ince, Phys. Rev. **184**, 574 (1969).
[8] E. A. Turov and M. P. Petrov, *Nuclear Magnetic Resonance in Ferro- and Antiferromagnets* [in Russian], Nauka, Moscow (1969), p.260.
[9] L. D. Landau and E. M. Lifshitz, *Theory of Elasticity*, Pergamon Press, New York (1986).
[10] M. I. Kurkin and E. A. Turov, *NMR in Magnetically Ordered Materials and its Applcations* [in Russian], Nauka, Moscow(1990), p. 244.
[11] K. Gezi, J. D. Axe, G. Shirane, and A. Linz, Phys. Rev. B **5**, 1933 (1972).
[12] Yu. A. Izyumov, F. A. Kassan-Ogly, and V. E. Naĭsh, Fiz. Met. Metalloved. **51**, 500 (1981).
[13] Kh. G. Bogdanova, V. A. Golenishchev-Kutuzov, V. E. Leont'ev *et al.*, Prib. Tekh. Eksp. **4**, 60 (1997).

Translated by M. E. Alferieff

Magnetic and structural phase transitions in the shape-memory ferromagnetic alloys $Ni_{2+x}Mn_{1-x}Ga$

A. D. Bozhko, A. N. Vasil'ev, and V. V. Khovaĭlo

M. V. Lomonosov Moscow State University, 119899 Moscow, Russia

I. E. Dikshteĭn, V. V. Koledov, S. M. Seletskiĭ, A. A. Tulaĭkova, A. A. Cherechukin, and V. G. Shavrov*)

Institute of Radio Engineering and Electronics, Russian Academy of Sciences, 103907 Moscow, Russia

V. D. Buchel'nikov

Chelyabinsk State University, 454021 Chelyabinsk, Russia
(Submitted 13 February 1998)
Zh. Éksp. Teor. Fiz. **115**, 1740–1755 (May 1999)

The results of an experimental investigation of the temperature dependences of the magnetic susceptibility and resistivity in the shape-memory ferromagnetic alloys $Ni_{2+x}Mn_{1-x}Ga$ ($x=0-0.20$) are reported. A $T-x$ phase diagram is constructed on the basis of these data. It is shown that partial substitution of Ni for Mn causes the temperatures of the structural (martensitic) T_M and magnetic T_C (Curie point) phase transitions to converge. In the region where $T_C=T_M$ the transition temperature increases linearly with magnetic field in the range from 0 to 10 kOe. The kinetics of a magnetic-field-induced martensitic phase transition is investigated, and the velocities of the martensite–austenite interphase boundary during direct and reverse transitions are measured. A theoretical model is proposed and the $T-x$ phase diagram is calculated. It is shown that there exist concentration ranges where the magnetic and martensitic transitions merge into a first-order phase transition. The theoretical results are in qualitative agreement with experiment. © *1999 American Institute of Physics.* [S1063-7761(99)01605-4]

1. INTRODUCTION

It is well known that some metal alloys undergo reversible, crystallographic, thermoelastic, martensitic transformations which are accompanied by a shape-memory effect. This effect is ordinarily manifested as follows: a deformed sample in the low-temperature martensitic phase returns to its initial shape after the stress is removed and the sample is heated. The restoration of the initial shape is attributed to a reversible transformation of the deformed martensitic phase to a high-temperature austenitic phase. These alloys can be trained by repeated deformation and heat cycling. In this manner, a bilateral shape-memory effect can be obtained. In this case the sample will spontaneously acquire a definite prescribed shape, wherein the austenite transforms to martensite and the reverse transformation returns the sample to its original shape.[1]

In most cases shape-memory alloys are nonmagnetic, and the methods for influencing their shape and size are limited to stresses and temperature. However, in Mn-bearing Heusler alloys an indirect exchange interaction between the magnetic moments of the atoms produces ferromagnetism. Among such alloys there is one compound, Ni_2MnGa, that undergoes a martensitic-type structural transformation to the ferromagnetic phase.[2–6] The combination of magnetic ordering and shape memory makes this alloy promising in the search for the possibility of controlling the shape of a sample not only by varying the temperature and pressure but also by varying the external magnetic field. For applications it is also helpful that the martensitic transition and the ferromagnetic properties appear in this alloy near room temperature.

For the stoichiometric composition of the alloy Ni_2MnGa the ferromagnetic transition temperature, $T_C=376$ K, and the structural transition temperature, $T_M=202$ K, differ strongly. The temperatures T_C and T_M can be changed by purposefully changing the composition of this compound. To realize this possibility some Mn atoms must be replaced by Ni atoms. Then the distance between the Mn atoms in the alloy will increase, and therefore the exchange integral and magnetic transition temperature T_C will decrease. On the other hand, an increase in the electron density accompanying the substitution of Ni atoms for some Mn atoms will be accompanied by an increase in the volume bounded by the Fermi surface and by an increase in the structural transition temperature. Thus in the alloys $Ni_{2+x}Mn_{1-x}Ga$, T_M can be increased and T_C decreased by partial substitution of Ni for Mn until these temperatures are the same.

One objective of the present work is to investigate experimentally the anomalies of the electrical conductivity and magnetic susceptibility of the alloys $Ni_{2+x}Mn_{1-x}Ga$ with $x=0-0.20$ near the martensitic and magnetic phase transitions, whose temperatures converge as x increases, and to study in greater detail the properties of the alloy with coin-

cident phase transition points, specifically, the influence of a magnetic field on the structural transition point. Another objective is to construct a theoretical model that describes all phase transformations, taking account of the interaction of the magnetic and elastic subsystems, to calculate the $T-x$ phase diagram, and to assess the influence of a magnetic field and pressure on the martensitic phase transition temperature.

2. EXPERIMENT

2.1. Samples and procedure

The polycrystalline samples of the alloys $Ni_{2+x}Mn_{1-x}Ga$ investigated in the present work were prepared by arc melting in an Ar atmosphere on a cold hearth. More than twenty ingots of these alloys with concentrations $x=0-0.20$ were obtained. The samples on which measurements of the resistivity ρ and the low-field magnetic susceptibility χ were performed were cut from ingots by the electric-spark method. The resistivity was measured by the four-point scheme, and the magnetic susceptibility was investigated by the induction method. These procedures measure the sample-averaged values of the transition parameters. The measurement accuracy is determined by the inhomogeneity of the polycrystalline sample.

Optical measurements were performed for local investigation of the martensitic domain structure and to determine the temperature of the structural phase transition with a real resolution over the sample. To this end, chips approximately $5\times5\times2$ mm^3 in size were polished at room temperature and then subjected to heat cycling. In an experiment a sample clamped to a substrate in a two-loop heat bath with transparent windows was placed between the poles of an electromagnet and observed under a microscope with oblique illumination. The mechanical stresses produced by a structural phase transition form a relief on the surface. The directions in which the variation of the relief of individual microsections of the sample surface yield the greatest optical contrast as a result of a martensitic transition can be found by adjusting the angle of incidence of the illumination. The formation and motion of a martensite–austenite boundary as well as the formation of martensitic (structural) domain walls can be observed under a microscope. The evolution of the motion of the walls was recorded with a video camera. A typical pattern on the surface of a sample in the austenitic state is displayed in the photomicrograph in Fig. 1a. In Fig. 1b the same section is shown in the martensitic state. The boundaries of microcrystals and martensitic domains in the form of plates or stripes differently oriented in different microcrystals are clearly seen. The observations were performed in 0–10 kOe fields, and the variations of the sample temperature were monitored with an accuracy of about 0.03 K.

As a result of the inhomogeneity of a polycrystalline sample, the processes leading to the formation of martensitic domains in each microcrystal proceed at their own individual temperature (a microcrystal is about 0.3 mm in size and the variance of the transition temperatures is approximately 1 K). Appreciable qualitative differences in the processes leading to the formation of martensitic domains in different microcrystals are also observed. Specifically, the velocity of the

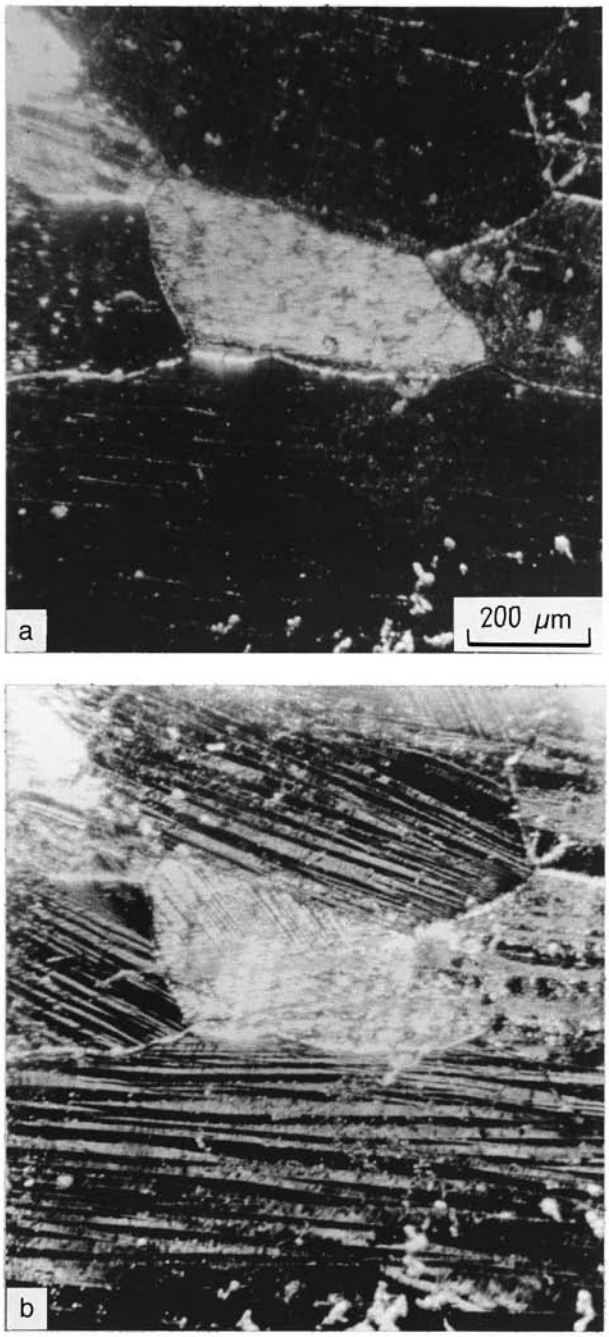

FIG. 1. Photomicrograph of the surface of a $Ni_{2.19}Mn_{0.81}Ga$ crystal in the austenitic (a) and martensitic (b) states.

walls and the temperature range where the phase boundary is mobile are different. Typical data characterizing the temperature variance of the direct and reverse structural transitions in different microcrystals in the same sample are displayed in Fig. 2. In this figure the arbitrary number N of a microcrystal in the field of view of the microscope is plotted along the ordinate in order of increasing transition onset temperature. It is evident from this plot that indifferent microcrystals not only the transition onset point but also the temperature range where the phase transition boundary passes through the entire microcrystal are somewhat different in different microcrystals, as is the width of the temperature hys-

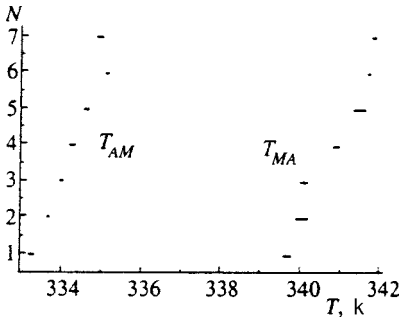

FIG. 2. Variance of the temperatures of the direct (T_{AM}) and reverse (T_{MA}) martensitic transitions for the sample $Ni_{2.19}Mn_{0.81}Ga$.

teresis loop. The magnetic field dependence of the phase transition parameters and of the characteristic features of the motion of the interphase boundaries was also studied locally in individual sections of the sample.

2.2. Temperature and concentration dependences of the magnetic and electric properties

The experimental temperature dependences of the resistivity $\rho(T)$ and the low-field magnetic susceptibility $\chi(T)$ obtained for some alloys are presented in Fig. 3. A kink was observed in the curves $\rho(T)$ at the point of a magnetic phase transition, and a jump occurred at the point of a structural phase transition. The increase in the slope of the curve $\rho(T)$ at the transition from the paramagnetic to the ferromagnetic state can be attributed to a decrease in the scattering of charge carriers by magnetic fluctuations. The jump in the curves $\rho(T)$ is due to the critical structural fluctuations in the region of formation of a martensitic phase. In the curves $\chi(T)$, as temperature decreases, the signal increases abruptly at the point of a magnetic phase transition and it decreases just as abruptly at the point of a structural phase transition. This decrease is due to an increase in the magnetic anisotropy of $Ni_{2+x}Mn_{1-x}Ga$ in the tetragonal phase. Near a structural phase transition $\rho(T)$ and $\chi(T)$ exhibit temperature hysteresis (which increases with x). The phase transition is

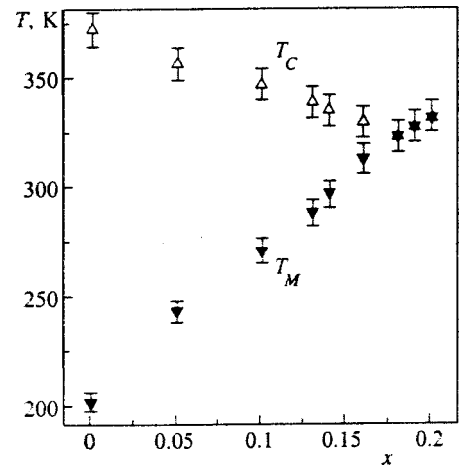

FIG. 4. Concentration dependences of the structural phase transition temperature T_M and the magnetic phase transition temperature T_C of the system of alloys $Ni_{2+x}Mn_{1-x}Ga$ (experiment).

characterized by the points T_{AM} and T_{MA} of the direct (austenite–martensite) and reverse (martensite–austenite) transitions ($T_{AM}<T_{MA}$), respectively. For simplicity, the curves obtained during cooling are displayed in Fig. 3. As the composition changed (nickel content increased at the expense of manganese), the magnetic and structural phase transition temperatures converged, and the values of T_M and T_C were essentially identical for $x=0.18$–0.20. Figure 4 shows $T_M(x)$ and $T_C(x)$ constructed from the current experimental data.

2.3. Martensitic domain structure and the effect of a magnetic field on the transition point

We investigated experimentally the effect of a magnetic field on the formation of structural defects accompanying a martensitic transition in polycrystalline samples with nonstoichiometric composition $Ni_{2.19}Mn_{0.81}Ga$. The martensitic transition temperature in a sample with this composition is close to the Curie point ($T_C \approx 338$ K, $T_{AM} \approx 337$ K, T_{MA}

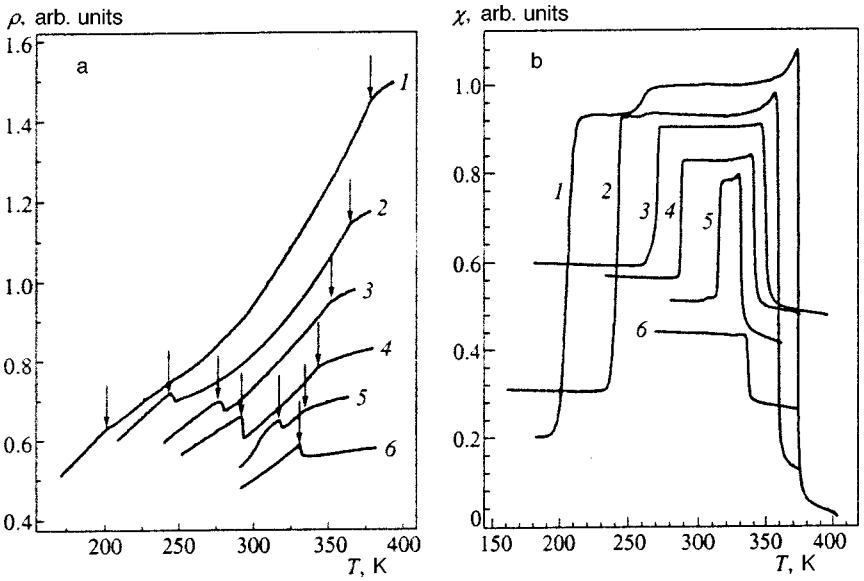

FIG. 3. Temperature dependences of the resistivity (a) and magnetic susceptibility (b) of the alloys $Ni_{2+x}Mn_{1-x}Ga$ with $x=0$ (*1*), 0.05 (*2*), 0.10 (*3*), 0.13 (*4*), 0.16 (*5*), and 0.19 (*6*). The arrows mark the phase transition points.

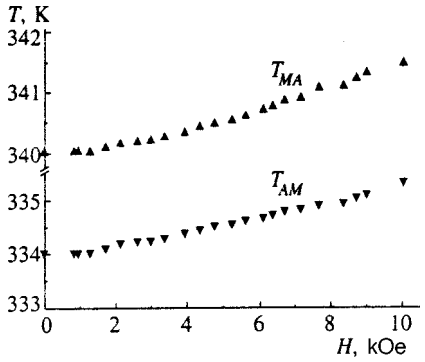

FIG. 5. Temperatures of the direct T_{AM} and reverse T_{MA} martensitic transitions versus the external magnetic field for one section of the sample $Ni_{2.19}Mn_{0.81}Ga$.

$\approx$342 K). The transition in temperature therefore occurs from the paramagnetic austenitic phase to the ferromagnetic martensitic phase and vice versa.

In the experiment T_{AM} and T_{MA} are observed to increase in a magnetic field. They increase linearly with the field with a slope of about 0.15 ± 0.02 K/kOe in 2–10 kOe fields, while slower growth is observed in fields below 2 kOe (Fig. 5). Data are presented for one of the crystallites. Since the increase in the transition temperature does not exceed 1.5 K even a 10 kOe field and remains of the same order of magnitude as the nonuniformity of the transition temperatures over the volume of the sample, it becomes understandable why the shift of the structural transition point in a magnetic field cannot be observed by other methods.

In Ref. 7 the shift of the martensitic transition temperature of a stoichiometric single-crystal Ni_2MnGa sample in a magnetic field was investigated by the dilatometric method. In this case the authors observed T_{AM} and T_{MA} to decrease with increasing field. This result might be due to the fact that in contrast to our experiments, the structural transition occurs from a ferromagnetic austenitic phase to a ferromagnetic martensitic phase.

We also studied the effect of a magnetic field on the kinetics of a martensitic transition, and we demonstrated that the motion of a phase boundary can be controlled by a magnetic field. Turning on a magnetic field in a certain temperature range, which depends on the magnitude of the field (somewhat above T_{AM}), during cooling of the sample gives rise to motion of the phase boundary with formation of martensite. Turning off the field in a certain temperature range (somewhat above T_{MA}) with the sample temperature increasing gives rise to motion of the phase boundary in the opposite direction with formation of austenite.

The velocity of the phase interface with decreasing temperature (austenite–martensite transition) is higher than with increasing temperature (martensite–austenite transition). Typical velocities are 10^{-2} and $5\cdot10^{-3}$ cm/s, respectively. It was also noted that there was a delay of about 1 s between the field being turned on or off and the phase boundary beginning to move.

3. THEORY

3.1. Phase diagram of a cubic ferromagnet

We employ Landau's phenomenological model of phase transitions to analyze the phase diagram of the ferromagnet $Ni_{2+x}Mn_{1-x}Ga$. We consider a cubic ferromagnet with the point symmetry group O_h and a magnetic phase transition, which is accompanied by the onset of spontaneous magnetization **M**, and a characteristic ferroelastic phase transition to a tetragonal phase with D_{4h} symmetry, with the onset of spontaneous deformations.[8–11] In this case the order parameters describing the structural transformations are components of the macroscopic deformation tensor e_{ik}. We describe the magnetic phase transitions by components of the macroscopic magnetization **M**. Then the expression for the thermodynamic potential of the ferromagnet can be written

$$\Phi=\frac{1}{2}(c_{11}+2c_{12})e_1^2+\frac{1}{2}a(e_2^2+e_3^2)+\frac{1}{2}c_{44}(e_4^2+e_5^2+e_6^2)$$
$$+\frac{1}{2}be_3(e_3^2-3e_2^2)+\frac{1}{4}c(e_2^2+e_3^2)^2+\frac{1}{\sqrt{3}}B_1e_1m^2$$
$$+B_2\left[\frac{1}{\sqrt{2}}e_2(m_1^2-\tfrac{2}{2})+\frac{1}{\sqrt{6}}e_3(3m_3^2-m^2)\right]$$
$$+B_3(e_4m_1m_2+e_5m_2m_3+e_6m_3m_1)+\frac{1}{2}\alpha_1(m_1^2$$
$$+m_2^2+m_3^2)+\frac{1}{4}\delta_1(m_1^2+m_2^2+m_3^2)^2+K_1(m_1^2m_2^2$$
$$+m_2^2m_3^2+m_3^2m_1^2)-\nu Te_1. \qquad (1)$$

Here the e_i are linear combinations of the components of the deformation tensor e_{ik}:

$$e_1=\frac{1}{\sqrt{3}}(e_{xx}+e_{yy}+e_{zz}),\quad e_2=\frac{1}{\sqrt{2}}(e_{xx}-e_{yy}),$$
$$e_3=\frac{1}{\sqrt{6}}(2e_{zz}-e_{xx}-e_{yy}),$$
$$e_4=e_{xy},\quad e_s=e_{yz},\quad e_6=e_{zx},$$

a, *b*, and *c* are linear combinations of the components of the elastic moduli of second, third, and fourth orders, respectively, and are given by

$$a=c_{11}-c_{12},\quad b=\frac{1}{6\sqrt{6}}(c_{111}-c_{112}+c_{123}),$$
$$c=\frac{1}{48}(c_{1111}+c_{1112}-3c_{1122}-8c_{1123}),$$

$\mathbf{m}=\mathbf{M}/M_0$, $\mathbf{M}_0$ is the saturation magnetization far from the Curie point, α_1 and δ_1 are exchange constants, B_1 and $B_{2,3}$ are, respectively, the exchange and relativistic magnetostriction constants, K_1 is the first cubic anisotropy constant, and ν is the thermal expansion coefficient.

As the point of the structural phase transition from the cubic to the tetragonal phase with a two-component order parameter (e_2,e_3) is approached, the elastic modulus $a=c_{11}-c_{12}$ approaches zero. In the expression for the free energy, only the terms that are responsible for this transition are retained in the terms containing the third and fourth powers of the components of the deformation tensor. The third-order terms present in the thermodynamic potential give rise to a first-order structural transition.

The expression (1) describes the case in which the magnetic and martensitic phase transition temperatures are close to one another, as well as the situation in which these temperatures differ substantially. In the first case, near the martensitic transition temperature the magnetic moment varies strongly with temperature, both in direction and magnitude. In the situation $T_C>T_M$, the structural phase transition is accompanied primarily by a change in the direction of the magnetization.

Minimizing the thermodynamic potential with respect to the components ofthe deformation tensor e_1, e_4, e_5, and e_6, which are not responsible for the structural phase transition, the expression (1) becomes

$$\Phi=\Phi_0+\frac{1}{2}a(e_2^2+e_3^2)+\frac{1}{3}be_3(e_3^2-3e_2^2)+\frac{1}{4}c(e_2^2+e_3^2)^2$$
$$+B_2\left[\frac{1}{\sqrt{2}}e_2(m_1^2-m_2^2)+\frac{1}{\sqrt{6}}e_3(3m_3^2-m^2)\right]$$
$$+\frac{1}{2}\alpha(m_1^2+m_2^2+m_3^2)+\frac{1}{4}\delta(m_1^2+m_2^2+m_3^2)^2$$
$$+K(m_1^2m_2^2+m_2^2m_3^2+m_3^2m_1^2), \qquad (2)$$

where

$$\alpha=\alpha_1+\frac{\nu B_1T}{\sqrt{3}\,(c_{11}+2c_{12})}, \quad \delta=\delta_1-\frac{B_1^2}{6(c_{11}+2c_{12})},$$
$$K=K_1-\frac{B_3^2}{2c_{44}}$$

are the magnetostriction-renormalized exchange constants and the first cubic-anisotropy constant.

To find all possible structural and magnetic phases, the potential (2) must be minimized with respect to the variables e_2, e_3, m_1, m_2, and m_3. As a result, we find the following states of the ferromagnet under study and their stability conditions, choosing for definiteness $b>0$, $c>0$, and $K<0$ (which corresponds to magnetization orientation in the [111] direction in the cubic phase; this state occurs in the stoichiometric alloy Ni_2MnGa).

1. Cubic paramagnetic phase

$$m_1=m_2=m_3=0, \quad e_2=e_3=0,$$

stable for $\alpha\geqslant 0$ and $a\geqslant 0$.

2. Tetragonal paramagnetic phase

$$m_1=m_2=m_3=0, \quad e_2=0, \quad e_3=-\frac{b+\sqrt{b^2-4ac}}{2c},$$

stable for

$$\alpha\geqslant\frac{2Bb}{\sqrt{6}\,c}, \quad a\leqslant\frac{b^2}{4c}, \quad a\geqslant\frac{b^2}{4c}-\left(\frac{\alpha\sqrt{6c}}{4B}-\frac{b}{2\sqrt{c}}\right)^2$$

(for simplicity we drop the subscript on B_2).

3. Cubic ferromagnetic phase

$$e_2=e_3=0, \quad m_1=m_2=m_3=\frac{m}{\sqrt{3}}, \quad m^2=-\frac{\alpha}{\delta-4q/3},$$

stable for $\alpha\leqslant 0$ and $a\geqslant B^2/q$, where $q=|K|$.

4. Tetragonal canted ferromagnetic phase

$$m_1^2=m_2^2=-\frac{1}{3}\frac{\alpha}{\delta-4q/3}+\frac{Be_3}{\sqrt{6}\,q},$$
$$m_3^2=-\frac{1}{3}\frac{\alpha}{\delta-4q/3}-\frac{2Be_3}{\sqrt{6}\,q},$$
$$e_2=0, \quad e_3=-\frac{b+\sqrt{b^2-4c(a-B^2/q)}}{2c},$$

stable for

$$a\leqslant\frac{b^2}{4c}+\frac{B^2}{q},$$
$$a\geqslant\frac{b^2}{4c}+\frac{B^2}{q}-\left(\sqrt{\frac{2}{3}}\frac{q}{B}\frac{\alpha\sqrt{c}}{\delta-4q/3}+\frac{b}{2\sqrt{c}}\right)^2,$$
$$\alpha\leqslant-\frac{\sqrt{6}\,bB}{4cq}\left(\delta-\frac{4q}{3}\right).$$

5. Tetragonal collinear ferromagnetic phase

$$m_1=m_2=0, \quad m_3^2=-\frac{1}{\delta}\left(\alpha+\frac{4Be_3}{\sqrt{6}}\right),$$
$$e_2=0, \quad ae_3+be_3^2+ce_3^3+\sqrt{\frac{2}{3}}\,Bm^2=0,$$

stable in the region bounded by the curves

$$a=\frac{b^2}{4c}+\frac{B^2}{q}-\left(\sqrt{\frac{2}{3}}\frac{q\alpha\sqrt{c}}{B(\delta-4q/3)}+\frac{b}{2\sqrt{c}}\right)^2, \quad \alpha\leqslant 0$$

and

$$a=\frac{b^2}{4c}-\left(\frac{\alpha\sqrt{6c}}{4B}-\frac{b}{2\sqrt{c}}\right)^2, \quad \alpha\geqslant 0$$

and the upper part of the discriminant curve of the cubic equation determining the deformation e_3 in this phase (see Eq. (5) below).

In the cubic ferromagnetic phase 3, the magnetization $\mathbf{M}\parallel[111]$. In the canted ferromagnetic tetragonal phase 4, the magnetization $\mathbf{M}$ changes direction from the [111] to the [001] axis as the temperature changes. Finally, $\mathbf{M}\parallel[001]$ in the tetragonal collinear phase 5.

It follows from symmetry considerations that besides these states, other tetragonal phases with the same energy

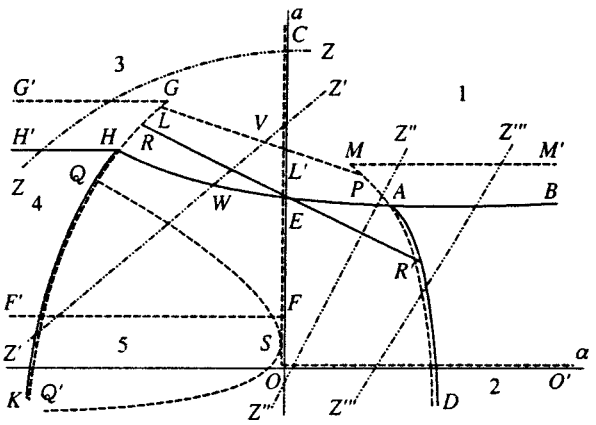

FIG. 6. Theoretical phase diagram of a cubic ferromagnet in the coordinates of the structural (a) and magnetic (α) order parameters (schematically): 1 — cubic paramagnetic phase; 2 — tetragonal paramagnetic phase; 3 — cubic ferromagnetic phase; 4 and 5 — isostructural tetragonal ferromagnetic phases. Solid lines — phase transition lines, dashed lines — phase-instability lines.

and domain of stability as those indicated above can also occur in a ferromagnet. States with $\mathbf{M} \| [100]$ and $\mathbf{M} \| [010]$ are energy-equivalent to phase 5. Canted states, which are similar to phase 4, and in which the magnetization changes direction from the [111] to the [100] and [010] axes, and in which the deformations are given by

$$e_3 = -\frac{b+\sqrt{b^2-4c(a-B^2/q)}}{2c}, \quad e_2^2 = 3e_3^2, \tag{3}$$

can arise in the same manner. The existence of several phases with the same energies and stability regions leads to the existence of structural and magnetic domains in ferromagnets. We note that from the standpoint of a tetragonal distortion of the lattice, phase 5 is identical in terms of symmetry to phase 4, and therefore transitions between these phases are isostructural.

The lines of phase transitions between possible states are determined by energy equality between the phases. The expressions for the phase transition lines are presented in the Appendix. The phase diagram of a ferromagnetic in the α–a plane is shown schematically in Fig. 6. The following phase transitions are possible from the paramagnetic cubic phase 1: along the line AB — a first-order structural phase transition to a tetragonal paramagnetic phase 2; along the line CE — a second-order isostructural transition to the cubic ferromagnetic phase 3 with $\mathbf{M} \| [111]$; and, along the line EA — first-order magnetic and structural transitions to the tetragonal ferromagnetic phase 5. Aside from the transition 1–2 indicated above, a second-order isostructural magnetic transition to the tetragonal ferromagnetic phase 5 (the line AD) can occur from the tetragonal paramagnetic phase 2. First-order structural and orientational transitions to tetragonal phases 4 (the line $H'H$) and 5 (the line HE) are also possible from the cubic ferromagnetic phase 3. A second-order isostructural orientational phase transition to phase 5 (the line HK) occurs from phase 4.

In summary, the domain of absolute stability of phase 1 lies in Fig. 6 in the first quadrant above the line $CEAB$. For phase 2 this region is bounded by the line BAD, and for phase 3 it is bounded by the line $H'HEC$. Phase 4 is absolutely stable in the region $H'HK$, and phase 5 is stable in the region $KHEAD$. The lines of instability in Fig. 6 are dashed: the lines CO and OO' for phase 1, $M'M$ and MD for phase 2, $F'F$ and FC for phase 3, $G'GHK$ for phase 4, and $KLL'PD$ for phase 5. The coordinates of the characteristic points of this phase diagram are presented in the Appendix.

In phase 5, the thermodynamic potential can have one or two minima as a function of the deformation. In the latter case there can exist two phases, differing with respect to the magnitude of the spontaneous deformation but identical with respect to the magnetic and crystallographic symmetries(the region of a double-well potential). The cubic equation

$$ce_3^3 + be_3^2 + \tilde{a}e_3 - \sqrt{\frac{2}{3}}\frac{\alpha}{\delta}B = 0, \tag{4}$$

which determines the deformation e_3 in phase 5, possesses three real roots, and its discriminant is negative. Here $\tilde{a} = a - 4B^2/2\delta$. Stability analysis of the solutions shows that only two of them correspond to an energy minimum. The discriminant curve of the cubic equation (4) can be expressed as

$$\alpha = \frac{\sqrt{6}\,\delta b^3}{27Bc^2}\left[1 - \frac{9c\tilde{a}}{2b^2} \pm \left(1 - \frac{3c\tilde{a}}{b^2}\right)^{3/2}\right]. \tag{5}$$

The region of the double-well potential lies in Fig. 6 inside the contour $Q'SQLL'PD$, the lines $Q'SQ$ and $LL'P$ being the discriminant curves (5).

The lines of isostructural phase transitions between the metastable tetragonal phases 5 (RE) and between the stable tetragonal phases 5 (ER'), with differing magnitudes of spontaneous deformation, are determined by the expression

$$a = \frac{4}{3}\frac{B^2}{\delta} + \frac{2}{9}\frac{b^2}{c} - \sqrt{6}\,B\frac{\alpha}{\delta}\frac{c}{b}. \tag{6}$$

We note that isostructural transitions in phase 5 are related to the ''magnetic pressure'' Bm^2 due to the magnetoelastic interaction.

When the diagram is traversed along the line ZZ, first a second-order magnetic transition of the order–disorder type occurs from the paramagnetic cubic phase to the ferromagnetic cubic phase. This transition is followed by a first-order orientational martensitic transition from a cubic ferromagnet with $\mathbf{M}$ along a body diagonal of the cube to a tetragonal ferromagnet with magnetization in a (110)-type plane.

When the phase diagram is traversed along the line $Z'Z'$, a cubic ferromagnetic phase with magnetization along a body diagonal of the cube arises in an ordering-type transition on the line CE. At the point W of the martensitic transition on the line HE, the cubic ferromagnetic phase transforms to a tetragonal martensitic phase with $\mathbf{M}$ along the [001] axis. Next, as temperature decreases, a transition from the tetragonal collinear phase to a canted tetragonal martensitic phase with $\mathbf{M}$ in a (110) type plane occurs on the line HK.

In the case of the thermodynamic path $Z''Z''$, first-order magnetic and structural phase transitions occur simulta-

neously on the line *EA* from the cubic paramagnetic phase to the tetragonal ferromagnetic phase with **M** along the [001] axis.

Finally, when the diagram is traversed along the line $Z'''Z'''$, first a martensitic transition from the cubic paramagnetic phase to the tetragonal martensitic paramagnetic phase occurs on the straight line *AB*, and then a second-order isostructural martensitic transition of the order–disorder type from phase 2 to phase 5 occurs on the line *AD*.

We note that all indicated sequences of phase transitions should lead to the behavior typical of the various magnetic characteristics of a ferromagnet. In particular, the indicated temperature phase transitions should be accompanied by corresponding kinks. It is this behavior of the susceptibility in stoichiometric samples that was observed in Refs. 2–5 in transitions to an ordered state and martensitic transitions.

It follows from the phase diagram that transitions between cubic paramagnetic and ferromagnetic phases (the line *CE*) and tetragonal paramagnetic and ferromagnetic phases (the line *AD*) are second-order phase transitions. This is a typical situation for phase transitions in magnets at the Curie point. The transition from the cubic (austenite) paramagnetic phase to the tetragonal (martensite) ferromagnetic phase is a first-order phase transition (the line *EA*). This effect is entirely due to the interaction of structural and magnetic order parameters. For similar values of T_M and T_C, the first-order phase transition should be accompanied by a latent heat of transition and hysteresis in the temperature dependences of various characteristics of the ferromagnet. This has been observed experimentally in nonstoichiometric alloys.[3–5]

To compare theory and experiment, the $a-\alpha$ diagram we have constructed can be represented in the $T-x$ plane. To do so, we expand the coefficients a and α in T and x near the temperatures T_{M0} and T_{C0} of the structural and magnetic transitions with stoichiometric composition $(x=0)$, truncating the series at the linear terms,

$$a(T,x)=a_0(T-T_1-\kappa x) \quad \text{and} \quad \alpha(T,x)=\alpha_0(T-T_{C0}+\gamma x), \tag{7}$$

where

$$T_1=T_{M0}-\frac{\sqrt{6}\,Bc}{a_0 b}\,\frac{\alpha_0}{\delta}(T_{C0}-T_{M0})-\frac{2b^2}{9ca_0}-\frac{4B^2}{3\delta a_0},$$

and a_0, α_0, κ, and γ are proportionality coefficients. The temperature T_1 was chosen such that in the $\alpha-a$ diagram the point T_{M0} (at $x=0$) corresponding to the martensitic transition lay within the range *HE* (the point *W*).

The phase transition lines in the $T-x$ plane have the form

$$1\leftrightarrow 2: \quad T=T_1+\frac{2b^2}{9ca_0}+\kappa x, \tag{8}$$

$$1\leftrightarrow 3: \quad T=T_{C0}-\gamma x. \tag{9}$$

Analytic expressions for all other transition lines are too complicated to present here. The coordinates of the characteristic points of this diagram are given in the Appendix.

The $T-x$ phase diagram constructed in this manner is shown in Fig. 7. Since many parameters of the problem are unknown at the present time, a comparison of this theoretical diagram with the experimental diagram in Fig. 4 will only be qualitative. The diagram presented in Fig. 7 was obtained with parameter values $T_{C0}=375$ K, $T_{M0}=200$ K, $a_0=b/T_{M0}$, $\alpha_0=\delta/T_{C0}$, $\delta\sim 10^8$ erg/cm^3, $b/c\sim 0.5$, $b\sim 10^{12}$ erg/cm^3, $B\sim 10^7$ erg/cm^3, $q\sim 10^4$ erg/cm^3, $\gamma=295$ K, and $\kappa=800$ K. The notation in Fig. 7 is the same as in Fig. 6.

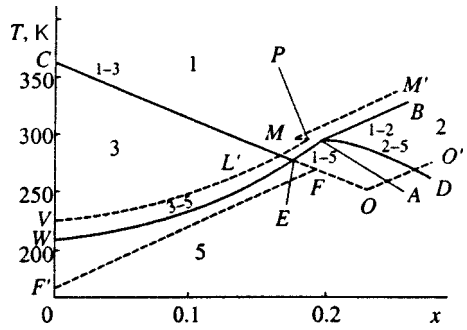

FIG. 7. Phase diagram of a cubic ferromagnet in the coordinates temperature T–concentration x (calculation).

3.2. Thermodynamic estimate of the effect of a magnetic field and pressure on the martensitic phase transition temperature

The effect of a magnetic field and pressure on the temperature of a first-order structural transition can be estimated using the Clausius–Clapeyron thermodynamic equation.[12]

In phase equilibrium the thermodynamic potentials Φ_M and Φ_A of martensitic and austenitic phases, which are functions of temperature T, magnetic field H, and pressure P, are equal:

$$\Phi_M(T,H,P)=\Phi_A(T,H,P). \tag{10}$$

The condition (10) determines the phase transition surface in the space (T,H,P). Near a fixed point with coordinates $H_0=0$, T_M, and P_0, the phase transition surface can be described by the equation

$$\left(\frac{\partial\Phi_M}{\partial T}-\frac{\partial\Phi_A}{\partial T}\right)\Delta T+\left(\frac{\partial\Phi_M}{\partial H}-\frac{\partial\Phi_A}{\partial H}\right)\Delta H + \left(\frac{\partial\Phi_M}{\partial P}-\frac{\partial\Phi_A}{\partial P}\right)\Delta P=0, \tag{11}$$

where $\Delta T=T-T_M$, $\Delta P=P-P_0$,

$$\frac{\partial\Phi_M}{\partial T}-\frac{\partial\Phi_A}{\partial T}=S_A-S_M=\frac{Q}{T_M}, \tag{12}$$

$$\frac{\partial\Phi_M}{\partial H}-\frac{\partial\Phi_A}{\partial H}=M_A V_A-M_M V_M, \tag{13}$$

$$\frac{\partial\Phi_M}{\partial P}-\frac{\partial\Phi_A}{\partial P}=V_M-V_A, \tag{14}$$

S is the entropy, Q is the latent heat of the phase transformation at temperature T_M, and M and V are the magnetization and volume of the corresponding phase.

Substituting the expressions (12)–(14) into Eq. (11) makes it possible to determine the temperature shift of the martensitic transition induced by a magnetic field at constant pressure,

$$\Delta T=(M_M V_M-M_A V_A)HT_M/Q, \tag{15}$$

and the pressure-induced shift of T_M in a constant magnetic field,

$$\Delta T=(T_M/Q)(V_A-V_M)\Delta P. \tag{16}$$

Equation (15) describes the shift of the martensitic transformation temperature when the martensitic and austenitic phases are ferromagnetic. According to the phase diagrams (Figs. 6 and 7), this situation prevails in $Ni_{2+x}Mn_{1-x}Ga$ on the lines $H'H$ and HE.

When the martensitic transition temperature is higher than the ferromagnetic-transition temperature, both phases are paramagnetic at a structural transformation and the change in the transition temperature can be cast in the form

$$\Delta T=(\chi_M V_M-\chi_A V_A)H^2T_M/2Q, \tag{17}$$

where χ is the magnetic susceptibility. This occurs in $Ni_{2+x}Mn_{1-x}Ga$ on the line AB (Figs. 6 and 7).

When a structural transition occurs from the austenitic paramagnetic phase to a martensitic ferromagnetic phase, the magnetization of the paramagnetic phase can be neglected in comparison with the ferromagnetic phase, and the temperature shift has the form

$$\Delta T=M_M V_M HT_M/Q. \tag{18}$$

This situation occurs in $Ni_{2+x}Mn_{1-x}Ga$ on the line EA of the first-order phase transition (Figs. 6 and 7). Therefore the magnetization M_M is nonzero on the phase transition line itself.

Equating the change in the temperature of the martensitic transition in a magnetic field (18) and the change (16) in T_M under external pressure, we obtain the estimated value of ΔP for which the same change would have occurred in T_M in a field H:

$$\Delta P=V_M M_M H/(V_A-V_M). \tag{19}$$

4. DISCUSSION

Comparison of Figs. 4 and 7 shows that the experimentally obtained T–x phase diagram in the concentration range $0<x<0.20$ agrees qualitatively with the computed T–x diagram. Specifically, the temperature of the magnetic phase transition from the cubic paramagnetic phase 1 to the cubic ferromagnetic phase 3 decreases linearly with concentration. The temperature of the phase transition from the cubic ferromagnetic phase 3 to the tetragonal ferromagnetic phase 5 increases with concentration.

We conclude above on the basis of a calculation using the Landau theory of phase transitions that on the line EA (Fig. 7), the martensitic and magnetic phase transitions merge to a single first-order transition (paramagnetic austenite–ferromagnetic martensite). Such a coincident section of the phase diagram was found experimentally near concentrations $x=0.18$–0.20 (Fig. 4). Another indirect confirmation of a first-order magnetic phase transition here is the linear dependence of ΔT on H implied by Eq. (18). Indeed, when a martensitic transition occurs in a paramagnetic state of the sample, according to Eq. (17) the shift of the martensitic transition temperature should depend quadratically on the external field. If the martensitic transition occurs in a ferromagnetic state of the sample, a strongly nonlinear dependence $\Delta T(H)$ should also be expected in fields ranging from 1 to 10 kOe, since, as one can see from Eq. (15), the slope of the linear dependence $\Delta T(H)$ is determined by the difference of the magnetizations of the martensitic and austenitic phases. It is known[5,7] that the magnetization curve for austenite first grows more rapidly than for martensite, and it saturates in fields near 2 kOe. The magnetically harder martensite saturates more slowly, only in fields 8–10 kOe, and the saturation magnetization is higher. Thus, in this case $\Delta T(H)$ should be nonlinear in fields 1–8 kOe. It also follows from Eq. (15) that the sign of the effect should be negative in this case, since the magnetization of austenite increases with field more rapidly than in martensite. Such behavior probably occurred in the experiments described in Ref. 7.

The shift of the martensitic transition temperature in a magnetic field can be estimated from Eq. (18). Assuming the saturation magnetization $M_s\simeq 300$ G,[5,7] martensitic transition temperature $T_M\simeq 340$ K, and latent heat of transition $Q\simeq 5\cdot 10^8$ erg/cm^3, we obtain $d\Delta T/dH\simeq 0.2$ K/kOe. This estimate agrees satisfactorily with the experimental value $\simeq 0.15$ K/kOe.

Assuming the relative change in volume at a martensitic transformation to be $(V_M-V_A)/V_M\sim 10^{-2}$, we estimate the pressure giving rise to the same change in the transition temperature as a saturating magnetic field. Using Eq. (19) we obtain $\Delta P\sim 10^8$ Pa. This estimate agrees qualitatively with the experimental data:[13] $\Delta T_M\simeq -1.5\cdot 10^{-8}\Delta P\simeq -1.5$ K.

The results obtained in this work also make it possible to estimate the compositional inhomogeneity of the experimental samples, assuming that the variance of the structural transition temperatures over the sample surface is entirely due to compositional inhomogeneity. Comparison of the data plotted in Figs. 2 and 4 shows that the inhomogeneity of the concentration x that yields a variance $\simeq 1$ K in the transition temperatures over the sample is $\simeq 0.001$.

We thank M. Matsumoto for providing the samples. This work was partially supported by the Russian Fund for Fundamental Research (Grant No. 96-02-19755) and ISSEP (Grant No. 615p).

APPENDIX

The phase transition lines determined by equating the energies of the phases are

$$1\leftrightarrow 2:\quad a=\frac{2b^2}{9c}\quad \text{for}\quad \alpha\geqslant\frac{8}{3}\frac{bB}{\sqrt{6}\,c};$$

$$1\leftrightarrow 3:\quad \alpha=0\quad \text{for}\quad a\geqslant B^2/q;$$

$$1\leftrightarrow 5:\quad 27c\delta\tilde{a}^4-(6b^2\delta-36cB^2)\tilde{a}^3-(8b^2B^2-60\sqrt{6}\,bcB\delta\tilde{\alpha}-54c^2\delta^2\tilde{\alpha}^2)\tilde{a}^2-(54b^2c\delta^2\tilde{\alpha}^2-324c^2B^2\delta\tilde{\alpha}^2-72\sqrt{6}\,bcB^3\tilde{\alpha}+12\sqrt{6}\,b^3B\delta\tilde{\alpha})\tilde{a}+27c^3\delta^3\tilde{\alpha}^4-36\sqrt{6}\,bc^2B\delta^2\tilde{\alpha}^3+324c^2B^4\tilde{\alpha}^2-12b^2cB^2\delta\tilde{\alpha}^2+9b^4\delta^2\tilde{\alpha}^2-128\sqrt{6}\,b^3B^3\tilde{\alpha}/9=0; \tag{3}$$

and for $0<\alpha<(8/3)bB/\sqrt{6}c$

$$2\leftrightarrow 5:\quad a=\frac{b^2}{4c}-\frac{1}{4c}\left(b-\sqrt{6}\,\frac{c\alpha}{2B}\right)^2 \quad \text{for} \quad \alpha\geqslant\frac{8}{3}\,\frac{bB}{\sqrt{6}\,c};$$

$$3\leftrightarrow 4:\quad a=\frac{2b^2}{9c}+\frac{B^2}{q} \quad \text{for} \quad \alpha<-\frac{3}{2}\left(\delta-\frac{4q}{3}\right)\frac{bB}{6cq};$$

$$3\leftrightarrow 5:\quad 9c\tilde{q}a^4-(2b^2\tilde{q}+9cB^2)\tilde{a}^3+(2b^2B^2+20\sqrt{6}\,bcB\tilde{q}\tilde{\alpha}-24c^2\tilde{q}^2\tilde{\alpha}^2)\tilde{a}^2+(24c\tilde{q}^2\tilde{\alpha}^2+108c^2B^2\tilde{q}\tilde{\alpha}^2-18\sqrt{6}\,bcB^3\tilde{\alpha}-4\sqrt{6}\,b^3B\tilde{q}\tilde{\alpha})\tilde{a}+16c^3\tilde{q}^3\tilde{\alpha}^4+16\sqrt{6}\,bc^2B\tilde{q}^2\tilde{\alpha}^3-81c^2B^4\tilde{\alpha}^2-4b^2cB^2\tilde{q}\tilde{\alpha}^2-4b^4\tilde{q}^2\tilde{\alpha}^2+32\sqrt{6}\,b^3B^3\tilde{\alpha}/9=0; \tag{4}$$

$$4\leftrightarrow 5:\quad a=\frac{b^2}{4c}+\frac{B^2}{q}-\left(\sqrt{\frac{2}{3}}\,\frac{q}{B}\,\frac{\alpha\sqrt{c}}{\delta-4q/3}+\frac{b}{2\sqrt{c}}\right)^2 \quad \text{for} \quad \alpha<-2\left(\delta-\frac{4q}{3}\right)\frac{bB}{\sqrt{6}\,cq},$$

where

$$\tilde{\alpha}=\frac{\alpha}{\delta},\quad \tilde{a}=a-\frac{4B^2}{3\delta},\quad \tilde{q}=\frac{\delta q}{\delta-4q/3}.$$

The coordinates of the characteristic points of this phase diagram (Fig. 6) are

$$A\left(\frac{8}{3}\,\frac{bB}{\sqrt{6}\,c},\frac{2b^2}{9c}\right),\quad M\left(\frac{2bB}{\sqrt{6}\,c},\frac{b^2}{4c}\right),$$

$$H\left(-\frac{2bB}{\sqrt{6}\,cq}\left(\delta-\frac{4q}{3}\right),\quad \frac{2b^2}{9c}+\frac{B^2}{q}\right),$$

$$F\left(0,\frac{B^2}{q}\right),\quad G\left(-\sqrt{\frac{3}{8}}\left(\delta-\frac{4q}{3}\right)\frac{bB}{cq},\quad \frac{b^2}{4c}+\frac{B^2}{q}\right),$$

$$S\left(0,\frac{4B^2}{3\delta}\right),\quad E\left(0,\frac{2b^2}{9c}+\frac{4B^2}{3\delta}\right),\quad L'\left(0,\frac{b^2}{4c}+\frac{4B^2}{3\delta}\right).$$

The characteristic points of the T–x phase diagram (Fig. 7) are

$$A\left\{X_A=\frac{1}{\kappa+\gamma}\left[T_{C0}-T_1+\frac{8bB}{3\sqrt{6}\,c\alpha_0}-\frac{2b^2}{9ca_0}\right],\right.$$
$$\left.T_A=\frac{1}{\kappa+\gamma}\left[\kappa\left(T_{C0}+\frac{8bB}{3\sqrt{6}\,c\alpha_0}\right)+\gamma\left(T_1+\frac{2b^2}{9ca_0}\right)\right];\right. \tag{5}$$
$$E\left\{X_E=\frac{1}{\kappa+\gamma}\left[T_{C0}-T_1-\frac{2b^2}{9ca_0}-\frac{4B^2}{3\delta a_0}\right],\right.$$
$$\left.T_E=\frac{1}{\kappa+\gamma}\left[\kappa T_{C0}+\gamma\left(T_1+\frac{2b^2}{9ca_0}+\frac{4B^2}{3\delta a_0}\right)\right].\right.$$

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

[1] C. M. Wayman, J. Met. **6**, 129 (1980).
[2] P. J. Webster, K. R. A. Ziebeck, S. L. Town, and M. S. Peak, Phil. Mag. **49**, 295 (1984).
[3] A. N. Vasil'ev, A. Kaĭper, V. V. Kokorin *et al.*, JETP Lett. **58**, 306 (1993).
[4] A. N. Vasil'ev, A. R. Keiper, V. V. Kokorin *et al.*, Int. J. Appl. Electromagn. Mater. **5**, 163 (1994).
[5] A. N. Vasil'ev, S. A. Klestov, V. V. Kokorin *et al.*, Zh. Éksp. Teor. Fiz. **109**, 973 (1996) [JETP **82**, 524 (1996)].
[6] A. Zheludev, S. M. Shapiro, and P. Wocher, Phys. Rev. B **51**, 11310 (1995).
[7] K. Ullakko, J. K. Huang, C. Kantner, R. C. O'Handley, and V. V. Kokorin, Appl. Phys. Lett. **69**, 1966 (1996).
[8] M. A. Fradkin, Phys. Rev. B **50**, 16326 (1994).
[9] Yu. M. Gufin, *Structural Phase Transitions* [inRussian], Nauka, Moscow (1982).
[10] Yu. A. Izyumov and V. N. Syromyatnikov, *Phase Transitions and Crystal Symmetry* [in Russian], Nauka, Moscow (1984).
[11] J.-C. Toledano and P. Toledano, *The Landau Theory of Phase Transitions*, World Science Publishers, Singapore (1987).
[12] M. A. Krivoglaz and V. D. Sadovskiĭ, Fiz. Met. Metalloved. **18**, 502 (1964).
[13] T. Kanomata, K. Shirakawa, and T. Kaneko, J. Magn. Magn. Mater. **65**, 76 (1987).

Translated by M. E. Alferieff

Multiple self-propagating high-temperature synthesis and solid-phase reactions in thin films

V. G. Myagkov*) and L. E. Bykova

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

G. N. Bondarenko

Institute of Chemistry and Chemical Technology, Siberian Branch of the Russian Academy of Sciences, 660036 Krasnoyarsk, Russia

(Submitted 3 July 1998)
Zh. Éksp. Teor. Fiz. **115**, 1756–1764 (May 1999)

A variety of self-propagating high-temperature synthesis in thin films has been found and investigated. This modification, called multiple self-propagating high-temperature synthesis, occurs in the solid phase and is a reversible phase transition. Multiple self-propagating high-temperature synthesis is similar in many respects to a metal–insulator phase transition. It is shown that for eutectic systems it is equivalent to a repeated transition through the eutectic temperature of bulk samples. It is inferred that multiple self-propagating high-temperature synthesis in bilayer films is governed by phase separation mechanisms that take place during eutectic solidification and eutectoid decomposition. © *1999 American Institute of Physics.* [S1063-7761(99)01705-9]

1. INTRODUCTION

Solid-phase reactions in thin films have been an object of investigationin the past,[1] and this activity continues unabated (see, for example, Refs. 2–4). Above all else, they are studied because thin layers are the foundation of modern microelectronics. Solid-phase reactions occur at much lower temperatures in thin films than in bulk samples. The products of solid-phase reactions can be not only compounds but also solid solutions of reagents that result from the mixing of layers.[1,5,6] Layer mixing has also been observed during the formation of quasicrystals[7] and in heterostructures.[8] Searches for optimal heat-treatment temperatures, and times at which these reactions occur, are exclusively empirical.

It is believed that the dominant mechanism of solid-phase reactions is diffusion along grain boundaries. However, such an analysis neglects the possibility that self-propagating high-temperature synthesis (SHS) is initiated in thin films. SHS in powders is well known.[9,10] The kinetics and propagation mechanism of an SHS front in bilayer thin films have been described only recently.[11,12] These papers also show that SHS is a modification of solid-phase reactions in thin films. Previously,[13,14] SHS had been observed in multilayers, where initiation was accomplished by a point heat source. Samples for investigating SHS in bilayer films consist of layers of reagents successively deposited on various substrates. SHS occurs between the layers of reagents if the sample (substrate) temperature T_s becomes equal to the initiation temperature T_0. A nucleus of reaction products forms on the sample, and the SHS front propagates along the sample surface.

Experiments show that SHS comes in two forms in thin films. The first is similar to SHS in powders, where reactions produce compounds with relatively high negative formation enthalpy .[9,10] Thus, for the Al/Ni film system an Ni_2A_3 phase is observed in the reaction products after passage of the SHS wave, leaving no trace of Ni or Al.[11,12] The second is characterized by the emergence of a second front following passage of the SHS wave along the sample. This is followed by phase stratification. SHS in Al/Ge films is of the second type, where the products of the reaction largely contain solid solutions of aluminum and germanium, and only a negligible quantity of metastable phases is formed.[15] After the first SHS cycle, the reagents are therefore mixed. Since the original reagents form after the first cycle, SHS can be reinitiated in the sample. Thus, SHS was initiated about 300 times in a single Al/Ge sample, and could be initiated further. This phenomenon, called multiple SHS, emerges only in type-II samples. The motion of the SHS front and phase separation front can be easy to observe visually. The present paper describes the basic characteristics of multiple SHS and its physical interpretation.

2. EXPERIMENTAL PROCEDURE AND EXPERIMENTAL RESULTS

The procedure for obtaining samples and the method of initiating SHS in bilayer thin films are presented in Refs. 11 and 12. In the experiments, Al/Ge samples obtained by successive deposition of germanium and aluminum films on glass or mica substrates were investigated. The thickness of the germanium and aluminum layers ranged from 40 to 200 nm. The phases formed in the synthesis process were identified with a DRON-4-07 diffractometer (Cu K_α radiation). After a sample is heated to the initiation temperature $T_0 = 550 - 600$ K in a uniform temperature field, a nucleus of

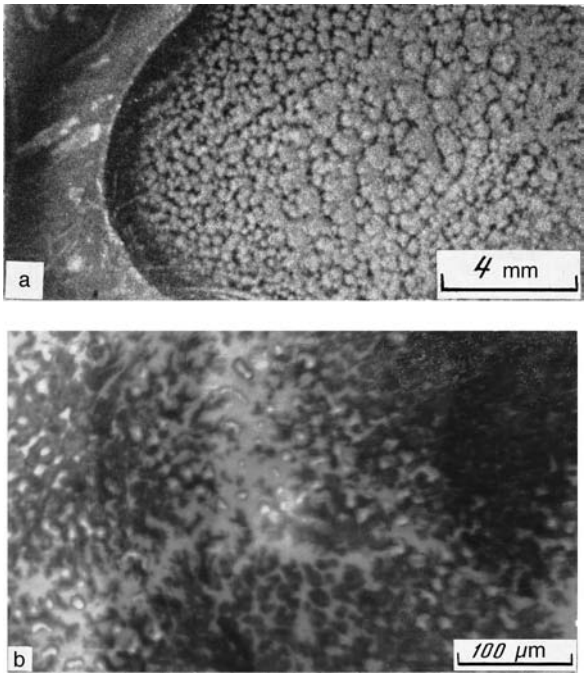

FIG. 1. a) Photograph of an SHS front and DBM clusters in the reaction products of the sample Al(100 nm)/Ge(100 nm). b) Microstructure of a bilayer Al(100 nm)/Ge(100 nm) film sample containing DBM clusters.

a new phase randomly forms on the surface of the sample (Fig. 1a). The SHS front moves along the surface and reflects the temperature topography of the film. The velocity v_f of the SHS front with initiation temperature T_0 is $v_f \simeq 3 \times 10^{-3}$ m/s, and increases with temperature approximately according to the Arrhenius law (Fig. 2). The SHS front can be stopped by reducing the film temperature below the initiation temperature T_0.

A decrease in sample temperature gives rise to a phase separation front, which starts from the boundary left behind by the SHS front. The velocity v_{ph} of the phase separation front increases strongly with decreasing substrate temperature T_s, and starts in the ''soft'' regime (Fig. 2). The emergence of a phase separation front after passage of an SHS wave was first noted in Ref. 15. The first SHS front has a sharp boundary, since there is a difference in reflection from the specular surface of the original film and the surface of the reacted sample, which produces diffuse scattering. The boundary of the subsequent SHS fronts is more diffuse, and it becomes much less appreciable after repeated SHS. Visual

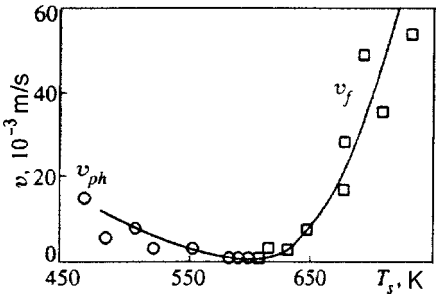

FIG. 2. SHS front velocity $v_f(T_s)$ and the phase separation front velocity $v_{ph}(T_s)$ as a function of substrate temperature T_s.

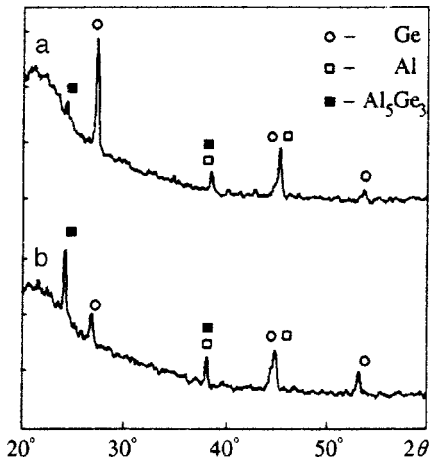

FIG. 3. Diffraction patterns of Al(100 nm)/Ge(120 nm) thin film samples: a) $n=1$, $v_{ph}=1\times10^{-3}$ m/s; $n=300$, $v_{ph}=1\times10^{-3}$ m/s; b) $n=1$, $v_{ph}=1\times10^{-1}$ m/s.

observations show that SHS fronts resulting from repeated initiation always start at the same point and mimic the morphological features of the motion of the first SHS front. The phase composition of Al/Ge films does not depend on the number n of SHS cycles (Fig. 3a), but on the thickness ratio of the reagent layers and the velocity of the phase separation front. For $v_{ph}\simeq 1\times10^{-3}$ m/s and various n, largely solid solutions of aluminum and germanium are produced; only negligible amounts of metastable phases exist (Fig. 3a). However, for $n=1$ and $v_{ph}\simeq 1\times10^{-1}$ m/s, the metastable phases stabilize: an Al_5Ge_3 phase emerges for Al(100 nm)/Ge(100) films (Fig. 3b), and an AlGe phase forms in Al(100 nm)/Ge(120 nm) samples.[15]

Investigations show that the temperature T_{ph} at which a phase separation front emerges is the same as the initiation temperature T_0, which is 100–150 K below the eutectic temperature T_E. In bulk sample, the phases separate after eutectic solidification. It is to be expected that the initiation temperature T_0 in Al/Ge films corresponds to the temperature T_E for a bulk Al–Ge alloy. Since the rate of heat removal from thin films is higher than for bulk samples, T_0 should be less than T_E. The initiation temperature T_0 does not depend on the thickness ratio of the reagent layers, just as the eutectic temperature T_E does not depend on the composition of the alloy.

All this suggests that type-II SHS should emerge in bilayer film systems for which the equilibrium phase diagram is of the eutectic type. This was checked for the Pb–Sn system, which has a simple phase diagram of the eutectic type with eutectic temperature $T_E = 456$ K. SHS in a Pb/Sn bilayer film can be repeatedly initiated at $T_0 = 440-450$ K. Stronger evidence can be gleaned from the initiation of multiple SHS in uniform films obtained by deposition of Pb_xSn_{1-x}, Al_xGe_{1-x} alloys ($0.4<x<0.7$). In this case the initiation temperatures does not depend on x, and are the same as the corresponding temperatures for Pb/Sn and Al/Ge bilayer films. Since the phase separation temperature $T_{ph} = T_0$ in thin films is different from the eutectic temperature T_E, the melting temperature of the film was expected to be the same as either T_0 or T_E.

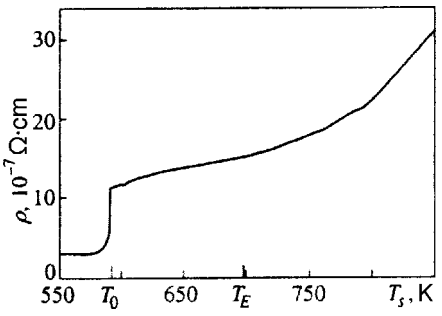

FIG. 4. Resistivity ρ of an Al(100 nm)/Ge(100 nm) bilayer film as a function of substrate temperature T_s.

Figure 4 shows the resistivity $\rho(T_s)$ of an Al(100 nm)/Ge(100 nm) film sample as a function of substrate temperature T_s. It follows from the dependence $\rho(T_s)$ that the Al(100 nm)/Ge(100 nm) sample does not undergo any phase transformations above the temperature T_0. This result was checked directly. The Al(100 nm)/Ge(100 nm) sample was obtained by deposition on a cleavage surface of NaCl, and was then transferred to a glass substrate. Microscopic and visual observations clearly confirm a lack of melting up to $T_s = 850\ \mathrm{K} > T_0$. This is surprising, since phase separation in bulk samples results from eutectic solidification. Nevertheless, multiple SHS in thin films occurs in the solid phase, and is similar to a repeated transition through the eutectic temperature in bulk samples.

The existence of phase separation is also confirmed by resistivity measurements on a film sample as a function of substrate temperature T_s and the number of SHS cycles. Figure 5 shows the resistivity ρ of an Al(100 nm)/Ge(100 nm) film as a function of temperature T_s for three SHS cycles. After initiation of SHS at $T_s > T_0$, the resistivity of the sample increases. At $T_s < T_0$ the resistivity of the sample returns to its original value. The fact that the resistivity of the sample is the same before and after SHS, and that it is the same as the resistivity of the aluminum layer, confirm that aluminum forms a percolation cluster in the film after phase separation. Repeated initiation of SHS increases the initial resistivity somewhat; this might be due to oxidation of the sample by residual oxygen. Multiple SHS is also observed in Al/Si ($T_0 \simeq 700$ K), Al/S ($T_0 \simeq 750$ K), Al/Zn ($T_0 \simeq 770$ K), Au/Ge ($T_0 \simeq 600$ K), and Al/Ti ($T_0 \simeq 770$ K) film systems.

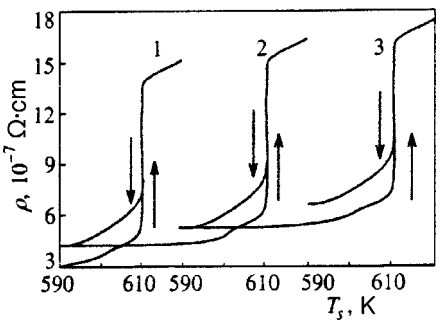

FIG. 5. Resistivity ρ of an Al(100 nm)/Ge(100 nm) film sample as a function of substrate temperature T_s near the initiation temperature T_0 for three SHS cycles. Arrows show the direction of variation of the resistivity.

3. DISCUSSION

The fact that the initiation temperature T_0 is the same as the temperature T_{ph} at which the phase separation front emerges provides a basis for considering the SHS front and the phase separation front to be a single phase-transformation wave. Since during multiple SHS the resistivity changes reversibly and SHS occurs in the solid phase, type-II SHS is a reversible structural phase transition resembling a metal–insulator phase transition. The microstructure of these samples is very diverse, but it consists primarily of clusters with a dense branching morphology (DBM clusters[15]). The microstructure of Al–Ge films was studied in connection with the emergence of fractal formations[16] as well as DBM clusters[17] within them. The intense interest in DBM clusters with fractal dimension $d_f = 2$ is due to the fact that they emerge in many physicochemical and biological systems. In Al–Ge films, DBM clusters comprise a nucleus of polycrystalline germanium, possessing dendritic structure, with single-crystal aluminum disposed among the branches of the dendrite.

The phase separation mechanism in Al–Ge films leading to the formation of DBM clusters was studied in Ref. 17. It is believed that DBM clusters form from an amorphous phase. This agrees with Ref. 18, where it is shown that in Al–Ge films with various concentrations there exists an amorphous phase that transforms into stable aluminum and germanium phases via intermediate metastable phases. The microstructure formed in Al/Ge films depends on the number n of SHS cycles and the velocity v_{ph} of the phase separation front. After the first SHS front has passed and as the phase separation front moves along the surface of the sample, circular nuclei emerge ahead of the front and subsequently merge with the phase separation front. Microscopic investigations show that these nuclei are DBM clusters (see Fig. 1b), which can range in diameter from 10 μm to several millimeters (Fig. 1a). DBM clusters of such sizes are observed during annealing in Bi/Al/Mn/SiO multilayers.[19] At low front velocities ($v_{ph} \simeq 1 \times 10^{-3}$ m/s), laminar microstructure forms perpendicular to the phase separation front. This structure resembles the cellular structure that emerges during directed crystallization.[20]

Investigations of sample surfaces on the substrate side show that SHS proceeds over the entire thickness of the film, even when the layer thickness is $\simeq 1.5\ \mu$m. A total thickness of 3–4 μm is probably the maximum for SHS in bilayer thin films. Subsequent SHS cycles do not alter the original microstructure; this confirms that multiple SHS proceeds in the solid phase.

With long-term initiation of SHS ($n \simeq 300$), the branches of DBM clusters break up and the microstructure of the film becomes uniform. If rapid mass transfer between layers in the first SHS cycle proceeds perpendicular to the film surface, then both SHS and phase separation occur in subsequent cycles along the surface on an interphase boundary between the branches of DBM clusters, which contain germanium, and the single-crystalline aluminum between them. During long-term initiation of SHS, therefore, DBM clusters break down and phase stratification becomes more subtle.

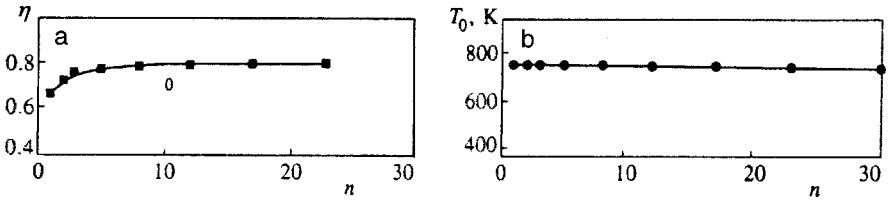

FIG. 6. Degree η of transformation (a) and SHS initiation temperature T_0 (b) of an Al(100 nm) / Fe_2O_3(200 nm) film sample as a function of number n of SHS cycles.

Assuming that diffusion between layers precedes to a depth equal to the thickness of the reagent layer, $d=2\times10^{-6}$ m, and its velocity equals the velocity of the phase separation front $v_{ph}=0.2\times10^{-1}$ m/s, we can estimate the diffusion coefficient during SHS in thin layers to be $D\simeq dv_{ph}=0.4\times10^{-7}$ m^2/s. This is 9–10 orders of magnitude greater than the diffusion coefficient along grain boundaries,[1] and 1–2 orders of magnitude greater than the diffusion coefficient in the liquid phase. The actual value of the diffusion coefficient during SHS can be even higher, so that SHS in thin films must be attributed to explosive chemical reactions.

Explosive chemical reactions in solids resulting from simultaneous uniaxial deformation and shear deformation are described in Refs. 21 and 22. The estimated diffusion coefficient in such reactions is 10^{10}–10^{15} times the value ordinarily observed in the solid phase. The proposed mechanism involves the avalanche-like emergence of structural defects at the instant elastic stresses relax. Such a solid is in a special state that is neither solid, liquid, nor gaseous. At that instant, the solid becomes permeable, and there is enough time for explosive reactions to occur.[21,22] An explosive reaction in the solid phase occurs between Al and Fe_2O_3.[22] The metallic–thermal reaction between Al and Fe_2O_3 has been studied in some depth; its initiation temperature is $T_0=1400-1500$ K, and the front temperature is approximately 2300 K.

In the present work, SHS was studied in Al/Fe_2O_3 bilayer films,which are type-II and have an initiation temperature $T_0=750-770$ K. The bulk abundance of iron in the sample produced after the reaction and the degree η of transformation were determined by the torque method.[6,7] Figure 6a shows the degree η of transformation as a function of the number n of SHS cycles. It is clear from the function $\eta(n)$ that the Fe abundance in the sample increases until $n=5$, due to preignition. For $n>5$ the degree of transformation is independent of n. This confirms that multiple SHS is not governed by the exothermal nature of the reaction between Al and Fe_2O_3.

The initiation temperature T_0 does not depend on the number n of SHS cycles (Fig. 6b). Such dependences $T_0(n)$ are observed in all bilayer film systems in which multiple SHS was obtained. It follows, then, that T_0 in Al/Fe_2O_3 films is analogous to the eutectic temperature of bulk Al–Fe_2O_3 samples. For eutectic solidification and eutectoid decomposition, phase separation often entails the formation of platelike structures, where the plate thicknesses can reach several microns. Multilayers (including bilayer films) are artificial analogs of such platelike microstructures. The kinetics and mechanism of multiple SHS in thin films should therefore be the same as in the case of the formation and decomposition of plate-like structures during transitions near the eutectic temperature or the temperature of eutectoid decomposition of bulk samples.

The following mechanism of multiple SHS in thin films can be inferred from the experimental data presented above. Solid-phase amorphization of the samples occurs after the first SHS front has passed. Below $T_0=T_{ph}$, the amorphous phase decomposes, depending on the velocity of the phase separation front and the thickness ratio. For eutectic systems, there is a correspondence between the phase equilibrium diagram of the reaction products of bulk samples and the characteristics of SHS. The eutectic temperature T_E and the percentage abundance of reaction products determine the initiation temperature T_0 and thickness ratio, respectively, while the liquid eutectic corresponds to the amorphous phase. For $n>2$, SHS does not encompass the entire sample, proceeding instead only at phase boundaries to a thickness of 3–4 μm. Large stresses are produced at the interphase boundary ahead of the SHS and phase separation fronts, producing zones of structural defects. Diffusion is greatly facilitated in these zones, and conditions for explosive reactions set in.[21,22] Metastable phases might play a significant role in the initiation of multiple SHS, since their formation enthalpy can be high.

4. CONCLUSIONS

In conclusion, we note that solid-phase reactions in thin films that are in fact SHS reactions occur only at temperature T_0. Rapid mass transfer and diffusion mixing at the atomic level occur only when SHS and phase separation fronts pass. After the passage of a phase separation front, diffusion in the film sample once again becomes negligible. This suggests that layer mixing, often observed in multilayer (bilayer) films subject to heat treatments and thermal influences,[1,5–8,20] occurs after type-II SHS in these samples. Multiple SHS is a reversible structural phase transition, similar to a metal–insulator phase transition, and can be used in microelectronics devices.

Multiple SHS corresponds to a transition through the eutectic temperature in bulk samples. At present the mechanisms of SHS and solid-phase reactions in thin films are not completely understood. However, the multiple-SHS phenomenon clearly indicates that these mechanisms are related to phase separation mechanisms observed during eutectic solidification and eutectoid decomposition. It is perhaps surprising, considering the many recent studies of solid-phase reactions by various methods, that SHS and multiple SHS in thin films went unnoticed, even though these pheonomena can be observed at atmospheric pressure (no vacuum is required), using straightforward experimental techniques.

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

[1] *Thin Films: Interdiffusion and Reactions*, J. M. Poate, K. Tu, and J. Meier (eds.), Wiley, New York (1978).
[2] L. A. Clavenger, B. Arcot, W. Ziegler *et al.*, J. Appl. Phys. **83**, 9099 (1998).
[3] J. S. Huang, S. S. Huang, K. N. Tu *et al.*, J. Appl. Phys. **82**, 644 (1997).
[4] L. Balzac, V. Freury, F. Duclos, and V. Van Herpen, Phys. Rev. E **54**, 599 (1996).
[5] A. F. Jankovski, L. R. Schrawyer, and M. A. Wall, J. Appl. Phys. **68**, 5162 (1990).
[6] H.-J. Voorma, E. Louis, N. B. Koster, and F. Biykerk, J. Appl. Phys. **83**, 4700 (1998).
[7] D. A. Lilienfeld, M. Nastasi, N. N. Johnson *et al.*, Phys. Rev. Lett. **55**, 1587 (1985); J. A. Knapp and D. M. Follstadt, Phys. Rev. Lett. **55**, 1591 (1985).
[8] D. G. Deppe and N. Holonyak, Jr., J. Appl. Phys. **64**, R93 (1988).
[9] A. G. Merzhanov, in *Physical Chemistry*, Kolotyrkin (ed.), Khimiya, Moscow (1983), p. 6.
[10] Z. A. Munir and U. Anselmi-Tamburini, Mater. Sci. Rep. **3**, 277 (1989).
[11] V. G. Myagkov and L. E. Bykova, Dokl. Ross. Akad. Nauk **354**, 777 (1997).
[12] V. G. Myagkov, V. S. Zhigalov, L. E. Bykova, and V. K. Mal'tsev, Zh. Tekh. Fiz. **68**(10), 58 (1998) [Tech. Phys. **43**, 1189 (1998)].
[13] T. S. Dyer, Z. A. Munir, and V. Ruth, Scr. Metall. Mater. **30**, 1281 (1994).
[14] E. Ma, C. V. Thompson, L. A. Clavenger, and K. N. Tu, Appl. Phys. Lett. **57**, 1262 (1990).
[15] V. G. Myagkov and L. E. Bykova, JETP Lett. **67**, 334 (1998).
[16] J. G. Hou and Z. Q. Wu, Phys. Rev. B **42**, 3271 (1990); B. Q. Li, B. Zheng, and Z. Q. Wu, Phys. Rev. B **47**, 3638 (1993).
[17] E. Ben Yakov, G. Deutscher, P. Garek *et al.*, Phys. Rev. Lett. **57**, 1903 (1986); S. Alexander, R. Bruisma, R. Hilfer *et al.*, Phys. Rev. Lett. **60**, 1514 (1988); G. Deutscher and Y. Lereah, Phys. Rev. Lett. **60**, 1510 (1988); Y. Lereah, G. Deutscher, and E. Grunbaum, Phys. Rev. A **44**, 8316 (1991); Y. Lereah, I. Zarudi, E. Grunbaum *et al.*, Phys. Rev. E **49**, 649 (1994).
[18] U. Koster, Acta Metall. **20**, 1361 (1972).
[19] J. S. Langer, Rev. Mod. Phys. **52**, 1 (1980).
[20] C. H. Shang, Phys. Rev. B **53**, 13759 (1996).
[21] N. S. Enikolopyan, A. A. Mkhitaryan, and A. S. Karagezyan, Dokl. Akad. Nauk SSSR **294**, 912 (1987).
[22] N. S. Enikolopyan, Zh. Fiz. Khim. **63**, 2289 (1989).

Translated by M. E. Alferieff

Models of the pseudogap state of two-dimensional systems

É. Z. Kuchinskiĭ*) and M. V. Sadovskiĭ†)

Institute of Electrophysics, Ural Branch of the Russian Academy of Sciences, 620049 Ekaterinburg, Russia
(Submitted 9 September 1998)
Zh. Éksp. Teor. Fiz. **115**, 1765–1785 (May 1999)

We analyze several almost exactly solvable models of the electronic spectrum of two-dimensional systems with well-developed short-range-order dielectric (e.g., antiferromagnetic) or superconducting fluctuations that give rise to an anisotropic pseudogap state in certain segments of the Fermi surface. We develop a recurrence procedure for calculating the one-electron Green's function that is equivalent to summing all Feynman diagrams. The procedure is based on an approximate ansatz for higher order terms in the perturbation series. We do detailed calculations of the spectral densities and the one-electron density of states. Finally, we analyze the limits of the adopted approximations and some important points concerning the substantiation of these approximations. © *1999 American Institute of Physics.* [S1063-7761(99)01805-3]

1. INTRODUCTION

In recent years there has been an upsurge of interest in observations of the pseudogap in the spectrum of elementary excitations of high-T_c superconductors in the range of current carrier concentrations below the optimum.[1,2] The corresponding anomalies were observed in a number of experiments, such as measurements of optical conductivity, NMR, inelastic neutron scattering, and angle-resolved photoemission (ARPES; see the review cited in Ref. 1). Probably the most striking evidence that such an unusual state exists was obtained in ARPES experiments,[3,4] which demonstrated the presence of essentially anisotropic changes in the current-carrier spectral density within a broad temperature range in the normal (nonsuperconducting) phase of these systems (see the review in Ref. 2). A remarkable feature observed in these experiments was the presence of a maximum of the corresponding anomalies close to the point $(\pi,0)$ in the Brillouin zone, while no such anomalies were observed in the direction of the zone diagonal [the point (π,π)], which actually means that near the point $(\pi,0)$ the Fermi surface is destroyed, while the Fermi-liquid behavior in the direction of the zone diagonal is retained. In this sense it is usually said that the pseudogap symmetry is of the *d*-wave type, which coincides with the symmetry of the superconducting energy gap in these compounds.[1,2] At the same time, the very fact that these anomalies exist at temperatures much higher than the superconducting transition temperature and at nonoptimal carrier concentrations could point to a different nature of these anomalies, not related directly to Cooper pairing.

There are many theoretical papers in which the authors attempt to explain the observed anomalies. Two main areas of such research can be identified. One is based on the idea that Cooper pairs form at temperatures higher than the superconducting transition temperature.[1,5–7] In the other it is assumed that pseudogap phenomena are due primarily to antiferromagnetic (AFM) short-range-order fluctuations.[8–12]

Some time ago one of the authors of the present paper (M.V.S.) proposed an exactly solvable model of pseudogap formation in a one-dimensional system due to well-developed short-range-order charge density wave (CDW) or spin density wave (SDW) fluctuations (see Refs. 13–17). Recently this model has attracted the attention of researchers in connection with attempts to explain the pseudogap state of high-T_c cuprates.[11,12,18–20] In particular, Schmalian *et al.*[11,12] made an important generalization of this model to the case of a two-dimensional system of electrons that is in the random field of well-developed spin fluctuations (short-range-order AFM fluctuations). In the model of hot spots on the Fermi surface developed in Refs. 11 and 12, the researchers obtained, via the formal scheme developed in Refs. 15–17, a detailed description of pseudogap anomalies at high temperatures (the weak-pseudogap region). Tchernyshyov[19] and Ren[20] used a simplified variant of the model developed in Refs. 13 and 14, which corresponds to the limit of very large correlation lengths of short-range-order fluctuations, to describe the pseudogap state determined by well-developed fluctuations of superconducting (SC) short range order. In a recent paper,[21] this simplified model was used to analyze the Ginzburg–Landau expansion (for different types of Cooper pairing) in a system with strong CDW (SDW, AFM) fluctuations using the model of hot patches on the Fermi surface proposed in the paper. At the same time, Tchernyshyov[22] reviewed in detail the model developed in Refs. 13–17 and found an error in the earlier papers[15–17] in the analysis of the case of finite correlation lengths of short-range-order fluctuations. In Ref. 12 it was suggested that this error is insignificant, especially in analyzing the two-dimensional hot-spot model, which is of the main interest to the physics of high-T_c systems.

The aim of the present paper is to analyze a number of important aspects of the almost exactly solvable model, mainly in the two-dimensional case. To this end we consider both the case of short-range-order CDW (SDW, AFM) fluctuations in the hot-spot model[11,12] and the possibility of using the model within the framework of fluctuation Cooper

pairing[7,19,20] (SC short-range-order fluctuations), in particular, in the most interesting case of *d*-wave pairing. In addition to a general analysis of the reliability of the formal scheme used in Refs. 11–17, we do detailed calculations of the spectral density and the one-electron density of states for the hot-spot model[11,12] and in the scenario of fluctuation Cooper pairing.

2. THE HOT-SPOT MODEL

2.1. Description of model and an "almost exact" solution for the Green's function

The model of a nearly ferromagnetic Fermi liquid[23,24] is based on the picture of well-developed fluctuations of AFM short-rang-order fluctuations within a wide region of the phase diagram of high-T_c systems. This model introduces the effective interaction of electrons and spin fluctuations that is described by the dynamic spin susceptibility $\chi_{\mathbf{q}}(\omega)$, which is determined mainly from the fit to the data of NMR experiments:[24]

$$V_{\text{eff}}(\omega,\mathbf{q})=g^2\chi_{\mathbf{q}}(\omega)\approx\frac{g^2\xi^2}{1+\xi^2(\mathbf{q}-\mathbf{Q})^2-i\omega/\omega_{\text{sf}}}, \tag{1}$$

where g is the coupling constant, ξ is the correlation length of the spin fluctuations, $\mathbf{Q}=(\pi/a,\pi/a)$ is the vector of antiferromagnetic ordering in the insulator phase, ω_{sf} is the characteristic frequency of spin fluctuations, and a is the lattice constant (of a square lattice).

Since the dynamic spin susceptibility $\chi_{\mathbf{q}}(\omega)$ has peaks at wave vectors that are in the vicinity of $(\pi/a,\pi/a)$, two types of quasiparticle arise in the system: "hot" quasiparticles with momenta in the vicinity of hot spots on the Fermi surface, and "cold" quasiparticles with momenta in the parts of the Fermi surface surrounding the diagonals of the Brillouin zone, $|p_x|=|p_y|$ (see Refs. 11 and 12). Such terminology is related to the fact that quasiparticles from the vicinity of hot spots are strongly scattered through a vector of order $\mathbf{Q}$ by spin fluctuations (1), while for particles with momenta far from hot spots this interaction is relatively weak.

In what follows we consider the case of high temperatures, $\pi T\gg\omega_{\text{sf}}$, which corresponds to the "weak pseudogap" region in the phase diagram.[11,12] In this case spin dynamics in irrelevant and we can limit ourselves to the static approximation:

$$V_{\text{eff}}(\mathbf{q})=\tilde{\Delta}^2\frac{\xi^2}{1+\xi^2(\mathbf{q}-\mathbf{Q})^2}, \tag{2}$$

where $\tilde{\Delta}$ is an effective parameter with the dimensions of energy, which in the model of AFM fluctuations can be written[12]

$$\tilde{\Delta}^2=g^2T\sum_{m\mathbf{q}}\chi_{\mathbf{q}}(i\omega_m)=g^2\langle\mathbf{S}_i^2\rangle/3\,, \tag{3}$$

with $\mathbf{S}_i$ the spin at a lattice site (Cu ions in the CuO_2 plane for high-T_c cuprates). Below we consider $\tilde{\Delta}$ (as well as ξ) a phenomenological parameter that determines the effective width of the pseudogap.

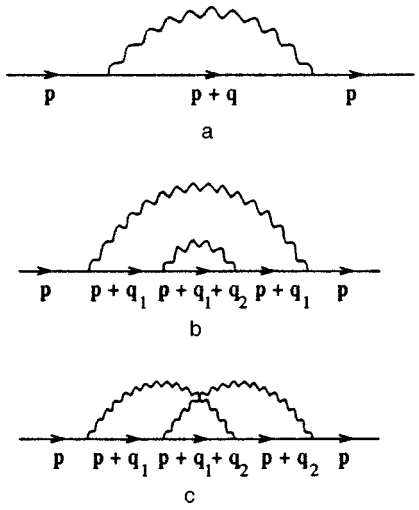

FIG. 1. First- and second-order self-energy diagrams for an electron interacting with short-range-order fluctuations.

Calculations can be simplified significantly if we replace (2) with a model interaction of the form (cf. a similar model in Ref. 8)

$$V_{\text{eff}}(\mathbf{q})=\Delta^2\frac{2\xi^{-1}}{\xi^{-2}+(q_x-Q_x)^2}\,\frac{2\xi^{-1}}{\xi^{-2}+(q_y-Q_y)^2}, \tag{4}$$

where $\Delta^2=\tilde{\Delta}^2/4$. Actually, Eq. (4) is quite similar to (2) and differs quantitatively very little in the most important region $|\mathbf{q}-\mathbf{Q}|<\xi^{-1}$.

Consider the first-order correction in V_{eff} to the electron self-energy, represented by the diagram in Fig. 1a:

$$\Sigma(\varepsilon_n,\mathbf{p})=\sum_{\mathbf{q}}V_{\text{eff}}(\mathbf{q})\frac{1}{i\varepsilon_n-\xi_{\mathbf{p}+\mathbf{q}}}. \tag{5}$$

The main contribution to the sum over $\mathbf{q}$ is provided by the region close to $\mathbf{Q}=(\pi/a,\pi/a)$. Then, writing

$$\xi_{\mathbf{p}+\mathbf{q}}=\xi_{\mathbf{p}+\mathbf{Q}+\mathbf{k}}\approx\xi_{\mathbf{p}+\mathbf{Q}}+\mathbf{v}_{\mathbf{p}+\mathbf{Q}}\cdot\mathbf{k}, \tag{6}$$

where $v^{\alpha}_{\mathbf{p}+\mathbf{Q}}=\partial\xi_{\mathbf{p}+\mathbf{Q}}/\partial p_{\alpha}$, and integrating over $\mathbf{k}$, we obtain[1)]

$$\Sigma(\varepsilon_n,\mathbf{p})=\frac{\Delta^2}{i\varepsilon_n-\xi_{\mathbf{p}+\mathbf{Q}}+(|v^x_{\mathbf{p}+\mathbf{Q}}|+|v^y_{\mathbf{p}+\mathbf{Q}}|)\kappa\,\text{sign}\,\varepsilon_n}, \tag{7}$$

with $\kappa=\xi^{-1}$.

The spectrum of bare (free) quasiparticles can be taken from Refs. 11 and 12:

$$\xi_{\mathbf{p}}=-2t(\cos p_xa+\cos p_ya)-4t'\cos p_xa\cos p_ya, \tag{8}$$

where t is the nearest-neighbor-hopping integral, t' is the next-nearest-neighbor-hopping integral for a square lattice, and μ is the chemical potential. When real high-T_c systems were analyzed in Refs. 11 and 12, it was assumed, e.g., for $YBa_2Cu_3O_{6+\delta}$, that $t=0.25$ eV and $t'=-0.45t$, and μ was fixed by hole concentration. Below we show that the analysis of the situation for different relationships between t and t' produces interesting results.

We now turn to second-order corrections to self-energy, which are depicted in Figs. 1b and 1c. Using (4) we obtain

$$\Sigma(\mathrm{b})=\Delta^4\int\frac{d\mathbf{k}_1}{\pi^2}\int\frac{d\mathbf{k}_2}{\pi^2}\frac{\kappa}{\kappa^2+k_{1x}^2}\frac{\kappa}{\kappa^2+k_{1y}^2}\frac{\kappa}{\kappa^2+k_{2x}^2}\times\frac{\kappa}{\kappa^2+k_{2y}^2}\frac{1}{i\varepsilon_n-\xi_{\mathbf{p}+\mathbf{Q}}-v^x_{\mathbf{p}+\mathbf{Q}}k_{1x}-v^y_{\mathbf{p}+\mathbf{Q}}k_{1y}}\times\frac{1}{i\varepsilon_n-\xi_{\mathbf{p}}-v^x_{\mathbf{p}}(k_{1x}+k_{2x})-v^y_{\mathbf{p}}(k_{1y}+k_{2y})}\times\frac{1}{i\varepsilon_n-\xi_{\mathbf{p}+\mathbf{Q}}-v^x_{\mathbf{p}+\mathbf{Q}}k_{1x}-v^y_{\mathbf{p}+\mathbf{Q}}k_{1y}},\tag{9}$$

$$\Sigma(\mathrm{c})=\Delta^4\int\frac{d\mathbf{k}_1}{\pi^2}\int\frac{d\mathbf{k}_2}{\pi^2}\frac{\kappa}{\kappa^2+k_{1x}^2}\frac{\kappa}{\kappa^2+k_{1y}^2}\frac{\kappa}{\kappa^2+k_{2x}^2}\times\frac{\kappa}{\kappa^2+k_{2y}^2}\frac{1}{i\varepsilon_n-\xi_{\mathbf{p}+\mathbf{Q}}-v^x_{\mathbf{p}+\mathbf{Q}}k_{1x}-v^y_{\mathbf{p}+\mathbf{Q}}k_{1y}}\times\frac{1}{i\varepsilon_n-\xi_{\mathbf{p}}-v^x_{\mathbf{p}}(k_{1x}+k_{2x})-v^y_{\mathbf{p}}(k_{1y}+k_{2y})}\times\frac{1}{i\varepsilon_n-\xi_{\mathbf{p}+\mathbf{Q}}-v^x_{\mathbf{p}+\mathbf{Q}}k_{2x}-v^y_{\mathbf{p}+\mathbf{Q}}k_{2y}},\tag{10}$$

where we have employed the spectrum (8), from which, in particular, it follows that $\xi_{\mathbf{p}+2\mathbf{Q}}=\xi_{\mathbf{p}}$ and $\mathbf{v}_{\mathbf{p}+2\mathbf{Q}}=\mathbf{v}_{\mathbf{p}}$ at $\mathbf{Q}=(\pi/a,\pi/a)$. If $v^x_{\mathbf{p}}$ and $v^y_{\mathbf{p}+\mathbf{Q}}$ are of the same sign, the integrals in (9) and (10) are determined solely by the poles of the Lorentzians determining the interaction with short-range-order fluctuations. Doing an elementary contour integration, we get[2)]

$$\Sigma(\mathrm{b})=\Sigma(\mathrm{c})=\Delta^4\frac{1}{[i\varepsilon_n-\xi_{\mathbf{p}+\mathbf{Q}}+i(|v^x_{\mathbf{p}+\mathbf{Q}}|+|v^y_{\mathbf{p}+\mathbf{Q}}|)\kappa]^2}\times\frac{1}{i\varepsilon_n-\xi_{\mathbf{p}}+i2(|v^x_{\mathbf{p}}|+|v^y_{\mathbf{p}}|)\kappa}.\tag{11}$$

Here and below we assume, for the sake of definiteness, that ε_n is positive. Clearly, when the velocity projections are of the same sign, we can use this approach to calculate the contributions of any higher-order diagrams. Accordingly, the contribution of an Nth-order diagram to the self-energy part in the interaction (4) is

$$\Sigma^{(N)}(\varepsilon_n,\mathbf{p})=\Delta^{2N}\prod_{j=1}^{2N-1}\frac{1}{i\varepsilon_n-\xi_j+in_jv_j\kappa},\tag{12}$$

where $\xi_j=\xi_{\mathbf{p}+\mathbf{Q}}$ and $v_j=|v^x_{\mathbf{p}+\mathbf{Q}}|+|v^y_{\mathbf{p}+\mathbf{Q}}|$ for odd j, and $\xi_j=\xi_{\mathbf{p}}$ and $v_j=|v^x_{\mathbf{p}}|+|v^y_{\mathbf{p}}|$ for even j. Here n_j is the number of interaction lines surrounding the jth Green's function in a given diagram.

In this case any diagram with crossing interaction lines is equal to a diagram of the same order with noncrossing interaction lines. Hence actually we may consider only diagrams with noncrossing interaction lines, taking into account the diagrams with crossing lines by introducing additional combinatorial factors into the interaction vertices. This method was first introduced (in another problem) by Elyutin[25] and was used in Refs. 15–17 for a one-dimension model of the pseudogap state.

As a result we arrive at the following expression for the one-electron Green's function in the form of a recurrence relation (the continued fraction representation; see Refs. 15–17):

$$G^{-1}(\varepsilon_n,\xi_{\mathbf{p}})=G_0^{-1}(\varepsilon_n,\xi_{\mathbf{p}})-\Sigma_1(\varepsilon_n,\xi_{\mathbf{p}}),\tag{13}$$

$$\Sigma_k(\varepsilon_n,\xi_{\mathbf{p}})=\Delta^2\frac{v(k)}{i\varepsilon_n-\xi_k+ikv_k\kappa-\Sigma_{k+1}(\varepsilon_n,\xi_{\mathbf{p}})},\tag{14}$$

where $\xi_k=\xi_{\mathbf{p}+\mathbf{Q}}$ and $v_k=|v^x_{\mathbf{p}+\mathbf{Q}}|+|v^y_{\mathbf{p}+\mathbf{Q}}|$ for odd k, and $\xi_k=\xi_{\mathbf{p}}$ and $v_k=|v^x_{\mathbf{p}}|+|v^y_{\mathbf{p}}|$ for even k. The combinatorial factor

$$v(k)=k\tag{15}$$

corresponds to our case of commensurate fluctuations with $\mathbf{Q}=(\pi/a,\pi/a)$ (see Ref. 15). Clearly, one can easily analyze the vase of incommensurate fluctuations, where $\mathbf{Q}$ is not locked to the period of the reciprocal lattice. In this case, diagrams with interaction lines surrounding an odd number of vertices are significantly smaller than diagrams with interaction lines surrounding an even number of vertices. Hence only the latter diagrams should be taken into account.[13–17] As a result, the recurrence relation (14) is retained, but the combinations of the diagrams and hence the combinatorial factor change:[15]

$$v(k)=\begin{cases}\dfrac{k+1}{2} & \text{for } k \text{ odd},\\[2mm] \dfrac{k}{2} & \text{for } k \text{ even}.\end{cases}\tag{16}$$

In Refs. 11 and 12, the spin structure of the interaction in the ''almost antiferromagnetic'' Fermi-liquid model (the spin-fermion model of Ref. 12) was taken into account. This leads to more complicated combinations in the commensurate case with $\mathbf{Q}=(\pi/a,\pi/a)$. More precisely, spin-conserving scattering yields formally commensurate combinations, while spin-flip scattering is described by diagrams of the incommensurate type (a ''charged'' random field, to use the terminology of Ref. 12). As result, the recurrence relation for the Green's function is still of the form (14), but the combinatorial factor $v(k)$ is now[11,12]

$$v(k)=\begin{cases}\dfrac{k+2}{3} & \text{for } k \text{ odd},\\[2mm] \dfrac{k}{3} & \text{for } k \text{ even}.\end{cases}\tag{17}$$

As noted earlier, the solution (14) can be obtained only if the signs of the velocity projections $v^x_{\mathbf{p}+\mathbf{Q}}(v^y_{\mathbf{p}+\mathbf{Q}})$ and $v^x_{\mathbf{p}}(v^y_{\mathbf{p}})$ are the same. Below we analyze the situation when this is really the case. When the signs are different, the integrals of the form (9) and (10), corresponding to higher-order corrections, cannot be calculated in such a simple form as

above because contributions from the poles of the electron Green's functions become important. Here instead of simple expressions of the form (11) we have much more complicated expressions and (even more importantly) the very fact that broad classes of diagrams with crossing and noncrossing interaction lines are equal is not true any more [the reader will recall that it was this fact that made it possible to classify higher-order contributions and to obtain the ''exact'' solution (14)]. This problem is important only for the case of finite correlation lengths $\xi=\kappa^{-1}$ of fluctuations, while in the limit $\xi\to\infty$ ($\kappa\to 0$) the exact solution for the Green's function is independent of the velocities $\mathbf{v_p}$ and $\mathbf{v_{p+Q}}$ and can easily be obtained in analytic form by the methods developed in Refs. 13 and 14 (see also Ref. 12). In the one-dimensional model considered in Refs. 13–17, the signs of the corresponding velocity projections are always different (they correspond to electrons travelling ''right'' and ''left''). This fact was stressed in a recent paper by Tchernyshyov.[22] In the Appendix we analyze these difficulties in detail for the one-dimensional case and show that the ansatz of the form (12) used in Refs. 15–17 for the contributions of higher-order diagrams and the solution (14) yield a very good approximation even when the velocity projections have opposite signs. Obviously, this solution is exact in the limits $\xi\to\infty$ ($\kappa\to 0$) and $\xi\to 0$ ($\kappa\to\infty$) and provides a fairly good (quantitative) description in the region of finite correlation lengths.

2.2. Analysis of the spectrum

For the energy spectrum (8) we can easily specify the conditions (the relationships between t, t', and μ) for the solution (14) to be exact. First, let us define the region of the parameters t, t', and μ where there are hot spots on the Fermi surface, i.e., the conditions for the existence of points connected by the vector $\mathbf{Q}=(\pi/a,\pi/a)$. If $\mathbf{p}=(p_x,p_y)$ specifies the position of a hot spot on the Fermi surface, the point $\mathbf{p}+\mathbf{q}=(p_x+\pi/a, p_y+\pi/a)$ must also belong to the Fermi surface, so that for the spectrum (8) we have

$$-2t(\cos p_x a+\cos p_y a)-4t'\cos p_x a\cos p_y a-\mu=0, \tag{18}$$

$$2t(\cos p_x a+\cos p_y a)-4t'\cos p_x a\cos p_y a-\mu=0.$$

This yields the conditions needed for hot spots to exist:

$$\cos p_y a=-\cos p_x a,\quad \cos^2 p_x a=\mu/4t'. \tag{19}$$

Thus, hot spots on the Fermi surface exist if

$$0\leqslant \mu/4t'\leqslant 1. \tag{20}$$

We now define the region of the parameters t, t', and μ where the solution (14) is exact by requiring that the products $v^x_{\mathbf{p}}v^x_{\mathbf{p+Q}}$ and $v^y_{\mathbf{p}}v^y_{\mathbf{p+Q}}$ be positive. We have

$$v^x_{\mathbf{p}}=\frac{\partial\xi_{\mathbf{p}}}{\partial p_x}=2ta\sin p_x a+4t'a\sin p_x a\cos p_y a,$$

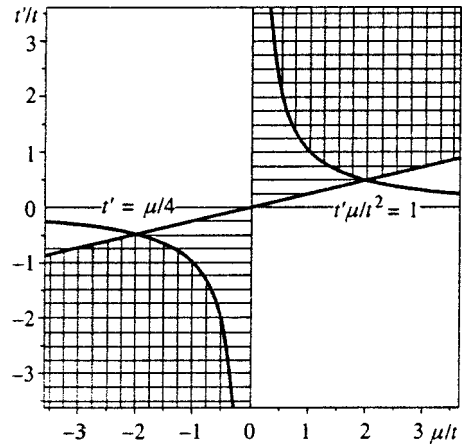

FIG. 2. The region of parameters where hot spots exist (hatched) and the region where such spots exist and the velocity projections have the same sign (doubly hatched).

$$v^y_{\mathbf{p}}=\frac{\partial\xi_{\mathbf{p}}}{\partial p_y}=2ta\sin p_y a+4t'a\sin p_y a\cos p_x a,$$

$$v^x_{\mathbf{p}}v^x_{\mathbf{p+Q}}=16t'^2a^2\sin^2 p_x a\left[\cos^2 p_y a-\left(\frac{t}{2t'}\right)^2\right],$$

$$v^y_{\mathbf{p}}v^y_{\mathbf{p+Q}}=16t'^2a^2\sin^2 p_y a\left[\cos^2 p_x a-\left(\frac{t}{2t'}\right)^2\right]. \tag{21}$$

Clearly, for the Fermi surface to have points where the projections of velocities have the same sign, $|t'/t|$ must be greater than 1/2. Here we are chiefly interested in the region surrounding the hot spots, where on account of (19) we have

$$v^x_{\mathbf{p}}v^x_{\mathbf{p+Q}}=v^y_{\mathbf{p}}v^y_{\mathbf{p+Q}}=4t^2a^2\left(1-\frac{\mu}{4t'}\right)\left(\frac{\mu t'}{t^2}-1\right). \tag{22}$$

Thus, the projections of velocities at hot spots have the same sign if

$$\mu t'/t^2>1. \tag{23}$$

Obviously, the same condition ensures that $\mathbf{v_p}\mathbf{v_{p+Q}}$ is positive (this is needed for the solution (14) to be valid in the model described in Refs. 11 and 12).

Figure 2 depicts the region of parameters where hot spots exist (the hatched area), or $0\leqslant\mu/4t'\leqslant 1$, and the region where such spots exist and the velocity projections have the same sign ($\mu t'>1$). Figure 3 depicts, for different values of the chemical potential μ (band filling), the Fermi surfaces specified by the spectrum (8) for which these conditions are either met or not met.

2.3. Spectral density and density of states

Let us examine the spectral density

$$A(E,\mathbf{p})=-\frac{1}{\pi}\operatorname{Im}G^R(E,\mathbf{p}), \tag{24}$$

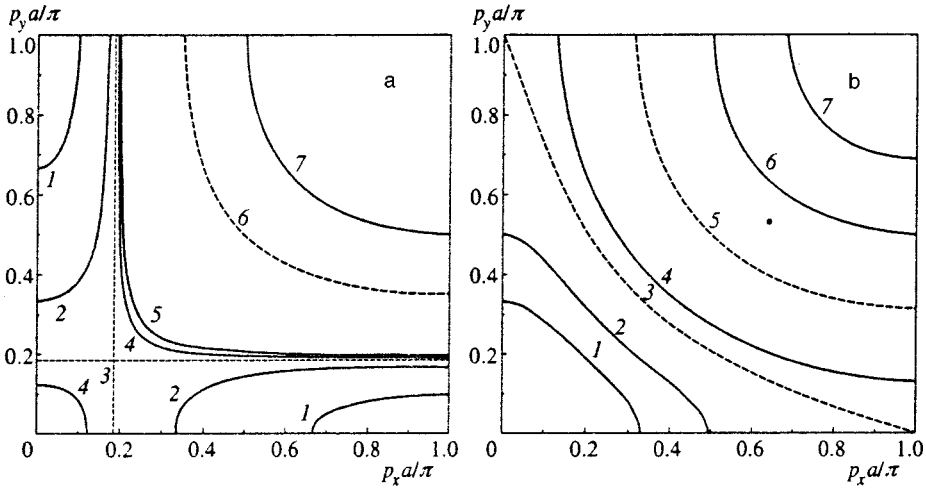

FIG. 3. Fermi surfaces defined by the spectrum (8) for different values of the chemical potential μ (band filling) and the parameter t'/t. (a) The case where $t'/t=-0.6$ and μ/t has the following values: curve *1*, -2.2; curve *2*, -1.8; curve *3*, -1.666; curve *4*, -1.63; curve *5*, -1.6; curve *6*, 0; and curve *7*, 2; the solution (14) is exact in the vicinity of hot spots (the velocity projections are of the same sign) for $\mu/t<-1.666\ldots$, and hot spots exist if μ/t is negative. (b) The case where $t'/t=-0.4$ (which is characteristic of high-T_c cuprates) and μ/t has the following values: curve *1*, -2.2; curve *2*, -2; curve *3*, -1.6; curve *4*, -1.3; curve *5*, 0; curve *6*, 2; and curve *7*, 4; hot spots exist if $-1.6<\mu/t<0$.

where $G^R(E,\mathbf{p})$ is the retarded Green's function obtained by ordinary analytic continuous of (13) into the real energy axis E. Figure 4 depicts the energy dependence of $A(E,\mathbf{p})$ obtained from (13) and (14) for different variants of the combinatorial factors (15) and (16). Since the energy dependence of the spectral density in the case of the combinations (17) for the spin-fermion model is qualitatively (and even quantitatively) very close to that obtained in the incommensurate case, Eq. (16), we have not displayed it in Fig. 4a so as to save space. For $t'/t=-0.6$ and $\mu/t=-1.8<t/t'=1.666$, the projections of the velocities at the hot spots have the same sign and the solution (14) defines the Green's function exactly. We see that in the incommensurate case (16) (Fig. 4a) as well as for the combinations (17) of the spin-fermion model, the spectral density at a hot spot clearly exhibits non-Fermi-liquid behavior (for large values of the correlation length ξ of the fluctuations). In the case of commensurate combinations, Eq. (15) (Fig. 4b), it is precisely at a hot spot that the spectral density has a single peak and, in this sense, is similar to the spectral density of an ordinary Fermi liquid even when ξ is large. However, even in the vicinity of a hot spot the spectral density acquires two non-Fermi-liquid peaks (the ''shadow'' band) for large values of ξ (see the inset in Fig. 4b).

Far from hot spots, the velocity projections have, in general, opposite signs, even if condition (23) is met. Accordingly, the recurrence relation (14) for the Green's function is not exact. At the same time, as ξ increases, the region with the hot spot in the momentum space narrows and the accuracy of our approximation grows. However, from a discussion in the Appendix it becomes clear that our ansatz (12) and the solution (14) only slightly overestimate the role of the finiteness of the correlation length ξ. There we also propose a slightly different variant of the solution, Eq. (A11), which somewhat underestimates this role. The insets in Fig. 4 depict the energy dependence of the spectral density far from a hot spot for different combinations, (15) and (16).

Figure 5 depicts the energy dependence of the spectral density for the combinations (15) and (16) at a hot spot with $t'/t=-0.4$, which, according to Schmalian *et al.*,[11,12] corresponds to the $YBa_2Cu_3O_{6+\delta}$ system. The spectral density in the case of the combinations (17) of the spin-fermion model is very close to that obtained in the incommensurate case (16). For such a value of t/t', even at hot spots the velocity projections have opposite signs. However, the spectral density (in the incommensurate case) obtained from the solution with ''alternating'' κ, Eq. (A11) (the dashed curve in Fig. 5a) is seen to be very close to that obtained from (14). This

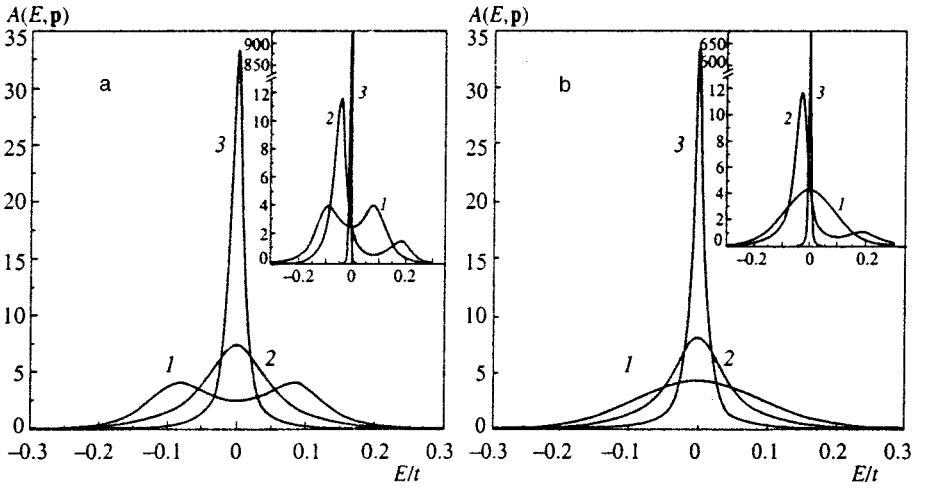

FIG. 4. Energy dependence of the spectral density at a hot spot ($p_x a/\pi=0.1666$ and $p_y a/\pi=0.8333$) for different diagram combinations at $t'/t=-0.6$ and $\mu/t=-1.8$, when the solution (14) is exact: (a) the incommensurate case, and (b) the commensurate case. The correlation length corresponds to the following values of κa: curve *1*, 0.01; curve *2*, 0.1; and curve *3*, 0.5; $\Delta=0.1t$. The insets depict the energy dependence of the spectral density for the corresponding diagram combinations at $\kappa a=0.01$: curve *1*, at the hot spot $p_x a/\pi=0.1666$ and $p_y a/\pi=0.8333$; curve *2*, near the hot spot $p_x a/\pi=0.1663$ and $p_y a/\pi=0.8155$; and curve *3*, far from the hot spot $p_x a/\pi=0.0$ and $p_y a/\pi=0.333$.

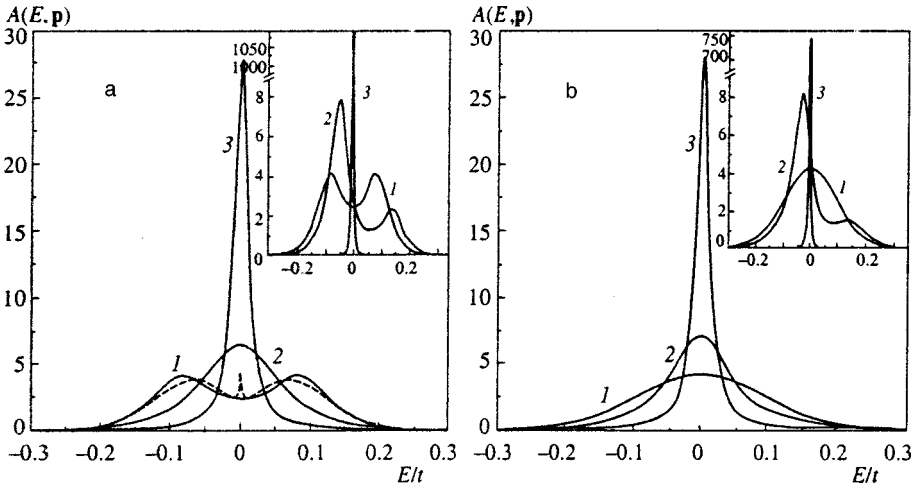

FIG. 5. Energy dependence of the spectral density at a hot spot ($p_x a/\pi = 0.142$ and $p_y a/\pi = 0.857$) for different diagram combinations at $t'/t = -0.4$ and $\mu/t = -1.3$, which approximately corresponds to high-T_c cuprates: (a) the incommensurate case [the dashed curve represents the spectral density for the incommensurate case obtained by (A11)], and (b) the commensurate case. The correlation length corresponds to the following values of κa: curve *1*, 0.01; curve *2*, 0.1; and curve *3*, 0.5; $\Delta = 0.1t$. The insets depict the energy dependence of the spectral density for the corresponding diagram combinations at $\kappa a = 0.01$: curve *1*, at the hot spot $p_x a/\pi = 0.142$ and $p_y a/\pi = 0.857$; curve *2*, near the hot spot $p_x a/\pi = 0.145$ and $p_y a/\pi = 0.843$; and curve *3*, far from the hot spot $p_x a/\pi = p_y a/\pi = 0.375$.

suggests that the ansatz (12) and solution (14) quantitatively are close to an exact solution. We stress once more that the solution (14) is exact in the limits $\xi \to \infty$ and $\xi \to 0$, while for finite ξ it provides a good interpolation between the two limits.

Now consider the one-electron density of states,

$$N(E) = \sum_{\mathbf{p}} A(E,\mathbf{p}) = -\frac{1}{\pi} \sum_{\mathbf{p}} \operatorname{Im} G^R(E,\mathbf{p}), \tag{25}$$

determined by the integral of the spectral density $A(E,\mathbf{p})$ over the entire Brillouin zone. Earlier we have seen that although for some topologies of the initial Fermi surface (band fillings) we can guarantee that near hot spots the signs of the velocity projections are the same, far from hot spots the signs are usually different, and the solution (14) based on the ansatz (12) is only an approximation. Correspondingly, using the solution (14) to calculate the density of states also yields an approximation, according to (25). Figure 6 depicts the densities of states obtained from (13), (15), and (25) with allowance for the spectrum (8), for different diagram combinations, Eqs. (15), (16), and (17), at $t'/t = -0.4$ (Fig. 6a) and $t'/t = -0.6$ (Fig. 6b). We see that at $t'/t = -0.4$ the density of states vs. energy curves acquire a dip (pseudogap). This decrease in the density of states is weakly dependent on the correlation length ξ (see the inset in Fig. 6a). If the band filling is such that the Fermi level μ lands in this energy interval, there are hot spots on the Fermi surface. At $t'/t = -0.6$, the region where the hot spots exist is rather wide, but nevertheless the pseudogap in the density of states is essentially unobservable. What can be seen is a smearing of the Van Hove singularity, a singularity that exists when there is no scattering by fluctuations.

3. MODEL OF "SUPERCONDUCTING" FLUCTUATIONS

3.1. Description of model and the solution for the Green's function

As noted earlier, pseudogap phenomena can probably be explained by employing the idea of fluctuation Cooper pairing at temperatures above the superconducting transition temperature T_c (see Refs. 1, 5–7). Consider the simplest possible model approach to this problem. Figure 7a depicts the self-energy diagram of first order in the fluctuation propagator of Cooper pairs for $T > T_c$. Bearing in mind that we wish to consider both ordinary *s*-wave pairing and *d*-wave pairing, which is a characteristic feature of high-T_c systems, we introduce the pairing interaction of the simplest (separable) form

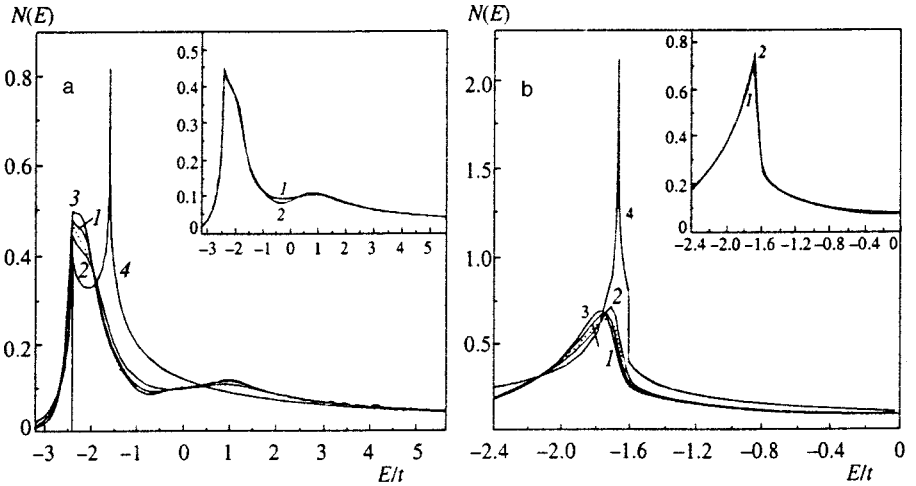

FIG. 6. One-electron density of states for different diagram combinations: (a) the case where $t'/t = -0.4$ and $\mu/t = -1.3$, and (b) the case where $t'/t = -0.6$ and $\mu/t = -1.8$. Curves *1* correspond to the incommensurate case, curves *2* to the commensurate case, curves *3* to the combinations of the spin-fermion model, and curves *4* to the case where there is no AFM fluctuations. The dotted curves represent the spectral density for the incommensurate case obtained by (A11), $\Delta/t = 1$, and the correlation length corresponds to $\kappa a = 0.1$. The insets depict the one-electron densities of states energy for the corresponding diagram combinations at $\kappa a = 0.1$ (curves *1*) and $\kappa a = 0.01$ (curves *2*).

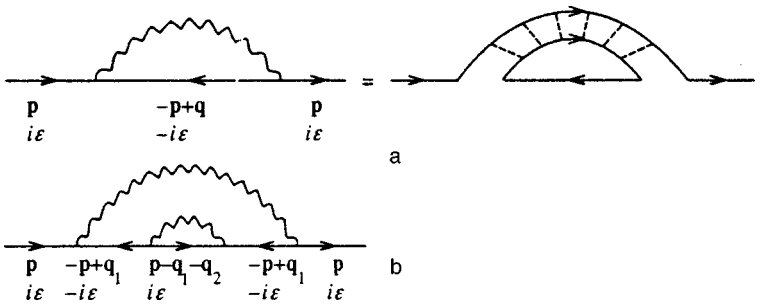

FIG. 7. Self-energy diagrams in the model of SC fluctuations: (a) the first-order diagram with an ''explanation'' of the meaning of the wave line, the fluctuation operator of Cooper pairs (the dashed lines correspond to pairing interaction), and (b) the second-order diagram.

$$V(\mathbf{p},\mathbf{p}')=-Ve(\phi)e(\phi'), \tag{26}$$

where ϕ is the polar angle specifying the direction of the electron momentum $\mathbf{p}$ in the plane, and for $e(\phi)$ we use the model dependence adopted in Refs. 26 and 27:

$$e(\phi)=\begin{cases}1 & \text{for } s-\text{wave pairing},\\ \sqrt{2}\cos 2\phi & \text{for } d-\text{wave pairing}.\end{cases} \tag{27}$$

As usual, the coupling constant V is assumed finite for electrons within an energy layer near the Fermi surface. Then the self-energy part corresponding to Fig. 7a takes the form

$$\Sigma(\varepsilon_n,\mathbf{p})=\sum_{m\mathbf{p}} V_{\rm eff}(i\omega_m,\mathbf{q})G(i\omega_m-i\varepsilon_n,-\mathbf{p}+\mathbf{q}), \tag{28}$$

where the effective interaction with SC fluctuations is given by the expression

$$V_{\rm eff}(i\omega_m,\mathbf{q})=-\frac{Ve^2(\phi)}{1-VT\sum_{n\mathbf{p}} G_0(i\varepsilon_n,\mathbf{p})G_0(i\omega_m-i\varepsilon_n,-\mathbf{p}+\mathbf{q})e^2(\phi)}. \tag{29}$$

Below we assume that the SC fluctuations are static, so that in (33) we can limit ourselves to the term with $\omega_m=0$. Here the static approximation is valid for $\pi T\gg\omega_{\rm SC}=8(T-T_c)/\pi$, which is formally similar to the condition $\pi T\gg\omega_{\rm sf}$ used in the hot-spot model. The closer the system is to the superconducting transition point, the better the condition is met. Then the effective interaction can be written

$$V_{\rm eff}(\mathbf{q})\approx-\frac{\tilde{\Delta}^2e^2(\phi)}{\xi^{-2}(T)+\mathbf{q}^2}, \tag{30}$$

where

$$\xi(T)=\frac{\xi_0}{\sqrt{(T-T_c)/T_c}},\quad \xi_0\approx 0.18v_F/T_c, \tag{31}$$

with ξ_0 the ordinary coherence length of the superconductor, and $\tilde{\Delta}^2=1/N(E_F)\xi_0^2$ (here $N(E_F)$ is the density of states at the Fermi level E_F). Of course, within the elementary BCS model considered here,

$$\tilde{\Delta}\approx 2\pi^2T_c\frac{T_c}{E_F}\sim\Delta_0\frac{\Delta_0}{E_F}\ll\Delta_0$$

(where Δ_0 is the energy gap of the superconductor at $T=0$), and so the obvious problem of explaining the scale of the anomalies observed in the experiments arises. However, below we again assume that ξ and $\tilde{\Delta}$ are phenomenological parameters of the theory, bearing in mind that in high-T_c systems these parameters should be found from experiments rather than from a simple BCS-type theory, which does not apply to this case anyway.

Reasoning in the same way as we did in passing from (2) to (4), instead of (30) we introduce the model interaction

$$V_{\rm eff}(\mathbf{q})=-\Delta^2e^2(\phi)\frac{2\xi^{-1}}{\xi^{-2}+q_x^2}\frac{2\xi^{-1}}{\xi^{-2}+q_y^2}, \tag{32}$$

where $\Delta^2=\tilde{\Delta}^2/4$. Quantitatively this is very close to Eq. (30) and simplifies calculations significantly by making it possible to classify the contributions of higher-order diagrams. In this case the first-order contribution of the diagram in Fig. 7a has the form

$$\Sigma^{(1)}(\varepsilon_n,\mathbf{p})=\frac{\Delta^2e^2(\phi)}{i\varepsilon_n+\xi_{\mathbf{p}}+i(|v_x|+|v_y|)\kappa\,\mathrm{sign}\,\varepsilon_n}, \tag{33}$$

where $v_x=v_F\cos\phi$, $v_y=v_F\sin\phi$, and $\kappa=\xi^{-1}$. The contribution of the second-order diagram in Fig. 7b is

$$\Sigma^{(2)}(\varepsilon_n,\mathbf{p})=(\Delta^2e^2(\phi))^2\times\int\frac{dq_{1x}}{\pi}\frac{\kappa}{\kappa^2+q_{1x}^2}\int\frac{dq_{1y}}{\pi}\frac{\kappa}{\kappa^2+q_{1y}^2}\times\int\frac{dq_{1x}}{\pi}\frac{\kappa}{\kappa^2+q_{2x}^2}\int\frac{dq_{1y}}{\pi}\frac{\kappa}{\kappa^2+q_{2y}^2}\times\frac{1}{(i\varepsilon_n+\xi_{\mathbf{p}}-\mathbf{v}_1\cdot\mathbf{q}_1)^2}\frac{1}{i\varepsilon_n-\xi_{\mathbf{p}}-\mathbf{v}_2\cdot\mathbf{q}_1-\mathbf{v}_2\cdot\mathbf{q}_2}, \tag{34}$$

where $\mathbf{v}_1=-\mathbf{v}_2=\mathbf{v}_F$. We can easily see that in the given problem we have essentially the same rules of the diagrammatic technique as in the hot-spot model with combinations corresponding to the incommensurate case. This becomes especially obvious if we study the topology of the interaction line (the fluctuation propagator of Cooper pairs) in the diagram of Fig. 7a: we see that in higher orders the only diagrams that exist are those in which the interaction line surrounds an even number of vertices. Equation (34) is similar to (9), but the signs of the velocity projections in the denominators of the Green's functions are always different, $\mathbf{v}_1=-\mathbf{v}_2$. Hence contributions to the integrals over momentum

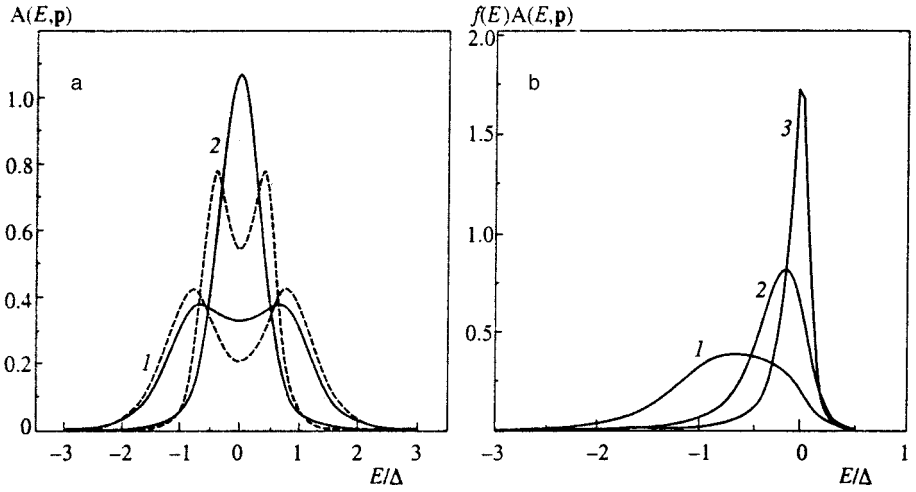

FIG. 8. (a) Energy dependence of the spectral density $A(E,\mathbf{p})$ for the case of *d*-wave fluctuation pairing at different values of the polar angle ϕ, which defines the direction of electron momentum in the plane: curve *1*, $\phi=0$; and curve *2*, $\phi=\pi/6$. The correlation length corresponds to $v_F\kappa/\Delta=0.5$ (solid curves). (b) Energy dependence of the product $f(E)A(E,\mathbf{p})$ ($f(E)$ is the Fermi function): curve *1*, $\phi=0$; curve *2*, $\phi=\pi/6$; curve *3*, $\phi=\pi/4.83$. The temperature (in the Fermi function) is $T=0.1\Delta$, and $v_F\kappa/\Delta=0.5$.

transfer in higher-order diagrams are provided not only by the poles of Lorentzians but also by the poles of the Green's functions. Nevertheless (bearing in mind the discussion in the Appendix) we can estimate the contribution of higher-order diagrams by using the ansatz (12), i.e., we calculate all the integrals in, say, (34) as if the velocity projections were of the same sign, and then in the answer we put $\mathbf{v}_1=-\mathbf{v}_2=\mathbf{v}_F$. We again arrive at a recurrence relation for the Green's function of the form (14):

$$\Sigma_k(\varepsilon_n,\xi_{\mathbf{p}})=\frac{\Delta^2 e^2(\phi)v(k)}{i\varepsilon_n-(-1)^k\xi_{\mathbf{p}}+ikv_F\kappa(|\cos\phi|+|\sin\phi|)-\Sigma_{k+1}(\varepsilon_n,\xi_{\mathbf{p}})}, \tag{35}$$

where $v(k)$ has been defined in (16). Of course, Eq. (35) is an approximation, but it gives the exact result in the limits $\kappa\to 0$ ($\xi\to\infty$) and $\kappa\to\infty$ ($\xi\to 0$) and provides a fairly good (quantitative) interpolation between these two limits for finite correlation lengths.

3.2. Spectral density and density of states

Figure 8a depicts the energy dependence of the spectral density $A(E,\mathbf{p})$ [Eq. (24)] of the one-particle Green's function calculated by (35) for different values of the polar angle ϕ determining the direction of electron momentum in the plane (here we assume that $|\mathbf{p}|=p_F$) for the case of *d*-wave fluctuation pairing. Clearly, in the vicinity of the point $(\pi/a,0)$ of the Brillouin zone, the spectral density exhibits non-Fermi-liquid (pseudogap) behavior. As the vector $\mathbf{p}$ rotates in the direction of the zone diagonal, the double-peak structure disappears and the spectral density becomes a typical Fermi-liquid spectral density with a single peak, and the closer the value of ϕ is to $\pi/4$ the narrower the peak. The spectral density undergoes a similar transformation as the correlation length ξ becomes smaller.

Figure 8b depicts the evolution of $f(E)A(E,\mathbf{p})$ (here $f(E)$ is the Fermi distribution), which is actually the parameter measured in ARPES experiments.[2] Note that the curves in Fig. 8b closely resemble the curves obtained in Refs. 11 and 12 in the hot-spot model. The picture of destruction of the Fermi surface suggested by these calculations is very similar to the one that follows from the experimental data obtained by Norman *et al.*[28] for $Bi_2Sr_2CaCu_2O_{8+\delta}$.

In the case of *s*-wave fluctuation pairing, the pseudogap appears isotropically on the entire Fermi surface, and the spectral density is of the non-Fermi-liquid type everywhere for large fluctuation lengths ξ of SC fluctuations.

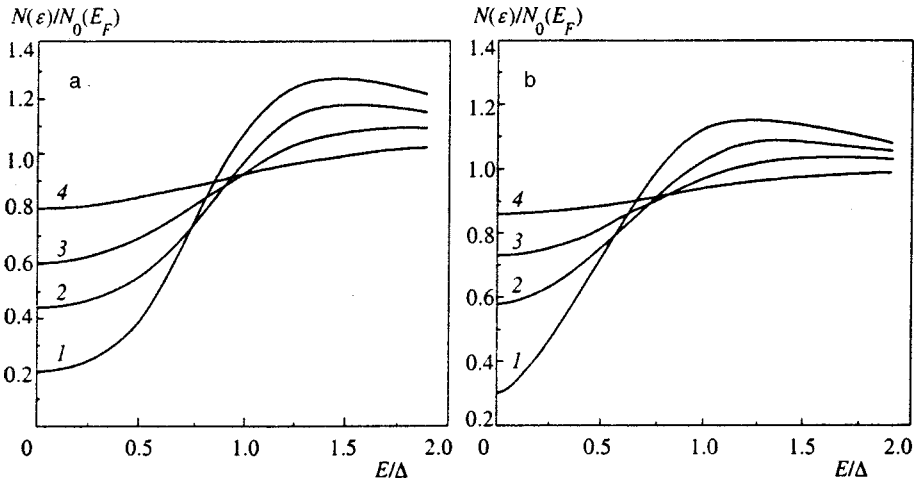

FIG. 9. One-electron density of states in the model of SC fluctuations: (a) in the case of *s*-wavepairing, and (b) in the case of *d*-wave pairing. The curves are built for the following values of the parameter $v_F\kappa/\Delta$, which determines the correlation lengths of short-range-order fluctuations: 0.1 (curve *1*), 0.5 (curve *2*), 1.0 (curve *3*), and 2.0 (curve *4*).

In Fig. 9 we present the results of calculations of the one-electron density of states using (35) for the case of *s*-wave pairing (Fig. 9a) and in the case of *d*-wave pairing (Fig. 9b) for different correlation lengths of the SC fluctuations. We see that for *d*-wave pairing the pseudogap in the density of states is not so pronounced as for *s*-wave pairing, even for large correlation lengths of the fluctuations. At the same, Fig. 9 clearly shows that in the model of SC fluctuations the pseudogap is more pronounced than the hot-spot model discussed earlier.

4. CONCLUSION

We have examined almost exactly solvable models of the pseudogap state of the electronic spectrum of two-dimensional systems. These models are based on alternative scenarios of the origin of these anomalies: the picture of ''dielectric'' (AFM, SDW, CDW) fluctuations, which gives rise to the hot-spot model, and the picture of fluctuational formation of Cooper pairs above T_c (SC fluctuations). The term ''almost exactly solvable'' means that in this approach it is possible to sum the entire series of Feynman diagrams for the one-electron Green's function (and actually also for the two-electron Green's function[16,17]), using for the higher-order diagrams the approximate ansatz (12). As shown in the Appendix and also by the numerical examples in the main body of the text, the ansatz guarantees a rather good approximation (speaking quantitatively) to the exact solution in the region of finite correlation lengths ξ of short-range-order fluctuations, while in the limits $\xi\to\infty$ and $\xi\to 0$ our solution is exact.

Our calculations of spectral densities have shown that in both scenarios we can obtain a rather appealing picture (from the standpoint of possible comparison with the experimental data in high-T_c cuprates) of destruction of the Fermi-liquid state in specific (hot) parts of the Fermi surface, with the Fermi-liquid state retained in the remaining (cold) part of the Fermi surface. Such non-Fermi-liquid behavior is due to the strong scattering of electrons by short-range-order fluctuations, and the larger the correlation length ξ the more pronounced the behavior. At the same time, there are certain differences between these two scenarios, which can, in principle, be utilized in the analysis of the situation in real systems. In particular, in the hot-spot model (AFM fluctuations), the pseudogap in the density of states is relatively small (see Fig. 6). In the model of SC fluctuations the pseudogap in the density of states is much more visible (see Fig. 9). At the same time, the model of dielectric AFM fluctuations appears to be more attractive even from a simple consideration of the phase diagram of a high-T_c system: pseudogap anomalies are observed in the underdoped region, and the closer the system is to a dielectric AFM state the more pronounced are the anomalies. It is in this region that we can expect the short-range-order dielectric (AFM) fluctuations to play a more important role, the correlation length ξ to increase, etc. It is rather difficult to understand why in this region of the phase diagram the fluctuational formation of Cooper pairs (SC fluctuations) may become more important. On the contrary, it would seem that such formation should manifest itself in the region close to optimal doping (corresponding to the maximum superconducting transition temperature). Moreover, an obvious problem inherent in this scenario is that of explaining the characteristic scales of the anomalies (in temperature and in energy). The problem cannot be resolved by using simple approaches based on the BCS theory—the solution requires new microscopic approaches.[5,7] The models considered in the present paper are useful in analyzing the pseudogap formation in both scenarios, since they are actually based on a fairly general (semiphenomenological) form of the correlation function of short-range-order fluctuations.

The authors would like to express their gratitude to Oleg Tchernyshyov for supplying the preliminary information on his analysis of the one-dimensional model. This was partially supported by the Russian Fund for Fundamental Research (Project 96-02-16065) and Project No. IX.1 of the Statistical Physics State Program and Project No. 96-051 of the High-T_c Superconductors State Program of the Russian Ministry of Science.

APPENDIX: ANALYSIS OF THE ONE-DIMENSIONAL MODEL

Let us examine in greater detail the use of the ansatz (12) in estimating the contributions of higher-order diagrams. We limit ourselves to the analysis of the one-dimensional model,[15–17] since in one dimension the problem is most serious.[22] We are interested in the vicinity of the Fermi points $+p_F$ and $-p_F$, with electrons scattered by Gaussian short-range-order fluctuations scattering by a momentum $Q\sim\pm-2p_F$, shifting them from one end of the Fermi line to the other with an accuracy of order $\xi^{-1}=\kappa$ (Refs. 13–17). We examine the electronic spectrum in the linearized approximation, $\xi_{p\pm p_F}=\pm-v_F p$, and assume, for the sake of brevity, that $v_F=1$. Here the system consists of two types of electron: those electrons that move to the left, and those that move to the right. It is convenient to do our analysis in a representation[22] in which the equation of motion for the electrons in the given model takes the form[18,22]

$$\left(i\hat{1}\frac{\partial}{\partial t}-i\hat{\sigma}_3\frac{\partial}{\partial x}\right)\hat{\Psi}(t,x)=\begin{pmatrix}0 & \Delta(x)\\ \Delta^*(x) & 0\end{pmatrix}\hat{\Psi}(t,x). \tag{A1}$$

We limit ourselves to incommensurate fluctuations, i.e., $\Delta^*(x)\neq\Delta(x)$. The spinor $\hat{\Psi}=\binom{\psi_+}{\psi_-}$ describes ''right'' and ''left'' electrons. The fluctuations $\Delta(x)$ are assumed Gaussian with $\langle\Delta(x)\rangle=0$ and $\langle\Delta^*(x)\Delta(x')\rangle=|\Delta|^2\exp(-\kappa|x-x'|)$. The free propagator in the frequency–coordinate representation is

$$G_0(\varepsilon x)=i\theta(\varepsilon\sigma_3 x)\,\mathrm{sign}(\varepsilon)\exp(i\varepsilon\sigma_3 x), \tag{A2}$$

with $\sigma_3=+1$ for right particles and $\sigma_3=-1$ for left particles. A particle traversing a path of length l produces a phase factor $e^{i\varepsilon l}$. When calculating specific diagrams, it is convenient to change the integration variables from the coordinates x_k of interaction vertices to the lengths l_k of paths traversed by particles from one scattering act to another.[22] Here it is important to account for the fact that these path lengths are not independent, since for a given diagram the

total particle displacement $x-x'$ is always fixed. The rules of the diagrammatic technique for calculating $G(\varepsilon,x-x')$ that result are as follows:[22]

1. A solid line of length l_k yields a factor $-ie^{il_k(\varepsilon-(-1)^k p)}$.

2. A wavy (interaction) line connecting vertices m and n gives a factor

$$|\Delta|^2\exp(-\kappa|x_m-x_n|)=|\Delta|^2\exp\left(-\kappa\left|\sum_{k=m}^{n-1}(-1)^k l_k\right|\right).$$

3. Integration over all l_k is done from 0 to ∞.

4. Integration over p is done with a weighting factor $e^{ip(x-x')}/2\pi$.

In calculating $G(\varepsilon,p)$ the last rule can simply be dropped. These rules show that allowing for the finiteness of the correlation length $\xi=\kappa^{-1}$ leads in each diagram to a damping of the corresponding transition amplitude with the displacement of the particle. Taking this effect into account exactly constitutes a complicated problem, but lower and upper bounds on this effect can be found. On the one hand, we have the obvious inequality

$$\exp\left(-\kappa\left|\sum_{k=m}^{n-1}(-1)^k l_k\right|\right)>\exp\left(-\kappa\sum_{k=m}^{n-1}l_k\right). \tag{A3}$$

By using the right-hand side of (A3) as the interaction line we overestimate the transition amplitude damping (i.e., effectively overestimate κ). We can easily see that the use of this approximation in calculating the Green's function in the momentum representation amounts to adding $i\kappa$ to the denominator in each Green's function surrounded by the interaction line and yields an expression for any higher-order correction of the form (12) (cf. Ref. 22). For instance, the following expression corresponds to the diagram in Fig. 1b (we assume that $\varepsilon>0$ and $\delta=0^+$):

$$\Delta G(\varepsilon,p)=\Delta^4\frac{1}{\varepsilon-p+i\delta}\left(\frac{1}{\varepsilon+p+i\kappa}\times\frac{1}{\varepsilon-p+2i\kappa}\,\frac{1}{\varepsilon+p+i\kappa}\right)\frac{1}{\varepsilon-p+i\delta}, \tag{A4}$$

which is similar to (9) and (11). On the other hand, we can employ the inequality

$$\exp\left(-\kappa\left|\sum_{k=m}^{n-1}(-1)^k l_k\right|\right)<\exp\left(-\kappa\sum_{k=m}^{n-1}(-1)^{k-m}l_k\right). \tag{A5}$$

By using the right-hand side of (A5) for the interaction line we underestimate the transition amplitude damping (i.e., effectively underestimate κ). It may seem that this choice of the expression for the interaction line can even increase the transition amplitude over its value at $\kappa=0$, but this is not so. Since we are considering the incommensurate case, where the interaction line surrounds only an even number of vertices (i.e., an odd number of l_k), the choice of a specific sign in the exponent after the absolute-value sign has been removed is determined by what number of l_k is greater, the odd or the even. This leads to a situation in which the effective transition amplitude of any higher-order diagram can only decrease. For the diagram in Fig. 1b in the coordinate representation the contribution of the interaction lines is

$$e^{-\kappa l_2}e^{-\kappa|l_1-l_2-l_3|}\to e^{-\kappa l_2}e^{-\kappa(l_1-l_2+l_3)}=e^{-\kappa(l_1+l_3)}. \tag{A6}$$

In the momentum representation this yields

$$\Delta G(\varepsilon,p)=\Delta^4\frac{1}{\varepsilon-p+i\delta}\left(\frac{1}{\varepsilon+p+i\kappa}\,\frac{1}{\varepsilon-p+i\delta}\,\frac{1}{\varepsilon+p+i\kappa}\right)\times\frac{1}{\varepsilon-p+i\delta}, \tag{A7}$$

An analysis of any higher-order diagram shows that in this case the contributions of all N-order diagrams are equal and in the momentum representation have the form (the alternating-κ ansatz)

$$G_N(\varepsilon,p)=|\Delta|^{2N}\frac{1}{(\varepsilon-p+i\delta)^{N+1}}\,\frac{1}{(\varepsilon+p+i\kappa)^N}. \tag{A8}$$

Then the entire series can easily be summed, much like the case with $\kappa=0$ (Refs. 13 and 14), and for the Green's function we obtain

$$G^R(\varepsilon,p)=\sum_{N=0}^{\infty}N!\,G_N(\varepsilon,p)=\int_0^\infty d\zeta\,e^{-\zeta}\,\frac{\varepsilon+p+i\kappa}{(\varepsilon-p+i\delta)(\varepsilon+p+i\kappa)-\zeta|\Delta|^2}. \tag{A9}$$

This expression can easily be used to calculate the corresponding spectral density or the one-particle density of states:

$$\frac{N(\varepsilon)}{N(E_F)}=\frac{v_F\kappa}{\pi}\int_{-\infty}^{\infty}d\xi_p\int_0^\infty d\zeta\times e^{-\zeta}\,\frac{\zeta|\Delta|^2}{(\varepsilon^2-\xi_p^2-\zeta|\Delta|^2)^2+(v_F\kappa)^2(\varepsilon-\xi_p)^2}, \tag{A10}$$

where we have restored v_F. In Fig. 10 we compare the densities of states for different values of κ (or correlation length) that we calculated by the alternating-κ ansatz and a recurrence relation of the form (12) in the one-dimensional model.[15–17] We see that the results are quantitatively close for almost all values of κ. Since, as noted earlier, our main ansatz (12) and (A4) somewhat overestimates the role of the finiteness of κ, while the alternating-κansatz (A7) underestimates it, we can easily see that the exact value of the density of states differs little from the these two approximation to the contributions of higher-order diagrams. The situation with the spectral densities is similar. Actually this means that the results for the main physical quantities determined by the one-electron Green's function are not strongly dependent on the way in which a finite κ enters the expressions for higher-order diagrams. What is important is that we must take into account (at least approximately) all perturbation-theory diagrams with allowance for their different combinations. This

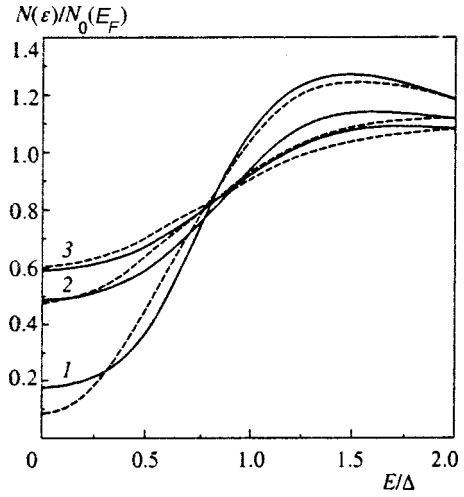

FIG. 10. One-electron density of states in the one-dimensional model for different values of the parameter $v_F\kappa/\Delta$: 0.1 (curve *1*), 0.8 (curve *2*), and 1.2 (curve *3*). The solid curves represent the results of calculations by formulas of the form (12) and (14) (Ref. 15), and the dashed curves represent the results of calculations by (A10).

should not come as a big surprise, since the main effect of pseudogap formation is due primarily to backward scattering by a vector $Q\sim 2p_F$, which is accounted for exactly in the limit $\xi\to\infty$, while the effect of a finite κ reduces to an additional weak modulation of the random field, which leads to damping of the field's correlator and smearing of the pseudogap.

Naturally, the alternating-κ ansatz can also be written in the form of a recurrence relation of the form (14) for two-dimensional models, which were discussed in the main body of the text. For instance, if the hot-spot model we have

$$\Sigma_k(\varepsilon_n,\xi_{\mathbf{p}})=\Delta^2\,\frac{v(k)}{i\varepsilon_n-\xi_k+i\alpha_k v_k\kappa-\Sigma_{k+1}(\varepsilon_n,\xi_{\mathbf{p}})}, \tag{A11}$$

where $\alpha_k=1$ for odd k, and $\alpha_k=0$ for even k. The other notation is explained in the main body of the text. The data on the density of states obtained via (A11) are depicted in Fig. 6 and corroborate our conclusions. For the model of SC fluctuations an expression similar to (A11) can also easily be written.

Note that the alternating-κ ansatz is formal and is used here only to show that this more or less arbitrary approximation (which underestimates the role of the finiteness of κ in higher-order diagrams) leads to results that are quantitatively very close to those obtained by the building-up-κ ansatz (12) and (A4) (which generally overestimates this role). The latter approximation was used in Refs. 15–18 and in the main part of the present paper and has a much deeper meaning. As noted earlier, this approximation is exact in the vicinity of hot spots for values of the parameters of the bare spectrum, t, t', and μ (topologies of the Fermi surface), that guarantee equal signs for the velocity projections at the hot spots connected by the vector **Q**. Reasoning along similar lines, in the one-dimensional model we can obtain an expression of the form (12) or (A4) for the higher-order contributions if we consider a model for the correlator of short-range-order fluctuations with its maximum at an arbitrary scattering vector **Q** much shorter than p_F. In this case, for large correlation lengths ξ, the electrons are scattered by fluctuations, staying always on one branch (right or left) of the spectrum. Here expressions if the form (A4) remain exact. After this is done, in the final expressions for the contributions of higher-order diagrams we perform a continuation to the region $Q\sim 2p_F$ of interest to us, since the only dependence on Q is already present via the bare electron spectrum. A similar result can be achieved by varying the chemical potential μ (band filling).

*)E-mail: kuchinsk@ief.uran.ru
†)E-mail: sadovski@ief.uran.ru

1)A model similar in meaning to the one used here but differing somewhat from (4) was employed by Schmalian *et al.*:[11,12]

$$V_{\rm eff}(\mathbf{k})=\Delta^2\frac{2\xi^{-1}}{\xi^{-2}+k_\parallel^2}\,\frac{2\xi^{-1}}{\xi^{-2}+k_\perp^2},$$

where $k_\parallel$ and $k_\perp$ are the projections of the vector **k** parallel and perpendicular to $v_{\mathbf{p}+\mathbf{Q}}$, so that a result similar to (7) is obtained:

$$\Sigma(\varepsilon_n,\mathbf{p})=\frac{\Delta^2}{i\varepsilon-\xi_{\mathbf{p}+\mathbf{Q}}+i|\mathbf{v}_{\mathbf{p}+\mathbf{Q}}|k\ \mathrm{sign}\ \varepsilon_n}.$$

2)In the model of $V_{\rm eff}$ employed by Schmalian *et al.*[11,12] for the case $\mathbf{v}_{\mathbf{p}}\cdot\mathbf{v}_{\mathbf{p}+\mathbf{Q}}>0$ the following expression can be derived in a similar way:

$$\Sigma(\mathrm{b})=\Sigma(\mathrm{c})=\Delta^4\frac{1}{[i\varepsilon_n-\xi_{\mathbf{p}+\mathbf{Q}}+i|\mathbf{v}_{\mathbf{p}+\mathbf{Q}}|\kappa]^2}$$
$$\times\frac{1}{i\varepsilon_n-\xi_{\mathbf{p}+\mathbf{Q}}+i2|\mathbf{v}_{\mathbf{p}}|(|\cos\phi|+|\sin\phi|)\kappa},$$

where ϕ is the angle between $\mathbf{v}_{\mathbf{p}}$ and $\mathbf{v}_{\mathbf{p}+\mathbf{Q}}$.

[1] M. Randeria, Varenna Lectures 1997, E-prints archive cond-mat/9710223.
[2] M. Randeria and J. C. Campuzano, Varenna Lectures 1997, E-prints archive cond-mat/9709107.
[3] H. Ding, T. Yokoya, J. C. Campuzano, T. Takahashi, M. Randeria, M. R. Norman, T. Mochiku, K. Kadowaki, and J. Giapintzakis, Nature (London) **382**, 51 (1996).
[4] H. Ding, M. R. Norman, T. Yokoya, T. Takeuchi, M. Randeria, J. C. Campuzano, T. Takahashi, T. Mochiki, and K. Kadowaki, Phys. Rev. Lett. **78**, 2628 (1997).
[5] V. B. Geshkenbein, L. B. Ioffe, and A. I. Larkin, Phys. Rev. B **55**, 3173 (1997).
[6] V. J. Emery, S. A. Kivelson, and O. Zachar, Phys. Rev. B **56**, 6120 (1997).
[7] J. Maly, B. Janko, and K. Levin, E-printsarchive cond-mat/9710187 (1997); cond-mat/9805018 (1998).
[8] A. P. Kampf and J. R. Schrieffer, Phys. Rev. B **41**, 6399 (1990); **42**, 7967 (1990).
[9] V. Barzykin and D. Pines, Phys. Rev. B **52**, 13 585 (1995).
[10] D. Pines, Tr. J. of Physics **20**, 535 (1996).
[11] J. Schmalian, D. Pines, and B. Stojković, Phys. Rev. Lett. **80**, 3839 (1998).
[12] J. Schmalian, D. Pines, and B. Stojković, E-prints archive cond-mat/9804129 (1998).
[13] M. V. Sadovskiĭ, Zh. Éksp. Teor. Fiz. **66**, 1720 (1974) [Sov. Phys. JETP **39**, 845 (1974)].
[14] M. V. Sadovskiĭ, Fiz. Tverd. Tela (Leningrad) **16**, 2504 (1974) [Sov. Phys. Solid State **16**, 1632 (1974)].
[15] M. V. Sadovskiĭ, Zh. Éksp. Teor. Fiz. **77**, 2070 (1979) [Sov. Phys. JETP **50**, 989 (1979)].
[16] M. V. Sadovskii and A. A. Timofeev, Sverkhprovodimost': Fiz., Khim., Tekhnol. **4**, 11 (1991) [Supercond., Phys. Chem. Technol. **4**, 9 (1991)].
[17] M. V. Sadovskii and A. A. Timofeev, J. Mosc. Phys. Soc. **1**, 391 (1991).
[18] R. H. McKenzie and D. Scarratt, Phys. Rev. B **54**, R12 709 (1996).
[19] O. Tchernyshyov, Phys. Rev. B **56**, 3372 (1997).
[20] H. C. Ren, E-prints archive cond-mat/9612184 (1996).
[21] A. I. Posazhennikova and M. V. Sadovskiĭ, E-prints archive cond-mat/9806199 (1998).

[22] O. Tchernyshyov, E-prints archive cond-mat/9804318 (1998).
[23] P. Monthoux, A. V. Balatsky, and D. Pines, Phys. Rev. B **46**, 14 803 (1992).
[24] P. Monthoux and D. Pines, Phys. Rev. B **47**, 6069 (1993); **48**, 4261 (1994).
[25] P. V. Elyutin, Opt. Spektrosk. **43**, 542 (1977) [Opt. Spectrosc. **43**, 318 (1977)].
[26] L. S. Borkowski and P. J. Hirschfeld, Phys. Rev. B **49**, 15 404 (1994).
[27] R. Fehrenbacher and M. R. Norman, Phys. Rev. B **50**, 3495 (1994).
[28] M. R. Norman, H. Ding, M. Randeria, J. C. Campuzano, T. Yokoya, T. Takeuchi, T. Takahashi, T. Mochiki, K. Kadowaki, P. Guptasarma, and D. G. Hinks, E-prints archive cond-mat/9710163 (1997).

Translated by Eugene Yankovsky

Superfluidity of indirect biexcitons in superlattices

Yu. E. Lozovik*) and O. L. Berman

Institute of Spectroscopy, Russian Academy of Sciences, 142092 Troitsk, Moscow Region, Russia

M. Willander

Göteborg University, Chalmers University of Technology, S-41296, Göteborg, Sweden
(Submitted 26 October 1998)
Zh. Éksp. Teor. Fiz. **115**, 1786–1798 (May 1999)

Instability in a system of interacting quasi-two-dimensional excitons in a type II superlattice of a finite thickness due to attraction between oppositely-directed excitonic dipoles in neighboring layers has been discovered. A stable system is that of indirect quasi-two-dimensional biexcitons formed by indirect excitons with dipole moments oriented in opposite directions. The radius and binding energy of indirect biexcitons has been calculated. A collective spectrum of a system of such biexcitons with a weak quadrupole interaction between them has been studied. Feasibility of Bose condensation, the density $n_s(T)$ of the superfluid component, and a phase transition to the superfliud state in a low-density system of indirect biexcitons have been analyzed. © *1999 American Institute of Physics.* [S1063-7761(99)01905-8]

1. INTRODUCTION

Intense interest in electron–hole systems in coupled quantum wells[1–5] has been stimulated by prediction of exciton superfluidity in such systems, which can be manifested by the existence of undamped electric currents in each well.[6,7] A number of other effects have been studied for such systems, namely, quasi-Josephson effects,[8] crystallization of indirect excitons,[9] and a number of effects in high magnetic fields.[10–13]

Of special interest are also collective properties of excitons in superlattices and layered structures.[14–16] The main subject of the reported investigation is superfluidity in a low-density electron–hole system in a superlattice. There are several plausible physical realizations of the model with spatially separated electrons (e) and holes (h) in superlattices. The holes can be in thermodynamic equilibrium in type II superlattices. In addition, spatially separated electrons and holes can be generated by laser pumping in superlattices in the form of periodic structures of coupled quantum wells (for example, based on AlAs/GaAs structures[1]) if the size-quantized level of carriers of a single type in one well is below the quantization level of the same carriers in another well (for carriers of the other type the relative positions of their levels is reversed). If the overlap between the wave functions of spatially separated electrons and holes in coupled quantum wells is sufficiently small, the recombination time can be much longer than the relaxation time τ_r. At low densities, indirect excitons (composed of spatially separated electrons and holes) exist at times when direct excitons have already recombined.[15] Then indirect excitons can come to quasi-equilibrium in a time shorter than the recombination time of indirect excitons, and photogenerated electrons and holes are characterized by different chemical quasipotentials. In this case, various quasi-equilibrium phases may turn up in a system of indirect excitons in superlattices, just as in coupled quantum wells.[6,7] If we have a structure of alternating electron and hole quantum wells (in the configuration described above), the excitons in one pair of wells have equal dipole moments, whereas the dipole moments of excitons in the couple including the neighboring well are oriented in the opposite direction (Fig. 1). It turns out that this property results in a notable difference between the electron–hole system in a superlattice and in coupled quantum wells (which even in three-layer e-h-e or h-e-h systems).

This paper considers a type II superlattice of a finite thickness. Instability of the system of indirect excitons with the dipole–dipole interaction at large separations between them has been discovered in such a structure, which is caused by the long-range attraction between excitons with oppositely-directed dipole moments. We will discuss quasi-two-dimensional indirect biexcitons formed by two indirect excitons with oppositely-directed dipole moments. The radius and binding energy of such biexcitons have been calculated. In contrast to the case of indirect excitons, there is a repulsive quadrupole interaction between these biexcitons at large distances. As a result, a system of indirect biexcitons is stable.

In the ladder approximation,[17,18] we will calculate a collective spectrum of two-dimensional indirect biexcitons formed by indirect excitons with opposite dipole moments due to the quadrupole interaction between biexcitons. The density $n_s(T)$ of the superfluid component of interacting indirect two-dimensional biexcitons has been calculated at low temperatures T. The temperature of the Kosterlitz–Thouless phase transition to the superfluid state[19] has been calculated. Probable coexistence of phases of direct and indirect excitons will be discussed in the Conclusion.

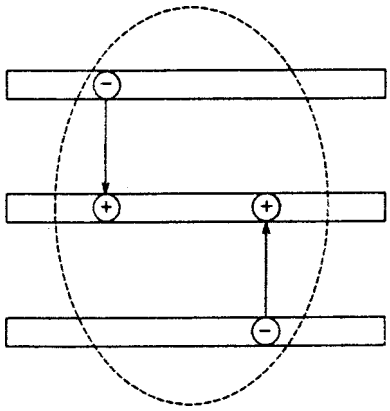

FIG. 1. Indirect two-dimensional biexcitons formed by indirect excitons with oppositely-directed dipole moments from neighboring pairs of quantum wells.

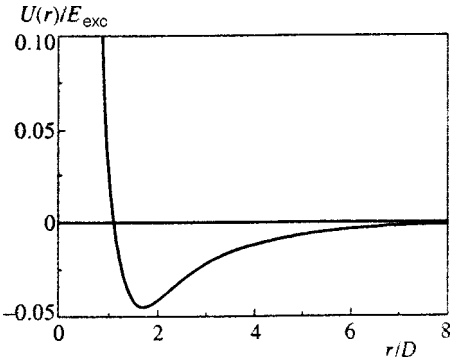

FIG. 2. Potential energy $U(r)$ of interaction (in units of indirect exciton binding energy $E_{\rm exc}=e^2/\epsilon D$) between indirect excitons with oppositely-directed dipole moments from neighboring pairs of quantum wells as a function of distance r between the indirect excitons (in units of D).

2. INSTABILITY OF A SYSTEM OF INDIRECT EXCITONS

Let us consider a low-density system of weakly interacting indirect two-dimensional excitons (their electrons and holes being located in neighboring layers) in a type II superlattice of a finite thickness L. Suppose that $n_{\rm exc}^{-1/2}>L$, where $n_{\rm exc}$ is the surface density of excitons. Let us prove that in this case, unlike in the case of a two-layered system (coupled quantum wells), the excitonic system is unstable. At low densities of indirect excitons $na^2\ll 1$ and low temperatures, the system is a quasi-two-dimensional Bose gas of indirect excitons[14] with dipole moments $\mathbf{d}$ ($d\sim eD$) perpendicular to the quantum well plane and increasing with the separation D between neighboring quantum wells ($a_{\rm exc}(D)$ is the exciton radius in the well plane: $a_{\rm exc}\approx a^*/4$, $a^*=\hbar^2\epsilon/\mu_{\rm exc}e^2$ for $D\ll a^*$ and $a_{\rm exc}\approx(a^*)^{1/4}D^{3/4}$ for $D\gg a^*$, the spectrum of the lowest excitonic levels being identical to that of a two-dimensional oscillator in the latter case, ϵ is the permittivity of the material, $\mu_{\rm exc}=m_em_h/(m_e+m_h)$). In contrast to the case of conventional excitons, the dipole–dipole energies U_- and U_+ of excitons with opposite and the same directions of dipole moments, respectively, contribute to the energy of low-density spatially indirect excitons. The interaction between two dipoles has the energy

$$U_+=-U_-=U=\frac{e^2D^2}{\epsilon R^3},$$

where R is the distance between the dipoles in the well plane. We assume that $D/R\ll 1$, which is the case in low-density systems. In what follows, we consider the case when the number k of quantum wells in the superlattice is limited, so that $k\ll R_0/D=1/D\sqrt{\pi n_{\rm exc}}$, which applies to the case of fairly large k but very low exciton densities (R_0 is the mean distance between neighboring excitons in the quantum well plane). Here we have the inequality $r/R\ll 1$, where r is the distance between dipoles in the direction normal to the well plane.

The difference between excitons and point-like bosons relates to exchange effects.[7,20] For excitons with spatially separated e and h, these effects are suppressed at large separations D between quantum wells, when $n_{\rm exc}a_{\rm exc}^2(D)\ll 1$, owing to the negligible overlap between wave functions of two excitons, if the potential barrier due to the direct dipole–dipole repulsion between two indirect excitons with the same directions of dipole moments and the repulsion between two indirect excitons with oppositely-directed dipole moments at distances of order of D is taken into account. The small tunneling parameter due to the barrier is given by the expression

$$\exp\left[-\frac{1}{\hbar}\int_{a(D)}^{r_0}\sqrt{2M[U(r)-\mu]}dr\right],$$

where $M=m_e+m_h$, μ is the chemical potential of the system, and r_0 is the classical turning point for repulsion at an energy equal to μ. The potential energy of interaction between two indirect excitons with the same directions of dipole moments at large distances is

$$U(r)=\frac{e^2D^2}{\epsilon r^3},$$

and for two indirect excitons with oppositely-directed dipole moments it has the form (Fig. 2)

$$U(r)=\frac{e^2}{\epsilon r}-\frac{2e^2}{\epsilon\sqrt{r^2+D^2}}+\frac{e^2}{\epsilon\sqrt{r^2+4D^2}}. \tag{1}$$

At large D, the small tunneling parameter due to the barrier for two indirect excitons with dipole moments in the same direction is

$$\exp\left[-\frac{2eD}{\hbar}\sqrt{\frac{M}{a_{\rm exc}(D)}}\right]\sim\exp\left[-\frac{\sqrt{2}D}{a_{\rm exc}(D)}\right]\sim\exp\left[-\left(\frac{D}{2r_{\rm exc}}\right)^{1/4}\right]$$

($r_{\rm exc}=a^*/4$ is the radius of the plane two-dimensional exciton). At large D, the small tunneling parameter for two indirect excitons with oppositely-directed dipole moments is $\exp(-0.33\sqrt{D/r_{\rm exc}})$.

Let us show that at a small overlap between excitonic wave functions, the exchange contribution to the chemical potential is small. For simplicity (only in this specific case) we consider the Hamiltonian of a two-layered electron–hole system:

$$\hat{H}=\sum_{p=0,i}^{\infty}\left(\frac{p^2}{2m_e}-\mu_e\right)a_{pi}^{+}a_{pi}+\left(\frac{p^2}{2m_h}-\mu_h\right)b_{pi}^{+}b_{pi}$$
$$+\frac{1}{2}\sum_{pp'k}\sum_{i\neq j}V(k)(a_{pi}^{+}a_{p'j}^{+}a_{p'+k,j}a_{p-k,i}$$
$$+b_{pi}^{+}b_{p'j}^{+}b_{p'+k,j}b_{p-k,i})$$
$$-2\sum_{i,j}\tilde{V}(k)a_{pi}^{+}b_{p'j}^{+}b_{p'+k,j}a_{p-k,i}, \qquad (2)$$

where the subscripts i and j label the excitons composed of the corresponding electrons and holes; a_{pi}^{+} and b_{pi}^{+} are the creation operators of the electron and hole included in the exciton with number i; $m_e=m_h=m$ are the electron and hole effective masses; $V(k)=2\pi e^2/\epsilon k$ is the Coulomb interaction within one layer; $\tilde{V}(k)=(2\pi e^2/\epsilon k)\exp(-kD)$ is the interaction between an electron and a hole in different layers; μ_e and μ_h are the chemical potentials determined by the normalization condition (the densities of e and h are assumed to be equal, $N_e=N_h$):

$$\sum_{p,i}\langle a_{pi}^{+}a_{pi}\rangle=\sum_{p,i}\langle b_{pi}^{+}b_{pi}\rangle=\frac{1}{2}N,$$

where $N=N_e+N_h$ is the total number of particles in the system; $n=N/S$ is the two-dimensional density of particles in the system, and S is the total area. In order to calculate the exchange energy due to coupling between electrons and holes, let us apply Bogolyubov's canonical transform described by unitary operator $\hat{S}$ to the electron and hole operators:[20]

$$\hat{S}=\exp\left[\sum_{p,i}\phi_{pi}(a_{pi}^{+}b_{-pi}^{+}-b_{-pi}a_{pi})\right], \qquad (3)$$

$$\hat{S}a_{pi}\hat{S}^{+}=u_p a_{pi}+v_p b_{-pi}^{+}, \qquad (4)$$

$$\hat{S}b_{pi}\hat{S}^{+}=u_p b_{pi}+v_p a_{-pi}^{+}, \qquad (5)$$

where

$$u_p=\cos\phi_p, \quad v_p=\sin\phi_p, \quad u_p^2+v_p^2=1. \qquad (6)$$

Using anticommutation relations for fermion operators, we derive in the standard manner the transformed Hamiltonian

$$\hat{H}=\hat{H}_0+\hat{H}'+U, \qquad (7)$$

where U is a numerical functional of u and v, and $\hat{H}_0$ and $\hat{H}'$ are Hamiltonians that are quadratic and quartic in the transformed operators, respectively.

Function v_p obeys the normalization conditions

$$\sum_p v_p^2=n/2, \qquad (8)$$

which derives from the relations

$$\sum_{p,i}\langle\hat{S}a_{pi}^{+}a_{pi}\hat{S}^{+}\rangle=\sum_{p,i}\langle\hat{S}b_{pi}^{+}b_{pi}\hat{S}^{+}\rangle=2\sum_p v_p^2=n,$$

where n is the dimensionless surface density of charge carriers $na_{\rm exc}^2$; $a_{\rm exc}=(8r_{\rm exc})^{1/4}D^{3/4}$ is the indirect exciton radius.[14]

In order to calculate the exchange energy, let us analyze some components of Hamiltonian H_0:

$$H_{\rm ex}=\sum_{p,i\neq j}\left[(\varepsilon_p-\mu)u_p v_p-u_p^2\sum_{p'}\tilde{V}_{p-p'}u_{p'}v_{p'}\right]$$
$$\times(a_{pi}^{+}b_{-pi}^{+}+b_{-pi}a_{pi})-\left(2\sum_{p'}V_{p-p'}v_{p'}^2u_p v_p\right.$$
$$\left.+v_p^2\sum_{p'}\tilde{V}_{p-p'}u_{p'}v_{p'}\right)(a_{pi}^{+}b_{-pj}^{+}+b_{-pj}a_{pi}), \qquad (9)$$

where $\varepsilon_p=p^2/2\mu_{\rm exc}$. For calculating the average in the ground state, let us use the relationship

$$\langle a_{pi}^{+}b_{-pj}^{+}\rangle+\langle b_{-pj}a_{pi}\rangle\sim\mathcal{T}_{ij}(\langle a_{pi}^{+}b_{-pi}^{+}\rangle+\langle b_{-pi}a_{pi}\rangle), \qquad (10)$$

where $\mathcal{T}_{ij}=\mathcal{T}\sim\exp[-\sqrt{2}D/a(D)]\sim\exp[-(D/2r_{\rm exc})^{1/4}]$ is the tunneling matrix element for electrons and holes of different excitons through the potential barrier of the dipole–dipole repulsion, which is determined by the exponentially small overlap between their wave functions. Let us use the condition that coefficient k_p of the average pair combinations of operators $\langle a_{pi}^{+}b_{-pi}^{+}\rangle+\langle b_{-pi}a_{pi}\rangle$ (which yield singular contributions to the energy) should equal zero in Hamiltonian H_0 in Eq. (7) transformed to take into account the e-h coupling. From condition $k_p=0$ follows

$$\left(\varepsilon_p-\mu-2\sum_{p'}\mathcal{T}_{ij}V_{p-p'}v_{p'}^2\right)u_p v_p-(u_p^2-\mathcal{T}_{ij}v_p^2)$$
$$\times\sum_{p'}\tilde{V}_{p-p'}u_{p'}v_{p'}=0. \qquad (11)$$

In a low-density system, u_p is of order of unity, whereas v_p is small in terms of a dimensionless parameter. In the approximation of the lowest order in v_p, i.e., with terms of order up to $\sqrt{n}$, Eq. (11) reduces to

$$(\varepsilon_p-\mu_0)v_p-\int\tilde{V}_{p-p'}v_{p'}\frac{d^2p'}{(2\pi)^2}=0. \qquad (12)$$

In combination with the normalization condition (8), Eq. (12) yields

$$v_p=\sqrt{n}\psi_0(p), \quad \mu_0=-\epsilon_0, \qquad (13)$$

where ϵ_0 and $\psi_0(p)$ are the binding energy and wavefunction of the ground state of an isolated exciton with spatially separated electron and hole. The wave function ψ_0 and energy ϵ_0 of indirect excitons were previously calculated for various separations D between layers.[14] The correction of the next order in v_p to Eq. (11) describes exchange effects. The correction to the exciton chemical potential $\mu_{\rm exc}$ can be derived from Eq. (11) using the conventional perturbation theory by substituting v_p in the zero-order approximation from Eq. (13) in terms of order $n^{3/2}$, which are omitted in Eq. (12) and proportional to small parameter $\mathcal{T}$, and treating them as a perturbation. As a result, we obtain a correction to

the chemical potential $\mu_{\rm exc}$ in the region of small exciton densities n and small separations D between layers:

$$\mu_{\rm exc}\sim\mathscr{T}na_{\rm exc}^2\epsilon_0. \tag{14}$$

Thus, the exchange energy in a system with spatial separation of carriers is reduced in comparison with a one-layer electron–hole system. This is caused by the small tunneling exponent $\mathscr{T}$ due to penetration through a barrier due to the dipole–dipole interaction. Therefore, one can neglect exchange effects associated with the nonboson statistical properties of indirect excitons.

In analyzing the stability of the ground state of a weakly nonideal Bose gas of indirect excitons in a superlattice, let us employ the Bogolyubov approximation. The total Hamiltonian $\hat{H}_{\rm tot}$ of a low-density system of layered excitons is the Hamiltonian for a tenuous gas of two-dimensional indirect excitons with parallel and oppositely-directed dipole moments:

$$\hat{H}_{\rm tot}=\hat{H}_0+\hat{H}_{\rm int}. \tag{15}$$

Here $\hat{H}_0$ is the Hamiltonian of a system of noninteracting excitons:

$$\hat{H}_0=\sum_{\mathbf{p}}\ \varepsilon_0(p)(a_{\mathbf{p}}^+a_{\mathbf{p}}+b_{\mathbf{p}}^+b_{\mathbf{p}}+a_{-\mathbf{p}}^+a_{-\mathbf{p}}+b_{-\mathbf{p}}^+b_{-\mathbf{p}}), \tag{16}$$

where $\varepsilon_0(p)=p^2/2M$ is the spectrum of an isolated two-dimensional indirect exciton in a superlattice; $a_{\mathbf{p}}^+$, $b_{\mathbf{p}}^+$, $a_{\mathbf{p}}$, and $b_{\mathbf{p}}$ are the creation and annihilation operators of excitons with different orientations of their dipole moments; $\hat{H}_{\rm int}$ is the Hamiltonian of interaction between excitons:

$$\hat{H}_{\rm int}=\frac{U}{2V}\sum_{\mathbf{p}_1+\mathbf{p}_2=\mathbf{p}_3+\mathbf{p}_4}(a_{\mathbf{p}_4}^+a_{\mathbf{p}_3}^+a_{\mathbf{p}_2}a_{\mathbf{p}_1}+b_{\mathbf{p}_4}^+b_{\mathbf{p}_3}^+b_{\mathbf{p}_2}b_{\mathbf{p}_1} -a_{\mathbf{p}_4}^+a_{\mathbf{p}_3}^+b_{\mathbf{p}_2}b_{\mathbf{p}_1}-b_{\mathbf{p}_4}^+b_{\mathbf{p}_3}^+a_{\mathbf{p}_2}a_{\mathbf{p}_1}-a_{\mathbf{p}_4}^+b_{\mathbf{p}_3}^+a_{\mathbf{p}_2}b_{\mathbf{p}_1}), \tag{17}$$

where V is the system volume.

We now consider the case $T=0$. Assuming that most particles are in the condensate $[(N-N_0)/N_0\ll 1$, where N and N_0 are the total number of particles and the number of particles in the condensate], we take into account, as usual, only the interactions among condensate particles and between the excited particles and condensate, thus neglecting interactions among excited particles. Then the total Hamiltonian has the form

$$\hat{H}_{\rm tot}=\frac{1}{2}\sum_{\mathbf{p}\neq 0}[\varepsilon_0(p)(a_{\mathbf{p}}^+a_{\mathbf{p}}+b_{\mathbf{p}}^+b_{\mathbf{p}}+a_{-\mathbf{p}}^+a_{-\mathbf{p}}+b_{-\mathbf{p}}^+b_{-\mathbf{p}}) -Un_{\rm exc}(a_{\mathbf{p}}^+b_{-\mathbf{p}}^+ +a_{\mathbf{p}}b_{-\mathbf{p}}+a_{-\mathbf{p}}^+b_{\mathbf{p}}^+ +a_{-\mathbf{p}}b_{\mathbf{p}}+a_{\mathbf{p}}^+b_{\mathbf{p}} +a_{-\mathbf{p}}^+b_{-\mathbf{p}}+a_{\mathbf{p}}b_{\mathbf{p}}^+ +a_{-\mathbf{p}}b_{-\mathbf{p}}^+)], \tag{18}$$

where $n_{\rm exc}=N/V$ is the exciton density. All terms due to the first and second summands in Hamiltonian (17), which describe the mutual repulsion of indirect excitons with parallel dipole moments, are cancelled by the rest of the terms in Hamiltonian (17), which represent the attraction between indirect excitons with oppositely-directed dipole moments. As a result, we have only the terms describing attraction (terms responsible for attraction in Eq. (17) are more numerous than those describing repulsion because attraction can be described using a greater number of combination of creation and annihilation operators). Let us diagonalize Hamiltonian $\hat{H}_{\rm tot}$, which is quadratic in these operators, using a unitary transform of the Bogolyubov type[17]:

$$a_{\mathbf{p}}=\frac{1}{\sqrt{1-A_{\mathbf{p}}^2-B_{\mathbf{p}}^2}}(\alpha_{\mathbf{p}}+A_{\mathbf{p}}\alpha_{-\mathbf{p}}^+ +B_{\mathbf{p}}\beta_{-\mathbf{p}}^+),$$
$$b_{\mathbf{p}}=\frac{1}{\sqrt{1-A_{\mathbf{p}}^2-B_{\mathbf{p}}^2}}(\beta_{\mathbf{p}}+A_{\mathbf{p}}\beta_{-\mathbf{p}}^+ +B_{\mathbf{p}}\alpha_{-\mathbf{p}}^+), \tag{19}$$

where coefficients $A_{\mathbf{p}}$ and $B_{\mathbf{p}}$, obtained by equating the coefficients of the off-diagonal terms of the Hamiltonian to zero, are

$$A_{\mathbf{p}}=B_{\mathbf{p}}=\frac{1}{Un}[-\varepsilon_0(p)+\sqrt{(\varepsilon_0(p))^2-(nU)^2}]. \tag{20}$$

As a result, we obtain the diagonalized Hamiltonian:

$$\hat{H}_{\rm tot}=\sum_{\mathbf{p}\neq 0}\ \varepsilon(p)(\alpha_{\mathbf{p}}^+\alpha_{\mathbf{p}}+\beta_{\mathbf{p}}^+\beta_{\mathbf{p}}+\alpha_{-\mathbf{p}}^+\alpha_{-\mathbf{p}}+\beta_{-\mathbf{p}}^+\beta_{-\mathbf{p}}), \tag{21}$$

where $\varepsilon(p)$ is the spectrum of modified quasiparticles:

$$\varepsilon(p)=\sqrt{\varepsilon_0^2(p)-(nU)^2}. \tag{22}$$

At small momenta $p<\sqrt{2MnU}$, the excitation spectrum is purely imaginary because of interaction between excitons at large distances. Therefore, the low-density system of weakly interacting indirect excitons in a superlattice is unstable.

3. RADIUS AND BINDING ENERGY OF AN INDIRECT BIEXCITON

Consider the ground state of the system to be a tenuous weakly nonideal gas of two-dimensional indirect biexcitons formed by excitons with oppositely-directed dipole moments. If electron and hole quantum wells alternate, indirect biexcitons are constructed from indirect excitons with oppositely-directed dipole moments in adjacent well pairs (Fig. 1). With the appropriate mutual alignment of dipoles, they attract at long distances between them and repel at short distances, and their interaction potential $U(r)$ has the form described by Eq. (1), where r is the distance between parallel dipoles of indirect excitons in the quantum well plane (Fig. 2). Here the separation D between coupled quantum wells is assumed to be greater than the indirect exciton radius $a_{\rm exc}$ (Ref. 14): $D\gg a_{\rm exc}\sim(a^*)^{1/4}D^{3/4}$. When $r>1.11D$, the indirect excitons attract, whereas at $r<1.11D$ they repel. Notwithstanding the fact that, unlike the case of atoms and molecules, electron and hole masses are comparable, the adiabatic approximation applies to the problem of biexcitons. The small parameter in this case is not the ratio between the electron and hole masses, but the numerically small ratio between the biexciton and exciton binding energies.[21] In terms of spatial scales, the small parameter is the ratio be-

tween the exciton and biexciton radii in the quantum well plane. Curiously enough, in the biexciton problem, these parameters are small even if $m_e = m_h$. Moreover, they are even smaller than similar small parameters for atoms and molecules.[21] Their smallness will be confirmed by calculations of parameters of the indirect biexciton. The Schrödinger equation for internal motion of indirect excitons in an indirect biexciton in the center-of-mass frame has the form

$$-\frac{\hbar^2}{M}\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\psi(r)}{\partial r}\right)+U(r)\psi(r)=E\psi(r), \tag{23}$$

where $\psi(r)$ and E are the wave function and energy of internal motion in the indirect biexciton. The potential energy minimum is at distance $r=r_0\approx 1.67D$ between the indirect excitons (Fig. 2). For large D we expand potential energy $U(r)$ in powers of $(r-r_0)/D\ll 1$ and retain terms up to the second order:

$$U(r)=-0.04\frac{e^2}{\epsilon D}+0.44\frac{e^2}{\epsilon D^3}(r-r_0)^2. \tag{24}$$

Thus, the biexciton levels at large D correspond to those of a two-dimensional harmonic oscillator with frequency $\omega=\sqrt{0.88e^2/M\epsilon D^3}$ and an equilibrium point at $r=r_0$. The spectrum of low-lying levels is, therefore, equidistant:

$$E_n=-0.04\frac{e^2}{\epsilon D}+2\sqrt{2}E_0\left(\frac{r^*}{D}\right)^{3/2}(n+1), \tag{25}$$

where

$$E_0=\frac{Me'^4}{\hbar^2\epsilon},\quad r^*=\frac{\hbar^2\epsilon}{2Me'^2},\quad e'^2=0.88e^2.$$

In the ground state the energy E_0 and the characteristic ''spread'' $a_{\rm biex}$ of the biexciton in the quantum well plane (near the biexciton mean radius r_0) are

$$E_0=-0.04\frac{e^2}{\epsilon D}+2\sqrt{2}E_0\left(\frac{r^*}{D}\right)^{3/2}, \tag{26}$$

$$a_{\rm biex}=\sqrt{\frac{2\hbar}{M\omega}}=(8r^*)^{1/4}D^{3/4}=1.03a_{\rm exc}. \tag{27}$$

It follows from the equations given above that the adiabatic condition holds, since the ratio between the exciton and biexciton binding energies is $E_{\rm biex}/E_{\rm exc}=0.04\ll 1$. The ratio between the exciton and biexciton radii is $a_{\rm exc}/r_0=0.67(8r_{\rm exc})^{1/4}D^{-1/4}\ll 1$ for $D\gg a_{\rm exc}$.

The mean dipole moment of this biexciton is zero. It is clear, however, that its quadrupole moment is nonzero and equal to $Q=3eD^2$ (the quadrupole major axis is perpendicular to the quantum well plane). Therefore, indirect biexcitons interact at large distances as similarly aligned quadrupoles in accordance with the formula $U(R)=9e^2D^4/R^5$.

4. COLLECTIVE PROPERTIES AND SUPERFLUIDITY OF INDIRECT BIEXCITONS

At large separations between indirect biexcitons, $R\gg D$, there is a weak quadrupole repulsion described by the formula $U(R)=9e^2D^4/R^5$. Inequality $R_0\gg D$ applies to low-density systems provided that $n\ll 1/\pi D^2$. The exchange effects due to the nonboson statistical properties of indirect biexcitons are reduced by the small overlap between wave functions of two biexcitons if we take into account the potential barrier due to the long-range quadrupole repulsion (see the similar reasoning for the case of dipole excitons). The small tunneling parameter associated with this barrier is

$$\exp\left[-\frac{1}{\hbar}\int_{r_0}^{R'}\sqrt{2M_{\rm biex}\left(\frac{9e^2D^4}{R^5}-\mu\right)}\,dR\right],$$

where μ is the chemical potential of the system (see below); R' is the classical turning point for the quadrupole interaction, and is determined by the condition $U(R')=\mu$. At large D this parameter has the form $\exp(-0.93\sqrt{D/r_{\rm exc}})$. Hence, exchange effects for indirect biexcitons can be neglected for $D\gg r_{\rm exc}$.

In order to take into account the biexciton–biexciton scattering, one can use results of the two-dimensional Bose gas theory.[18] The chemical potential μ of two-dimensional bosons, with repulsion described by the formula $U(R)=AR^{-k}/M$ and the interparticle interaction taken into account by summing ladder diagrams,[17] is given by (hereafter $\hbar=1$)[18]

$$\mu=\frac{4\pi n}{M}\ln^{-1}\frac{1}{8\pi nA^{2/(k-2)}}. \tag{28}$$

In the case of biexcitons repelling like quadrupoles, the relations $A=9e^2D^4M_{\rm biex}$ and $k=5$ apply, and the chemical potential of this system is

$$\mu=\frac{4\pi n_{\rm biex}}{M_{\rm biex}}\ln^{-1}\frac{1}{8\pi(9e^2D^4M_{\rm biex})^{2/3}n_{\rm biex}}, \tag{29}$$

where $n_{\rm biex}=n/2$ is the biexciton density and $M_{\rm biex}=2M$ is the biexciton mass.

At small momenta the collective spectrum of biexcitons is acoustic: $\varepsilon(p)=c_s p$, where $c_s=\sqrt{\mu/M_{\rm biex}}$ is the speed of sound and μ is determined by Eq. (29). This spectrum satisfies the Landau superfluidity criterion. The local density $n_s(T)$ of the superfluid component for a two-dimensional Bose gas with an acoustic spectrum is given by

$$n_s=n_{\rm biex}-\frac{3\zeta(3)}{2\pi}\frac{T^3}{c_s^4M_{\rm biex}}, \tag{30}$$

where the second term on the right-hand side is the temperature dependent normal component density due to noninteracting excitations with the spectrum $\varepsilon(p)=\sqrt{\mu/M_{\rm biex}}p$ (cf. Ref. 7). The estimate of the local superfluid component density by Eq. (30) applies to low temperatures and low densities of excitations, which are assumed to be noninteracting, whereas the renormalization of n_s due to a contribution from vortices at temperatures below the Kosterlitz–Thouless transition is deemed negligible.

Superfluidity emerges in a two-dimensional system at temperatures below the Kosterlitz–Thouless transition T_c (Ref. 19):

$$T_c=\frac{\pi n_s}{2M_{\rm biex}}, \tag{31}$$

where bound vortices dissociate. Substituting the estimate for the superfluid component density n_s from Eq. (30) into Eq. (31), we obtain an equation for T_c, which then yields

$$T_c=\left[\left(1+\sqrt{\frac{16}{(6\cdot 0.45)^3\pi^4}\left(\frac{M_{\rm biex}T_c^0}{n_{\rm biex}}\right)^3+1}\right)^{1/3}+\left(1-\sqrt{\frac{16}{(6\cdot 0.45)^3\pi^4}\left(\frac{M_{\rm biex}T_c^0}{n_{\rm biex}}\right)^3+1}\right)^{1/3}\right]\frac{T_c^0}{(4\pi)^{1/3}}. \tag{32}$$

Here T_c^0 is an auxiliary parameter equal to the temperature at which the superfluid component density vanishes in the mean-field approximation, $n_s(T_c^0)=0$:

$$T_c^0=\left[\frac{2\pi n_{\rm biex}c_s^4M_{\rm biex}}{3\zeta(3)}\right]^{1/3}=\left[\frac{32}{3\zeta(3)}\ln^{-2}\frac{1}{8\pi n_{\rm biex}M_{\rm biex}^2D^4}\right]^{1/3}\frac{\pi n_{\rm biex}}{M_{\rm biex}}. \tag{33}$$

Equation (33) can be used as a rough estimate of the characteristic crossover temperature, at which the local superfluid component density becomes nonzero over distances smaller than the mean separation between vortices. The local superfluid component density can show up in local optical properties or local transport properties of the system.[7] In a low-density two-dimensional system and in the ladder approximation (when $\ln^{-1}[8\pi(9e^2D^4M_{\rm biex})^{2/3}n_{\rm biex}]\gg 1$), the Kosterlitz–Thouless temperature derived from Eqs. (32) and (33) is

$$T_c=(2\pi)^{-1/3}T_c^0. \tag{34}$$

As follows from Eq. (30), the normal component density at the maximum temperature of superfluidity (Kosterlitz–Thouless temperature) is given by

$$n_n(T_c)=\frac{3\zeta(3)}{2\pi}\frac{T_c^3}{c_s^4M_{\rm biex}}. \tag{35}$$

Substituting Eq. (32) into (35), we obtain

$$n_n(T_c)=\frac{n_{\rm biex}}{4\pi}\left\{\left[1+\sqrt{\frac{16}{(6\cdot 0.45)^3\pi^4}\left(\frac{M_{\rm biex}T_c^0}{n_{\rm biex}}\right)^3+1}\right]^{1/3}+\left[1-\sqrt{\frac{16}{(6\cdot 0.45)^3\pi^4}\left(\frac{M_{\rm biex}T_c^0}{n_{\rm biex}}\right)^3+1}\right]^{1/3}\right\}^3. \tag{36}$$

In the ladder approximation, we have for a low-density two-dimensional system

$$\frac{n_n(T_c)}{n_{\rm biex}}=\frac{1}{2\pi}. \tag{37}$$

Note that Eqs. (34) and (37) apply to any tenuous two-dimensional Bose gas. The dimensionless quantity $n_n(T_c)/n_{\rm biex}$ can be treated as a small parameter. Consequently, the approximation of an ideal Bose gas of acoustic excitations can be used in calculating n_s and T_c.

5. CONCLUSION

Spatially indirect excitons can transform to direct excitons via two processes: tunneling of electrons and holes between quantum wells and interaction between indirect excitons. The second effect was detected in experiments[15] and should be negligible at low densities. Interaction between direct excitons in different wells can lead to their transformation to indirect excitons. This process is also very slow in a low-density excitonic system. Transformations of direct excitons to indirect should fix the phase of the superfluid order parameter. This process should lead to various quasi-Josephson effects in an excitonic system in a superlattice (cf. Ref. 8). At sufficiently low temperatures, superfluid phases of direct and indirect excitons can coexist. In this case, the difference between phases of order parameter of direct and indirect excitons can be ascertained. As a result, a gap should open in oscillation spectra of the phase difference, which is proportional to the matrix element of the direct exciton transformation into the indirect exciton. In addition, various soliton-like excitations should be generated in a system.

This work was supported by the grants from the Russian Fund for Fundamental Research, INTAS, and Solid-State Nanostructures program sponsored by the Ministry of Science and Technology of Russia. One of the authors (O.L.B.) was supported by the Soros Postgraduate program financed by the G. Soros ISSEP fund, and by the International Center of Fundamental Physics program in Moscow (ICFPM).

*)E-mail: lozovik@isan.troitsk.ru

[1] L. V. Butov, A. Zrenner, G. Abstreiter, G. Bohm, and G. Weimann, Phys. Rev. Lett. **73**, 304 (1994).
[2] T. Fukuzawa, E. E. Mendez, and J. M. Hong, Phys. Rev. Lett. **64**, 3066 (1990); J. A. Kash, M. Zachau, E. E. Mendez, J. M. Hong, and T. Fukuzawa, Phys. Rev. Lett. **66**, 2247 (1991).
[3] U. Sivan, P. M. Solomon, and H. Strikman, Phys. Rev. Lett. **68**, 1196 (1992).
[4] J. P. Cheng, J. Kono, B. D. McCombe, I. Lo, W. C. Mitchel, and C. E. Stutz, Phys. Rev. Lett. **74**, 450 (1995).
[5] M. Bayer, V. B. Timofeev, F. Faller, T. Gutbrod, and A. Forchel, Phys. Rev. B **54**, 8799 (1996).
[6] Yu. E. Lozovik and V. I. Yudson, JETP Lett. **22**, 274 (1975); Zh. Éksp. Teor. Fiz. **71**, 738 (1976) [Sov. Phys. JETP **44**, 389 (1976)]; Solid State Commun. **18**, 628 (1976); **21**, 211 (1977); Yu. E. Lozovik, in *Proceedings of the 1st All-Union Conference on Dielectric Electronics*, Tashkent (1973).
[7] Yu. E. Lozovik and O. L. Berman, JETP Lett. **64**, 573 (1996); Zh. Éksp. Teor. Fiz. **111**, 1879 (1997) [JETP **84**, 1027 (1997)]; Fiz. Tverd. Tela **39**, 1654 (1997) [Phys. Solid State **39**, 1476 (1997)]; Phys. Scr. **55**, 491 (1997).
[8] A. V. Klyuchnik and Yu. E. Lozovik, Zh. Éksp. Teor. Fiz. **76**, 670 (1979) [Sov. Phys. JETP **49**, 335 (1979)]; Yu. E. Lozovik and A. V. Kluchnik, J. Low Temp. Phys. **38**, 761 (1980); J. Phys. C **11**, L483 (1978); Yu. E. Lozovik and V. I. Yudson, JETP Lett. **25**, 14 (1977); I. O. Kulik and S. I. Shevchenko, Solid State Commun. **21**, 409 (1977); Yu. E. Lozovik and V. I. Yudson, Solid State Commun. **22**, 117 (1977).
[9] Yu. E. Lozovik and O. L. Berman, Fiz. Tverd. Tela **40**, 1350 (1998) [Phys. Solid State **40**, 1228 (1998)]; Phys. Scr. **58**, 86 (1998).
[10] I. V. Lerner and Yu. E. Lozovik, Zh. Éksp. Teor. Fiz. **78**, 1167 (1980) [Sov. Phys. JETP **51**, 588 (1980)]; Zh. Éksp. Teor. Fiz. **80**, 1488 (1981) [Sov. Phys. JETP **53**, 763 (1981)]; Zh. Éksp. Teor. Fiz. **82**, 1188 (1982) [Sov. Phys. JETP **55**, 691 (1982)]; A. B. Dzyubenko and Yu. E. Lozovik, Fiz. Tverd. Tela **25**, 1519 (1983) [Sov. Phys. Solid State **25**, 874 (1983)]; **26**, 1540 (1984); J. Phys. A **24**, 415 (1991).
[11] D. Paquet, T. M. Rice, and K. Ueda, Phys. Rev. B **32**, 5208 (1985).

[12] S. M. Dikman and S. V. Iordanskiĭ, JETP Lett. **63**, 50 (1996).
[13] Yu. E. Lozovik, O. L. Berman, and V. G. Tsvetus, JETP Lett. **66**, 355 (1997); submitted to Phys. Rev. B.
[14] Yu. E. Lozovik and V. N. Nishanov, Fiz. Tverd. Tela **18**, 3267 (1976) [Sov. Phys. Solid State **18**, 1905 (1976)].
[15] A. I. Filin, V. B. Timofeev, S. I. Gubarev, D. Birkedal, and J. M. Hvam, JETP Lett. **65**, 656 (1997).
[16] M. Hermann, *Semiconductor Superlattices*, Academic Verlag, Berlin (GDR) (1986).
[17] A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskiĭ, *Methods of Quantum Field Theory in Statistical Mechanics* [in Russian], Fizmatgiz, Moscow (1962).
[18] Yu. E. Lozovik and V. I. Yudson, Physica A **93**, 493 (1978).
[19] J. M. Kosterlitz and D. J. Thouless, J. Phys. C **6**, 1181 (1973); D. R. Nelson and J. M. Kosterlitz, Phys. Rev. Lett. **39**, 1201 (1977).
[20] L. V. Keldysh and A. N. Kozlov, Zh. Éksp. Teor. Fiz. **54**, 978 (1968) [Sov. Phys. JETP **27**, 521 (1968)]; L. V. Keldysh and Yu. V. Kopaev, Fiz. Tverd. Tela **6**, 2791 (1964) [Sov. Phys. Solid State **6**, 221 (1964)]; A. N. Kozlov and L. A. Maksimov, Zh. Éksp. Teor. Fiz. **48**, 1184 (1965) [Sov. Phys. JETP **21**, 790 (1965)]; B. I. Halperin and T. M. Rice, Solid State Phys. **21**, 115 (1968).
[21] L. N. Ivanov, Yu. E. Lozovik, and D. R. Musin, J. Phys. C **11**, 2527 (1978).

Translation provided by the Russian Editorial office

Nonadiabatic effects in the phonon spectra of superconductors

A. E. Karakozov

L. F. Vereshchagin Institute of High-Pressure Physics, Russian Academy of Sciences

E. G. Maksimov*)

P. N. Lebedev Physics Institute, Russian Academy of Sciences

(Submitted 19 November 1998)
Zh. Éksp. Teor. Fiz. **115**, 1799–1817 (May 1999)

The nonadiabatic corrections to the self-energy part $\Sigma_s(\mathbf{q},\omega)$ of the phonon Green's function are studied for various values of the phonon vectors $\mathbf{q}$ resulting from electron–phonon interactions. It is shown that the long-range electron–electron Coulomb interaction has no direct influence on these effects, aside from a possible renormalization of the corresponding constants. The electronic response functions and $\Sigma_s(\mathbf{q},\omega)$ are calculated for arbitrary vectors $\mathbf{q}$ and energy ω in the BCS approximation. The results obtained for $q=0$ agree with previously obtained results. It is shown that for large wave numbers q, vertex corrections are negligible and $\Sigma_s(\mathbf{q},\omega)$ possesses a logarithmic singularity at $\omega=2\Delta$, where Δ is the superconducting gap. It is also shown that in systems with nesting, $\Sigma_s(\mathbf{Q},\omega)$ (where $\mathbf{Q}$ is the nesting vector) possesses a square-root singularity at $\omega=2\Delta$, i.e., exactly of the same type as at $q=0$. The results are used to explain the recently published experimental data on phonon anomalies, observed in nickel borocarbides in the superconducting state, at large $\mathbf{q}$. It is shown, specifically, that in these systems nesting must be taken into account in order to account for the emergence of a narrow additional line in the phonon spectral function $S(\mathbf{q},\omega) \approx -\pi^{-1}\mathrm{Im}\, D_s(\mathbf{q},\omega)$, where $D_s(\mathbf{q},\omega)$ is the phonon Green's function, at temperatures $T < T_c$. © *1999 American Institute of Physics.* [S1063-7761(99)02005-3]

1. INTRODUCTION

The phonon spectra of metals are ordinarily calculated in the adiabatic approximation,[1] i.e., the restructuring of the electronic subsystem in response to a shift of the ions is assumed to be instantaneous. Formally, this means that in the calculation of the lattice dynamics all electron response functions drawn into this process are taken into account in the static approximation. The corrections due to the dynamic characteristics of the electronic system (i.e., the frequency dependence of the electron response functions) are ordinarily small (of order $\sqrt{m/M}$, where m is the electron mass and M is the ion mass). Engelsberg and Schrieffer were the first to note[2] that nonadiabatic effects may not be small for optical phonons with small wave numbers $qv_F < \omega_0$, where v_F is the Fermi velocity of the electrons and ω_0 is the corresponding phonon frequency. In this case nonadiabatic effects can lead to a large renormalization of the phonon frequencies (not proportional to $\sqrt{m/M}$), strong dispersion of phonon frequencies for small q, and finite damping of such phonons.[3] This phenomenon has been observed in the Raman scattering of light by certain metals.[4,5] This problem has recently been investigated in detail in Ref. 6.

Nonadiabatic effects can be even stronger in the superconducting state for phonon frequencies $\omega \approx 2\Delta$, where Δ is the superconducting gap. The change in the phonon frequencies and dampings at a transition to the superconducting state has been observed[7] in the standard superconducting metal Nb. The theory of nonadiabatic effects for acoustic phonons in superconductors has been developed by Schuster[8] in the BCS approximation. In this work it was shown that the phonon frequencies and phonon linewidths (i.e., damping times) change at the transition of the metal to a superconducting state. Specifically, for phonon frequencies $\omega_{ph} > 2\Delta$ the phonon frequency itself increases (becomes harder) and the linewidth increases. For frequencies $\omega_{ph} < 2\Delta$ nonadiabatic effects decrease (soften) the phonon frequencies and the linewidths. Moreover, it was predicted in this work that a new, narrow line can appear in the phonon spectral function at frequencies $\omega \approx 2\Delta$. However, such lines have not been observed for acoustic phonons.[7]

Subsequently, the corresponding behavior of the phonon spectral function was observed in $NbSe_2$ for low-frequency ($q \approx 0$) optical phonons with $\omega \approx 2\Delta$.[9] A theoretical interpretation of this phenomenon, quite close to Schuster's interpretation,[8] was given in Ref. 10. In fact, in Refs. 8 and 10 the standard Fröhlich Hamiltonian, in which there is no direct Coulomb interelectron interaction, was used to describe nonadiabatic effects in the phonon spectra of superconductors. On this basis the results in Ref. 10 were criticized by Littlewood and Varma.[11] They asserted that the long-range Coulomb interaction leads at small wave numbers to complete screening of all contributions of the electron–phonon interaction (including nonadiabatic contributions) to the phonon frequencies. To explain the experimentally observed phenomena they invoked the specific nature of the low-frequency optical mode in $NbSe_2$, treating it as an amplitude mode of the charge-density wave in this system.

Note, however, that there is nothing special about this optical mode, associated with the structural transition to an incommensurate phase. Its only distinctive feature is that its frequency is much lower than the frequencies of all other optical phonons in this metal, and is comparable in magnitude to 2Δ. This mode has another property that can give rise to the observed phenomena, but we discuss it below. Moreover, we show that in the absence of low-frequency collective excitations of the electron charge and spin densities, the long-range Coulomb interaction does not introduce any substantial modifications in nonadiabatic effects.

A large number of investigations of phonons in high-temperature superconductors have been conducted in the last few years.[12] Shifts of the phonon frequencies and a change in the linewidths at a transition to the superconducting state have been observed. No new lines in the phonon spectral functions have been observed. The theory of nonadiabatic effects in the superconducting state of metals with a strong electron–phonon interaction has been proposed by Zeyher and Zwicknagl.[13] This theory made it possible to explain a substantial part of the experimental data on the shift of the phonon frequencies and the change in the linewidths. However, some data that do not fit the theory developed in Ref. 13 show changes in the phonons in the superconducting state.[14–16]

Very interesting observations of phonon spectra have recently been performed using inelastic neutron scattering.[17–19] The new superconducting compounds $LnNi_2B_2C$, where Ln is either Lu or Y, were studied. The behavior of the low-energy acoustic and optical vibrational branches in the direction $(\xi,0,0)$ with $\xi\approx 0.55$, i.e., close to the center of the Brillouin zone, were studied. The frequencies of both branches decreased with temperature; this in itself is atypical of ordinary metals. The phonon spectrum radically changed below the superconducting transition temperature T_c. A narrow peak appeared at energies ≈ 4 meV, which is somewhat less than 2Δ, and a wide peak corresponding to the energy of the phonons studied also appeared.

There exist at least two published theoretical works attempting to explain the observed facts. One[20] employs the conventional BCS approach for a three-dimensional quasi-isotropic system. In fact, the numerical results are very close to those obtained by Schuster[8] and, as follows from Schuster's work itself, for an appropriate choice of system parameters they can explain the changes in the phonon spectrum at temperatures $T<T_c$. However, the corresponding model cannot describe the behavior of phonon spectra at temperatures $T>T_c$. Moreover, the choice of the parameters in this model that are necessary to describe the situation at temperatures $T<T_c$ is in our opinion clearly unrealistic. In Ref. 21 it was suggested that the electron spectrum of the experimental compounds contains a certain fraction of nesting with momenta $\mathbf{Q}=(0.55,0,0)$. This conjecture is also confirmed by a detailed first-principles calculation[22] of the electronic polarization of the electronic band structure of these compounds. However, the analytic expressions in Ref. 21 for the changes in the phonon spectra in the superconducting state differ substantially from those that can actually be obtained for systems with nesting. The changes in the frequencies of the phonon modes and linewidths for $q\neq 0$ have also been calculated elsewhere.[23,24] Thus far, however, the role and magnitude of the nonadiabatic effects in superconducting metals are still not completely understood, even in the simple BCS model. We discuss these problems in detail in the present paper.

2. DERIVATION OF THE GENERAL RELATIONS IN THE BCS MODEL

We begin our analysis of nonadiabatic effects with the simple system described by the Fröhlich Hamiltonian

$$H=H_{ee}+H_{ph}+H_{\rm int}, \tag{1}$$

where H_{ee} is the Hamiltonian of the superconducting electrons, H_{ph} is the Hamiltonian of the starting phonons, and $H_{\rm int}$ describes the electron–phonon interaction

$$H_{\rm int}=\sum_{\mathbf{k},\mathbf{q},\sigma,s} g_s(\mathbf{k},\mathbf{q})a^+_{\mathbf{k},\sigma}a_{\mathbf{k}+\mathbf{q},\sigma}(b_{\mathbf{q},s}+b_{-\mathbf{q},s}). \tag{2}$$

Here $a_{\mathbf{q},\sigma}$ and $b_{\mathbf{q},s}$ are, respectively, the electron and phonon operators, and s characterizes the vibrational branches. The single-particle phonon Green's function can be represented in the form

$$D_s^{-1}(\mathbf{q},\omega)=\frac{\omega^2-\omega_0^2(\mathbf{q},s)}{\omega_0^2(\mathbf{q},s)}-\Sigma_s(\mathbf{q},\omega), \tag{3}$$

Here $\Sigma_s(\mathbf{q},\omega)$ is the self-energy part of the phonon Green's function and $\omega_0(\mathbf{q},s)$ is the starting frequency of the phonons. In the Matsubara representation the function $\Sigma_s(\mathbf{q},\omega)$ satisfies the integral equation[13]

$$\Sigma_s(\mathbf{q},i\omega_n)=T\sum_{\mathbf{k},m} g_s(\mathbf{k},\mathbf{q})\mathrm{Tr}\,\tau_3 G(\mathbf{k}+\mathbf{q},i\omega_m+i\omega_n)\times\Gamma_s(\mathbf{k}+\mathbf{q},\mathbf{k},i\omega_m+i\omega_n,i\omega_m)G(\mathbf{k},i\omega_m), \tag{4}$$

where the matrix vertex functions Γ_s satisfy the Bethe–Salpeter equation

$$\Gamma_s(\mathbf{k}+\mathbf{q},\mathbf{k},i\omega_m+i\omega_n,i\omega_m)=g_s(\mathbf{k},\mathbf{q})\tau_3-T\sum_{\mathbf{k}',m'}\tau_3 G(\mathbf{k}',i\omega_{m'})\Gamma_s(\mathbf{k}'+\mathbf{q},\mathbf{k}',\times i\omega_{m'}+i\omega_n,i\omega_{m'})G(\mathbf{k}'+\mathbf{q},i\omega_{m'}+i\omega_n)\tau_3 V(\mathbf{k}',\mathbf{k}), \tag{5}$$

where $V(\mathbf{k},\mathbf{k}')$ is the electron–electron interaction leading to superconductivity. In the Fröhlich model, generally speaking, there is no need to introduce any special electron–electron interaction leading to superconductivity, since this is simply the electron–phonon interaction itself.

In this case, the function $V(\mathbf{k},\mathbf{k}')$ can be expressed in terms of the phonon Green's function, and it depends not only on the momenta $\mathbf{k}$ and $\mathbf{k}'$ but also the energy $i\omega_m$. For superconductors with weak coupling, it is well known that the Eliashberg theory[25] for systems with an electron–phonon interaction reduces to the BCS model. At present we confine ourselves to this approximation. The effect of strong

electron–phonon coupling will be examined later. We represent the electronic Green's function in the BCS model as

$$G(\mathbf{k},i\omega_n)=\frac{i\omega_n\tau_0+\Delta_{\mathbf{k}}\tau_1+\epsilon_{\mathbf{k}}\tau_3}{(i\omega_n)^2+\epsilon_{\mathbf{k}}^2+\Delta_{\mathbf{k}}^2}. \tag{6}$$

Here $\epsilon_{\mathbf{k}}$ is the spectrum of single-particle excitations and $\Delta_{\mathbf{k}}$ is the superconducting gap, described by the BCS equation

$$\tau_2\Delta_{\mathbf{k}}=-T\sum_{\mathbf{k}',n}\tau_3G(\mathbf{k}',i\omega_n)\tau_2\Delta_{\mathbf{k}'}\tau_3V(\mathbf{k}',\mathbf{k}). \tag{7}$$

To simplify the solution of the resulting equations we represent the electron–electron interaction $V(\mathbf{k},\mathbf{k}')$ in the factorized form

$$V(\mathbf{k}',\mathbf{k})=-\sum_{s,L}g_{sL}^2\theta(\omega_c-|\epsilon_{\mathbf{k}}|)\theta(\omega_c-|\epsilon_{\mathbf{k}}'|)\Psi_L^*(\mathbf{k}')\Psi_L(\mathbf{k}). \tag{8}$$

Here ω_c is the cutoff energy, equal in order of magnitude to the characteristic phonon frequencies, and $\Psi_L(\mathbf{k})$ is a complete set of orthogonal functions. The latter can be chosen, for example, to be Fermi-surface harmonics.[26] Their specific form is of no consequence for our purposes.

We seek a solution of the Bethe–Salpeter equation (5) in the form

$$\Gamma_s=\Gamma_s^{(2)}\tau_2+\Gamma_s^{(3)}\tau_3, \tag{9}$$

$$\Gamma_s^{(2,3)}(\mathbf{k}',\mathbf{k})=\sum_L\Gamma_{s,L}^{(2,3)}\Psi_L^*(\mathbf{k}')\Psi_L(\mathbf{k}). \tag{10}$$

All other functions in Eqs. (3) and (4) can also be expanded in the harmonics $\Psi_L(\mathbf{k})$. Here we confine ourselves to isotropic s pairing. This means that the energy gap $\Delta_{\mathbf{k}}$ will possess only the harmonic with $L=0$

$$\Delta_{\mathbf{k}}=\Psi_{L=0}(\mathbf{k})\Delta_{L=0}. \tag{11}$$

After long but simple calculations, which we do not present here, we obtain for the analytically continued vertex functions

$$\Gamma_{sL}^{(2)}(\mathbf{q},\omega)=g_{sL}(\mathbf{q})\frac{ig_L^2I_2^L(\mathbf{q},\omega)}{[1+g_L^2I_3^L(\mathbf{q},\omega)][1-g_L^2I_1^L(\mathbf{q},\omega)]+[g_L^2I_2^L(\mathbf{q},\omega)]^2}, \tag{12}$$

$$\Gamma_{sL}^{(3)}(\mathbf{q},\omega)=g_{sL}(\mathbf{q})\left\{1-\frac{g_L^2I_3^L(\mathbf{q},\omega)-g_L^4[I_1^L(\mathbf{q},\omega)I_3^L(\mathbf{q},\omega)-(I_2^L(\mathbf{q},\omega))^2]}{[1+g_L^2I_3^L(\mathbf{q},\omega)][1-g_L^2I_1^L(\mathbf{q},\omega)]+[g_L^2I_2^L(\mathbf{q},\omega)]^2}\right\}, \tag{13}$$

where

$$g_L^2=\sum_s g_{sL}^2 \tag{14}$$

and the functions $I_j^L(\mathbf{q},\omega)$ can be represented using the functions $I_{\alpha,\beta,\gamma}^L(\mathbf{q},\omega)$:

$$I_{\alpha,\beta,\gamma}^L(\mathbf{q},\omega)=\frac{1}{4}\sum_{\mathbf{k}}|\Psi_L(\mathbf{k})|^2\theta(\omega_c-|\epsilon_{\mathbf{k}}|)\tanh\frac{E_{\mathbf{k}}}{2T}\left\{M_{\alpha,\beta,\gamma}^+\times\left(\frac{1}{E_{\mathbf{k}+\mathbf{q}}-E_{\mathbf{k}}+\omega+i\delta}+\frac{1}{E_{\mathbf{k}+\mathbf{q}}-E_{\mathbf{k}}-\omega-i\delta}\right)+M_{\alpha,\beta,\gamma}^-\left(\frac{1}{E_{\mathbf{k}+\mathbf{q}}+E_{\mathbf{k}}+\omega+i\delta}+\frac{1}{E_{\mathbf{k}+\mathbf{q}}+E_{\mathbf{k}}-\omega-i\delta}\right)\right\}. \tag{15}$$

Here $M_{\alpha,\beta,\gamma}^+$ and $M_{\alpha,\beta,\gamma}^-$ are the corresponding coherence factors

$$M_{\alpha,\beta,\gamma}^{\pm}=\alpha\pm\frac{\beta\Delta^2+\gamma\epsilon_{\mathbf{k}+\mathbf{q}}\epsilon_{\mathbf{k}}}{E_{\mathbf{k}+\mathbf{q}}E_{\mathbf{k}}}, \tag{16}$$

and $E_{\mathbf{k}}$ is the energy of the superconducting quasiparticles

$$E_{\mathbf{k}}=\sqrt{\epsilon_{\mathbf{k}}^2+\Delta^2}. \tag{17}$$

The functions $I_{j=1,2,3}^L(\mathbf{q},\omega)$ are determined by the function $I_{\alpha,\beta,\gamma}^L(\mathbf{q},\omega)$ with the following values of α, β, and γ:

$$\begin{aligned}&\alpha=1,\quad \beta=-1,\quad \gamma=-1,\quad j=1,\\ &\alpha=0,\quad \beta=\omega/\Delta,\quad \gamma=0,\quad j=2,\\ &\alpha=1,\quad \beta=-1,\quad \gamma=1,\quad j=3.\end{aligned} \tag{18}$$

The form of the resulting expressions for the vertex functions is very similar to that obtained in Refs. 27–29 in a calculation of the dielectric response of superconducting electrons. There is only one substantial difference, associated with the matrix elements of the electron–electron interaction. In our analysis the long-range Coulomb interaction is completely absent. As we will show, this interaction does not directly affect the phenomenon which we are studying.

We now write an expression for the Lth harmonic of the phonon self-energy part,

$$\Sigma_{sL}(\mathbf{q},\omega)=-2g_{sL}^2\tilde{I}_3^L(\mathbf{q},\omega)-2g_{sL}^2\times\frac{I_3^L(\mathbf{q},\omega)[1-g_L^2I_1^L(\mathbf{q},\omega)]+g_L^2(I_2^L(\mathbf{q},\omega))^2}{[1+g_L^2I_3^L(\mathbf{q},\omega)][1-g_L^2I_1^L(\mathbf{q},\omega)]+[g_L^2I_2^L(\mathbf{q},\omega)]^2}. \tag{19}$$

The function $\tilde{I}_3^L(\mathbf{q},\omega)$ is determined by the same equation (15) but with $\theta(|\epsilon_{\mathbf{k}}|-\omega_c)$ instead of $\theta(\omega_c-|\epsilon_{\mathbf{k}}|)$. This func-

tion describes in the Fröhlich model the contribution of the part of the electron–phonon interaction due to regions far from the Fermi surface to the phonon self-energy. As a result, the function $I_3^L(\mathbf{q},\omega)$ is completely independent of the superconductivity parameters, and to calculate it Δ can be set to zero.

The equations obtained above completely describe the situation in systems with s pairing for any wave vectors $\mathbf{q}$ and frequencies ω. They are suitable for quasi-isotropic systems and systems with nesting. We consider the case $q=0$ and $T=0$ first. Then the functions $I_j^L(\mathbf{q},\omega)$ have the form

$$I_1^L(0,\omega)=\frac{1}{g_0^2}+\left(\frac{\omega}{2\Delta}\right)^2\frac{N(0)I(\omega)}{2}, \tag{20}$$

$$I_2^L(0,\omega)=-\frac{\omega}{2\Delta}\frac{N(0)I(\omega)}{2}, \tag{21}$$

$$I_3^L(0,\omega)=\frac{N(0)I(\omega)}{2}, \tag{22}$$

$$\tilde{I}_3^L=0. \tag{23}$$

Here $N(0)$ is the density of states at the Fermi surface

$$I(\omega)=\frac{1}{2}\int_{-\epsilon_F}^{\epsilon_F}d\epsilon_{\mathbf{k}}\frac{\Delta^2}{E_{\mathbf{k}}[E_{\mathbf{k}}^2-(\omega/2)^2-i\delta]}. \tag{24}$$

Using expressions (20)–(23), it is easy to verify that

$$\Sigma_{s,0}(\mathbf{q}=0,\omega)=0. \tag{25}$$

This equality is a consequence of a Ward identity, which exists in the Fröhlich model, just as in any system with a gradient-invariant Hamiltonian.[27] It appears only when vertex corrections are rigorously taken into account. The starting value of the phonon self-energy part neglecting vertex corrections (i.e., for $\Gamma_s=g_s$) does not satisfy the identity (25). As shown in Refs. 11–13, the long-range Coulomb interaction is directly related to the component $L=0$ of the self-energy part $\Sigma_{sL}(\mathbf{q},\omega)$ and causes the latter to vanish at $q=0$. In systems with a gradient-invariant Hamiltonian this quantity, as we have verified, is zero in itself. For this reason, phonons with $q=0$, which have a representation with $L=0$ (or, from the standpoint of the theory of crystal groups, a unitary representation with the complete symmetry of the point group), do not interact with the electrons in either the normal or superconducting states and do not manifest any nonadiabatic effects. In contrast to the assertions made in Refs. 11 and 13 this result is in no way related to the long-range Coulomb interaction. Of course, just as for the dielectric response,[27–29] the Coulomb interaction can renormalize the corresponding polarization operators and the constants characterizing them, such as, for example, $N(0)$. But this does not at all alter the physics of nonadiabatic effects so long as there are no low-energy collective excitations of the electron charge or spin densities.

We now consider the $L\neq 0$ component of the phonon self-energy part $\Sigma_{sL}(\omega)$, which has the form

$$\Sigma_{sL}(\omega)=-2\lambda_{sL}\frac{I(\omega)/2}{1-\lambda_L\beta_L(\omega/2\Delta)I(\omega)/2}, \tag{26}$$

where λ_L is the coupling constant in the Lth harmonic

$$\lambda_L=g_L^2N(0) \tag{27}$$

and

$$\beta_L\left(\frac{\omega}{2\Delta}\right)=\left(\frac{\omega}{2\Delta}\right)^2\frac{\lambda_0}{\lambda_0-\lambda_L}-1. \tag{28}$$

The function $I(\omega)$ can easily be calculated and is given by

$$I(\omega)=\frac{1}{2}\int_{-\infty}^{\infty}\frac{dx}{\sqrt{1+x^2}[1+x^2-(\omega/2\Delta)^2-i\delta]}$$

$$=\frac{4\Delta/\omega}{\sqrt{1-(\omega/2\Delta)^2}}\arcsin\frac{\omega}{2\Delta},\quad \omega<2\Delta,$$

$$\frac{-2\cdot 2\Delta/\omega}{\sqrt{(\omega/2\Delta)^2-1}}\left(\operatorname{arcsinh}\sqrt{\left(\frac{\omega}{2\Delta}\right)^2-1}-\frac{i\pi}{2}\right),\quad \omega>2\Delta. \tag{29}$$

The expression (26) for $\Sigma_{sL}(\omega)$ demonstrates the possible existence, well-known in the theory of dielectric response,[26] of a pole in the polarization operator of superconducting electrons in the $L\neq 0$ channel if

$$1-\lambda_L\frac{I(\omega_b)}{2}\beta_L\left(\frac{\omega_b}{2\Delta}\right)=0. \tag{30}$$

This expression yields the energy ω_b of collective oscillations,

$$\omega_b\approx 2\Delta-\Delta\lambda_L^2\left(\frac{\lambda_L}{\lambda}\right)^2. \tag{31}$$

As Maki and Tsuneto showed long ago,[30] impurity scattering of electrons can substantially modify the expressions obtained above for the electronic response functions. Specifically, this can cause the pole in the polarization operator and therefore the collective electronic excitations to vanish. This problem was recently discussed in detail in Ref. 31 as part of a study of the Raman scattering spectra of superconductors. This work confirmed the results in Ref. 30 and showed that vertex corrections can be neglected for strongly disordered systems. These results can easily be extended to our case of nonadiabatic effects in the phonon spectra of superconductors. For optical phonons with $q=0$, this was done in Ref. 32, where isotropic s pairing and anisotropic d pairing were considered. Specifically, it was shown that in the weak-coupling approximation, vertex corrections do not significantly alter the phonon spectral functions compared with those calculated in their absence, even for systems without impurities.

To conclude this section we examine the function $\Sigma(\mathbf{q},\omega)$ for momenta $q\neq 0$ satisfying

$$k_F>q>\frac{\omega}{v_F},\frac{2\Delta}{v_F}, \tag{32}$$

where v_F is the Fermi velocity of the electrons and k_F is the Fermi momentum. It is easy to show that the functions $I_{j=1,2,3}(\mathbf{q},\omega)$ in this case reduce to the expressions

$$\tilde{I}_3(\mathbf{q},\omega)\approx N(0), \tag{33}$$

$$I_1(\mathbf{q},\omega)\approx I_3(\mathbf{q},\omega), \tag{34}$$

$$I_2(\mathbf{q},\omega)\approx -\frac{\omega}{2\Delta}I_3(\mathbf{q},\omega). \tag{35}$$

It follows from Eqs. (33)–(35) and (19) that vertex corrections for large phonon momenta q do not lead to a pole in the vertex functions, and they can be neglected in the weak-coupling approximation under study. The self-energy part of the phonon Green's function in this case is given by

$$\Sigma_s(\mathbf{q},\omega)=-2g_s^2N(0)-2g_s^2I_3(\mathbf{q},\omega), \tag{36}$$

where

$$I_3(\mathbf{q},\omega)\approx\frac{\pi\omega N(0)}{2qv_F}\ln\frac{2\Delta}{|2\Delta-\omega|}+i\frac{\pi\omega N(0)}{2qv_F}\theta(\omega-2\Delta), \tag{37}$$

which is identical to the result obtained in Refs. 8 and 20. Formally, the experimentally observed[16–18] change in the phonon spectral density can be explained, as shown in Ref. 20, using these formulas. In fact, however, Eqs. (36) and (37), as we have already noted in the introduction and as we show below, cannot serve to explain these data.

3. SYSTEMS WITH NESTING

Theoretical calculations of the electronic polarizability of the compound $LuNi_2B_2C$ show the existence of so-called nesting in the electron spectrum for sufficiently large sections of the Fermi surface. Nesting is usually understood to mean that

$$\epsilon_{\mathbf{k}}\approx -\epsilon_{\mathbf{k}+\mathbf{Q}} \tag{38}$$

for single-electron energies in some finite phase volume near the Fermi surface. The vector $\mathbf{Q}$ for which Eq. (38) holds is the nesting vector. As is well known,[33] for ideal nesting [i.e., exact satisfaction of Eq. (38)] the static electron polarizability $\Pi(\mathbf{q},0)$ diverges at $\mathbf{q}=\mathbf{Q}$. In turn, this can lead to waves of spin density and/or structural transitions. Magnetic structure with an incommensurate period characterized by the vector $\mathbf{Q}=(0.55,0,0)$ is indeed observed in a number of nickel borocarbide compounds, for example, $EuNi_2B_2C$ and $HoNi_2B_2C$.[34] In $LuNi_2B_2C$ and YNi_2B_2C, no magnetic structure is observed, but as mentioned in the introduction substantially softening of the acoustic and optical modes occurs for wave vectors $\mathbf{q}$ close to the nesting vector $\mathbf{Q}$. Both phonon lines are broadened in this case. In the superconducting state, i.e., at $T<T_c$, the phonon spectral function discussed in the introduction changes abruptly. These circumstances together show that nesting could play an important role in the modification of the phonon lines in the normal and superconducting states.

Proceeding now to the self-energy part of the phonon Green's function for systems with nesting, we first write its expression in zeroth order, i.e., neglecting vertex functions. Using the nesting condition (38) and the equations for $\Sigma_s(\mathbf{q},\omega)$ obtained in the preceding section, it is easy to obtain for $\mathbf{q}=\mathbf{Q}$

$$\Sigma_0(\mathbf{Q},\omega)=-\frac{g^2(\mathbf{Q})}{2}\sum_{\mathbf{k}}\tanh\left(\frac{E_{\mathbf{k}}}{2T}\right)\left(\frac{1}{2E_{\mathbf{k}}+\omega+i\delta}+\frac{1}{2E_{\mathbf{k}}-\omega-i\delta}\right). \tag{39}$$

For the normal state, i.e., $\Delta=0$, the well-known expression[33] for $\Sigma_0(\mathbf{Q},\omega)$ follows from Eq. (39):

$$\Sigma_0(\mathbf{Q},\omega)=-2g^2(\mathbf{Q})N(0)\left(\ln\frac{\epsilon_F}{\max\{\omega,T\}}+i\frac{\pi}{4}\tanh\frac{\omega}{4T}\right). \tag{40}$$

The equation (40) shows that as temperature decreases, the phonon frequencies at $\mathbf{q}=\mathbf{Q}$ soften and the phonon lines are broadened as a result of an increase in the imaginary part of $\Sigma_0(\mathbf{Q},\omega)$. This behavior of the phonon lines of the function agrees quite reasonably with that observed in borocarbides.[17,19] Of course, a direct comparison of experimental data with the theoretical results in the zeroth approximation is hardly reasonable. It is obvious that in calculating the function $\Sigma_0(\mathbf{Q},\omega)$, even for the normal state, vertex corrections as well as the changes induced in the single-particle electron Green's functions by the electron–phonon and the electron–electron Coulomb interactions must be taken into account self-consistently.

The corresponding analysis of the electronic susceptibility for metallic systems with nesting has been carried out in Ref. 35 on the basis of a semiphenomenological self-consistent method taking account of only the electron–electron Coulomb interaction. We do not consider this problem here, since our main concern here is to clarify the specific nature of the change induced in the phonon spectra by a transition to the superconducting state.

The expression (39) for $\Sigma_0(\mathbf{Q},\omega)$ at $T=0$ can be cast in the form

$$\Sigma_0(\mathbf{Q},\omega)=-g^2(\mathbf{Q})N(0)\int_{-\epsilon_F}^{\epsilon_F}\frac{d\epsilon_{\mathbf{k}}}{E_{\mathbf{k}}}-g^2(\mathbf{Q})N(0)\times\left(\frac{\omega}{2\Delta}\right)^2\int_{-\epsilon_F}^{\epsilon_F}\frac{d\epsilon_{\mathbf{k}}}{E_{\mathbf{k}}}\frac{\Delta^2}{E_{\mathbf{k}}^2-(\omega/2)^2}. \tag{41}$$

The first term in this equation

$$-2g^2(\mathbf{Q})N(0)\ln(\epsilon_F/\Delta), \tag{42}$$

is quite well known[33,36] and describes the elimination of the singularity of the static response functions in systems with nesting when the system passes to a superconducting state. In particular, this terminates the softening of the phonon modes in the adiabatic approximation and leads to the absence of structural transformations at temperatures $T<T_c$. The second term describes the singular behavior of the function $\Sigma_0(\mathbf{Q},\omega)$ at energies $\omega\approx 2\Delta$, and is due to a transition to the superconducting state. Comparing Eq. (41) and the expression (24) for the function $I(\omega)$, we see that $\Sigma(\mathbf{Q},\omega)$ at $\omega\approx 2\Delta$ has exactly the same singularity as at $q=0$, i.e., $\approx[1-(\omega/2\Delta)^2]^{-1/2}$. This result, just like the actual expression for the function $\Sigma_0(\mathbf{Q},\omega)$, completely contradict the recently published expression in Ref. 21. A logarithmic sin-

gularity of $\Sigma_0(\mathbf{Q},\omega)$ at $\omega\approx 2\Delta$ was obtained incorrectly there. In reality, as shown above, this occurs in a quasi-isotropic system at large momenta $qv_F \gg \Delta$.

Using the expressions (15)–(18) for the function $I_{j=1,2,3}(\mathbf{Q},\omega)$ it is easy to show that

$$I_1(\mathbf{Q},\omega)=N(0)I(\omega), \tag{43}$$

$$I_2(\mathbf{Q},\omega)=-(\omega/2\Delta)N(0)I(\omega), \tag{44}$$

$$I_3(\mathbf{Q},\omega)=1/\lambda+(\omega/2\Delta)^2N(0)I(\omega), \tag{45}$$

$$\tilde{I}_3(\mathbf{Q},\omega)=N(0)\ln(\epsilon_F/\omega_c). \tag{46}$$

Accordingly, $\Sigma_s(\mathbf{Q},\omega)$ at $T=0$ can be written in the form

$$\Sigma_s(\mathbf{Q},\omega)=-2\lambda_{\mathbf{Q}}\ln\frac{\epsilon_F}{\Delta}-2\lambda_{\mathbf{Q}}\frac{I(\omega)/2-1/\lambda}{2-\lambda I(\omega)/2[2-(\omega/2\Delta)^2]}. \tag{47}$$

Here $\lambda_{\mathbf{Q}}=g^2(\mathbf{Q})N(0)$ and λ is the total electron–phonon coupling constant determined by the condition

$$1=\lambda\ln(2\omega_c/\Delta). \tag{48}$$

It is evident from the expression (46) that $\Sigma_s(\mathbf{Q},\omega)$ possesses a pole corresponding to collective excitations that is determined by the second term in Eq. (47):

$$2-\frac{\lambda I(\omega_p)}{2}\left[2-\left(\frac{\omega_p}{2\Delta}\right)^2\right]=0. \tag{49}$$

Hence it follows that

$$(\omega_p)^2=(2\Delta)^2\left[1-\left(\frac{\lambda}{2}\right)^2\right]. \tag{50}$$

The corresponding residue r is

$$r=\lambda_{\mathbf{Q}}\lambda\Delta^2. \tag{51}$$

Using the expressions obtained for $\Sigma_s(\mathbf{Q},\omega)$ and Eq. (3), the expression for $D_s^{-1}(\mathbf{Q},\omega)$ can be rewritten as

$$D_s^{-1}(\mathbf{Q},\omega)=\frac{\omega^2-\tilde{\omega}_0^2(\mathbf{Q},s)}{\omega_0^2(\mathbf{Q},s)}-\tilde{\Sigma}_s(\mathbf{Q},\omega), \tag{52}$$

where $\tilde{\Sigma}_s(\mathbf{Q},\omega)$ is the second term in Eq. (47). Notation was also introduced for the phonon frequencies renormalized as a result of the first term in Eq. (47):

$$\tilde{\omega}_0^2(\mathbf{Q},s)=\omega_0^2(\mathbf{Q},s)\left(1-2\lambda_{\mathbf{Q}}\ln\frac{\epsilon_F}{\Delta}\right). \tag{53}$$

Using the equations (49)–(51) obtained above, the expression (52) can be rewritten as

$$D_s^{-1}(\mathbf{Q},\omega)=\frac{\omega^2-\tilde{\omega}_0^2(\mathbf{Q},s)}{\omega_0^2(\mathbf{Q},s)}-\frac{r}{\omega^2-\omega_p^2}. \tag{54}$$

It is easy to show that the equation

$$D_s^{-1}(\mathbf{Q},\omega)=0 \tag{55}$$

possesses two solutions for any $\tilde{\omega}_0(\mathbf{Q},s)>2\Delta$. One solution corresponds to a phonon with energy

$$\omega_{ph}^2\approx\omega_0^2+\frac{r\omega_0^2(\mathbf{Q},s)}{\tilde{\omega}_0^2(\mathbf{Q},s)-\omega_p^2}>(2\Delta)^2. \tag{56}$$

The other solution exists for $\omega<2\Delta$ and describes the existence of an additional mode, mentioned several times in this paper, with energy

$$\omega_b^2\approx\omega_p^2-\frac{r\omega_0^2(\mathbf{Q},s)}{\tilde{\omega}_0^2(\mathbf{Q},s)-\omega_p^2}<(2\Delta)^2. \tag{57}$$

The spectral density of this mode can also easily be calculated:

$$S_s(\mathbf{Q},\omega)=-\frac{2}{\pi\omega_0^2}\,\mathrm{Im}D_s(\mathbf{Q},\omega)$$

$$=\frac{r\omega_0^2(\mathbf{Q},s)}{[\tilde{\omega}_0^2(\mathbf{Q},s)-\omega_p^2]^2+r\omega_0^2(\mathbf{Q},s)}\,\frac{\delta(\omega-\omega_b)}{\pi\omega_b}. \tag{58}$$

Although two solutions of Eq. (55) exist for any phonon energy $\omega\approx\tilde{\omega}_0(\mathbf{q},s)$, the intensity of the additional mode is found to be very low for $\tilde{\omega}_0(\mathbf{q},s)\gg 2\Delta$. The intensities of the additional mode and of the phonon are found to be comparable only for phonon energies $\tilde{\omega}_0(\mathbf{q},s)\gtrsim 2\Delta$.

Formally, our solutions for the energy ω_p of collective excitations and the energy ω_b of the additional mode and its intensity are very close to those obtained in Ref. 21. However, there are substantial qualitative differences. First and foremost, this is due to the fact that the energy ω_p of the collective modes differs from both 2Δ and the residue r, which were obtained in Ref. 21, by an exponentially small quantity

$$\frac{2\Delta-\omega_p}{2\Delta}\approx e^{-\epsilon_F/\lambda\Delta},$$

where ϵ_F is the Fermi energy. An exponentially small difference of the same type,

$$\frac{2\Delta-\omega_b}{2\Delta}\approx e^{-qv_F/\lambda\Delta},$$

between the energy ω_b of the additional mode and 2Δ also arises in a quasi-isotropic system, as follows from Eqs. (36) and (37). It is evident from the expressions (50), (51), and (57) that no exponential smallness in Δ/ϵ_F and Δ/qv_F exists for systems with nesting. This makes it much easier for an additional mode to appear in such systems. From the experimental data of Ref. 18, which show that ω_b ranges from 4 to 6 meV, and especially from the wave-number dependence of the energy of the additional mode, it clearly follows that the difference between this energy and 2Δ does not have any special exponential smallness. We do not give here a detailed discussion of the experimental data using the expressions obtained above, since it is unlikely that ideal nesting exists in real physical systems. Instead, we present below the results of numerical calculations of $\mathrm{Im}\,D_s(\mathbf{q},\omega)$ performed using a model where nesting can vary from ideal to complete absence.

To this end we employed a two-dimensional model of a square lattice of strongly coupled electrons with overlap between nearest neighbors. The electron spectrum of such a system can be written in the form

$$\epsilon_{\mathbf{k}} = -2t(\cos k_x + \cos k_y) - \mu, \tag{59}$$

where μ is the chemical potential of the system, characterizing the degree of filling of the band, and t is the overlap integral. For a half-filled band ($\mu = 0$), the Fermi surface is square, i.e., it corresponds to a system with ideal nesting. A sparsely filled Fermi surface is nearly circular, i.e., it is a standard quasi-isotropic system. By varying the degree of filling of the band it is possible to investigate the entire transitional region from ideal nesting to an isotropic system. This model was used in Ref. 24 to calculate the self-energy parts of the phonon Green's functions in superconductors in the zeroth approximation, i.e., neglecting vertex corrections. However, the photon Green's functions themselves and the quantities $S_s(\mathbf{q},\omega)$ were not calculated in this work.

We calculated the function $S_s(\mathbf{q},\omega)$ numerically in a model with isotropic pairing, taking account of vertex corrections completely. To this end, the functions $I_{\alpha,\beta,\gamma}(\mathbf{q},\omega)$ were calculated for various values of the filling of the band, i.e., the values of μ, and for wave vectors $\mathbf{q}$ equal and close to the nesting vector. For ideal nesting the choice of nesting vector is unique, and is governed by the geometry of the Fermi surface. In our model with $\mu = 0$, the nesting vector $\mathbf{Q}_0$ is directed along the diagonals of the square Brillouin zone, and equals in magnitude half the corresponding reciprocal lattice vector. By changing the filling of the band the ''nesting'' vector $\mathbf{Q}$ can be determined, for example, according to the position of the maximum of the static response function in the normal state, i.e., the function $\epsilon_0(\mathbf{q},0)$.

In our simple case the ''nesting'' vector can also be found from the following considerations. In a geometric sense, the vector $\mathbf{Q}$ connects congruent sections of the Fermi surface. For a simply-connected Fermi surface, as in our case, these are ''almost flat'' sections whose relative arrangement is determined by the lattice symmetry, so that it can be inferred that the ''nesting surface,'' i.e., the ''surface'' of local minima of the static electron response function, is formed by the tips of the vectors $2\mathbf{k}_F$ and the vectors $\mathbf{q}$, which are equivalent to them, obtained by symmetry transformations of a square lattice from the vectors $2\mathbf{k}_F$ (Fig. 1). The vectors $\mathbf{Q}$ are determined by the position of the singular ''lines'' on the ''nesting surface,'' which are ''lines'' of self-intersection of this ''surface'' (and, naturally, ''lines'' of its intersection with the boundaries of the Brillouin zone, and for this reason one of the coordinates of the nesting vectors is exactly $\pm\pi/a$).

Figure 1 shows the corresponding construction for $\mu = -0.8t$ (henceforth μ is considered to be a dimensionless quantity in units of the overlap integral t). The nesting lines are identical to the numerical results obtained by Marsiglio.[24] The ''nesting'' vectors corresponding the maxima of the static electron response function are also shown. Thus, as the filling varies, the maximum corresponding to the nesting vector $\mathbf{Q}_0$ at $\mu = 0$ splits into four equivalent maxima. The parameter $\mu = -0.8$, which for clarity we chose to construct

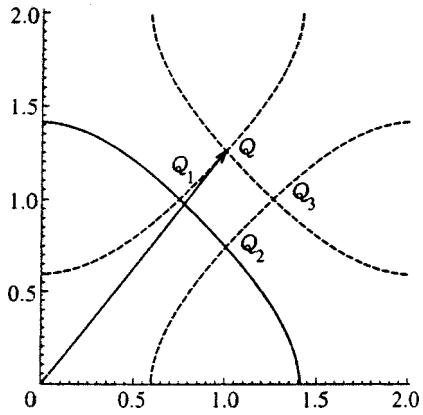

FIG. 1. Nesting vectors and lines for a square lattice ($\mu = -0.8$) at the coordinates π/a. The ''surface'' $2\mathbf{k}_F$ (solid curve) and equivalent surfaces (dashed curves).

Fig. 1, is quite far from the ideal nesting parameter, so the nesting effect is quite weak here. Below, we examine more suitable examples for band filling factors corresponding to 1.11 electrons percenter ($\mu = 0.2$) and 0.89 electrons per center ($\mu = -0.2$).

Figure 2 shows the behavior of the spectral function of phonon modes 1 and 2 at four different temperatures. It was assumed that the initial frequencies of these modes, $\omega_0^{(1)}(\mathbf{q})$ and $\omega_0^{(2)}(\mathbf{q})$ are degenerate, i.e., $\omega_0^{(1)}(\mathbf{q},s) = \omega_0^{(2)}(\mathbf{q},s)$. The electron–phonon coupling constants of these modes were chosen to be $\lambda_1 = 0.1$ and $\lambda_2 = 0.085$, respectively, so that $\lambda_1 > \lambda_2$. The total coupling constant $\lambda = 0.372$ was found from the condition (48), so that $\Delta \approx 3$ meV for $\omega_c \approx 30$ meV. This choice of Δ and ω_c approximately corresponds to the values observed in borocarbides. The results shown in Fig. 2 were obtained in a calculation with $\mu = 0.2$, i.e., for a more than half-filled band. The same results are obtained from $\mu = -0.2$.

It is evident from Fig. 2 that at $T > T_c$ the phonon frequencies soften with decreasing temperature and the phonon lines broaden, as also happens for an ideal nesting. At first glance it may appear that intensity is transferred from the peak with high energy $\omega^{(2)}(\mathbf{q})$ to the peak with $\omega^{(1)}(\mathbf{q})$. Of course, in our approach there can be no transfer of intensity from one peak to another, since mode interaction is neglected. In reality, the amplitude of the spectral density increases for all modes whose energy decreases with temperature. This can easily be verified by writing the function $S_s(\mathbf{q},\omega)$ in the form

$$S_s(\mathbf{q},\omega) \approx \frac{\delta(\omega - \tilde{\omega}(\mathbf{q},s))}{\tilde{\omega}(\mathbf{q},s)}. \tag{60}$$

Of course, such an increase in $S_s(\mathbf{q},\omega)$ has no effect on the sum rule for the Green's function $D_s(\mathbf{q},\omega)$, which expresses the conservation of the total number of phonon states,

$$\int_{-\infty}^{\infty} d\omega\, \omega\, S_s(\mathbf{q},\omega) = 1. \tag{61}$$

This example demonstrates that to interpret the experimental data of Ref. 19 care must be taken in determining the param-

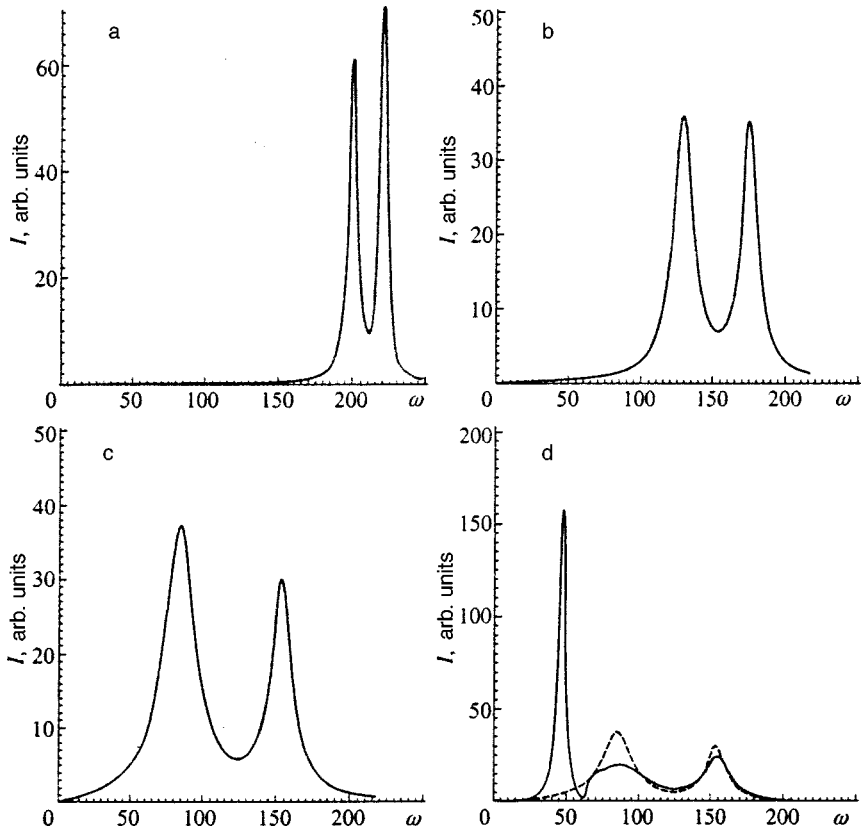

FIG. 2. Intensities of phonon modes 1 and 2 as a function of frequency ω, expressed in kelvins, for $T=500$ K (a), $T=120$ K (b), $T=T_c$ (c) and $T=0$ (solid curve), $T=T_c$ (dashed curve) (d).

eters of the model of coupled modes and the intensity transfer, arising in the model, from one mode to another at temperatures $T>T_c$. Figure 2d shows the phonon spectral density at $T\ll T_c$. The existence of a narrow additional peak is clearly seen in this figure, and it is also evident, as compared with Fig. 2c, that there exists an additional transfer of intensity from mode 1 to the additional mode.

We have studied variations in the energy and intensity of the additional mode as a function of the deviation of the phonon wave vector from the ''nesting'' vector. Figure 3 shows the numerical results obtained with the vectors $\mathbf{q}=\mathbf{Q}\pm\delta\mathbf{q}$, which vary in absolute magnitude over the range $1-\xi\equiv 1-|\mathbf{q}|/|\mathbf{Q}|\approx\pm 0.15$ on the ''nesting'' lines, i.e., on the lines of local maxima of the static electron response function, and in the direction $\mathbf{Q}$. The ξ dependences of the energy and intensity of the additional mode which have been presented above agree to some extent with the experimental data of Ref. 19, despite the existence of obvious discrepancies. Thus, the experimental data demonstrate only the existence of a large anisotropy with respect to an increase or decrease of the vector $\mathbf{q}$. It is clear from Fig. 1 that as the filling of the band approaches 50%, the equivalent nesting vectors converge, which produces a large nesting phase volume and increases the probability of observing the additional mode in practice. Depending on the degree of filling, the peaks associated with the wave vectors $\mathbf{Q}$ and $\mathbf{Q}_1$ are due either to direct phonon scattering or Umklapp processes. Thus, when less than half the band is filled ($\mu<0$), the peak at the wave vector $\mathbf{Q}_1$ is due to direct scattering and the peak at a larger vector $\mathbf{Q}$ is due an Umklapp process. This is easily verified from Fig. 1, whence it is obvious that the difference of the vectors $-\mathbf{Q}_2$ and $\mathbf{Q}$ is precisely a reciprocal lattice vector directed along the diagonal of the Brillouin zone. The opposite situation occurs for a more than half-filled band ($\mu>0$). Generally speaking, this can lead to a large difference of the matrix elements of the electron–phonon interaction which are responsible for the corresponding processes. Specifically, the matrix element of the electron–phonon interaction for the direct scattering process can be very small

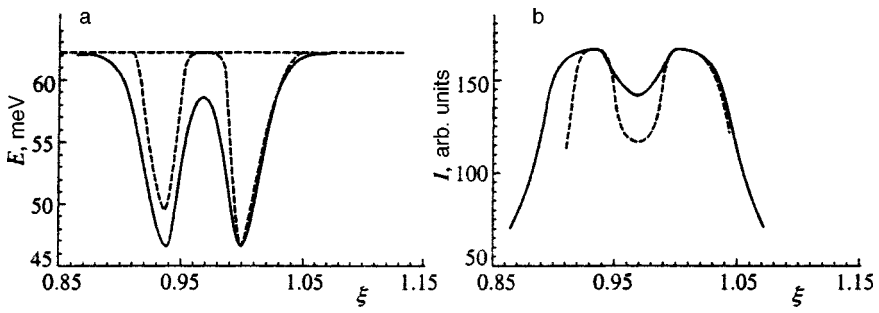

FIG. 3. Variation of the energy (a) and intensity (b) of the additional phonon mode as a function of the deviation of the phonon wave vector $\mathbf{q}$ from the nesting vector $\mathbf{Q}$ on the nesting line (solid curve) and in the direction of the vector $\mathbf{Q}$ (dashed curves); $\xi=|\mathbf{q}|/|\mathbf{Q}|$.

for a transverse phonon. In our calculations we neglected this, since otherwise it would be necessary to perform direct microscopic calculations of the corresponding matrix elements, which falls outside the scope of the present work.

The last problem that we wish to discuss here briefly is the temperature dependence of the characteristics of the additional mode. In the BCS model studied earlier, the temperature dependences of the energy and intensity of this mode are governed by the T-dependence of 2Δ. The experimental data[17–19] show an obvious deviation from this behavior. Thus the energy of the additional mode is essentially temperature-independent, and its intensity decreases with increasing temperature much more rapidly than $2\Delta(T)$. As we show below, this can be explained completely naturally on the basis of the tight-binding theory based on the Eliashberg equations.[37]

Just the fact, for example, that the intensity of the additional mode decreases abruptly as the temperature approaches T_c can easily be understood even without any numerical calculations. Indeed, as is well known for the case $q=0$,[30] impurities sharply reduce the intensity of collective excitations in superconductors. This happens because the electrons acquire a finite lifetime $1/\gamma_{\rm imp}$ ($\gamma_{\rm imp}$ is the relaxation rate of electrons on impurities), which results both in a decrease in the amplitude of the response functions near singularities and a change (weakening) in the type of singularities from square-root to logarithmic for $\Delta/\gamma_{imp}>1$. The dynamic electron–phonon interaction also results in electrons acquiring a finite lifetime, even in the superconducting state, which also depends on energy and temperature, and in the elimination of singularities in the response functions and in the electronic density of states.[38] The electron relaxation time (the reciprocal of the lifetime) increases rapidly as temperature approaches T_c, and this results in a sharp decrease in the intensity of the additional mode.

We calculated these effects numerically for the ideal-nesting model ($\mu=0$) in the zeroth approximation, i.e., neglecting vertex corrections. Similar calculations for systems without nesting with $q=0$ were performed in Ref. 13 at a temperature close to zero. Omitting the lengthy but simple calculations, the expression for the phonon self-energy part $\Sigma_{0s}(\mathbf{Q},\omega)$ can be written in the form

$$\frac{\Sigma_{0s}(\mathbf{Q},\omega)}{2\lambda_s N(0)}=\int_0^{\omega}d\omega'\tanh\frac{\omega'}{2T}\left\{\frac{-1+n(\omega-\omega')n(\omega')+a(\omega-\omega')a(\omega')}{\varepsilon(\omega-\omega')+\varepsilon(\omega')+2i\gamma_{\rm imp}}\right.$$
$$\left.-\frac{-1-n(\omega-\omega')n^*(\omega')-a(\omega-\omega')a^*(\omega')}{\varepsilon(\omega-\omega')-\varepsilon^*(\omega')+2i\gamma_{\rm imp}}\right\}-\int_0^{\infty}d\omega'\left(\tanh\frac{\omega+\omega'}{2T}-\tanh\frac{\omega'}{2T}\right)$$
$$\times\left\{\frac{-1-n(\omega+\omega')n(\omega')+a(\omega+\omega')a(\omega')}{\varepsilon(\omega+\omega')+\varepsilon(\omega')+2i\gamma_{\rm imp}}-\frac{-1+n(\omega+\omega')n^*(\omega')-a(\omega+\omega')a^*(\omega')}{\varepsilon(\omega+\omega')-\varepsilon^*(\omega')+2i\gamma_{\rm imp}}\right\}$$
$$+2{\rm Re}\left\{\int_0^{\omega_c}d\omega'\tanh\frac{\omega+\omega'}{2T}\,\frac{-1-n(\omega+\omega')n(\omega')+a(\omega+\omega')a(\omega')}{\varepsilon(\omega+\omega')+\varepsilon(\omega')+2i\gamma_{\rm imp}}\right\}-\ln\frac{\epsilon_F}{\omega_c}. \qquad (62)$$

Here

$$n(\omega)=\frac{\omega}{\sqrt{\omega^2-\Delta^2(\omega)}},\quad a(\omega)=\frac{\Delta(\omega)}{\sqrt{\omega^2-\Delta^2(\omega)}}, \qquad (63)$$

$$\varepsilon(\omega)=Z(\omega)\sqrt{\omega^2-\Delta^2(\omega)}, \qquad (64)$$

and the functions $\Delta(\omega)$ and $Z(\omega)$ are determined by the Eliashberg equations. The main quantity determining the solution of the Eliashberg equations is the spectral density $\alpha^2(\omega)F(\omega)$ of the electron–phonon interaction. We chose this function in accordance with the function obtained from tunneling experiments.[39] We normalized it so that the total coupling constant

$$\lambda=2\int_0^{\infty}\frac{d\omega}{\omega}\alpha^2(\omega)F(\omega) \qquad (65)$$

was approximately 0.7. This value of λ makes it possible to obtain T_c and Δ in good agreement with experiment. (The renormalized coupling constant $\lambda/(1+\lambda)\approx 0.4$ also agrees well with the coupling constant that we used in the BCS calculations above.)

Our numerical results for the energy and intensity of the additional mode are displayed in Fig. 4. It is evident from Fig. 4a that the energy of the mode is essentially independent of T over a sizable temperature range. Even in the tight-binding theory, the energy of the mode naturally starts to decrease rapidly as $T\to T_c$, but as follows from Fig. 4b, this decrease occurs mainly in the temperature range where the intensity of this mode is low and where the mode is in fact unobservable.

4. CONCLUSIONS

In summary, our analytic and numerical calculations demonstrate that the experimentally observed behavior of the phonon spectral density in superconducting nickel borocarbides can be reasonably explained qualitatively on the basis of a model that takes account of the existence of a sizable

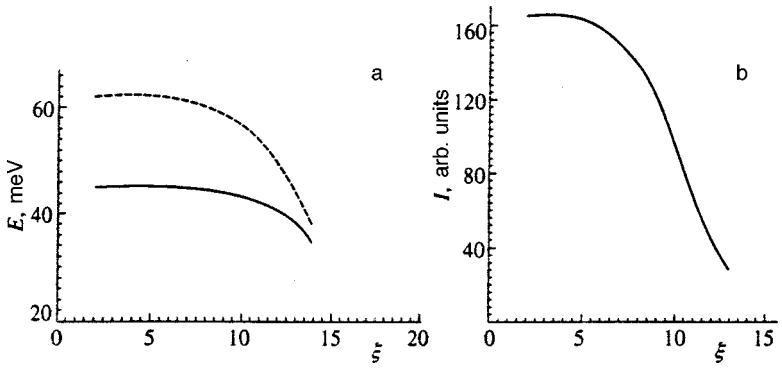

FIG. 4. Variation of the energy (a) and intensity (b) of the additional phonon mode in the superconducting state as a function of temperature. The dashed curve shows the function $2\Delta(T)$.

''nesting'' fraction in the electronic spectrum of these compounds. This model makes it possible to explain simultaneously the change in the phonon spectrum (its softening) in the normal state and the emergence and properties of the additional mode in the superconducting state. A quantitative comparison of the numerical results with experimental data would be premature at the moment. This would require more detailed experimental data, along with numerical calculations of the electronic response functions that take account of the specific crystalline and electronic structure of borocarbide compounds.

This work was partially supported by the Russian Fund for Fundamental Research (Grants 96-02-16134 and 98-02-17187) and by the high-temperature superconductivity program VTSP-64.

*)E-mail: maksimov@tp.lpi.ac.ru

[1] E. G. Brovman and Yu. M. Kagan, Usp. Fiz. Nauk **112**,369 (1974) [Sov. Phys. Usp. **17**, 125 (1974)].
[2] S. Engelsberg and J. R. Schriffer, Phys. Rev. **131**, 953 (1963).
[3] M. Cardona and I. P Ipatova, in *Elementary Excitation in Solids*, L. J. Birman, C. Sebenal, and R. F. Wallis (eds.), Elsevier Science Publishers, Amsterdam (1992).
[4] Yu. S. Ponosov and G. A. Bolotin, Fiz. Tverd. Tela (Leningrad) **27**, 2636 (1985) [Sov. Phys. Solid State **27**, 1581 (1985)].
[5] Yu. S. Ponosov, G. A. Bolotin, C. Thomsen *et al.*, submitted to Phys. Stat. Sol.
[6] E. G. Maksimov and S. V. Shulga, Solid State Commun. **97**, 553 (1996).
[7] J. D. Axe and G. Shirane, Phys. Rev. B **11**, 1965 (1972).
[8] H. G. Schuster, Solid State Commun. **13**, 1559 (1973).
[9] S. Sooryakumar and M. V. Klein, Phys. Rev. Lett. **45**, 660 (1980).
[10] C. A. Balsiero and L. M. Falikov, Phys. Rev. Lett. **45**, 662 (1980).
[11] P. B. Littlewood and C. M. Varma, Phys. Rev. B **26**, 4883 (1982).
[12] C. Thomsen, in *Light Scattering in Solids*, Vol. 6, M. Cardona and G. Guntherodt (eds.), Springer-Verlag, Heidelberg (1995), p. 285.
[13] R. Zeyher and G. Zwicknagl, Z. Phys. B **78**, 175 (1990).
[14] H. S. Obhi and E. K. H. Salje, Physica C **171**, 547 (1990).
[15] Y. Yagil and E. K. H. Salje, Physica C **229**, 152 (1994).
[16] N. N. Zhigadlo, M. Zhang, and E. K. H. Salje, Supercond. Sci. Technol. **10**, 209 (1997).
[17] C. Stassis, M. Bullock, P. Canfield, A. I. Goldman, G. Shirane, and S. M. Shapiro, Phys. Rev. B **55**, R8678 (1997).
[18] H. Kawano, H. Yoshizawa, H. Takeya, and K. Kadowaki, Phys. Rev. Lett. **77**, 4628 (1996).
[19] M. Bullock, J. Zarestky, C. Stassis, A. I. Goldman, P. Canfield, G. Shirane, and S. M. Shapiro, Phys. Rev. B **57**, 7916 (1998).
[20] P. B. Allen, V. N. Kostur, N. Takesue, and G. Shirane, Phys. Rev. B **56**, 5552 (1997).
[21] H.-Y. Kee and C. M. Varma, Phys. Rev. Lett. **79**, 4250 (1997).
[22] J. Y. Rhee, X. Wang, and B. N. Harmon, Phys. Rev. B **51**, 15585 (1995).
[23] R. Zeyher, Phys. Rev. B **44**, 9596 (1991).
[24] F. Marsiglio, Phys. Rev. B **47**, 5419 (1993).
[25] G. M. Éliashberg, Zh. Éksp. Teor. Fiz. **38**, 966 (1960) [Sov. Phys. JETP **11**, 696 (1960)].
[26] P. B. Allen, Phys. Rev. B **13**, 1416 (1976).
[27] J. Schrieffer, *Theory of Superconductivity*, Benjamin, New York (1964).
[28] R. E. Prange, Phys. Rev. **129**, 2495 (1963).
[29] V. G. Vaks, V. M. Galitskiĭ, and A. I. Larkin, Zh. Éksp. Teor. Fiz. **41**, 1655 (1961) [Sov. Phys. JETP **14**, 1177 (1962)].
[30] K. Maki and T. Tsuneto, Prog. Theor. Phys. **28**, 163 (1962).
[31] T. P. Devereaux, Phys. Rev. B **47**, 5230 (1993).
[32] T. P. Devereaux, Phys. Rev. B **50**, 10277 (1994).
[33] Yu. V. Kopaev, Trudy FIAN **86**, 3 (1975).
[34] J. Zarestky, C. Stassis, A. I. Goldman *et al.*, Phys. Rev. B **51**, 678 (1995).
[35] A. Virosztek and J. Ruvalds, Phys. Rev. B **42**, 4064 (1990).
[36] T. V. Ramakrishnan and C. M. Varma, Phys. Rev. B **24**, 137 (1981).
[37] G. M. Éliashberg, Zh. Éksp. Teor. Fiz. **39**, 1437 (1960) [Sov. Phys. JETP **12**, 1000 (1961)].
[38] A. E. Karakozov, E. G. Maksimov, and A. A. Mikhailovsky, Solid State Commun. **79**, 329 (1991); A. E. Karakozov, E. G. Maksimov, and A. A. Mikhaĭlovskiĭ, Zh. Éksp. Teor. Fiz. **102**, 132 (1992) [Sov. Phys. JETP **75**, 70 (1992)].
[39] I. K. Yanson, V. V. Fizun, A. G. M. Jansen *et al.*, Phys. Rev. Lett. **78**, 935 (1997).

Translated by M. E. Alferieff

Thermal stabilization of anomalies in inhomogeneous conducting structures

A. M. Satanin*) and V. V. Skuzovatkin

N. I. Lobachevskiĭ Nizhniĭ Novgorod State University, 603600 Nizhniĭ Novgorod, Russia
(Submitted 25 November 1998)
Zh. Éksp. Teor. Fiz. **115**, 1818–1832 (May 1999)

We study the thermal mechanism of suppression of the anomalies in the nonlinear characteristics of inhomogeneous media. A generalized expression for the effective nonlinear conductivity is derived that allows for heat transfer from hot regions. We study the nature of the divergences in two- and three-dimensional inhomogeneous structures as depending on the local parameters and the microgeometry of the system. Finally, we show that in the critical region the effective nonlinear conductivity may be much higher than the conductivity of the components. © *1999 American Institute of Physics.* [S1063-7761(99)02105-8]

1. INTRODUCTION

The nonlinear electrical properties of randomly inhomogeneous and periodic structures may differ substantially from the properties of the components. This difference occurs because the effective characteristics are depend not only on the properties of the microcomponents but also on the local distribution of fields and currents in the medium. Lately it has been discovered that the effective nonlinear characteristics may be anomalously sensitive to the microstructure of substances.[1] Earlier this was demonstrated in a series of experiments. For instance, Dubson *et al.*[2] measured the third-harmonic amplitude $V_{3\omega}$ on metal films near the metal–insulator transition. They found that the critical exponents of the concentration dependence of $V_{3\omega}$ are not universal—they depend on the method of film preparation, i.e., on the microstructure of the substance. A sharp increase in the strength of the linear effects and the nonuniversal behavior of these effects near the percolation threshold were demonstrated in Refs. 3–5.

The occurrence of anomalies is related to the heating of the electron gas in local regions, in which the current percolates through narrow bridges formed by a poorly conducting material. It is these regions where there is concentration of current and field that provide the main contribution to the effective nonlinear characteristics, such as the nonlinear conductivity, the third-harmonic amplitude, and the coefficient of $1/f$ noise.[2,3,6,7] Obviously, local temperature distribution and conditions of heat exchange can strongly influence the nonlinear response of the medium.

In the present paper we study the effect of heat transfer (thermal conduction and heat exchange) on the nonlinear properties of periodic structures. As an example of a two-dimensional structure we study a periodic ''checkerboard'' lattice. In the three-dimensional case, we study the packing of alternating cubes in which the neighborhoods of contact of adjacent cubes are approximated by two highly conducting cones. We generalize the results of Refs. 8–11 by allowing for heat transfer from ''hot'' regions. We also find the conditions under which the nonlinear response is determined by small singular regions and reveal the role of the thermal mechanism of stabilization of anomalies of the medium.

2. STRUCTURE MODEL AND BASIC EQUATIONS

The geometry of the two-dimensional structure is depicted in Fig. 1a. The hatched sections correspond to an electrical conductivity σ_1 and a thermal conductivity κ_1 (metal), while the light sections correspond to σ_2 and κ_2 (insulator). In three-dimensional media, the regions of greatest singularity are those where pyramidal or conical regions touch each other. We discuss the singularity by using an exactly solvable model whose geometry is depicted in Fig. 1b, with the internal regions of the cones having an electrical conductivity σ_1 and a thermal conductivity κ_1 (metal) and the external regions, an electrical conductivity σ_2 and a thermal conductivity κ_2 (insulator). We assume that the structure in Fig. 1b is only a fragment of a three-dimensional lattice. We denote the characteristic size of the unit cell by L_0 in both two- and three-dimensional cases.

The electric current in the structures is described by the formula

$$\mathbf{j}=\sigma(T)\mathbf{e}, \tag{1}$$

where σ is temperature-dependent. In the absence of heating at $T=T_0$ (T_0 is the equilibrium temperature of the medium), $\sigma(T_0)$ is a periodic function, with values σ_1 and σ_2 in adjacent regions. The current and field obey the equations

$$\operatorname{div}\mathbf{j}=0, \quad \operatorname{curl}\mathbf{e}=0 \tag{2}$$

and the following conditions at the boundaries of the regions with different conductivity:

$$(\mathbf{j}\cdot\mathbf{n})_1=(\mathbf{j}\cdot\mathbf{n})_2, \quad (\mathbf{e}\cdot\boldsymbol{\tau})_1=(\mathbf{e}\cdot\boldsymbol{\tau})_2, \tag{3}$$

where $\mathbf{n}$ and $\boldsymbol{\tau}$ are vectors normal and tangential to the interface separating the two media.

We can determine the effective characteristics of an inhomogeneous medium from the expression for the volume-averaged energy dissipation:

$$\langle\mathbf{j}\cdot\mathbf{e}\rangle=\sigma_e\mathbf{E}^2+\chi_e\mathbf{E}^4+\cdots, \tag{4}$$

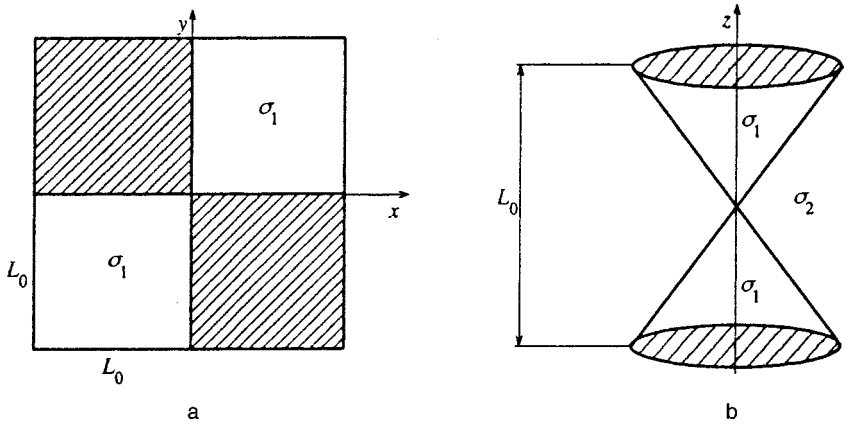

FIG. 1. Geometry of the structure considered in this paper: (a) a portion of a two-dimensional two-component structure, and (b) example of contact of two conical sections in a three-dimensional medium.

which is written in the form of an expansion in powers of the mean field $\mathbf{E}=\langle\mathbf{e}\rangle$. The angle brackets stand for averaging over the volume (or surface area) of the system. Below we limit ourselves to cubic nonlinearities. The expansion (4) is valid if the nonlinear corrections are small in a certain sense. For this to be the case, σ_e must be much larger than $\chi_e E^2$, and this can be achieved by reducing the mean field in the sample. This limitation means that we are in the weak-nonlinearity regime, and an increase in the effective nonlinear conductivity means that the linear region narrows. When there is a local relationship between current and field, $\mathbf{j}=\sigma\mathbf{e}+\chi\mathbf{e}^2\mathbf{e}$, we can use (4) to show (basing our reason on Tellegen's theorem[12]) that the effective nonlinear conductivity χ_e is determined by the fourth-order correlator of the field in a linear medium:

$$\chi_e\mathbf{E}^4=\langle\chi\mathbf{e}^4\rangle \tag{5}$$

(justification for this can be found in Refs. 13 and 14). Equation (5) suggests that the effective conductivity is determined by the regions that provide the main contribution to the integral $\int\chi\mathbf{e}^4\,d\mathbf{r}$. As shown in Refs. 8–15, the solutions of Eqs. (1)–(3) in the linear case have singularities near microconstrictions, which enhances the release of Joule heat. Since the function $\mathbf{j}\cdot\mathbf{e}$ is spatially inhomogeneous, large temperature gradients develop near the singularities. Here we must allow for heat exchange and an inhomogeneous temperature distribution in the system. To calculate the effective nonlinear conductivity we must in turn generalize Eq. (5), which earlier was obtained for a local temperature–field relationship. The nonlocal relationship between current and field can dramatically change the nature of the singularities.

We can describe the steady-state temperature distribution in the medium by the equation

$$-\operatorname{grad}\kappa\cdot\operatorname{grad}T=-\alpha(T-T_0)+\mathbf{j}\cdot\mathbf{e}, \tag{6}$$

where κ is the thermal conductivity coefficient, which takes on values κ_1 and κ_2 in the different components, and α is the heat-exchange coefficient, which can differ from sector to sector.

In determining the temperature in a two-dimensional medium we additionally allow for heat exchange between film and substrate. To do this we use a model according to which the heat flux into the substrate is proportional to the temperature difference between film and substrate. It is assumed that the temperature of the massive substrate is constant and equal to T_0. Within these assumptions, the equation for the temperature formally has the appearance of (6), where α is the total heat-exchange coefficient.

At the boundary of the sectors, T satisfies the equations

$$T_1=T_2,\quad (\kappa\mathbf{n}\cdot\operatorname{grad}T)_1=(\kappa\mathbf{n}\cdot\operatorname{grad}T)_2. \tag{7}$$

For a given temperature dependence of $\sigma(T)$, Eqs. (1)–(7) determine the distribution of the electric field and temperature in an inhomogeneous medium.

The nonlinear field terms in the expression for the current arise because of the temperature dependence of $\sigma(T)$, which according to (6) is determined by the electric field. We limit ourselves to the case of cubic nonlinearity. In the adopted approximation we can put

$$\sigma(T)=\sigma(T_0)+\beta\,\delta T, \tag{8}$$

where $\beta=\partial\sigma(T_0)/\partial T$, and $\delta T=T-T_0$. The solution of Eq. (6) can be written

$$\delta T(\mathbf{r})=\int G(\mathbf{r},\mathbf{r}')\mathbf{e}^2(\mathbf{r}')\sigma(\mathbf{r}')\kappa^{-1}(\mathbf{r}')\,d\mathbf{r}', \tag{9}$$

where $G(\mathbf{r},\mathbf{r}')$ is the Green's function of the heat equation (6). Using (8) and (9), we can write the left-hand side of Eq. (4) in the form

$$\langle(\sigma(T_0)+\beta\,\delta T)\mathbf{e}^2\rangle=\sigma_e\mathbf{E}^2+\frac{1}{V}\int\beta(\mathbf{r})\mathbf{e}^2(\mathbf{r})G(\mathbf{r},\mathbf{r}')\times\mathbf{e}^2(\mathbf{r}')\sigma(\mathbf{r}')\kappa^{-1}(\mathbf{r}')\,d\mathbf{r}\,d\mathbf{r}', \tag{10}$$

where V is the volume (or surface area) occupied by the system. Comparing Eqs. (10) and (4) and using Tellegen's theorem, we arrive at a generalized relationship for the effective nonlinear conductivity in the form

$$\chi_e=\frac{1}{\mathbf{E}^4V}\int\beta(\mathbf{r})e^2(\mathbf{r})G(\mathbf{r},\mathbf{r}')e^2(\mathbf{r}')\times\sigma(\mathbf{r}')\kappa^{-1}(\mathbf{r}')\,d\mathbf{r}\,d\mathbf{r}', \tag{11}$$

where, as in (5), the right-hand side contains the electric field in a linear medium. Equation (9) implies that if we allow for thermal conduction, the relationship between temperature and electric field becomes nonlocal. Obviously, the nonlocality reduces the contribution of singular regions to the effec-

tive nonlinear conductivity. However, the effect of thermal conduction can be suppressed by intense heat exchange, when the term $-\text{grad}\,\kappa\cdot\text{grad}\,T$ in Eq. (6) is small compared to the term that allows for heat exchange. In this case $G(\mathbf{r},\mathbf{r}')$ behaves as $\delta(\mathbf{r}-\mathbf{r}')$ and Eq. (11) becomes Eq. (5). The limiting cases are discussed in greater detail in Sec. 3.

3. THERMAL STABILIZATION OF ANOMALIES

Equations (2) and (3) show that microconstrictions in a inhomogeneous medium cause focusing of the electric-field and current lines.[8–11] A consequence of this is an increase in heat release near singular regions. We can easily show that the asymptotic behavior of the temperature distribution cannot strongly depend on the discontinuities in the thermal conductivity and heat-exchange coefficients. This is demonstrated by solving Eq. (6) exactly in Appendix B. Since we are interested in the nature of the divergences and in the mechanism of stabilization of anomalies, we assume that the thermal conductivity coefficient κ, and the heat-exchange coefficient α in each region are the same, and calculate the effective nonlinear conductivity χ_e in the first order in the nonlinearity.

Equation (6) shows that when thermal conduction and heat transfer are taken into account, the problem acquires a new spatial scale, $L=\sqrt{\kappa/\alpha}$, which is assumed to be much smaller than the unit cell size L_0. Here we are interested in the nature of cutoff of the anomalies in the effective characteristics. We demonstrate that if the system has a small parameter $L\ll L_0$, then in the critical region we can find the main contribution to the effective nonlinear conductivity, determined only by small singular regions near the microinhomogeneities of the medium.

3.1. Nonlinear conductivity of a film

The electric field and current in the linear case for a checkerboard lattice has been calculated in Refs. 15 and 16. Here it is convenient to write the starting system of equations (1)–(3) in a complex-valued representation. To simply matters, we limit ourselves to the vase where the external field **E** is directed along a diagonal of the squares. If this is the case, the square of the electric field in adjacent cells is determined by the relationships

$$|e_1(z)|^2=\frac{AE^2}{\sigma_1}K^{4\gamma}|X(z)|^2,$$
$$|e_2(z)|^2=\frac{AE^2}{\sigma_2}K^{4\gamma}|X(iz^*)|^2. \tag{12}$$

Here

$$X(z)=\left[\frac{\text{cn}(Kz/L_0,k)}{\text{sn}(Kz/L_0,k)\ \text{dn}(Kz/L_0,k)}\right]^{2\gamma},\quad z=x+iy, \tag{13}$$

is the complete elliptic integral with modulus k (for a square, $k=1/\sqrt{2}=1.8541$), where $\text{sn}(\cdot)$, $\text{cn}(\cdot)$, and $\text{dn}(\cdot)$ are Jacobi elliptic functions, and L_0 is the length of a side of the square cell. The parameter γ is linked to h by the relationship

$$\tan\pi\gamma=\frac{1-h}{2\sqrt{h}},\quad 0\leqslant\gamma\leqslant\frac{1}{2}. \tag{14}$$

The constant A is defined as follows:

$$A=\frac{(1+h)\sigma_1}{2I^2(\gamma)K^{4\gamma}}, \tag{15}$$
$$I(\gamma)=\frac{\pi^{3/2}}{2K\cos\pi\gamma}\left[\Gamma\left(\frac{3}{4}+\frac{\gamma}{2}\right)\Gamma\left(\frac{3}{4}-\frac{\gamma}{2}\right)\right]^{-1},$$

where $\Gamma(\cdot)$ is the gamma function. Equations (12) show that the electric field has singularities at the corners of the lattice, where the field is concentrated. As noted earlier, we are interested in the contribution of singular regions, and we assume that the thermal length L is small compared to the lattice constant L_0. In this case the expression for the field simplifies. Expanding (12) in power series in the small parameters r/L_0, we obtain

$$e_i^2(r,\vartheta)=\frac{AE^2}{\sigma_i}\left(\frac{L_0}{r}\right)^{4\gamma}\left(1-\frac{2}{5}\gamma\left(\frac{Kr}{L_0}\right)^4\cos 4\vartheta\right), \tag{16}$$

where r and ϑ are polar coordinates, and $i=1,\ 2$. We can easily show that the quadratic correlator of the field, which determines the effective linear conductivity, converges as $r\to 0$ despite the fact that the field diverges. However, the local expression for the effective nonlinear conductivity [Eq. (5)] diverges for certain values of the system parameters. This property is an indication of an anomalous increase in the nonlinear conductivity. Note that the nature of the divergence is determined by the parameter γ, which according to (14) depends on the ratio of linear conductivities, $h=\sigma_2/\sigma_1$. If we neglect nonlocal effects and calculate the nonlinear conductivity by the local expression (5) with the expansion (16), χ_e will diverge for $\gamma\geqslant 1/4$, or $h\leqslant h_c=(\sqrt{2}-1)^2$ (see Refs. 8–11).

How does heat transfer affect this conductivity? If we substitute (16) in the right-hand side of Eq. (6), we arrive at an equation for the temperature in the first approximation:

$$\left(-\Delta+\frac{1}{L^2}\right)\delta T=\frac{AE^2}{\alpha L^2}\left(\frac{L_0}{r}\right)^{4\gamma}, \tag{17}$$

where $L=\sqrt{\kappa/\alpha}$ is the thermal length. It is assumed that in the event of heat exchange $L\ll L_0$ holds. The Green's function of Eq. (17) is well known:

$$G(\mathbf{r}-\mathbf{r}')=\frac{1}{2\pi}K_0\left(\frac{|\mathbf{r}-\mathbf{r}'|}{L}\right), \tag{18}$$

where $K_0(x)$ is the zeroth-order modified Hankel function. In integrating in (9) with respect to the angular variable it is convenient to use the expansion

$$K_0\left(\frac{|\mathbf{r}-\mathbf{r}'|}{L}\right)=I_0\left(\frac{r}{L}\right)K_0\left(\frac{r'}{L}\right)$$
$$+2\sum_{n=1}^{\infty}\cos(n\vartheta)I_n\left(\frac{r}{L}\right)K_n\left(\frac{r'}{L}\right) \tag{19}$$

for $r<r'$, where $I_n(x)$ and $K_n(x)$ are modified Bessel and modified Hankel functions (for $r>r'$ the respective expansion can be obtained from (19) by interchanging r and r' on the right-hand side). Integrating with respect to the angular variable, we arrive at an expression for the temperature:

$$\delta T(r)=\frac{AE^2}{\alpha}\frac{L_0^{4\gamma}}{L^2}\left[K_0\left(\frac{r}{L}\right)\int_0^r I_0\left(\frac{t}{L}\right)t^{1-4\gamma}\,dt + I_0\left(\frac{r}{L}\right)\int_r^{L_0}K_0\left(\frac{t}{L}\right)t^{1-4\gamma}\,dt\right]. \tag{20}$$

Consider the asymptotic behavior of this solution. To find the expression for δT at small distances, where $r\ll L$, we write the second integral on the right-hand side of Eq. (20) in the form

$$\int_r^{L_0}K_0\left(\frac{t}{L}\right)t^{1-4\gamma}\,dt=\int_0^{L_0}K_0\left(\frac{t}{L}\right)t^{1-4\gamma}\,dt - \int_0^{r}K_0\left(\frac{t}{L}\right)t^{1-4\gamma}\,dt. \tag{21}$$

Then in the first integral we can let L_0 go to infinity (bearing in mind the properties of Bessel functions and the fact that $L_0/L\gg 1$), while in the remaining integrals in (20) we use the asymptotic expansion of Bessel functions for small values of the argument,

$$I_0(x)\sim 1,\quad K_0(x)\sim -\ln x.$$

As a result

$$\delta T(r)=\frac{AE^2}{\alpha}\left(\frac{L_0}{L}\right)^{4\gamma}\left[\frac{\Gamma^2(1-2\gamma)}{2^{4\gamma}} - \frac{1}{4(1-2\gamma)^2}\left(\frac{r}{L}\right)^{2(1-2\gamma)}\right]. \tag{22}$$

We see that $\delta T(r)$ diverges only in the limit $h\to 0$, i.e., $\gamma\to 1/2$.

At distances $r\gg L$ the temperature distribution is due primarily to heat exchange. In this region the temperature is proportional to the release of Joule heat: $T\propto(\sigma/\alpha)e^2$. This result can be obtained from (20) by using the asymptotic expansion of Bessel functions for $r\gg L$:

$$K_n\left(\frac{r}{L}\right)\sim\sqrt{\frac{\pi L}{2r}}\exp\left(-\frac{r}{L}\right),\qquad I_n\left(\frac{r}{L}\right)\sim\sqrt{\frac{L}{2\pi r}}\exp\frac{r}{L}. \tag{23}$$

Equation (20) reduces to

$$\delta T(r)\approx\frac{AE^2}{\alpha}\left(\frac{L_0}{r}\right)^{4\gamma}. \tag{24}$$

We see that the temperature rapidly tends to T_0, which justifies the use of the asymptotic expansion for the field when calculating the correlator at distances $L_0\gg r\gg L$.

We now examine the behavior of the effective nonlinear conductivity. Using the solution (20), we can write (11) as follows:

$$\chi_e=\frac{\pi A^2 L_0^{8\gamma-2}}{\alpha L^2}\left(\frac{\beta_1}{\sigma_1}+\frac{\beta_2}{\sigma_2}\right)\times\left[\int_0^{L_0}dr\,r^{1-4\gamma}\left(K_0\left(\frac{r}{L}\right)\int_0^r I_0\left(\frac{t}{L}\right)t^{1-4\gamma}\,dt + I_0\left(\frac{r}{L}\right)\int_r^{L_0}K_0\left(\frac{t}{L}\right)t^{1-4\gamma}\,dt\right)\right]. \tag{25}$$

Next we represent the effective nonlinear conductivity by a sum of two terms, $\chi_e^<$ and $\chi_e^>$, reflecting the contributions of the regions $r<L$ and $r>L$, respectively, to χ_e. We estimate the contribution of the region $r<L$. The expression in parentheses in (25) is calculated for small r in the same way we calculated (20), and integration with respect to r in (25) in the interval $0\leqslant r\leqslant L$ yields

$$\chi_e^<\sim\frac{A^2}{16\alpha(1-2\gamma)}\left(\frac{\beta_1}{\sigma_1}+\frac{\beta_2}{\sigma_2}\right)\left(\frac{L}{L_0}\right)^{2(1-4\gamma)}\times\left[2^{3-4\gamma}\Gamma^2(1-2\gamma)-\frac{1}{(1-2\gamma)^2}\right]. \tag{26}$$

To estimate the contribution of the region $r>L$, we substitute the asymptotic expansion of Bessel functions for large values of the argument [Eqs. (23)]. Integrating over r in (25) in the limit $L\leqslant r\leqslant L_0$, the result is

$$\chi_e^>\sim\frac{A^2}{\alpha}\left(\frac{\beta_1}{\sigma_1}+\frac{\beta_2}{\sigma_2}\right)\frac{1}{2(1-4\gamma)}\left[1-\left(\frac{L}{L_0}\right)^{2(1-4\gamma)}\right]. \tag{27}$$

Note that (27) can be derived directly from (11), since in integrating in (11) over the region $r\gg L$ the argument $|\mathbf{r}-\mathbf{r}'|$ of the Green's function (18) always remains larger than L and hence the function $K_0(|\mathbf{r}-\mathbf{r}'|/L)$ behaves as $2\pi L^2\delta(\mathbf{r}-\mathbf{r}')$. If in (11) we replace the Green's function with $-L^2\delta(\mathbf{r}-\mathbf{r}')$ and integrate over $r>L$, we again arrive at (27).

How does the effective nonlinear conductivity depend on heat exchange, i.e., on the parameter L? To answer this question, we begin with the case where the ratio of the cell linear conductivities, h, is greater than h_c, or when $\gamma<1/4$. If the thermal length tends to zero, the contribution of (26) to the linear conductivity vanishes and (27) becomes

$$\chi_e^>\sim\frac{A^2}{\alpha}\left(\frac{\beta_1}{\sigma_1}+\frac{\beta_2}{\sigma_2}\right)\frac{1}{2(1-4\gamma)}. \tag{28}$$

This coincides with the result obtained in Refs. 9 and 10. Formula (28) implies that in this case we have a power-law divergence of the effective nonlinear conductivity as $h\to h_c$ $(\gamma\to 1/4)$.

We now assume that the thermal length is fixed and $h\to h_c$. Clearly, as $h\to h_c$, the function (27) behave as

$$\chi_e^>\sim\frac{A^2}{\alpha}\left(\frac{\beta_1}{\sigma_1}+\frac{\beta_2}{\sigma_2}\right)\ln\frac{L_0}{L}. \tag{29}$$

For $h<h_c$

$$\chi_e^{>}\sim\frac{A^2}{\alpha}\left(\frac{\beta_1}{\sigma_1}+\frac{\beta_2}{\sigma_2}\right)\frac{1}{2(4\gamma-1)}\left(\frac{L_0}{L}\right)^{2(4\gamma-1)} \tag{30}$$

is anomalously large, since $L_0\gg L$.

If $L\neq 0$ holds, Eq. (26) implies that the contribution of the region $r<L$ to the effective nonlinear conductivity diverges as a power function only as $h\to 0$.

Thus, in calculating the nonlinear conductivity we have used the terms of greatest singularity in the expansion (16). Let us estimate the contribution of the other (discarded) terms. We denote the correction to χ_e due to the second term in (16) by $\delta\chi_e$, with $\delta\chi_e=\delta\chi_e^{<}+\delta\chi_e^{>}$. For $r\ll L_0$ the correction to the square of the electric field is small in parameter r/L_0. When we calculate $\delta\chi_e^{<}$, integration is over the region $r<L$, and in this case $\delta\chi_e^{<}$ is small in the parameter L/L_0. The expression for $\delta\chi_e^{>}$ can be found in the same way as we found (27):

$$\delta\chi_e^{>}\sim\frac{9\pi\gamma^2K^8A^2}{25\alpha}\left(\frac{\beta_1}{\sigma_1}+\frac{\beta_2}{\sigma_2}\right)\frac{1-(L/L_0)^{2(5-4\gamma)}}{5-4\gamma}$$
$$\sim\frac{9\pi\gamma^2K^8A^2}{25\alpha(5-4\gamma)}\left(\frac{\beta_1}{\sigma_1}+\frac{\beta_2}{\sigma_2}\right). \tag{31}$$

Equations (27) and (31) imply that the behavior of the effective nonlinear conductivity in the critical region is determined by γ. Here we must distinguish two cases: $\gamma<1/4$ and $\gamma>1/4$. When $\gamma<1/4$ $(h>h_c)$, we have the inequality $\chi_e^{<}\ll\chi_e^{>}$, and $\chi_e\sim A^2\beta_1/\sigma_1\alpha$. The correction $\delta\chi_e$ turns out to be of order χ_e. This means that when Eq. (11) is used to calculate χ_e, in the general case we must integrate over the entire surface of the unit cell and use the exact expression (12) for the electric field. In the critical region, however, where $\gamma<1/4$ but $\gamma\to 1/4$ or $\gamma>1/4$, we have, respectively,

$$\chi_e\sim\frac{A^2\beta_2}{\sigma_2\alpha}\ln\frac{L_0}{L},\quad \chi_e\sim\frac{A^2\beta_2}{\sigma_2\alpha}\left(\frac{L_0}{L}\right)^{2(4\gamma-1}},$$

where to simplify matters we assume that the second component provides the main contribution to the nonlinear conductivity. Here the domain of integration $r\sim L$ provides the main contribution to $\chi_e^{>}$. Thus, in the critical region, the corrections to the nonlinear conductivity that result if we allow for the next terms in the expansion of the squares of the field prove to be small.

3.2. Nonlinear conductivity of a three-dimensional medium

For the three-dimensional case we limit ourselves to the analysis of a medium with conical singular regions (see Fig. 1b). The electric field near the point of contact of the vertices of two cones is calculated in Appendix A. The square of the electric field strength, which determines dissipation inside and outside the cones, is given by the formula

$$\mathbf{e}^2(r,\vartheta)=\frac{AE^2}{\sigma(\vartheta)}\left(\frac{L_0}{r}\right)^{4\gamma}s(\vartheta),$$
$$s(\vartheta)=(f'(\vartheta))^2+\lambda^2f^2(\vartheta), \tag{32}$$

where $\sigma(\vartheta)$ is equal to σ_1 in the region $\vartheta<\vartheta_0$ and to σ_2 in the region $\vartheta_0<\vartheta<\pi/2$, E is the amplitude of the external electric field, A is a constant depending on the parameters of the structure, and the function $f(\vartheta)$ and its derivative $f'(\vartheta)$ describe the angular dependence of the solution (see Appendix A), and $\lambda=1-2\gamma$.

We begin by calculating the contribution to the correlator $\langle\chi e^4\rangle$ due to the singularities of the current in the conical region without allowing for thermal conduction. Plugging the above solution for the field into Eq. (5), we obtain

$$\langle\chi\mathbf{e}^4\rangle=\frac{\text{const}}{L_0^3}\int_0^{L_0}drr^2r^{4(\lambda-1)}\int_0^{\pi}d\vartheta\,\sin\,\vartheta\,\chi(\vartheta)s^2(\vartheta). \tag{33}$$

We see that the integral diverges at $\lambda_c=1/4$, or $\gamma_c=3/8$. Knowing λ_c, we can use (45) to find h_c as a function of ϑ_0. To make the picture complete, we note that a numerical solution of Eq. (45) yields $h_c=0.094$ and $\vartheta_{0c}=55.50$ for the maximum values of the critical parameters.

Now we allow for thermal conduction. The Green's function of Eq. (6) is

$$G(|\mathbf{r}-\mathbf{r}'|)=\frac{\exp(-|\mathbf{r}-\mathbf{r}'|/L)}{|\mathbf{r}-\mathbf{r}'|}.$$

In calculating T by formula (9) we use an expansion for the Green's function:

$$\frac{\exp(-|\mathbf{r}-\mathbf{r}'|/L)}{|\mathbf{r}-\mathbf{r}'|}=\frac{1}{\sqrt{rr'}}\sum_{n=0}^{\infty}(2n+1)$$
$$\times P_n(\cos\,\vartheta)I_{n+1/2}(r/L)K_{n+1/2}(r'/L) \tag{34}$$

for $r<r'$ [when $r>r'$, we must interchange r and r' in (34)]. Plugging (34) and (32) into (9) and integrating with respect to the angular variable yields

$$\delta T(r,\vartheta)=\frac{AE^2}{\alpha L^2}\frac{L_0^{4\gamma}}{\sqrt{r}}\sum_{n=0}^{\infty}P_n(\cos\,\vartheta)Q_n$$
$$\times\left[K_{n+1/2}\left(\frac{r}{L}\right)\int_0^{r}I_{n+1/2}\left(\frac{t}{L}\right)t^{3/2-4\gamma}dt\right.$$
$$\left.+I_{n+1/2}\left(\frac{r}{L}\right)\int_r^{L_0}K_{n+1/2}\left(\frac{t}{L}\right)t^{3/2-4\gamma}dt\right], \tag{35}$$

where $I_{n+1/2}(x)$ and $K_{n+1/2}(x)$ are modified Bessel and modified Hankel functions, and

$$Q_n=\frac{2n+1}{2}\int_0^{\pi}d\vartheta\,\sin\,\vartheta\,P_n(\cos\,\vartheta)s(\vartheta).$$

What is the asymptotic behavior of this solution? At small distance, $r\ll L$, δT can be calculated in the same way we calculated (22):

$$\delta T(r,\vartheta)=\frac{AE^2}{\alpha}\left(\frac{L_0}{L}\right)^{4\gamma}\left(\frac{r}{L}\right)^{2-4\gamma}F(r,\vartheta), \tag{36}$$

where

$$F(r,\vartheta)=\sum_{n=0}^{\infty} P_n(\cos\vartheta)Q_n[n(n+1)-(3-4\gamma)(2-4\gamma)]^{-1}$$
$$+\frac{\Gamma((3-4\gamma)/2)\Gamma(1-2\gamma)}{\Gamma(3/2)}2^{2(1-2\gamma)}Q_0$$
$$+\cos\ \vartheta\, Q_1\frac{\Gamma(2-2\gamma)\Gamma((1-4\gamma)/2)}{\Gamma(5/2)}$$
$$\times 2^{2(1-2\gamma)}\left(\frac{r}{2L}\right)+\cdots.$$

At large distances $r\gg L$ we use the asymptotic expansion of Bessel functions in (35) and obtain

$$\delta T(r,\vartheta)\sim\frac{AE_0^2}{\alpha}s(\vartheta)\left(\frac{L_0}{r}\right)^{4\gamma}. \tag{37}$$

This expression (37) corresponds to the local limit, since for $r\gg L$ the Green's function behaves approximately as $4\pi L^2\delta(\mathbf{r}-\mathbf{r}')$, which can easily be shown to be true if we use (34) together with (23). As in the two-dimensional case, in this region the relationship between the nonlinear current and the electric field becomes local (because of heat transfer).

Let us estimate the value of the effective nonlinear conductivity. As in the case of a film, we write χ_e as a sum of contributions due to the regions $r<L$ and $r>L$. The main contribution to $\chi_e^<$ due to the region $r<L$ can be calculated by inserting the asymptotic expressions for the temperature [Eq. (36)] and the electric field [Eq. (32)] in the left-hand side of Eq. (10) and integrating over the volume within the specified region. In the three-dimensional case, the square of the electric field and the temperature depend in a complicated manner on the angular variable, which makes it impossible to calculate the integral with respect to ϑ in the expression for χ_e exactly. However, the anomalous increase in the effective conductivity is determined by the dependence of the electric field and the temperature on r. Hence to calculate χ_e approximately we only need to evaluate the integral with respect to r, since the angular integral yields a nonsingular factor of order unity. The result is

$$\chi_e^<\sim\frac{\beta_2A^2Q_0}{\sigma_2\alpha(5-8\gamma)}\left(\frac{L}{L_0}\right)^{3-8\gamma}. \tag{38}$$

Using Eq. (45) from Appendix A, we can easily show that the denominator in (38) cannot vanish when the cone parameters meet the following conditions: $0<h<1$ and $0<\vartheta_0<\pi/2$. The contribution of the region $r>L$ can be found by substituting (37) and (32) into (10) and integrating within $L<r<L_0$:

$$\chi_e^>\sim\frac{\beta_2A^2Q_0}{\sigma_2\alpha(3-8\gamma)}\left[1-\left(\frac{L}{L_0}\right)^{3-8\gamma}\right]. \tag{39}$$

We see that when L is zero, χ_e diverges at a finite value of the parameter h_c, which is specified by the condition $\gamma_c=3/8$. If the thermal length is fixed, there is logarithmic cutoff of the divergence in (39). Thus, as in the case of a film, heat transfer cuts off the divergence.

4. DISCUSSION

As shown in Refs. 8–11, when there is a local relationship between current and field, the effective nonlinear conductivity of two-component periodic lattices is anomalously high near the percolation threshold. At the threshold the conductivity diverges at a finite value of the linear-conductivity ratio $h=\sigma_2/\sigma_1$.

In the present paper we have studied the thermal mechanism of divergence cutoff and have shown that allowing for heat transfer from the hot regions may remove the singularities in the effective nonlinear characteristics at finite values of the parameter h. Note that the geometric cutoff effect were discussed in Ref. 17.

Our results show that the values of h_c determined in Refs. 8–11 for two-dimensional media and the similar quantity h_c for a three-dimensional medium (discussed in the present paper) largely determine the behavior of the effective conductivity χ_e if we also allow for heat transfer. Allowance for thermal conduction and heat transfer gives rise to a new characteristic scale, the thermal length L, which acts as the cutoff parameters for the singularities. At the same time, in the nonlinear regime, the parameter L determines the spatial size of the critical region that is the source of anomalous increase in the effective nonlinear conductivity χ_e. At a fixed value of L, as $h\to h_c$, the effective nonlinear conductivity χ_e exceeds the characteristic nonlinear conductivity of the components, χ_0:

$$\chi_e\sim\chi_0\ \ln\ (L_0/L),$$

where L_0 is the lattice constant. In the critical region $h<h_c$ the parameter of increase of the effective nonlinear conductivity becomes large, $L_0/L\gg1$. For instance, in the two-dimensional case with $\gamma<1/4$, according to (26) and (27), the effective nonlinear conductivity is finite, but its value exceeds the nonlinear conductivity of the components:

$$\chi_e\sim\chi_0(L_0/L)^{2(4\gamma-1)}.$$

A similar property appears in the three-dimensional case for $\gamma<3/8$ [see Eqs. (38) and (39)]; in the case of conical singularities,

$$\chi_e\sim\chi_0(L_0/L)^{8\gamma-3}.$$

As mentioned earlier, the physical mechanism of the effective increase of the nonlinearity is related to the focusing of field and current lines by microconstrictions (''microbridges'') in the nonlinear medium. Thus, the effective nonlinear conductivity in the critical region (near the metal–insulator transition) is not only determined by the values of the nonlinear conductivities of the components of the medium but also strongly depends on the distribution of fields in the medium. This property must be accounted for in designing artificial nonlinear media.

The work was supported by the Russian Fund for Fundamental Research (Grant No. 97-02-16923a) and the Universities of Russia Basic Research Program (Project No. 2480).

APPENDIX A: ELECTRIC FIELD NEAR CONICAL SINGULARITIES

Let us find the electric field near the point of contact of the vertices of two cones. We seek the solution of Eqs. (2) in the form

$$\mathbf{j}=-\sigma\ \mathrm{grad}\ \varphi, \tag{40}$$

where it is convenient to write the potential as

$$\varphi=EL_0\left(\frac{L_0}{r}\right)^{4\gamma} f(\vartheta),$$

and the function $f(\vartheta)$ can be found by solving the equation

$$\frac{1}{\sin\ \vartheta}\frac{\partial}{\partial\vartheta}\sin\ \vartheta\frac{\partial}{\partial\vartheta}f+\lambda(\lambda+1)f=0. \tag{41}$$

For $\vartheta<\vartheta_0$ we seek the solution of Eq. (41) in the form

$$f_1=aP_\lambda(\cos\ \vartheta), \tag{42}$$

where $P_\lambda(x)$ is the Legendre function of the first kind, while for $\vartheta_0<\vartheta<\pi/2$ we seek the solution in the form

$$f_2=b\left[\frac{1}{2}\,(1-\cos\ \pi\lambda)\,P_\lambda(\cos\ \vartheta)+\frac{\sin\ \pi\lambda}{\pi}Q_\lambda(\cos\ \vartheta)\right], \tag{43}$$

where $Q_\lambda(x)$ is the Legendre function of the second kind.

The functions f_1 and f_2 on the surface of a cone obey the boundary conditions

$$f(\vartheta_0)_1=f(\vartheta_0)_2,\quad f'(\vartheta_0)_1=hf'(\vartheta_0)_2, \tag{44}$$

where $f'=\partial f/\partial\vartheta$. Note that the solution for $\vartheta>\pi/2$ can be obtained from (42) and (43) via continuation, as an odd function, in the angle (in view of the boundary conditions in the external field). Plugging (42) and (43) into the boundary conditions (44), we can easily see that there is a general solution if

$$\begin{aligned}hP_\lambda(x)&\left[\frac{1}{2}\,(1-\cos\ \pi\lambda)\,P_{\lambda+1}(x)\right.\\&\left.+\frac{\sin\ \pi\lambda}{\pi}Q_{\lambda+1}(x)-xF_\lambda(x)\right]\\&=F_\lambda(x)[P_{\lambda+1}(x)-xP_\lambda(x)],\end{aligned} \tag{45}$$

where

$$x=\cos\vartheta_0,\quad F_\lambda(x)=\frac{1}{2}\,(1-\cos\ \pi\lambda)\,P_\lambda(x)+\frac{\sin\ \pi\lambda}{\pi}Q_\lambda(x).$$

Equation (45) makes it possible to find the dependence of the parameter λ on h and ϑ_0.

APPENDIX B: TEMPERATURE DISTRIBUTION IN A INHOMOGENEOUS LATTICE

Let us give the solution of Eq. (6) with the boundary conditions (7) for the case where the thermal conductivity and heat exchange coefficients take different values in adjacent sectors: κ_1 and α_1, and κ_2 and α_2. We assume that $\alpha_1/\kappa_1=\alpha_2/\kappa_2$ (this is true for metals). As in the main body of the text, we introduce the length $L=\sqrt{\kappa_1/\alpha_1}$. The equation for δT is similar to Eq. (17), but now the parameter α on the right-hand side of Eq. (17) depends on the angular variable ϑ. To solve Eq. (17) with the boundary conditions (7), here it is convenient to represent δT by a sum of two terms: $\delta T=\delta T_a+\delta T_b$, where δT_a is the temperature distribution due to release of Joule heat inside the sectors, and δT_b is the temperature distribution due to the appearance of heat sources at the boundary of the media with different thermal conductivity coefficients. The Green's function of Eq. (17) in homogeneous space is known [see Eq. (18)], so that

$$\delta T_a=\frac{L_0^{4\gamma}}{2\pi}\int K_0\left(\frac{|\mathbf{r}-\mathbf{r}'|}{L}\right)\frac{AE^2}{\alpha(\vartheta')L^2}(r')^{-4\gamma}r'\,dr'\,d\vartheta', \tag{46}$$

$$\delta T_b=-\frac{1}{2\pi}\sum_{j=1}^{4}\int K_0\left(\frac{r}{L},\vartheta;\frac{r'}{L},\vartheta_j\right)\rho(r')r'\,dr', \tag{47}$$

where $\rho(r)$ is the number density of heat sources at the boundaries $\vartheta_1=\pi/4$, $\vartheta_2=3\pi/4$, $\vartheta_3=5\pi/4$, and $\vartheta_4=7\pi/4$. In writing (47) we have allowed for the symmetry of reflection with respect to the x and y axes for the function δT. Using the properties of Bessel functions, we can employ Eq. (47) to obtain the value of the derivative $\partial\delta T_b/\partial\vartheta$ at, say, the boundary $\vartheta=\pi/4$:

$$\frac{\partial\delta T_b(r,\pi/4+\delta)}{\partial\vartheta}=\frac{1}{2}\rho(r)r^2\ \mathrm{sgn}\ \delta, \tag{48}$$

where $\delta\to 0$. As expected, at the boundaries the normal derivative of the function δT_b has a discontinuity, while δT_a, δT_b, and $\partial\delta T_a/\partial\vartheta$ are continuous. The heat-source number density $\rho(r)$ can be found from the boundary conditions (7). If we write these conditions for δT_a and δT_b and use (48), we obtain

$$\rho(r)=2\,\frac{\alpha_1-\alpha_2}{\alpha_1+\alpha_2}\,\frac{1}{r^2}\,\frac{\partial\delta T_a(r,\pi/4)}{\partial\vartheta}. \tag{49}$$

Plugging this into (47), we arrive at an expression for the heat-source number density:

$$\begin{aligned}\rho(r)=-\frac{AE^2L_0^{4\gamma}}{\pi L^2}\,\frac{(\alpha_1-\alpha_2)^2}{\alpha_1\alpha_2(\alpha_1+\alpha_2)}\,\frac{1}{r^2}\int_0^{L_0}\frac{t\,dt}{t^{4\gamma}}\\\times\left[K_0\left(\frac{r+t}{L}\right)-K_0\left(\frac{|r-t|}{L}\right)\right].\end{aligned} \tag{50}$$

Thus, formulas (46), (47), and (50) constitute the exact solution of Eq. (17) with the boundary conditions (7). Equation (46) can be simplified if we use the addition theorem for Bessel functions and integrate with respect to the angular variable in (46):

$$\delta T_a=\frac{1}{2L^2}AE^2L_0^{4\gamma}\left(\frac{1}{\alpha_1}+\frac{1}{\alpha_2}\right)\left[K_0\left(\frac{r}{L}\right)\int_0^r\frac{t\,dt}{t^{4\gamma}}I_0\left(\frac{t}{L}\right)+I_0\left(\frac{r}{L}\right)\int_r^\infty\frac{t\,dt}{t^{4\gamma}}K_0\left(\frac{t}{L}\right)\right]+\frac{1}{\pi L^2}AE^2L_0^{4\gamma}\left(\frac{1}{\alpha_1}-\frac{1}{\alpha_2}\right)$$
$$\times\sum_{n=1}^{\infty}\frac{\sin(2n(\pi/4-\vartheta))+\sin(2n(\pi/4+\vartheta))}{n}$$
$$\times\left[K_{2n}\left(\frac{r}{L}\right)\int_0^r\frac{t\,dt}{t^{4\gamma}}I_{2n}\left(\frac{t}{L}\right)+I_{2n}\left(\frac{r}{L}\right)\int_r^\infty\frac{t\,dt}{t^{4\gamma}}K_{2n}\left(\frac{t}{L}\right)\right]. \tag{51}$$

Consider the asymptotic behavior of the solution. At small distances $r\ll L$, the main contribution to the integrals in (51) [and in similar expansions (47) and (50)] is provided by the region where r is small. Hence, when calculating the integrals with Bessel functions, we can use the approximate relationships

$$I_n(x)\sim(1/n!)x^n,\quad K_n(x)\sim((n-1)!/2)(2/x)^n,$$
$$K_0(x)\sim-\ln x.$$

Integrating and then summing in (51) [and in the corresponding representations (47) and (50)], we get

$$\delta T^{(k)}(r,\vartheta)=\frac{AE_0^2}{2}\left(\frac{1}{\alpha_1}+\frac{1}{\alpha_2}\right)\frac{\Gamma^2(1-2\gamma)}{2^{4\gamma}}\left(\frac{L_0}{L}\right)^{4\gamma}-\left(\frac{r}{L}\right)^{2-4\gamma}\left(\frac{L_0}{L}\right)^{4\gamma}\left(\frac{AE_0^2}{(2-4\gamma)^2\alpha_k}+C_k\cos((2-4\gamma)(\vartheta_k-\vartheta))\right), \tag{52}$$

where $k=1$ corresponds to the region $-\pi/4<\vartheta<\pi/4$, $k=2$ corresponds to the region $\pi/4<\vartheta<3\pi/4$, $\vartheta_1=0$, $\vartheta_2=\pi/2$, and

$$C_1=AE_0^2\,\frac{\alpha_1-\alpha_2}{\alpha_1(\alpha_1+\alpha_2)}\,\frac{1}{(2-4\gamma)^2\sin(\pi\gamma)},$$
$$C_2=-(\alpha_1/\alpha_2)\,C_1.$$

The solution in the other sectors can easily be obtained by employing the symmetry of the problem. At distance $r\gg L$, the term $-\Delta\delta T$ in Eq. (17) is negligible, so that

$$\delta T(r,\vartheta)=\frac{AE^2}{\alpha(\vartheta)}\left(\frac{L_0}{r}\right)^{4\gamma}. \tag{53}$$

Note that, strictly speaking, we cannot use Eq. (53) over distances of order L near the boundaries. Due to thermal conduction, the temperature in this region smoothly changes from $(AE^2L^2/\kappa_1)(L_0/r)^{4\gamma}$ to $(AE^2L^2/\kappa_2)(L_0/r)^{4\gamma}$. However, analysis of the asymptotic expressions (52) and (53) shows that the difference in the values of α and in the values of κ in adjacent sectors does not change the dependence of δT on the coordinate r. Analysis also shows that the asymptotic behavior of δT depends only on the nature of the divergence of the field and the relationship between heat transfer and thermal conduction. These arguments justify the use of the same values of α and κ in the qualitative analysis done in the main body of the text.

*)E-mail: satanin@phys.unn.runnet.ru

[1] *Proceedings of the Fourth International Conference on Electrical Transport and Optical Properties of Inhomogeneous Media*, Physica A **241** (1–2) (1997).
[2] M. A. Dubson, Y. C. Hui, M. B. Weissman, and J. C. Garland, Phys. Rev. B **39**, 6807 (1989).
[3] Y. Yagil and G. Deutscher, Phys. Rev. B **46**, 16 115 (1992).
[4] Y. Gefen, W.-H. Shih, R. B. Laibowitz, and J. M. Viggiano, Phys. Rev. Lett. **57**, 3097 (1986).
[5] R. K. Chakrabarty, K. K. Bardhan, and A. Basu, Phys. Rev. B **44**, 6773 (1991).
[6] D. J. Bergman, Phys. Rev. B **39**, 4598 (1989).
[7] A. M. Satanin, A. A. Snarskiĭ, K. V. Slichenko, and I. V. Bezsudnov, Zh. Tekh. Fiz. **68**(5), 132 (1998) [Tech. Phys. **43**, 602 (1998)].
[8] A. M. Dykhne, Phystech J. **3**, 2 (1996).
[9] A. M. Satanin, S. V. Khor'kov, and V. V. Skuzovatkin, JETP Lett. **64**, 538 (1996).
[10] A. M. Satanin, S. V. Khor'kov, and V. V. Skuzovatkin, Zh. Éksp. Teor. Fiz. **112**, 643 (1997) [JETP **85**, 351 (1997)].
[11] A. M. Satanin and V. V. Skuzovatkin, JETP Lett. **66**, 126 (1997).
[12] P. Penfield Jr., R. Spence, and S. Duinker, *Tellegen's Theorem and Electrical Networks*, MIT Press, Cambridge, Mass. (1970).
[13] A. Aharony, Phys. Rev. Lett. **58**, 2726 (1987).
[14] D. Stroud and P. M. Hui, Phys. Rev. B **37**, 8719 (1988).
[15] Yu. P. Emets, *Electrical Characteristics of Composites with a Regular Structure* [in Russian], Naukova Dumka, Kiev (1986).
[16] Yu. P. Emets, Zh. Éksp. Teor. Fiz. **96**, 701 (1989) [Sov. Phys. JETP **69**, 397 (1989)].
[17] A. M. Satanin, S. V. Khor'kov, and V. V. Skuzovatkin, JETP Lett. **65**, 544 (1997).

Translated by Eugene Yankovsky