Two-photon axion decay in an external electromagnetic field

V. V. Skobelev

Moscow State Industrial University, 109280 Moscow, Russia
(Submitted 19 February 1998)
Zh. Éksp. Teor. Fiz. **115**, 385–391 (February 1999)

The amplitude of two-photon axion decay in a crossed field is calculated, taking into account the polarization states of the photons. The decay probability and amplitude in an external electromagnetic field exceed the decay probability and amplitude in vacuum even for values of the invariant parameter $\chi=\sqrt{e^2(kF^2k)}/m_e^3$ greater than the small quantity m_a/m_e. Astrophysical aspects associated with this circumstance are discussed. © *1999 American Institute of Physics.* [S1063-7761(99)00102-X]

The decay processes of elementary particles with mass near or equal to zero have been of continuing interest for the last several decades. As a rule, this is associated with astrophysical and cosmological aspects such as the missing-mass problem, the ''aging'' of photons, and other problems. For example, the process of photon decay (or splitting) $\gamma\to 2\gamma$ in an external constant and homogeneous electromagnetic field, which is forbidden in vacuum because of considerations of CP invariance, has been intensively discussed in the literature.[1–4] In fields $F\ll F_0=m_e^2/e=4.41\times 10^{13}$ G, the splitting amplitude is proportional to F^3 (Ref. 1), and in a purely magnetic field with induction $F\gg F_0$ goes to a constant.[3],[1)] Without answering the questions associated with the red shift (because of the smallness of the interstellar fields), this effect can cause the observed linear polarization of the hard radiation of pulsars.[2] A feature of the kinematics is that the momenta of the ''decay'' photons and the initial photons are parallel, so that the phase volume can be written as

$$\int \frac{d^3k_1}{2k_{10}} \int \frac{d^3k_2}{2k_{20}} \delta(k-k_1-k_2)=\frac{\pi}{2}\int_0^1 dx, \tag{1}$$

where $k_1=kx$, $k_2=k(1-x)$, and k is the momentum of the decaying photon. When the mass m_i of the initial particle is nonzero, the contributions to the matrix element that describe the deviation from ''parallelness'' and the decay probability will contain $k^2=m_i^2$. For small mass, this allows them to be neglected by comparison with the contribution from the other invariant parameters (for example, the field parameters) that enter into the problem (this circumstance will be used below).

Another decay that is of interest, of a massive neutrino to less massive ones, $\nu_i\to\nu_j\gamma$, is possible in models with mixing.[6–8] The massive relict neutrinos formed at the time of the Big Bang very probably owe their origin to precisely this channel, and this has a direct relationship to the missing-mass problem. The inhibiting factors are associated here with the smallness of m_i, the phase volume $(m_j\to m_i)$, and with the GIM mechanism.[9] As shown in our paper,[10] these limitations are largely removed in superstrong magnetic fields $F\gg F_0$, and the decay probability in a magnetic field exceeds that in vacuum by many orders of magnitude.

That paper discusses the effect of an electromagnetic field on two-photon axion decay, $a\to 2\gamma$, and shows that when the actual values of the invariant field parameter

$$\chi=\frac{\sqrt{e^2(kF^2k)}}{m_e^3}, \tag{2}$$

are much less than unity the decay probability in a field exceeds the decay probability in vacuum. A similar circumstance was pointed out earlier only in beta decay with a small energy yield[11] (of course, we are not thinking of processes that are in general forbidden in the absence of an external field).

An axion is a pseudoscalar Goldstone boson that appears in the spontaneous breakdown of the global symmetry introduced by Peccei and Quinn.[12] Its existence, although not yet confirmed experimentally, would be extremely desirable for a natural explanation of the observed CP invariance of strong interactions.[13,14] The coupling constants of an axion with ordinary particles are inversely proportional to f—the energy scale of the breakdown of global symmetry—with the lower limit of f being extremely large, $f\gtrsim 10^{10}$ GeV (an invisible axion), but the axion mass m_a is a fraction of an electron volt or less (in this connection, an axion can be a candidate for the role of the carrier of the missing mass). It is assumed that the only decay channel is two-photon decay, with an effective interaction Lagrangian of

$$\mathscr{L}_{a\gamma}=-\frac{1}{16\pi}g_\gamma F^{\mu\nu}\tilde{F}_{\mu\nu}a. \tag{3}$$

Here we have put $g_\gamma=-e^2c_\gamma/2\pi f$, where $c_\gamma\sim 1$ is a model-dependent dimensionless parameter, and $\tilde{F}_{\mu\nu}=(1/2)\varepsilon_{\mu\nu\alpha\beta}F^{\alpha\beta}$ is the dual tensor of the radiation field. The corresponding decay probability per unit time equals[15]

$$W_0=\frac{\alpha^2c_\gamma^2m_a^4}{32(2\pi)^3k_0f^2}, \tag{4}$$

while the quantity W_0^{-1} is greater than the lifetime of the universe.

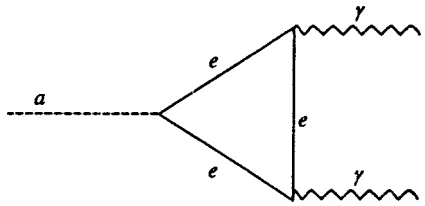

FIG. 1.

The most ''economical'' version of the implementation of the Peccei–Quinn mechanism is expressed in a direct axion–fermion coupling of the form[16]

$$\mathscr{L}_{af}=\frac{c_f}{2f}(\bar{\Psi}\gamma_\mu\gamma^5\Psi)\frac{\partial a}{\partial x_\mu}, \tag{5}$$

while the interaction with the other particles has an effective character.

In the presence of an external magnetic field F, diagrams with an electron loop ($f=e$) contribute both to the axion mass, which is estimated to be considerably less than m_a, and to the field-dependent part of the two-photon decay amplitude (see Fig. 1). The corresponding Lagrangian is the sum of Eq. (5) and the electrodynamic part $e(\bar{\Psi}\gamma_\mu\Psi)A^\mu$.

Further calculations are carried out in the crossed-field configuration, in which both invariants, $F_{\mu\nu}F^{\mu\nu}$ and $F_{\mu\nu}\tilde{F}^{\mu\nu}$, equal zero. Such an approach is fairly general and, moreover, significantly simplifies the calculations. The computations are largely analogous to those in Ref. 4, which discussed the photon-splitting process in a crossed field, also described by an electron tripole.

The Green's function of the Dirac equation for an electron in a crossed field in the proper-time representation has the form

$$G(x_2,x_1)=\exp\left[-ie\int_{x_1}^{x_2}(dx'A(x'))\right]S(x_2-x_1), \tag{6}$$

$$S(x)=-\frac{1}{(4\pi)^2}\int_0^\infty\frac{ds}{s^2}\exp\left[-\frac{ix^2}{4s}-ism_e^2-\frac{ise^2}{12}(xF^2x)\right]\left[m_e+\frac{\hat{x}}{2s}-\frac{se^2}{3}(\gamma F^2x)+\frac{m_ese}{2}(\gamma F\gamma)-\frac{ie}{2}\gamma^5(\gamma\tilde{F}x)\right]. \tag{6a}$$

As a result of certain transformations, the decay matrix element, defined in the usual way,[17] can be written as

$$M=-\frac{2\pi i\alpha c_e}{f}\int d^4x\int d^4x_2\mathrm{Tr}[S(x_2)\hat{e}_2^*S(x_1)\hat{e}_1^*S(x)\hat{k}\gamma^5]\times\exp\left[-\frac{ie}{2}(x_2Fx)-i(k_1x)+i(k_2x_2)\right]+(\gamma_1\leftrightarrow\gamma_2), \tag{7}$$

where e_i are the polarization vectors of the photons, $x+x_1+x_2=0$, $k=k_1+k_2$, and the notation $(\gamma_1\leftrightarrow\gamma_2)$ denotes photon permutation. Using Eq. (6a) and integrating over the spatial coordinates, we get

$$M=-\frac{i\alpha c_e}{2\pi f}\int_0^\infty\frac{ds\,ds_1ds_2}{(s+s_1+s_2)^2}(Q^{\nu_1\nu_2}e^{-i\Phi})e_{1\nu_1}^*e_{2\nu_2}^*+(\gamma_1\leftrightarrow\gamma_2). \tag{8}$$

Here we have introduced the notation

$$\Phi=m_e^2(s+s_1+s_2)-\beta\left(\frac{k^2}{s}+\frac{k_1^2}{s_1}+\frac{k_2^2}{s_2}\right)-2\beta e(k_2Fk_1)+\frac{\beta e^2}{3}[\rho(kF^2k)+\rho_1(k_1F^2k_1)+\rho_2(k_2F^2k_2)], \tag{8a}$$

$$\beta=\frac{ss_1s_2}{s+s_1+s_2},\quad \rho=s_1+s_2-s+\frac{s_1s_2}{s},$$

$$Q^{\nu_1\nu_2}=\frac{1}{4}\mathrm{Tr}[(m_e+\hat{V}_1+T_1+\gamma^5\hat{A}_1)\gamma^{\nu_2}(m_e+\hat{V}+T+\gamma^5\hat{A})\gamma^{\nu_1}(m_e+\hat{V}_2+T_2+\gamma^5\hat{A}_2)\hat{k}\gamma^5], \tag{8b}$$

$$V_\alpha=\frac{X_\alpha}{2s}-\frac{se^2}{3}(XF^2)_\alpha,\quad T=\frac{m_ese}{2}(\gamma F\gamma),$$

$$A_\alpha=\frac{ie}{2}(X\tilde{F})_\alpha,$$

the form of $\rho_{1,2}$ is obtained from ρ by cyclic permutation, and $V_{1,2}$, $T_{1,2}$, and $A_{1,2}$ are obtained from V, T, and A by obvious extensions of the symbols, with the following definition of the differential operators acting on $e^{-i\Phi}$:

$$X_{1\alpha}=-i\frac{\partial}{\partial k_2^\alpha},\quad X_{2\alpha}=i\frac{\partial}{\partial k_1^\alpha},\quad X_\alpha=-X_{1\alpha}-X_{2\alpha}. \tag{8c}$$

We note the following circumstances as important for the subsequent transformations:

(a) It is easy to show that the $(\gamma_1\leftrightarrow\gamma_2)$ term in Eq. (7) reduces to the replacement $F\to-F$ in the first term, so that only even powers of the field remain in the total amplitude.

(b) The field-independent contribution can be related to the effective Lagrangian of the interaction of the axion with the radiation field, Eq. (3), so that we are interested for now in the ''field'' contribution

$$M_F=M-M|_{F=0}. \tag{9}$$

(c) The terms $\sim k_{1,2}^2$ remain in Eq. (8a) for Φ and can be set equal to zero only after carrying out differential operations; the same holds true of the terms $\sim k^2$ and (k_1Fk_2), which can then be dropped as inserting a power of an additional small parameter m_a^2/m_e^2 into M_F.

(d) As in Ref. 4, it should be recalled that the result of the action of the differential operators on $e^{-i\Phi}$ in linear combinations has the form

$$X_\alpha\to\tilde{X}_\alpha=2\beta\left[\frac{k_{1\alpha}}{s_1}-\frac{k_{2\alpha}}{s_2}+e(kF)_\alpha+\frac{e^2}{3}(-\rho_1(k_1F^2)_\alpha+\rho_2(k_2F^2)_\alpha)\right], \tag{10a}$$

and so on, and, in cubic combinations,

$$X_{1\alpha}X_\beta X_{2\sigma}\to\tilde X_{1\alpha}\tilde X_\beta\tilde X_{2\sigma}+2i\beta\left[\tilde X_{1\alpha}\left(\frac{g_{\beta\sigma}}{s_1}-eF_{\beta\sigma}-\frac{e^2}{3}\rho_1F^2_{\beta\sigma}\right)+\tilde X_{2\sigma}\left(\frac{g_{\alpha\beta}}{s_2}-eF_{\alpha\beta}-\frac{e^2}{3}\rho_2F^2_{\alpha\beta}\right)+\tilde X_\beta\left(\frac{g_{\alpha\sigma}}{s}+eF_{\alpha\sigma}-\frac{e^2}{3}\rho F^2_{\alpha\sigma}\right)\right]. \tag{10b}$$

We shall not show the resulting form of the tensor $Q^{\nu_1\nu_2}$ because it is too complicated.

Further simplifications are associated with the introduction of two independent states of linear polarization of the photons of the form

$$e^{(V)}_{1\alpha}=\frac{(k_1F)_\alpha}{\sqrt{(k_1F^2k_1)}},\quad e^{(P)}_{1\alpha}=\frac{(k_1\tilde F)_\alpha}{\sqrt{(k_1\tilde F^2k_1)}} \tag{11}$$

and new dimensionless variables

$$\eta=m_e^2(s+s_1+s_2),\quad u=\frac{s_1+s_2}{s+s_1+s_2},\quad v=\frac{s_1}{s_1+s_2},$$

$$\int_0^\infty\frac{ds\,ds_1ds_2}{(s+s_1+s_2)^2}\to\frac{1}{m_e^2}\int_0^1du\,u\int_0^1dv\int_0^\infty d\eta. \tag{12}$$

In this notation, we get

$$M_F^{PV}=-\frac{ic_e\alpha m_e^2}{6\pi f}\chi^2\int_0^1du\,u\int_0^1dv\int_0^\infty d\eta\,\eta\times\left(\varphi_1+i\eta\varphi_2+\frac{i}{3}\chi^2\eta^3\varphi_4\right)e^{-i\Phi}, \tag{13}$$

where

$$\Phi=1+\frac{1}{3}\eta^3\Omega,$$

$$\Omega=\chi^2u^2\{[v(1-uv)-x(1-u)]^2+4xuv(1-u)(1-v)^2\}, \tag{13a}$$

the variable x and the parameter χ are given by Eqs. (1) and (2) and the functions $\varphi_{1,2,4}(u,v,x)$ are given in the Appendix. The form of M_F^{VP} is obtained from M_F^{PV} by the replacements $v\to1-v$ and $x\to1-x$, while $M_F^{VV}=M_F^{PP}=0$, which is associated with the CP invariance of the theory. Thus, in the approximation used here, an axion in a crossed field decays to photons with orthogonal polarizations. Introducing the complex Hardy–Stokes function

$$f(z)=i\int_0^\infty dt\exp\left[-i\left(zt+\frac{t^3}{3}\right)\right], \tag{14}$$

we rewrite the final general result in the form

$$M_F^{PV}=-\frac{ic_e\alpha m_e^2}{6\pi f}\chi^2\int_0^1du\,u\int_0^1dv\,z^2\left[\varphi_1f'(z)-z\varphi_2f''(z)+\frac{1}{3}z^3\chi^2\varphi_4f''''(z)\right],\quad z=\Omega^{-1/3}. \tag{15}$$

In the limit having practical significance, $\chi\ll1$, the value of z is everywhere large in the region of integration, and, using the relationship $f(z)|_{z\to\infty}\simeq1/z$ and Eqs. (A1) and (A2), it is easy to find that

$$M_F^{PV}\simeq-\frac{ic_e\alpha m_e^2}{240\pi f}\chi^2(1+3x), \tag{16}$$

which corresponds to a double interaction with the field in the electron loop.

This expression is actually valid even in the case of arbitrary constant fields $F=\text{const}$, since the possible difference would consist of an addition to χ^2 of a term of the form $(F^{\mu\nu}F_{\mu\nu}/F_0^2)(m_a^2/m_e^2)$, which is much less than χ^2 in the limit $k_0\gg m_a$. Moreover, as follows from the form of the effective Lagrangian in Eq. (3), the matrix element for ''free'' decay differs in order of magnitude from Eq. (16) by the replacement of $m_e^2\chi^2$ by $10^2m_a^2$. Thus, the field contribution begins to dominate in the limit $\chi\gg\chi_0=10(m_a/m_e)$, and this, for example, for a value of $m_a\sim0.1$ eV, gives a lower limit $\chi_0\sim10^{-6}$.

Assuming that this is satisfied, we get the total decay probability into unpolarized photons in a constant and homogeneous field $F=\text{const}$:

$$W_F=\frac{1}{32\pi k_0}\int_0^1dx[|M_F^{PV}|^2+|M_F^{VP}|^2], \tag{17}$$

i.e.,

$$W_F=\frac{7c_e^2\alpha^2m_e^4\chi^4}{2\times240^2(2\pi)^3k_0f^2}, \tag{18}$$

where we have allowed for the indistinguishability of the photons by dividing by 2. Of course, when the axion mass is small, the effect of the dominance of the field contribution to the decay probability increases for small χ.

In the asymptotic limit $\chi\gg1$, it is impossible in Eq. (15) to simply take the derivatives of the Hardy–Stokes function as zero (as was done, for example, in Ref. 4 in studying the photon-splitting process), since the integral from the third term will diverge as $u,v\to1$ (this singularity occurs when the mass operator of an axion in a crossed field is computed). In this connection, the power of χ in the amplitude can be greater than 2/3. In this version, the field contribution is the determining contribution in any case.

Our analysis thus allows us to conclude that, even in comparatively weak fields, the decay of a ''stiff'' axion is completely controlled by the value of this field, as follows from the condition $\chi\gg m_a/m_e$ and the form of χ given by Eq. (2). However, this has hardly any significance in astrophysical terms, since interstellar fields are rather weak, and the axion lifetime relative to two-photon decay remains as

before much longer than the lifetime of the universe because of the smallness of the coupling constant $1/f$. The presence of strong magnetic fields in collapsed objects such as neutron stars, however, cannot substantially reduce their assumed axion luminance[18] because of small spatial size.

In conclusion, we note that our result does not coincide in the asymptotic limit $\chi \ll 1$ with the result of Ref. 19, in which, using pseudoscalar $\otimes$ pseudoscalar coupling in the Lagrangian $\mathscr{L}_{af}$, it was found that $W_F \sim \chi^8$.

APPENDIX A

The form of the functions $\varphi_i(u,v,x)$ in the expression for the amplitude, Eq. (15), is the following:

$$\varphi_1 = uv[4-3u+3u^2-3u(7+u)v+24u^2v^2] + ux[8-21u+9u^2+3(-4+7u+u^2) \times v+12u(2-3u)v^2], \tag{A1}$$

$$\varphi_2 = 4uv(1-3uv+2u^2v^2)+4ux[5-6u+u^2-3(2-u)(1-uv)v], \tag{A2}$$

$$\varphi_4 = 2u^3\{2v^3(1-2uv)(1-uv)(1-u^2v^2)+xv^2[6(1-u)^2+2u^3(3+u)+3v(-4+2u+8u^2-7u^3-3u^4)+v^2u(24-52u+15u^2+13u^3)+6v^3u^2(2+3u-4u^2)+2v^4u^3(-12+11u)]+x^2v(1-u) \times[2(3-15u+20u^2-4u^3)+3v(-8+34u-38u^2+3u^3+u^4)+4v^2(6-21u+14u^2+5u^3)+v^3u^2(1-u)(36-19v)]+2x^3(1-u)^2[-7+10u-u^2+9v(2-3u)+3v^2(-4+6u+u^2)-3v^3u(1-u)]\}. \tag{A3}$$

[1)]This result was obtained again in a recently published paper[5] without referring to our paper.[3]

[1] Z. Bialynicka-Birula and I. Bialynicki-Birula, Phys. Rev. D **2**, 2341 (1970).
[2] S. L. Adler, J. N. Bahcall, C. G. Callan, and M. N. Rosenbluth, Phys. Rev. Lett. **25**, 1061 (1970).
[3] D. V. Gal'tsov, Yu. M. Loskutov, and V. V. Skobelev, Vestn. Mosk. Gos. Univ. Fiz. Astron. No. 5, 601 (1972).
[4] V. O. Papanyan and V. I. Ritus, Zh. Éksp. Teor. Fiz. **61**, 2231 (1971) [Sov. Phys. JETP **34**, 1195 (1972)].
[5] V. N. Baĭer, A. K. Mil'shteĭn, and R. Zh. Shaĭsultanov, Zh. Éksp. Teor. Fiz. **111**, 52 (1997) [JETP **84**, 29 (1997)].
[6] S. M. Bilenky, S. T. Petcov, and B. Pontecorvo, Phys. Rev. Lett. **67**, 309 (1977).
[7] W. J. Marciano and A. I. Sanda, Phys. Lett. B **67**, 303 (1977).
[8] J. Liu, Phys. Rev. D **44**, 2879 (1991).
[9] S. L. Glashow, J. Iliopoulos, and L. Maiani, Phys. Rev. D **2**, 1285 (1970).
[10] V. V. Skobelev, Zh. Eksp. Teor. Fiz. **108**, 3 (1995) [JETP **81**, 1 (1995)].
[11] A. I. Nikishov and V. I. Ritus, Trudy Fiz. Inst. Akad. Nauk SSSR **168**, 232 (1986).
[12] R. D. Peccei and H. R. Quinn, Phys. Rev. Lett. **38**, 1440 (1977).
[13] R. D. Peccei and H. R. Quinn, Phys. Rev. D **16**, 1791 (1977).
[14] K. Choi, K. Kang, and J. E. Kim, Phys. Rev. Lett. **62**, 849 (1989).
[15] G. Raffelt, Phys. Rev. D **33**, 897 (1986).
[16] G. Raffelt, Phys. Rep. **198**, 1 (1990).
[17] V. B. Berestetskiĭ, E. M. Lifshitz and L. P. Pitayevsky, *Relativistic Quantum Theory*, Pergamon Press, Oxford (1971); Russian original, Nauka, Moscow (1971), 285 pp.
[18] V. V. Skobelev, Zh. Eksp. Teor. Fiz. **112**, 25 (1997) [JETP **85**, 13 (1997)].
[19] L. A. Vasilevskaya, N. V. Mikheev, and A. Ya. Parkhomenko, Yad. Fiz. **60**, 2223 (1997) [Phys. At. Nucl. **60**, 2041 (1997)].

Translated by W. J. Manthey

Violation of the factorization theorem in large angle radiative Bhabha scattering

A. B. Arbuzov,*) E. A. Kuraev, and B. G. Shaikhatdenov

Joint Institute for Nuclear Research, 141980 Dubna, Moscow Region, Russia
(Submitted 22 June 1998)
Zh. Éksp. Teor. Fiz. **115**, 392–403 (February 1999)

The lowest order QED radiative corrections to the radiative large angle Bhabha scattering process in the region where all kinematic invariants are large compared to the electron mass are considered. We show that the leading logarithmic corrections do not factorize before the Born cross section, contrary to the picture assumed in the renormalization group approach. The leading and non leading contributions for typical kinematics of the hard process at the energy of the Φ factory are estimated. © *1999 American Institute of Physics.* [S1063-7761(99)00202-4]

1. INTRODUCTION

The large angle Bhabha scattering process LABS plays an important role in e^+e^- colliding beam physics.[1] First, it is traditionally used for calibration, because it has a large cross section and can be recognized easily. Second, it might provide essential background information in a study of quarkonia physics. The result obtained below can also be used to construct Monte Carlo event generators for Bhabha scattering processes.

In our previous papers we considered the following contributions to the large angle Bhabha cross section: pair production (virtual, soft,[2] and hard[3]) and two hard photons.[4] This paper is devoted to the calculation of radiative corrections to a single hard-photon emission process. We consider the kinematics essentially of type $2\to3$, in which all possible scalar products of 4-momenta of external particles are large compared to the electron mass squared.

Considering virtual corrections, we identify gauge invariant sets of Feynman diagrams. Loop corrections associated with emission and absorption of virtual photons by the same fermionic line are called as Glass-type (G) corrections. The case in which a loop involves exchange of two virtual photons between different fermionic lines is called Box-type (B) Feynman diagrams. The third class includes the vertex function and vacuum polarization contributions ($\Gamma\Pi$-type). We see explicitly that all terms that contain the square of large logarithms $\ln(s/m^2)$, as well as those that contain the infrared singularity parameter (fictitious photon mass λ), cancel out in the total sum, where the emission of an additional soft photon is also considered.

We note here that the part of the general result associated with scattering-type diagrams [see Fig. 1 (*1, 5*)] was used to describe radiative deep inelastic scattering with radiative corrections taken into account in Ref. 5 (we labeled it the Compton tensor with heavy photon). A similar set of Feynman diagrams can be used to describe the annihilation channel.[3]

The problem of virtual radiative corrections calculations at the one-loop level is cumbersome for the process

$$e^+(p_2)+e^-(p_1)\to e^+(p_2')+e^-(p_1')+\gamma(k_1). \tag{1}$$

Specifically, if at the Born level we need to consider eight Feynman diagrams, then at the one-loop level we have as many as 72. Furthermore, performing loop momentum integration, we introduce scalar, vector, and tensor integrals up to the third rank with 2, 3, 4, and 5 denominators (a set of relevant integrals is given in our preprint[6]). A high degree of symmetry of Feynman diagrams for a cross section can be exploited to calculate the matrix element squared. Using it, we can restrict ourselves to the consideration of interferences of the Born-level amplitudes [Fig. 1 (*1–4*)] with those that contain one-loop integrals [Fig. 1 (*5–16*)]. Our calculation is simplified since we omit the electron mass m in evaluating the corresponding traces due to the kinematic region under consideration:

$$s\sim s_1\sim -t_1\sim -t\sim -u\sim -u_1\sim \chi_{1,2}\sim \chi'_{1,2}\gg m^2,$$

$$s=2p_1p_2,\quad t=-2p_2p_2',\quad u=-2p_1p_2',$$

$$s_1=2p_1'p_2',\quad t_1=-2p_1p_1',\quad u_1=-2p_2p_1',$$

$$\chi_{1,2}=2k_1p_{1,2},\quad \chi'_{1,2}=2k_1p'_{1,2},$$

$$s+s_1+t+t_1+u+u_1=0,\quad s+t+u=\chi_1',$$

$$s_1+t+u_1=-\chi_1,\quad t+\chi_1=t_1+\chi_1'. \tag{2}$$

We found that some kind of local factorization took place both for the G- and B-type Feynman diagrams: the leading logarithmic contribution to the matrix element squared, summed over spin states, arising from interference of one of the four Feynman diagrams at the Born level [Fig. 1 (*1–4*)] with some one-loop-corrected Feynman diagrams [Fig. 1 (*5–16*)], turns out to be proportional to the interference of the corresponding amplitudes at the Born level. The latter has the form

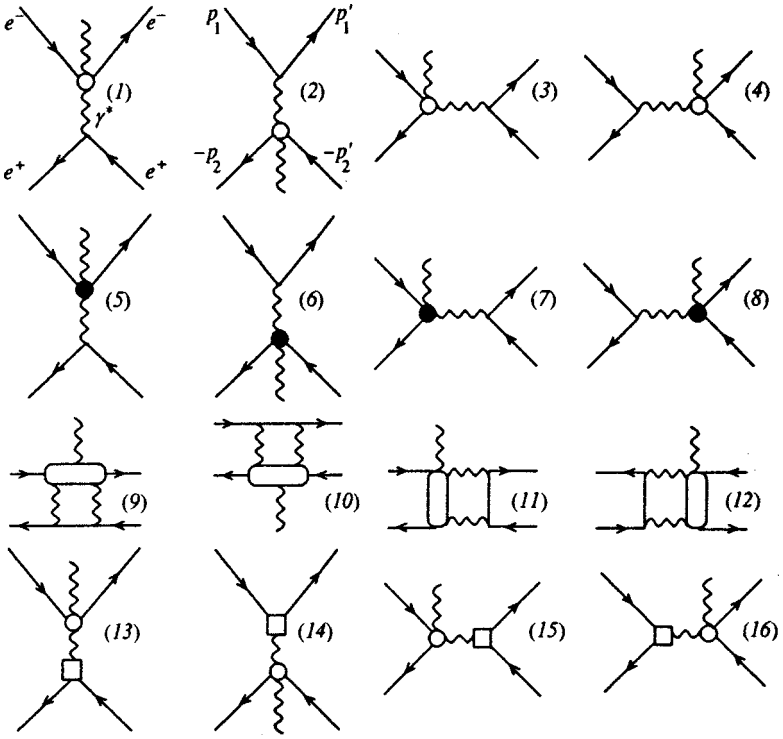

FIG. 1. G- and B-type Feynman diagrams for radiative Bhabha scattering.

$$E_0=(4\pi\alpha)^{-3}\sum |M_1|^2$$

$$=-\frac{16}{t^2}\frac{1}{4}\mathrm{Tr}(\hat{p}_1'\tilde{O}_{11'}\hat{p}_1 O_{11'})\frac{1}{4}\mathrm{Tr}(\hat{p}_2\gamma_\sigma\hat{p}_2'\gamma_\rho)$$

$$=-\frac{16}{t\chi_1\chi_1'}(u^2+u_1^2+s^2+s_1^2),$$

$$O_0=(4\pi\alpha)^{-3}\sum M_1M_2^*$$

$$=\frac{8}{tt_1}\left(\frac{s}{\chi_1\chi_2}+\frac{s_1}{\chi_1'\chi_2'}+\frac{u}{\chi_1\chi_2'}+\frac{u_1}{\chi_2\chi_1'}\right)$$
$$\times(u^2+u_1^2+s^2+s_1^2),$$

$$I_0=(4\pi\alpha)^{-3}\sum M_1(M_3^*+M_4^*)$$

$$=-(1+\hat{Z})\frac{4}{ts_1}\left\{-\frac{4u_1\chi_2'}{\chi_1}+\frac{4u(s_1+t_1)(s+t)}{\chi_2\chi_1'}\right.$$
$$-\frac{2}{\chi_1\chi_2}[2suu_1+(u+u_1)(uu_1+ss_1-tt_1)]$$
$$\left.+\frac{2}{\chi_1\chi_1'}[2t_1uu_1+(u+u_1)(uu_1+tt_1-ss_1)]\right\},$$

$$O_{11'}=\gamma_\rho\frac{\hat{p}_1'+\hat{k}_1}{\chi_1'}\gamma_\mu-\gamma_\mu\frac{\hat{p}_1-\hat{k}_1}{\chi_1}\gamma_\rho,$$

$$\tilde{O}_{11'}=O_{11'}(\rho\leftrightarrow\mu), \tag{3}$$

where the $\hat{Z}$-operator acts as follows:

$$\hat{Z}=\begin{vmatrix} p_1\leftrightarrow p_1' & s\leftrightarrow s_1 \\ p_2\leftrightarrow p_2' & u\leftrightarrow u_1 \\ k_1\to -k_1 & t,t_1\to t,t_1 \end{vmatrix}.$$

It can be shown that the total matrix element squared, summed over spin states, can be obtained using symmetry properties realized by means of the permutation operations:

$$\sum |M|^2=(4\pi\alpha)^3F,$$

$$F=(1+\hat{P}+\hat{Q}+\hat{R})\Phi$$

$$=16\frac{ss_1(s^2+s_1^2)+tt_1(t^2+t_1^2)+uu_1(u^2+u_1^2)}{ss_1tt_1}$$
$$\times\left(\frac{s}{\chi_1\chi_2}+\frac{s_1}{\chi_1'\chi_2'}-\frac{t}{\chi_2\chi_2'}-\frac{t_1}{\chi_1\chi_1'}+\frac{u}{\chi_1\chi_2'}+\frac{u_1}{\chi_2\chi_1'}\right),$$

$$\Phi=E_0+O_0-I_0. \tag{4}$$

The explicit form of the $\hat{P},\hat{Q},\hat{R}$ operators is

$$\hat{P}=\begin{vmatrix} p_1\leftrightarrow -p_2' & s\leftrightarrow s_1 \\ p_2\leftrightarrow -p_1' & t\leftrightarrow t_1 \\ k_1\to k_1 & u,u_1\to u,u_1 \end{vmatrix},$$

$$\hat{Q}=\begin{vmatrix} p_2\leftrightarrow -p_1' & s\leftrightarrow t_1 \\ p_2'\to p_2' & s_1\leftrightarrow t \\ p_1,k_1\to p_1,k_1 & u,u_1\to u,u_1 \end{vmatrix},$$

$$\hat{R}=\begin{vmatrix} p_1\leftrightarrow -p_2' & s\leftrightarrow t \\ p_1'\to p_1' & s_1\leftrightarrow t_1 \\ p_2,k_1\to p_2,k_1 & u,u_1\to u,u_1 \end{vmatrix}. \tag{5}$$

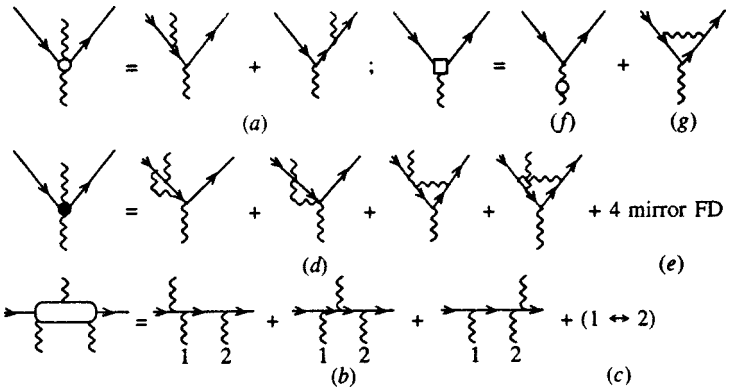

FIG. 2. Content of the notation for Fig. 1.

The differential cross section at the Born level in the case of large angle kinematics (2) was found in Ref. 7:

$$d\sigma_0(p_1,p_2)=\frac{\alpha^3}{32s\pi^2}F\frac{d^3p_1'd^3p_2'd^3k_1}{\varepsilon_1'\varepsilon_2'\omega_1}\times\delta^{(4)}(p_1+p_2-p_1'-p_2'-k_1), \tag{6}$$

where ε_1, ε_2, and ω_1 are the energies of the outgoing fermions and photon, respectively. The collinear kinematics regions (real photon emitted in the direction of one of the charged particles) corresponding to the case in which one of the invariants χ_i,χ_i' is of order m^2 yields the main contribution to the total cross section. These require separate investigation, and will be considered elsewhere.

Our paper is organized as follows. In Sec. 2 we consider the contribution due to the set of Feynman diagrams Fig. 1 (*5–8*) called *glasses* here (G-type diagrams). Using crossing symmetry, we construct the whole G-type contribution from the gauge-invariant set of Feynman diagrams in Fig. 1 (*5*). Moreover, only the set of Feynman diagrams depicted in Fig. 2(d) can be considered in practical calculations, due to an additional mirror symmetry in the diagrams of Fig. 2 (d and e). We therefore start by checking the gauge invariance of the Compton tensor described by the Feynman diagrams of Fig. 2(d and e) for all fermions and one of the photons on the mass shell. In Sec. 3 we consider the contribution of amplitudes containing vertex functions and the virtual photon polarization operator shown in Fig. 1 (*13–16*) and Fig. 2 (f and g). In Sec. 4 we take into account the contribution of Feynman diagrams with virtual two-photon exchange, shown in Fig. 1 (*9–12*), called boxes here (B-type diagrams). Again, using the crossing symmetry of Feynman diagrams, we show how to use only the Feynman diagrams of Fig. 1 (*9*) in calculations. We show that the terms containing infrared singularities, as well as these containing large logarithms, can be written in simple form, related to certain contributions to the radiative Bhabha cross section in the Born approximation (3). We also control terms in the matrix element squared that do not contain large logarithms and are infrared-finite. Thus our considerations permit us to calculate the cross section in the kinematic region (2), in principle, to power-law accuracy, i.e., neglecting terms that are

$$\mathcal{O}\left(\frac{\alpha}{\pi}\frac{m^2}{s}L_s^2\right), \tag{7}$$

as compared to $\mathcal{O}(\alpha/\pi)$ terms calculated in this paper. Note that the terms in (7) are less than 10^{-4} for typical moderately high energy colliders (DAΦNE, VEPP-2M, BEPS). Unfortunately, the non leading terms are too complicated to be presented analytically, so we have only estimated them numerically. In Sec. 5 we consider emission of an additional soft photon in our radiated Bhabha process. To conclude, we note that the expression for the total correction, taking into account virtual and real soft photon emission in the leading logarithmic approximation, has a very elegant and handy form, although it differs from what one might expect in the approach based on renormalization group ideas. Besides analytic expressions, we also give numerical values, along with the non leading terms for a few points under typical experimental conditions.

2. CONTRIBUTION OF G-TYPE DIAGRAMS

We begin by explicitly checking the gauge invariance of the tensor

$$\overline{u}(p_1')R_{1,1'}^{\sigma\mu}u(p_1). \tag{8}$$

This was done indirectly in Ref. 5, where the Compton tensor for a heavy photon was written in terms of explicitly gauge invariant tensor structures. We use the expression

$$R_{1,1'}^{\sigma\mu}=R^{\chi_1}+R^{\chi_1'}, \tag{9}$$

$$R^{\chi_1}=A_2\gamma_\sigma\hat{k}_1\gamma_\mu+\int\frac{d^4k}{i\pi^2}\times\left\{\frac{\gamma_\lambda(\hat{p}_1'-\hat{k})\gamma_\sigma(\hat{p}_1-\hat{k}_1-\hat{k})\gamma_\lambda(\hat{p}_1-\hat{k}_1)\gamma_\mu}{-\chi_1(0)(2)(q)}+\frac{\gamma_\lambda(\hat{p}_1'-\hat{k})\gamma_\sigma(\hat{p}_1-\hat{k}_1-\hat{k})\gamma_\mu(\hat{p}_1-\hat{k})\gamma_\lambda}{(0)(1)(2)(q)}\right\}, \tag{10}$$

where

$$(0)=k^2-\lambda^2,\quad (2)=(p_1'-k)^2-m^2,\quad (1)=(p_1-k)^2-m^2,$$

$$(q)=(p_1-k_1-k)^2-m^2,\quad A_2=\frac{2}{\chi_1}\left(L_{\chi_1}-\frac{1}{2}\right),$$

$$L_{\chi_1}=\ln\frac{\chi_1}{m^2}. \tag{11}$$

The quantity R^{χ_1} corresponds to the Feynman diagrams depicted in Fig. 2d, while $R^{\chi_1'}$ corresponds to those in Fig. 2(e). The first term on the right-hand side of Eq. (10) corresponds to the first two of Fig. 2d under conditions (2). The gauge invariance condition $R^{\sigma\mu}_{1,1'}k_\mu=0$ is clearly satisfied. The gauge invariance condition regarding the heavy photon Lorentz index provides some check of the loop momentum integrals, which can be found in Ref. 6:

$$\overline{u}(p_1')R^{\sigma\mu}_{1,1'}u(p_1)q_\sigma e_\mu(k_1)=Ak_1^\mu e_\mu(k_1),$$

$$A=-2\frac{L_{\chi_1}-2}{\chi_1}-6\frac{L_{\chi_1'}-1}{\chi_1'}. \tag{12}$$

The gauge invariance is thus satisfied due to the Lorentz condition for the on shell photon, $e(k_1)k_1=0$. As stated above, the use of crossing symmetries of amplitudes permits us to consider only R^{χ_1}. For interference of amplitudes at the Born level [see Fig. 1 (*1–4*) and Fig. 1 (*5–8*)], we obtain in terms of the replacement operators

$$(\Delta|M|^2)_G=2^5\alpha^4\pi^2(1+\hat P+\hat Q+\hat R)(1+\hat Z)\times[E^{\chi_1}_{15}+O^{\chi_1}_{25}-I^{\chi_1}_{35}-I^{\chi_1}_{45}], \tag{13}$$

with

$$E^{\chi_1}_{15}=\frac{16}{t^2}\frac14\mathrm{Tr}(\hat p_1'R^{\chi_1}\hat p_1Q_{11'})\frac14\mathrm{Tr}(\hat p_2\gamma_\rho\hat p_2'\gamma_\sigma),$$

$$O^{\chi_1}_{25}=\frac{16}{tt_1}\frac14\mathrm{Tr}(\hat p_1'R^{\chi_1}\hat p_1\gamma_\rho)\frac14\mathrm{Tr}(\hat p_2\gamma_\sigma\hat p_2'O_{22'}),$$

$$I^{\chi_1}_{35}=\frac{4}{ts_1}\frac14\mathrm{Tr}(\hat p_1'R^{\chi_1}\hat p_1O_{12}\hat p_2\gamma_\sigma\hat p_2'\gamma_\rho),$$

$$I^{\chi_1}_{45}=\frac{4}{ts}\frac14\mathrm{Tr}(\hat p_1'R^{\chi_1}\hat p_1\gamma_\rho\hat p_2\gamma_\sigma\hat p_2'O_{1'2'}),$$

$$O_{11'}=\gamma_\rho\frac{\hat p_1'+\hat k_1}{\chi_1'}\gamma_\mu-\gamma_\mu\frac{\hat p_1-\hat k_1}{\chi_1}\gamma_\rho,$$

$$O_{22'}=\gamma_\mu\frac{-\hat p_2'-\hat k_1}{\chi_2'}\gamma_\rho-\gamma_\rho\frac{-\hat p_2+\hat k_1}{\chi_2}\gamma_\mu,$$

$$O_{12}=-\gamma_\mu\frac{\hat p_1-\hat k_1}{\chi_1}\gamma_\rho-\gamma_\rho\frac{-\hat p_2+\hat k_1}{\chi_2}\gamma_\mu,$$

$$O_{1'2'}=\gamma_\rho\frac{\hat p_1'+\hat k_1}{\chi_1'}\gamma_\mu+\gamma_\mu\frac{-\hat p_2'-\hat k_1}{\chi_2'}\gamma_\rho. \tag{14}$$

In the logarithmic approximation, the G-type amplitude contribution to the cross section has the form

$$d\sigma_G=\frac{d\sigma_0}{F}\frac{\alpha}{\pi}(1+\hat P+\hat Q+\hat R)\Phi\times\left[-\frac12L_{t_1}^2+\frac32L_{t_1}+2L_{t_1}\ln\frac{\lambda}{m}\right],$$

$$L_{t_1}=\ln\frac{-t_1}{m^2}. \tag{15}$$

3. VACUUM POLARIZATION AND VERTEX INSERTION CONTRIBUTIONS

Let us examine a set of $\Gamma\Pi$-type Feynman diagrams. The contribution of the Dirac form factor of fermions and vacuum polarization [see Fig. 1 (*13–16*)] can be parametrized as $(1+\Gamma_t)/(1-\Pi_t)$, while the contribution of the Pauli form factor is proportional to the fermion mass, and is omitted here. We obtain

$$d\sigma_{\Gamma\Pi}=\frac{d\sigma_0}{F}\frac{\alpha}{\pi}2(1+\hat P+\hat Q+\hat R)(\Gamma_t+\Pi_t)\Phi, \tag{16}$$

where

$$\Gamma_t=\frac{\alpha}{\pi}\left\{\left(\ln\frac{m}{\lambda}-1\right)(1-L_t)-\frac14L_t-\frac14L_t^2+\frac12\zeta_2\right\},$$

$$\Pi_t=\frac{\alpha}{\pi}\left(\frac13L_t-\frac59\right),\quad L_t=\ln\frac{-t}{m^2}. \tag{17}$$

In realistic calculations, the vacuum polarization due to hadrons and muons can be taken into account in a very simple fashion,[8] just by adding it to Π_t.

4. CONTRIBUTION OF THE B-TYPE SET OF FEYNMAN DIAGRAMS

A procedure resembling the one used in the previous section, applied to the B-type set of Feynman diagrams [Fig. 1 (*9–12a*)], enables us to use only certain one-loop diagrams in practical calculations, specifically three of those in the scattering channel with uncrossed exchanged photon legs:

$$(\Delta|M|^2)_B=2^5\alpha^4\pi^2\,\mathrm{Re}(1+\hat P+\hat Q+\hat R)[(1-\hat P_{22'})I^{\chi_1}_{19}+(1+\hat P_{22'})I^{\chi_1}_{29}-I], \tag{18}$$

where

$$\hat P_{22'}=\begin{vmatrix} p_2\leftrightarrow-p_2' & s\leftrightarrow u\\ p_1\leftrightarrow p_1 & s_1\leftrightarrow u_1\\ p_1',k_1\rightarrow p_1',k_1 & t,t_1\rightarrow t,t_1\end{vmatrix}, \tag{19}$$

and

$$I^{\chi_1}_{19}=\int\frac{d^4k}{i\pi^2}\frac{1}{(0)(q)((p_2+k)^2-m^2)}\frac{16}{t}\frac14\times\mathrm{Tr}(\hat p_1'B^{\chi_1}\hat p_1O_{11'})\frac14\mathrm{Tr}(\hat p_2\gamma_\sigma(-\hat p_2-\hat k)\gamma_\lambda\hat p_2'\gamma_\rho),$$

$$I^{\chi_1}_{29}=\int\frac{d^4k}{i\pi^2}\frac{1}{(0)(q)((p_2+k)^2-m^2)}\frac{16}{t_1}\frac14\times\mathrm{Tr}(\hat p_1'B^{\chi_1}\hat p_1\gamma_\rho)\frac14\mathrm{Tr}(\hat p_2\gamma_\sigma(-\hat p_2-\hat k)\gamma_\lambda\hat p_2'O_{22'}),$$

$$I=\int\frac{d^4k}{i\pi^2}\frac{1}{(0)(q)}\left\{\frac{4}{s_1}\frac14\mathrm{Tr}\left(\hat p_2'\gamma_\rho\hat p_1'B^{\chi_1}\hat p_1O^{12}\hat p_2(\hat A+\hat B)+\frac{4}{s}\frac14\mathrm{Tr}(\hat p_2'O_{1'2'}\hat p_1B^{\chi_1}\hat p_1\gamma_\rho\hat p_2(\hat A+\hat B))\right)\right\},$$

$$\hat{A}=\frac{\gamma_\sigma(-\hat{p}_2-\hat{k})\gamma_\lambda}{(p_2+k)^2-m^2},\quad \hat{B}=\frac{\gamma_\lambda(-\hat{p}_2'+\hat{k})\gamma_\sigma}{(-p_2'+k)^2-m^2}. \tag{20}$$

Here

$$B^{\chi_1}=\frac{\gamma_\lambda(\hat{p}_1-\hat{k}_1-\hat{k})\gamma_\sigma(\hat{p}_1-\hat{k}_1)\gamma_\mu}{-\chi_1(d)}+\frac{\gamma_\lambda(\hat{p}_1-\hat{k}_1-\hat{k})\gamma_\mu(\hat{p}_1-\hat{k})\gamma_\sigma}{(d)(1)}+\frac{\gamma_\mu(\hat{p}_1'+\hat{k}_1)\gamma_\lambda(\hat{p}_1-\hat{k})\gamma_\sigma}{\chi_1'(1)},$$

$$(q)=(p_2-p_2'+k)^2-\lambda^2,\quad (d)=(p_1-k_1-k)^2-m^2,$$

$$(1)=(p_1-k)^2-m^2,\quad (0)=k^2-\lambda^2. \tag{21}$$

Analytic evaluations divulge a lack of both double logarithmic ($\propto L_s^2$) and infrared logarithmic ($\propto\ln(\lambda/m)L$) terms in the box contribution. In spite of the explicit proportionality of the individual contributions to the structures E_0, O_0, and I_0, the overall expression turns out to be somewhat convoluted, despite it has a factorized form in each gauge-invariant subset of diagrams. We parametrize the correction coming from the B-type Feynman diagrams as follows:

$$d\sigma_B=d\sigma_0\frac{\alpha}{\pi}L_s\Delta_B,\quad \Delta_B=2\ln\frac{ss_1}{uu_1}+\frac{2}{F}(\Phi_Q+\Phi_R)\ln\frac{tt_1}{ss_1}. \tag{22}$$

The total virtual correction to the cross section has the form

$$d\sigma^{\rm virt}=d\sigma_G+d\sigma_{\Gamma\Pi}+d\sigma_B=d\sigma_0\frac{\alpha}{\pi}\left[-L_s^2+L_s\left(\frac{11}{3}+4\ln\frac{\lambda}{m}+\Delta_G+\Delta_{\Gamma\Pi}+\Delta_B\right)+\mathcal{O}(1)\right],$$

$$\Delta_G+\Delta_{\Gamma\Pi}=\frac{1}{F}\left(\Phi\ln\frac{s^2}{tt_1}+\Phi_R\ln\frac{t^2}{ss_1}+\Phi_Q\ln\frac{t_1^2}{ss_1}+\Phi_P\ln\frac{s_1^2}{tt_1}\right), \tag{23}$$

where $\Phi_P=\hat{P}\Phi$, $\Phi_Q=\hat{Q}\Phi$, and $\Phi_R=\hat{R}\Phi$.

5. CONTRIBUTION FROM ADDITIONAL SOFT PHOTON EMISSION

Consider now radiative Bhabha scattering accompanied by emission of an additional soft photon in the center of mass reference frame. By soft we mean that its energy does not exceed some small quantity $\Delta\varepsilon$, compared to the energy ε of the initial beams. The corresponding cross section has the form

$$d\sigma^{\rm soft}=d\sigma_0\delta^{\rm soft},$$

$$\delta^{\rm soft}=-\frac{4\pi\alpha}{16\pi^3}\int\frac{d^3k_2}{\omega_2}\left(-\frac{p_1}{p_1k_2}+\frac{p_1'}{p_1'k_2}+\frac{p_2}{p_2k_2}-\frac{p_2'}{p_2'k_2}\right)^2\Bigg|_{\omega_2<\Delta\varepsilon}. \tag{24}$$

The soft photon energy does not exceed $\Delta\varepsilon\ll\varepsilon_1=\varepsilon_2\equiv\varepsilon\sim\varepsilon_1'\sim\varepsilon_2'$. In order to calculate the right-hand side of Eq. (24), we use the master equation:[9]

$$-\frac{4\pi\alpha}{16\pi^3}\int\frac{d^3k}{\omega}\frac{(q_i)^2}{(q_ik)^2}\Bigg|_{\omega<\Delta\varepsilon}=-\frac{\alpha}{\pi}\ln\left(\frac{m\Delta\varepsilon}{\lambda\varepsilon_i}\right),\quad \omega=\sqrt{k^2+\lambda^2}, \tag{25}$$

$$\frac{4\pi\alpha}{16\pi^3}\int\frac{d^3k}{\omega}\frac{2q_1q_2}{(kq_1)(kq_2)}\Bigg|_{\omega<\Delta\varepsilon}=\frac{\alpha}{\pi}\left[L_q\ln\left(\frac{m^2(\Delta\varepsilon)^2}{\lambda^2\varepsilon_1\varepsilon_2}\right)+\frac{1}{2}L_q^2-\frac{1}{2}\ln^2\left(\frac{\varepsilon_1}{\varepsilon_2}\right)-\frac{\pi^2}{3}+\mathrm{Li}_2\left(\cos^2\frac{\theta}{2}\right)\right]. \tag{26}$$

Here we used the notation

$$L_q=\ln\frac{-q^2}{m^2},\quad q_1^2=q_2^2=m^2,\quad -q^2=-(q_1-q_2)^2\gg m^2,$$

$$q_{1,2}=(\varepsilon_{1,2},\mathbf{q}_{1,2}),\quad \theta=\widehat{\mathbf{q}_1\mathbf{q}_2}, \tag{27}$$

where ε_1, ε_2, and θ are the energies and angle between the 3-momenta $\mathbf{q}_1$, $\mathbf{q}_2$, respectively, and λ is the fictitious photon mass (all defined in the center of mass system).

The contributions of each possible term on the right-hand side of Eq. (24) are

$$\frac{\pi}{\alpha}\delta^{\rm soft}=-\Delta_1-\Delta_2-\Delta_1'-\Delta_2'+\Delta_{12}+\Delta_{1'2'}+\Delta_{11'}+\Delta_{22'}-\Delta_{12'}-\Delta_{1'2},$$

$$\Delta_1=\Delta_2=\ln\frac{m\Delta\varepsilon}{\varepsilon\lambda},\quad \Delta_1'=\ln\frac{m\Delta\varepsilon}{\varepsilon_1'\lambda},\quad \Delta_2'=\ln\frac{m\Delta\varepsilon}{\varepsilon_2'\lambda},$$

$$\Delta_{12}=2L_s\ln\frac{m\Delta\varepsilon}{\varepsilon\lambda}+\frac{1}{2}L_s^2-\frac{\pi^2}{3},$$

$$\Delta_{1'2'}=L_{s_1}\ln\left(\frac{(m\Delta\varepsilon)^2}{\varepsilon_1'\varepsilon_2'\lambda^2}\right)+\frac{1}{2}L_{s_1}^2-\frac{1}{2}\ln^2\left(\frac{\varepsilon_1'}{\varepsilon_2'}\right)-\frac{\pi^2}{3}+\mathrm{Li}_2\left(\cos^2\frac{\theta_{1'2'}}{2}\right),$$

$$\Delta_{11'}=L_{t_1}\ln\left(\frac{(m\Delta\varepsilon)^2}{\varepsilon_1'\varepsilon\lambda^2}\right)+\frac{1}{2}L_{t_1}^2-\frac{1}{2}\ln^2\left(\frac{\varepsilon_1'}{\varepsilon}\right)-\frac{\pi^2}{3}+\mathrm{Li}_2\left(\cos^2\frac{\theta_{1'}}{2}\right),$$

$$\Delta_{22'}=L_t\ln\left(\frac{(m\Delta\varepsilon)^2}{\varepsilon\varepsilon_2'\lambda^2}\right)+\frac{1}{2}L_t^2-\frac{1}{2}\ln^2\left(\frac{\varepsilon_2'}{\varepsilon}\right)-\frac{\pi^2}{3}+\mathrm{Li}_2\left(\sin^2\frac{\theta_{2'}}{2}\right),$$

TABLE I. Numerical estimates of Δ_L and Δ versus y_1, y_2, c_1, c_2.

$\mathcal{N}$	y_1	y_2	c_1	c_2	Δ_L	Δ
1	0.36	0.89	−0.70	−0.10	10.70	−24.53
2	0.59	0.66	0.29	−0.06	4.86	−11.41
3	0.67	0.67	0.50	0.30	5.82	−35.58
4	0.68	0.65	0.60	−0.50	4.10	−10.45

$$\Delta_{1'2}=L_{u_1}\ln\left(\frac{(m\Delta\varepsilon)^2}{\varepsilon\varepsilon_1'\lambda^2}\right)+\frac{1}{2}L_{u_1}^2-\frac{1}{2}\ln^2\left(\frac{\varepsilon_1'}{\varepsilon}\right)-\frac{\pi^2}{3}+\mathrm{Li}_2\left(\sin^2\frac{\theta_{1'}}{2}\right),$$

$$\Delta_{12'}=L_u\ln\left(\frac{(m\Delta\varepsilon)^2}{\varepsilon\varepsilon_2'\lambda^2}\right)+\frac{1}{2}L_u^2-\frac{1}{2}\ln^2\left(\frac{\varepsilon}{\varepsilon_2'}\right)-\frac{\pi^2}{3}+\mathrm{Li}_2\left(\cos^2\frac{\theta_{2'}}{2}\right),$$

$$L_u=\ln\frac{-u}{m^2},\quad L_{u_1}=\ln\frac{-u_1}{m^2},$$

$$\mathrm{Li}_2(z)\equiv-\int_0^z\frac{dx}{x}\ln(1-x), \tag{28}$$

where $\varepsilon_1',\varepsilon_2'$ are the center of mass energies of the scattered electron and positron, respectively; $\theta_{1'},\theta_{2'}$, are their scattering angles (measured from the initial electron momentum direction); and $\theta_{1'2'}$ is the angle between the scattered electron and positron momenta.

Separating out large logarithms, we obtain

$$\delta^{\mathrm{soft}}=\frac{\alpha}{\pi}\left\{4(L_s-1)\ln\frac{m\Delta\varepsilon}{\lambda\varepsilon}+L_s^2+L_s\ln\frac{tt_1}{uu_1}+L_s\ln\frac{1-c_{1'2'}}{2}+\mathcal{O}(1)\right\},$$

$$c_{1'2'}=\cos\theta_{1'2'}. \tag{29}$$

This can be written in another form, using experimentally measurable quantities, the relative energies of the scattered leptons and the scattering angles (see Table I):

$$y_i=\frac{\varepsilon_i'}{\varepsilon},\quad c_i=\cos\theta_i',\quad \frac{1}{2}(1-c_{1'2'})=\frac{y_1+y_2-1}{y_1y_2},$$

$$-\frac{t}{s}=y_2\frac{1+c_2}{2},\quad -\frac{u}{s}=y_2\frac{1-c_2}{2},\quad -\frac{t_1}{s}=y_1\frac{1-c_1}{2},$$

$$\frac{s_1}{s}=y_1+y_2-1,\quad -\frac{u_1}{s}=y_1\frac{1+c_1}{2}. \tag{30}$$

6. CONCLUSIONS

The double logarithmic terms of type L_s^2 and those proportional to $L_s\ln(\lambda/m)$ cancel in the overall sum with the corresponding terms from the soft photon contribution (29). Omitting vacuum polarization, we obtain in the logarithmic approximation

$$d\sigma^{\mathrm{soft+virt}}=d\sigma_0\frac{\alpha}{\pi}\left[L_s\left(4\ln\frac{\Delta\varepsilon}{\varepsilon}+\Delta_L\right)+\Delta(y_1,y_2,c_1,c_2)\right],$$

$$\Delta_L=3+\ln\frac{(1-c_1)(1-c_2)}{(1+c_1)(1+c_2)}+\ln\frac{y_1+y_2-1}{y_1y_2}+\frac{1}{F}\left[\Phi\ln\frac{s^2}{tt_1}+\Phi_P\ln\frac{s_1^2}{tt_1}+\Phi_Q\ln\frac{t_1^2}{ss_1}+\Phi_R\ln\frac{t^2}{ss_1}\right]+2\ln\frac{ss_1}{uu_1}+\frac{2}{F}(\Phi_Q+\Phi_R)\ln\frac{tt_1}{ss_1}. \tag{31}$$

The function $\Delta(y_1,y_2,c_1,c_2)$ is quite complicated. To compare it with Δ_L, we give their numerical values (omitting vacuum polarization) for a certain set of points from physical regions (32) and $y_1+y_2>1$, $D>0$ (see Table I). Considering the kinematics typical of large angle inelastic Bhabha scattering, we show the lowest-order contribution previously obtained[10] and the radiative corrections calculated in this work.

After performing loop integration and shifting logarithms ($L_i=L_s+L_{is}$), one can see that the terms containing infrared singularities and double logarithmic terms $\sim L_s^2$, are associated with a factor equal to the corresponding Born contribution. This is true of all types of contributions.

The phase volume

$$d\Gamma=\frac{d^3p_1'd^3p_2'd^3k_1}{\varepsilon_1'\varepsilon_2'\omega_1}\,\delta^{(4)}(p_1+p_2-p_1'-p_2'-k_1)$$

can be transformed in various ways.[10] We introduce the variables [see Eq. (30)]

$$y_i=\frac{\varepsilon_i'}{\varepsilon},\quad c_i=\cos\theta_i',\quad \theta_i'=\widehat{\mathbf{p}_1,\mathbf{p}_i'},$$

$$0<y_i<1,\quad -1<c_{1,2}<1, \tag{32}$$

which parametrize the kinematics of the outgoing particles (these do not include a common degree of freedom, a rotation about the beam axis). The phase volume then takes the form

$$d\Gamma=\frac{\pi s\,dy_1dy_2dc_1dc_2}{2\sqrt{D(y_1,y_2,c_1,c_2)}}\Theta(y_1+y_2-1)\times\Theta[D(y_1,y_2,c_1,c_2)],$$

$$D(y_1,y_2,c_1,c_2)=\rho^2-c_1^2-c_2^2-2c_{1'2'}c_1c_2,$$

$$\rho^2=2(1-c_{1'2'})\frac{(1-y_1)(1-y_2)}{y_1y_2}. \tag{33}$$

The allowed region of integration is a triangle in the y_1,y_2 plane and the interior of the ellipse $D>0$ in the c_1,c_2 plane.

We now discuss the relation of our result to the renormalization group approach. The dependence on $\Delta\varepsilon/\varepsilon$ in (31) disappears when one takes into account hard two-photon emission. The leading contribution arises from the kinemat-

ics when the second hard photon is emitted close to the direction of motion of one of the incoming or outgoing particles:

$$d\sigma^{\rm hard}=\frac{\alpha}{2\pi}L_s\left[\frac{1+z^2}{1-z}(d\sigma_0(zp_1,p_2,p_1',p_2')+d\sigma_0(p_1,zp_2,p_1',p_2'))dz+\frac{1+z_1^2}{1-z_1}d\sigma_0\left(p_1,p_2,\frac{p_1'}{z_1},p_2'\right)dz_1+\frac{1+z_2^2}{1-z_2}d\sigma_0\left(p_1,p_2,p_1',\frac{p_2'}{z_2}\right)dz_2\right],$$

$$z=1-x_2,\quad z_i=\frac{y_i}{y_i+x_2},\quad x_2=\frac{\omega_2}{\varepsilon}. \tag{34}$$

The fractional energy of the additional photon varies within the limits $\Delta\varepsilon/\varepsilon<x_2=\omega_2/\varepsilon<1$. This formula agrees with the Drell–Yan form of radiative Bhabha scattering (with switched-off vacuum polarization)

$$d\sigma(p_1,p_2,p_1',p_2')=\int dx_1dx_2\mathscr{D}(x_1)\mathscr{D}(x_2)d\sigma_0\times\left(x_1p_1,x_2p_2,\frac{p_1'}{z_1},\frac{p_2'}{z_2}\right)\times\mathscr{D}(z_1)\mathscr{D}(z_2)dz_1dz_2, \tag{35}$$

where the non-singlet structure functions $\mathscr{D}$ are[11]

$$\mathscr{D}(z)=\delta(1-z)+\frac{\alpha}{2\pi}L\mathscr{P}^{(1)}(z)+\left(\frac{\alpha}{2\pi}L\right)^2\frac{1}{2!}\mathscr{P}^{(2)}(z)+\ldots,$$

$$\mathscr{P}^{(1)}(z)=\lim_{\Delta\to0}\left[\frac{1+z^2}{1-z}\Theta(1-z-\Delta)+\delta(1-z)\left(2\ln\Delta+\frac{3}{2}\right)\right]. \tag{36}$$

In our calculations we see explicitly a factorization of the terms containing double logarithmic contributions and infrared single logarithmic ones, which arise from G- and ΓΠ-type Feynman diagrams. To be precise, the corresponding contributions to the cross section have the structure of the Born cross section (6). But the above claim fails to be true for terms containing single logarithms. Hence, the Drell–Yan form (35) is not valid in this case, and the factorization theorem breaks down, because the mass singularities (large logarithms) do not factorize before the Born structure. That is because of plenty of different type amplitudes and kinematic variables, which describe our process. The reason for the violation of a naive usage of factorization in the Drell–Yan form has presumably the same origin with that found in Ref. 12, where the authors claimed that it is necessary to study independently the renormalization group behavior of leading logarithms before different amplitudes of the same process. Note that in the $e\mu\to e\mu\gamma$ reaction, which can easily be extracted from our results, factorization does take place. We also see from (31) that factorization will take place if all the logarithmic terms become equal, i.e., $\ln(s_1/m^2)=\ln(s/m^2)=\ldots$. The source for the violation of the factorization theorem, we found, might have a relation to some of those found in other problems.[13]

Numerical estimates (see Table I) for the Φ factory energy range ($\sqrt{s}\simeq1$ GeV) shows that the contribution of the non leading terms coming from virtual and soft real photon emission might reach 35%. Additional hard photon emission will also contribute to Δ_L and Δ. To get an explicit form of that correction, one has to take into account a definite experimental setup.

Obviously, an analogous phenomenon of the factorization theorem violation takes place in QCD in processes like $q\bar{q}\to q\bar{q}g$ and $q\bar{q}\to q\bar{q}\gamma$. A consistent investigation of the latter processes, taking into account the phenomenon found, can give a certain correction to predictions for large angle jet production and direct hard photon emission at proton–antiproton colliders.

We are grateful to D. V. Shirkov for useful discussions and pointing Ref. 12 out to us. We also thank A. Belitsky, P. Ferro, N. Merenkov and L. Trentadue for participating at the very beginning of this investigation. We are also indebted to A. Belitsky for assistance in the creation of the tables of integrals[14] used in this paper. This work was supported in part by INTAS (Grant No. 93-1867 ext).

*)E-mail: arbuzov@thsun1.jinr.ru

[1] S. I. Dolinsky, V. P. Druzhinin, M. S. Dubrovin *et al.*, Phys. Rep. **202**, 99 (1991).
[2] A. B. Arbuzov, E. A. Kuraev, N. P. Merenkov *et al.*, Yad. Fiz. **60**, 673 (1997) [Phys. At. Nucl. **60**, 591 (1997)].
[3] A. B. Arbuzov, E. A. Kuraev, N. P. Merenkov *et al.*, Nucl. Phys. B **474**, 271 (1996).
[4] A. B. Arbuzov, V. A. Astakhov, E. A. Kuraev *et al.*, Nucl. Phys. B **483**, 83 (1997).
[5] E. A. Kuraev, N. P. Merenkov, and V. S. Fadin, Yad. Fiz. **45**, 782 (1987) [Sov. J. Nucl. Phys. **45**, 486 (1987)].
[6] A. B. Arbuzov, E. A. Kuraev, and B. G. Shaikhatdenov, E-prints archive hep-ph/9805308.
[7] F. A. Berends *et al.*, Phys. Lett. B **103**, 124 (1981); Nucl. Phys. B **206**, 59 (1982).
[8] A. B. Arbuzov, G. V. Fedotovich, E. A. Kuraev *et al.*, JHEP **10**, 001 (1997).
[9] G. t'Hooft and M. Veltman, Nucl. Phys. B **153**, 365 (1979).
[10] S. Eidelman, E. Kuraev, and V. Panin, Nucl. Phys. B **148**, 245 (1979).
[11] E. A. Kuraev and V. S. Fadin, Yad. Fiz. **41**, 466 (1985) [*sic*].
[12] I. F. Ginzburg and D. V. Shirkov, Zh. Éksp. Teor. Fiz. **49**, 335 (1965) [Sov. Phys. JETP **22**, 234 (1965)].
[13] J. Collins, L. Frankfurt, and M. Strikman, Phys. Lett. B **307**, 161 (1993); E. Gotsman, E. Levin, and V. Maor, Phys. Lett. B **406**, 89 (1997); E. L. Berger, X. Guo, and J. Qiu, Phys. Rev. Lett. **76**, 2234 (1996); A. Duncan and A. H. Mueller, Phys. Rev. D **21**, 1636 (1980); Phys. Lett. B **90**, 159 (1980); A. Milshtein and V. S. Fadin, Yad. Fiz. **33**, 1391 (1981) [Sov. J. Nucl. Phys. **33**, 746 (1981)].
[14] A. B. Arbuzov, A. V. Belitsky, E. A. Kuraev, and B. G. Shaikhatdenov, JINR Communications E2-98-53, Dubna (1998).

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

Investigation of the passage of 100–1000-MeV/nucleon superheavy ions through homogeneous media

S. D. Bogdanov,*) S. S. Bogdanov, E. E. Zhurkin, and V. F. Kosmach

St. Petersburg State Technical University, 195251 St. Petersburg, Russia
(Submitted 2 July 1998)
Zh. Éksp. Teor. Fiz. **115**, 404–415 (February 1999)

This work is a continuation of the experimental and theoretical investigations of the effect of the Z_1^3 correction to the stopping power of ions on the passage of heavy ions ^{40}Ar, ^{56}Fe, ^{197}Au, ^{131}Xe, and ^{238}U with energies of about 1 GeV/nucleon through a homogeneous medium. The previously observed systematic deviations of the calculations based on the first Born approximation to the scattering of a particle by the atomic electrons in the medium from the experimental values of the total ionization ranges of the nuclei and their stopping powers is confirmed. The discrepancy increases with the atomic number of the projectile nucleus. It is shown that the Z_1^3 correction in the form proposed by Jackson and McCarthy eliminates, especially for ions with $Z_1>50$, the systematic discrepancy between the computed and experimental values. For the experimental energy range relativistic Mott scattering of a particle by the atomic electrons in the target makes the dominant contribution to the observed Z_1^3 effect. © *1999 American Institute of Physics.* [S1063-7761(99)00302-9]

1. INTRODUCTION

Despite the apparent abundance of experimental information, data on the passage of heavy nuclei through homogeneous media remains sporadic and incomplete. Most models describing the overall characteristics of the passage of heavy nuclei with energies of about 1 GeV/nucleon, such as the ionization ranges of heavy nuclei, the longitudinal and transverse stragglings of these ranges, and the stopping powers of heavy ions, were developed and tested more than 20 years ago on charged particles with small mass and low energy.[1–8]

In recent years, indications of appreciable deviations of the ranges of these nuclei[9–16] and their stopping powers[17] from the simple Z_1^2 dependence, prescribed by the Bethe–Bloch theory,[1,2] on the atomic number of the ion have been obtained in a number of experimental works investigating the passage of nuclei with atomic number greater than 75.

This work is a continuation of the experimental and theoretical investigations of the effect of the Z_1^3 correction to the stopping power of ions when heavy nuclei with energies of about 1 GeV/nucleon pass from neon to uranium through a homogeneous medium under the conditions of a complete experiment allowed by the nuclear-emulsion method.[12–15] In this work, new experimental data were obtained on the total ionization ranges of nuclei with the BR-2 emulsion, prepared at the NIKFI, and on the distributions of these ranges. The effect of the projectile energy, charge, and mass and of the atomic number of the target on the overall characteristics of the process is analyzed.

2. EXPERIMENT

Five chambers, consisting of 30–35 layers of the BR-2 emulsion with the standard composition[18] (number of nuclei, $n_i \cdot 10^{-22}$ per cm^3: 3.148 H, 1.412 C, 0.396 N, 0.956 O, 0.004 S, 0.002 I, 1.031 Br, and 1.036 Ag), were exposed in the Bevalac accelerator (Berkeley, CA USA) in beams of 505-MeV/nucleon ^{40}Ar nuclei, 500 and 1000-MeV/nucleon ^{56}Fe nuclei, 1147.2-MeV/nucleon ^{197}Au nuclei, and 927.6-MeV/nucleon ^{238}U nuclei. The average fluence in the irradiations was 500–1500 particles/cm^2. The dimensions of an individual layer were close to $10\times10\times0.05$ cm^3. The beam entered a stack of photoemulsion plates approximately parallel to the surface of the layers in all irradiations. The angle made by the beams of nuclei in an undeveloped emulsion was less than 2°. Two 220 μm thick layers of black paper and one 190 μm thick layer of polyethylene were present in front of the emulsion during irradiation. The photographic processing of the chambers was performed in the high-energy laboratory at the Joint Institute of Nuclear Research (Dubna, Russia). Before the chambers were developed, the thickness of each layer of emulsion was measured at four symmetrically arranged points. It was found that the thickness of an undeveloped layer fluctuates with standard deviation close to 4 μm. All plates were marked with a photographically applied millimeter grid.

The search for stopped primary ions was conducted by double (fast and slow) scanning along the primary track in MBI-9 microscopes with 10×15, 20×15, and 40×15 magnifications. Ion tracks showing indications of inelastic or elastic interaction with the emulsion nuclei were excluded. The choice of very low magnification of the experimental objects for the primary scanning was dictated by the specific nature of the tracks studied. The visible thickness of a track due to a uranium nucleus varied from 20 to 50 μm along the track. Tracks due to gold nuclei had similar thicknesses. For comparison, we note that a millimeter grid is close to 15 μm thick.

A coordinate method was used to determine the charac-

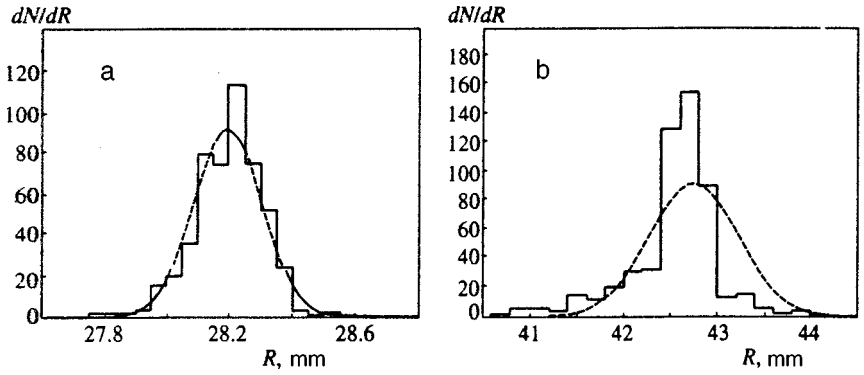

FIG. 1. FIG. 1. Ionization range distribution of 924.3-MeV/nucleon ^{238}U ions (1a, 0.05 mm cell) and 1144.4-MeV/nucleon ^{197}Au (1b, 0.2 mm cell) stopped in a BP-2 emulsion. The parameters of the Gaussian distributions (for ^{238}U $N=518$ tracks, $\langle R\rangle=28.197$ mm, $\sigma=0.116$ mm; for ^{197}Au $N=537$ tracks, $\langle R\rangle=42.75$ mm, $\sigma=0.471$ mm) were determined from experiment. The fitting criteria with 16 degrees of freedom are $\chi^2/k=2.19$ (a) and $\chi^2/k=9.43$ (b).

teristics of an individual track.[19] The method used a three-dimensional coordinate system, including the number of the plate in the chamber, the coordinates of a square in the millimeter grid of a plate, and the coordinates of an event within a square. Each track was characterized by the entrance coordinate of a nucleus in the plate and the coordinate of the event (stopping point or interaction). Multiple measurements were performed on specific tracks to estimate the accuracy of the coordinate method. Measurements were performed every week for three months, and the standard deviation of an individual measurement of the length of the same track was 30 μm on a 30000–80000 μm baseline.

The numbers of the observed and measured tracks of ^{40}Ar, ^{56}Fe, ^{197}Au, and ^{238}U ions which stopped without undergoing nuclear interactions were 267, 756 (for two energies), 537, and 518 tracks, respectively, which is more than two times larger than the data sample in Ref. 12. Figure 1 shows the experimental distributions, obtained in the present work, of the track lengths of uranium nuclei with initial energy 924.3 MeV/nucleon (Fig. 1a) and 1144.4-MeV/nucleon gold nuclei (Fig. 1b) which stopped in the BR-2 emulsion. In determining the energy of the nuclei, the energy losses in the material traversed by the beam of nuclei before entering the emulsion were taken into account. One can see that these unimodal distributions resemble Gaussian distributions, but the χ^2/k fitting criteria are far from unity. The straggling of the ranges of uranium nuclei was estimated to be $\sigma=0.017$ mm on the basis of data on the straggling of protons with the same velocity,[20] taking account of the inverse proportionality between the relative straggling and the square root of the uranium mass. This value of σ is approximately 10 times smaller than the experimental value. The observed deviations of the experimental distributions of the ion range lengths from the theoretical Gaussian distributions are probably due to the energy spread of the beam of incident nuclei.

3. MODELS AND TESTING PROGRAMS

The foundations of the theory of stopping of high-energy charged particles were laid by Bohr, Bethe, and Bloch.[1,2] The stopping power (S_{e2}) can be written in the form

$$S_{e2}=\frac{4\pi Z_1^2 e^4}{m_e v^2} Z_2 L(\beta,Z_2), \tag{1}$$

where e and m_e are the electron charge and mass, Z_1 and v are the atomic number and velocity of the ion, $\beta=b/c$, c is the speed of light in vacuum, Z_2 is the atomic number of the atoms in the medium, $L(\beta,Z_2)$ is the ''logarithm'' in the Bethe–Bloch formula, which is often written in the form

$$L(\beta,Z_2)=\ln\frac{2m_e v^2}{\langle I\rangle}+\ln\frac{1}{(1-\beta^2)^2}-\beta^2-\frac{C}{Z_2}, \tag{2}$$

where $\langle I\rangle$ is the average ionization potential of the target. The first term is the standard form of the logarithm in the Bethe formula, the second and third terms are the relativistic corrections, and the fourth term is a correction for the effect of the binding of the electrons in an atom. Quite often, charge effects due to the capture of electrons from the medium by a moving low-velocity ion are taken into consideration.

It should be noted that only the first Born approximation is used in the classical Bethe–Bloch theory for describing the scattering of a particle by the atomic electrons in a medium. The second Born approximation contributes a small positive term, proportional to Z_1^3, in the expression for the total stopping power S_{e3}:

$$S_{e3}=S_{e2}[1+Z_{1\text{eff}}(J/I)], \tag{3}$$

where $Z_{1\text{eff}}$ is the effective ion charge, the term S_{e2} corresponds to the standard Bethe–Bloch models, and J/I is a small positive quantity. Indeed, an effect of the same kind was discovered experimentally by Barkas and his colleagues in measurements of the ranges of π^+ and π^- mesons with the same energy in a nuclear photoemulsion. It was found that their ranges differ by several percent.[21] In addition, this difference was present even at nonrelativistic velocities, where the correction following directly from the second Born approximation in the Mott electron scattering is negligible.[22] Andersen and Ziegler also established such an effect.[23,24] They showed on the basis of a compilation of several thousands of experimental data points that the electronic energy losses of α particles are greater than the fourfold losses for protons at the same velocity. This is at variance with the classical dependence $S_{e2}\propto Z_1^2$.

Quantitative estimates of the Z_1^3 correction were made by Jackson and McCarthy in Ref. 22. Their formula is

$$\frac{J}{I}=\left(\frac{J}{I}\right)_{\text{rel}}+\left(\frac{J}{I}\right)_{\text{nonrel}}=\frac{\pi\alpha\beta}{2L(\beta,Z_2)}+\frac{F(\nu)}{Z_2^{1/2}}, \tag{4}$$

where $\alpha=1/137$ is the fine structure constant, $L(\beta,Z_2)$ is the logarithm in the Bethe–Bloch formula, $\nu=137\gamma\beta Z_2^{-1/2}$, and $\gamma=(1-\beta^2)^{-1/2}$. The function $F(\nu)$ is tabulated in Ref. 22 and can be expressed as $F(\nu)=0.3\nu^{-2}$ for $\nu<1.5$, $F(\nu)=0.5\nu^{-2.3}$ for $1.5<\nu<4$, and $F(\nu)=0.5\nu^{-2.5}$ for $\nu>4$.

The form of the terms in the expression for J/I is due to the fact that physics of the Z_1^3 correction is different in different velocity ranges. In the relativistic case (first term) it is due to the second Born approximation in the scattering of a particle by the atomic electrons in the target. This correction is appreciable at relativistic velocities of the incident ion and is approximately proportional to the particle velocity: $(J/I)_{\text{rel}}\propto\beta$. The second term is an extension of the Ashley, Anderson, Ritchie, and Brandt theory[25–28] (the so-called ARB approximation for the Z_1^3 correction in the weakly-relativistic region), and it is due to the additional incorporation of the displacement of an electron over the collision time. The nonrelativistic correction has a large effect at low incident-particle velocities, where $(J/I)_{\text{nonrel}}\propto\beta^{-2}$.

The calculations of the passage of fast nuclei through a photoemulsion were performed using the PRAL-96 program, which is based on the well-known Ziegler–Biersack–Littmark (ZBL) model[23] of continuous electronic stopping and takes account of the elastic collisions of an ion with the atomic nuclei in the medium, and on the basis of an improved ZBL model, which is incorporated in the RANGE-96 program package,[29] which makes it possible to perform calculations, both neglecting the Z_1^3 correction (i.e., in accordance with the ZBL model) and taking this correction into account via the relations (3) and (4).

An additional check was required in order to use these programs to calculate the passage of relativistic particles in a photoemulsion. This check was performed using experimental data[3–7] on the ranges and stragglings of H and He in an Inford G5 emulsion. For an emulsion (a multicomponent medium), the ionization stopping power was calculated using the Bragg rule. The results of the testing showed that the PRAL and RANGE programs give the same results within less than 0.1% of the range of H and He nuclei in an emulsion. The particle ranges computed with the PRAL and RANGE programs were found to be systematically less (approximately by 2%) than Barkas' experimental and computational data. Taking account of the indicated systematic discrepancies, the largest discrepancy between the theoretical and experimental data in both the Barkas calculations and in the calculations performed in the present work is the same and is less than 1.5–2%.

The RANGE and PRAL algorithms were further tested by calculations of the passage of heavier nuclei through a photoemulsion. In most works, Barkas' semiempirical formula is used to determine the ionization range of a particle in a nuclear emulsion.[8] In this formula, the RANGE $R(\beta)$ of a heavy ion with mass M_1 and atomic number Z_1 is related to the range of a proton with the same velocity ($\lambda(\beta)$) by

$$R(\beta)=(M_1/Z_1^2)\lambda(\beta)+C_z(\beta/Z_1)M_1Z_1^{2/3}, \tag{5}$$

where the additional correction $C_z(\beta/Z_1)$ is due to electron capture at low ion velocities of heavy ions. Ordinarily, it is considered to be a universal function of β/Z_1 for a given type of emulsion and does not depend on the atomic number of the projectile.

Using the RANGE program, the function $C_z(\beta/Z_1)$ was obtained for a series of ions by a numerical method (proceeding from precalculated values of $R(\beta)$ for heavy ions and the corresponding proton ranges λ at the same velocity β). The results of the calculations showed that in the ZBL model the function $C_z(\beta/Z_1)$ can be considered to be ''universal'' only to within 10%, since for heavy nuclei (Au, U) the correction $C_z(\beta/Z_1)$ reaches a plateau appreciably more rapidly with increasing ion velocity, and the magnitude of this plateau is somewhat less than for the lighter nuclei Ne, Ar, and Fe. The universality of the curve $C_z(\beta/Z_1)$ breaks down due to the different nature of the velocity dependence of the effective charge of the ions undergoing stopping—ions with a higher atomic number reach the value $Z_{\text{eff}}/Z=1$ more rapidly as their velocity increases. We note that the universality of the correction $C_z(\beta/Z_1)$ was discussed in Ref. 8 only for ions $Z_1\leqslant 18$ with possible 10% errors.

Data on the maximum correction to the range of heavy nuclei as a result of electron capture at low heavy-ion velocities are of special interest. These maximum corrections ($C_z(\beta/Z_1)M_1Z_1^{2/3}$) to the range of heavy nuclei with initial energy 1 GeV/nucleon were 0.08 mm, 0.1 mm, 0.6 mm, and 0.7 mm, respectively, for ^{40}Ar, ^{56}Fe, ^{137}Au, and ^{238}U nuclei (with a total range of 146 mm, 100 mm, 38.8 mm, and 34.5 mm, respectively), i.e., in the gigaelectron volt energy range this correction, even for ^{238}U, increases the range by no more than 2% and has little effect on $R(\beta)$ for heavy ions.

Thus, the results of the additional check showed that both the universal ZBL model of continuous electronic stopping, including elastic collisions of an ion with the atomic nuclei in the medium, that is the basis of the PRAL and RANGE program package, as well as the programs themselves, make it possible to describe with good accuracy the known stopping parameters for light nuclei with relativistic velocities in homogeneous media.

4. AVERAGE IONIZATION RANGES OF HEAVY IONS

The experimental data on the total ranges of nuclei and the computed values of these quantities are given for comparison in Tables I and II. Table I gives the ranges of ^{40}Ar, ^{56}Fe, ^{197}Au, and ^{238}U nuclei measured in the present work in a BR-2 emulsion and the data of Ref. 16 on the ranges of ^{197}Au and ^{238}U nuclei in lexan ($C_{16}H_{14}O_3$ with density 1.2 g/cm^3). The energy of the incident nuclei, given in column 2 of the table, takes account of the thickness of the material traversed by the beams of nuclei before entering the emulsion. The results of direct measurements are in column 3 of Table I. One can see that the computational accuracy in the present work is quite high. This is due to the quite large data sample of measured tracks. The next three columns in this table give the computed values of the ion ranges obtained using the semiempirical relation (5), the PRAL program, and the RANGE program neglecting the Z_1^3 correction. It follows

TABLE I. Ionization ranges of heavy ions in a BP-2 emulsion and lexan.

Type of ion and target	Energy, MeV/nucleon	Ionization range R, mm				
			Calculation			
			no Z_1^3 correction			with Z_1^3-correction, RANGE
		Experiment	(5)	PRAL	RANGE	
Ar	504.2	53.11	54.71	54.25	54.23	53.26
Br-2		±0.05	(3%)	(2.14%)	(2.11%)	(+0.28%)
Fe	498.7	34.84	36.26	35.82	35.85	34.94
Br-2		±0.01	(4%)	(2.81%)	(2.90%)	(+0.29%)
Fe	980.8	95.16	98.0	96.10	96.05	93.38
Br-2		±0.04	(3%)	(0.99%)	(0.94%)	(−1.87%)
Au	1144.4	42.75	46.23	45.81	45.76	42.11
Br-2		±0.05	(8.1%)	(7.15%)	(7.04%)	(−1.49%)
U	924.3	28.197	31.35	30.68	30.66	27.90
Br-2		±0.02	(11%)	(8.80%)	(8.80%)	(−1.05%)
Au	975.0	79.49	–	87.79	87.75	81.84
Lexan		±2.44		(10.4%)	(10.4%)	(2.95%)
U	901.0	63.50	–	70.50	70.48	650.9
Lexan		±1.90		(11.0%)	(11.0%)	(2.50%)

Note. The deviations δ of the computed from the experimental ranges are given in parentheses: $\delta=[(R_{calc}-R_{exp})/R_{exp}]\cdot 100\%$. The errors in R_{exp} were determined taking account of the experimental variance of the range length distribution. The experimental data on the ranges in lexan were taken from Ref. 16.

from these data that the PRAL and RANGE programs give very close computed values of the ranges, and the calculation using the relation (5) predicts somewhat larger (on the average by 1%) ion ranges.

Comparing the experimental data and the computed values, it is evident that the relative discrepancy $\delta=[(R_{calc}-R_{exp})/R_{exp}]\cdot 100\%$ between the computed (neglecting the Z_1^3 correction) values and the experimental data increases systematically with the atomic number of the projectile nucleus from 2% for Ar ($Z_1=18$) up to 9–11% for U ($Z_1=92$) for virtually all calculations performed neglecting the Z_1^3 correction. The observed deviations are systematic and fall outside the bounds of the possible experimental and computational errors, as well as possible differences of the charge states of ions with different values of Z_1 at low ion velocities. The correction $C_z(\beta/Z_1)M_1Z_1^{2/3}$, as shown above, contributes no more than 2% to the increase in the range of a heavy ion. In this connection, we inferred that the observed discrepancies in the experimental and computed ion ranges are mainly due to the fact that only the first Born approximation is used in the calculations to describe the scattering of a particle by the atomic electrons in the medium and the second approximation was neglected.

TABLE II. Ranges of 900-MeV/nucleon gold ions in various media.

Target type $\langle Z_2\rangle$	Range R, g/cm^2			
		Calculation		
		no Z_1^3 correction		with Z_1^3 correction, RANGE
	Experiment[11]	Ref. 11	RANGE	
(CH)	8.57±0.01	9.154	9.14	8.54
3.5		(6.81%)	(6.64%)	(−0.35%)
(C)	9.67±0.07	10.070	10.06	9.38
6		(4.14%)	(4.03%)	(−3.00%)
(Al)	10.54±0.05	11.364	11.33	10.50
13		(7.72%)	(7.39%)	(−0.47%)
(Cu)	12.10±0.04	13.027	13.07	12.05
29		(7.66%)	(8.02%)	(−0.41%)
(Sn)	13.98±0.08	14.877	15.06	13.80
50		(6.42%)	(7.73%)	(−1.29%)
(Au)	15.59±0.01	16.636	16.74	15.26
79		(6.71%)	(7.38%)	(−2.12%)
(Pb)	15.72±0.04	16.863	17.02	15.51
82		(7.27%)	(8.27%)	(−1.34%)

Note. The deviations δ of the computed from experimental ranges are indicated in parentheses: $\delta=[(R_{calc}-R_{exp})/R_{exp}]\cdot 100\%$.

The calculations for estimating the effect of the Z_1^3 correction on the particle ranges were performed using the RANGE program and the relations (3) and (4). The computational results for the total ranges with the Z_1^3 correction are given in column 7 in Table I. It was found that the relative decrease of the ranges due to the Z_1^3 effect in the stopping powers of ions for the cases considered in Table I are about 10% and 8%, respectively, for U and Au and less than 3% for all other ions. It is evident that taking account of the second Born approximation appreciably improved the agreement between the experimental and computed values for the heavy nuclei (Au, U) and that the systematic increase in the discrepancies between experiment and theory, which was observed with increasing atomic number of the ion, essentially vanished.

For an additional check of our supposition that the Z_1^3 correction in the ionization energy losses of the particles strongly influences on the characteristics of the process, we performed calculations of the ranges of 900-MeV/nucleon gold nuclei in CH, C, Al, Cu, Sn, Au, and Pb (experiment of Ref. 11), both with and without the Z_1^3 effect. The results are presented in Table II. It follows from these data that the calculations based on the conventional Z_1^2 dependence of the electronic stopping power (the first Born approximation) us-

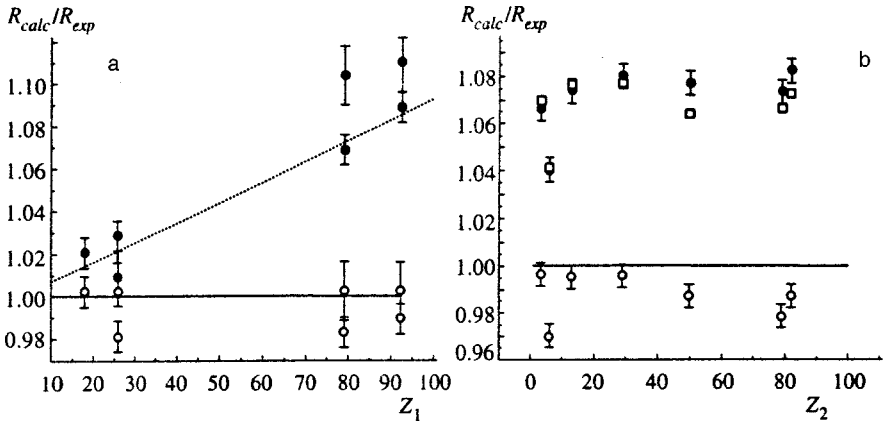

FIG. 2. Ratios of the calculated and experimental ranges of Ar, Fe, Au, and U ions in a BP-2 emulsoin and in lexan versus the atomic number Z_1 of the ion (a — experiment of this work and Ref. 16) and of 900-MeV/nucleon Au ions versus the atomic number Z_2 of the target (b — experiment of Ref. 11). Filled circles — calculation (R_{calc}) of this work neglecting the Z_1^3 correction, open circles — calculation (R_{calc}) of this work taking account of the Z_1^3 correction, rectangles — calculation (R_{calc}) of Ref. 11 (standard Bethe–Bloch model). The solid lines show the relation $R_{\text{calc}}/R_{\text{exp}}=1$ and the dotted lines show the relation $R_{\text{calc}}/R_{\text{exp}}=1+0.001Z_1$.

ing the RANGE program agree within the limits of the computational error (about 1%) with the calculations performed in Ref. 11 and differ from the experimental data by an average of 7.5%. The second Born approximation (the Z_1^3 correction) eliminates the discrepancies between the theory and the experimental data (see Table II). The remaining small differences ($\delta\sim(1-2)\%$) could be due to possible deviations from the standard values of the target densities used in the experiments.

The changes indicated are illustrated in Fig. 2. It follows from Fig. 2a that the ratio $R_{\text{calc}}/R_{\text{exp}}$ (neglecting the Z_1^3 effect) increases with the atomic number Z_1 of the ion essentially linearly as $1+10^{-3}Z_1$, i.e., the range computed on the basis of the first Born approximation for ionization energy losses is much greater than the experimental ion range, and this difference grows continuously with coefficient 10^{-3} up to the superheavy ions, which agrees with the theoretical estimates of the Z_1^3 correction according to Eqs. (3) and (4). One can see that this correction in the calculation removes the observed systematic change. It follows from Fig. 2b that the average ratio of the computed to the experimental ranges for 900-MeV/nucleon gold nuclei, determined according to the classical scheme, is virtually independent of the atomic number of the medium and is close to 1.07. Taking account of the second approximation decreases this ratio to 1.

In the present work, the effect of the Z_1^3 correction in the ionization losses of fast multiply charged particles on their range in three homogeneous media—in hydrogen (H, $\rho=0.0715$ g/cm^3), in BR-2 emulsion, and in silver (Ag, $\rho=10.5$ g/cm^3)—was analyzed. The range calculations were performed with the PRAL and RANGE programs for ^{56}Fe and ^{238}U nuclei with energies from 300 to 1000 MeV/nucleon. We note that the contribution of $(J/I)_{\text{rel}}$ to the stopping power exceeds the corresponding contribution from (J/I_{rel}) by more than a factor of 10 starting at energies above 100 MeV/nucleon, i.e., the correction $F(\nu)/Z_2^{1/2}$ in Eq. (4) plays a very weak role in this case.

To distinguish the effect of the atomic number of the medium and the particle energy, we introduced the coefficient $\alpha_R(E,Z_2)$ of the relative decrease of the particle ranges due to the Z_1^3 correction:

$$\alpha_R(E,Z_2)=\frac{R_2(E)-R_3(E)}{Z_1R_2(E)}, \tag{6}$$

where $R_2(E)$ is the range of an ion with energy E in the target Z_2, calculated neglecting the Z_1^3 correction, and $R_3(E)$ is the range calculated taking account of the Z_1^3 correction.

The calculation showed that in the energy range 300–1000 MeV/nucleon $\alpha_R(E,Z_2)$ increases weakly with the ion energy (in hydrogen from 0.062 to 0.079 for ^{56}Fe and from 0.056 to 0.075 for ^{238}U). As the atomic number of the target increases, the changes become greater—in the passage from a hydrogen to a silver target, $\alpha_R(E,Z_2)$ increases to 0.113 for 1000-MeV/nucleon ^{56}Fe and 0.104 for ^{238}U. Thus, experiments with the highest primary-ion energies and a target with a high atomic number are recommended for reliable experimental determination of the coefficient $\alpha_R(E,Z_2)$ of the relative decrease of particle ranges due to the Z_1^3 correction. Unfortunately, however, the accuracy of the experimental data (Table II, Fig. 2b) is still too low to distinguish this effect.

A very unexpected result is the dependence of the relative decrease of the particle ranges not only on the velocity of the incident ion and the atomic number of the medium, which follows in a natural manner from the physics of the process, but also on Z_1 (the atomic number of the incident ion): the larger Z_1, the smaller the coefficient $\alpha_R(E,Z_2)$. Thus, for 300 MeV/nucleon ^{56}Fe and ^{238}U in silver, the coefficient $\alpha_R(E,Z_2)$ was 0.097 and 0.088, respectively. Analysis showed that this effect is due mainly to the increase as a function of Z_1 of the correction $C_zM_1Z_1^{2/3}$ to the range of the nuclei as a result of electron capture when the velocity of the heavy ion is low and the consequent decrease in $\alpha_R(E,Z_2)$.

5. STOPPING POWER OF HEAVY IONS

The data examined above show that the relativistic Z_1^3 correction in the stopping powers of ions must be taken into account in order to calculate adequately the ranges of heavy ions in matter. However, it would be important to perform direct experimental measurements of the stopping powers of heavy relativistic particles with the same velocities. In Ref. 17, the stopping powers of 985-MeV/nucleon relativistic Ar nuclei and 780-MeV/nucleon Xe nuclei in Be, C, Al, Zn, and Bi were measured in experiments on the Bevalac accelerator. The authors indicated a systematic difference ($\approx 6\%$) between the measured values of the stopping powers S_e(Xe)

TABLE III. Electronic stopping powers of 985-MeV/nucleon Ar nuclei and 780-MeV/nucleon Xe nuclei in various media (Z_2).

	S_e(Ar), meV/(mg/cm^2)			S_e(Xe), meV/(mg/cm^2)		
		RANGE			RANGE	
Z_2	Experiment[17]	no Z_1^3 correction	with Z_1^3 correction	Experiment[17]	no Z_1^3 correction	with Z_1^3 correction
4	0.58±0.02	0.583	0.593	5.8±0.1	5.57	5.86
Be		(0.52%)	(2.24%)		(−3.97%)	(1.03%)
6	0.64±0.02	0.641	0.653	6.5±0.1	6.12	6.45
C		(0.16%)	(2.03%)		(−5.85%)	(−0.77%)
13	0.58±0.02	0.574	0.585	5.8±0.1	5.46	5.78
Al		(−1.03%)	(0.86%)		(−5.86%)	(−0.34%)
30	0.50±0.02	0.506	0.517	5.2±0.1	4.80	5.12
Zn		(1.20%)	(3.40%)		(−7.50%)	(−1.54%)
83	0.39±0.01	0.393	0.402	4.0±0.05	3.72	3.99
Bi		(0.77%)	(3.08%)		(−7.00%)	(−0.25%)

Note. The deviations of the computed values of the electronic stopping power from experiment are given in parentheses: $\delta=[(S_e^{\text{calc}}-S_e^{\text{exp}})/S_e^{\text{calc}}]\cdot 100\%$.

for xenon ions in various media and the results of the corresponding Bethe–Bloch calculations, while for argon ions such differences are not observed.

The results of Ref. 17 are presented in Table III together with the corresponding calculations which we performed both on the basis of the standard ZBL model and taking account of the Z_1^3 effect. It is evident that the calculations correctly describe the experimental dependence of the stopping powers of ions in various media on the atomic number Z_2 of the target. On the average, it is a decreasing function of Z_2. In addition, Z_2 oscillations of the stopping powers are clearly observed in the experiment and in both computational variants. Thus, S_e(Ar) and S_e(Xe) for beryllium and aluminum targets are less than for a carbon target. At the same time, the stopping powers calculated for the Xe ions neglecting the Z_1^3 effect are systematically 6% less than the experimentally measured values. The Z_1^3 correction gives numerical agreement between the experimental data and the calculations.

Figure 3 displays the ratios of the stopping powers of 780 MeV/nucleon Xe nuclei to the stopping powers of 985-MeV/nucleon Ar nuclei for various media from beryllium to bismuth. The ratio of the charges of the relativistic nuclei is 3, and in the classical approach the ratio of the stopping powers of these ions with equal velocities should be 10. Taking account of the small difference of the velocities of Ar and Xe ions gives $S_e(\text{Ar})/S_e(\text{Xe})=9.5$. The available experimental ratio $S_e(\text{Ar})/S_e(\text{Xe})$ is 10.157 ± 0.25.[17] The stopping powers of Ar and Xe ions computed taking account of the Z_1^3 effect give a different ratio, 9.9, which improves agreement with the experiment.

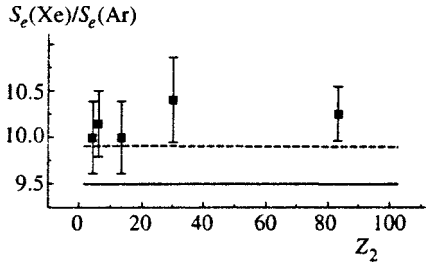

FIG. 3. Ratio of the electronic stopping power of 780-MeV/nucleon Xe ions to the electronic stopping power of 985-MeV/nucleon Ar ions versus the atomic number Z_2 of the target (experiment of Ref. 17). Lines—results of the calculation performed in the present work using the standard ZBL model on the basis of the Z_1^2 dependence (solid line) and taking account of the Z_1^3 correction (dashed line).

6. CONCLUSIONS

In summary, in the present work, a comprehensive investigation of the effect of the atomic number and energy of the projectile nucleus and the atomic number of the target nuclei on the basic characteristics of the passage of heavy nuclei with energies near 1 GeV/nucleon through homogeneous media was performed on an enlarged data sample.

The systematic deviation, observed in previous work, of the calculations based on the first Born approximation for the scattering of a particle by the atomic electrons of the medium (the Bethe–Bloch formula, the Barkas formula, the Z_1^2 scaling theory) from the experimental values of the total ionization ranges and stopping powers of nuclei has been confirmed. It increases with the atomic number of the projectile nucleus.

The effect of the second Born approximation (the Z_1^3 correction) in the ionization losses of fast multiply charged particles on the ranges and the stopping power of these particles in homogeneous media was investigated. It was shown that introducing the Z_1^3 correction in the form proposed by Jackson and McCarthy eliminates the systematic discrepancy between the computed and experimental values. For the experimental energy range the relativistic Mott scattering of particles by the atomic electrons in the target makes the decisive contribution to the observed Z_1^3 effect.

This work was supported by the Russian Fund for Fundamental Research (Project 98-02-18173).

*)E-mail: bogdanov@adron.hop.stu.neva.ru

[1] H. A. Bethe, Ann. Phys. **5**, 325 (1930).
[2] F. Bloch, Ann. Phys. **16**, 287 (1933).
[3] W. H. Barkas, P. H. Barrett, P. Cuer *et al.*, Phys. Rev. **102**, 583 (1956).
[4] W. H. Barkas, P. H. Barrett, P. Cuer *et al.*, Nuovo Cimento **8**, 185 (1958).
[5] W. H. Barkas, P. H. Barrett, P. Cuer *et al.*, Nuovo Cimento **8**, 201 (1958).
[6] W. H. Barkas, F. M. Smith, and W. Birnbaum, Phys. Rev. **98**, 605 (1955).
[7] W. H. Barkas, *Nuclear Research Emulsion* (Academic Press, New York, London, 1963, Vol. 1; 1973, Vol. 2).
[8] H. H. Heckmann, B. L. Perkins, W. G. Simon *et al.*, Phys. Rev. **117**, 544 (1960).
[9] C. J. Waddington, P. S. Freier, and D. J. Fixsen, Phys. Rev. A **28**, 464 (1983).
[10] E. M. Friedlander, H. H. Heckman, and Y. J. Karant, Phys. Rev. C **27**, 2436 (1983).
[11] C. J. Waddington, D. J. Fixseen, H. J. Crawford *et al.*, Phys. Rev. A **34**, 3700 (1986).
[12] S. D. Bogdanov, E. E. Zhurkin, V. F. Kosmach, and D. Khassan, JETP Lett. **58**, 682 (1993).
[13] S. D. Bogdanov, S. S. Bogdanov, V. E. Dudkin *et al.*, Radiat. Meas. **25**, 111 (1995).
[14] V. A. Bakaev, S. D. Bogdanov, S. S. Bogdanov *et al.*, Poverkhnost', No. 2, 101 (1997).
[15] V. A. Bakaev, S. D. Bogdanov, S. S. Bogdanov *et al.*, Poverkhnost', No. 6, 31 (1997).
[16] J. Mompart, C. Domingo, C. Baixeras *et al.*, Nucl. Instrum. Methods Phys. Res. B **107**, 56 (1996).
[17] C. Scheidenberger, H. Geissel, Th. Stohlker *et al.*, Nucl. Instrum. Methods Phys. Res. B **98**, 36 (1994).
[18] É. Skzhipchak, A. Yakholkovska, K. Karabova *et al.*, JINR Report No. R1-9364, Dubna (1975).
[19] D. Kh. Khassan, Candidate's Dissertation in Physicomathematical Sciences, St. Petersburg State Technical University, St. Petersburg (1996).
[20] A. Bonetti *et al.*, *Nuclear Emulsions* (Russian translation, Fizmatgiz, Moscow, 1961).
[21] W. H. Barkas, N. J. Dyer, and H. H. Heckman, Phys. Rev. Lett. **11**, 26 (1963).
[22] J. D. Jackson and R. L. McCarthy, Phys. Rev. B **6**, 4131 (1972).
[23] J. F. Ziegler, J. P. Biersack, and U. Littmark, *The Stopping and Range of Ions in Solid* (Pergamonn Press, New York, 1985).
[24] H. H. Andersen and J. F. Ziegler, *Hydrogen Stopping Powers and Ranges in all Elements* (Pergamon Press, New York, 1977).
[25] J. C. Ashley, V. E. Anderson, R. H. Ritchie, and W. Brandt, Z_1^3-*effect in the stopping power of matter for charged particles: Tables function* (National Auxiliary Publication Service (Doc. No. 02195), New York, 1974).
[26] J. C. Ashley, W. Brandt, and R. H. Ritchie, Phys. Rev. B **5**, 2393 (1972).
[27] J. C. Ashley, W. Brandt, and R. H. Ritchie, Phys. Rev. A **8**, 2402 (1973).
[28] R. H. Ritchie and W. Brandt, Phys. Rev. A **17**, 2102 (1978).
[29] E. E. Zhurkin, Candidate's Dissertation in Physicomathematical Science, St. Petersburg State Technical University, St. Petersburg (1995).

Translated by M. E. Alferieff

Equation for the envelope of the field in a linear array of semiconductor lasers and an analysis of nonlinear effects

D. V. Vysotskiĭ*) and A. P. Napartovich

Troitsk Institute of Innovative and Thermonuclear Research, 142092 Troitsk, Moscow Region, Russia
(Submitted 26 March 1998)
Zh. Éksp. Teor. Fiz. **115**, 416–430 (February 1999)

An equation is derived for the smooth envelope of a wave field propagating in a plane stratified medium possessing the resonance small-scattering effect on one period of the structure. Rectilinear propagation of the envelope of the field is detected in the xz plane, where x is normal to the layers. The conditions under which this equation can be used are indicated. For a linear array of semiconductor diodes with a stepwise dependence of the permittivity on the coordinates, the conditions for stable propagation of the radiation are analyzed and an explicit solution of the isolated-soliton type is found. © *1999 American Institute of Physics.* [S1063-7761(99)00402-3]

1. INTRODUCTION

An enormous number of monographs,[1,2] and other publications have been devoted to the problem of radiation propagation in stratified periodic media. Synthesized periodic structures that essentially reconstruct the spectrum of propagating waves (so-called photonics-band structures) have been actively developed in recent years. Bragg structures, used as distributed feedback[3] or naturally appearing during scattering and four-wave mixing, play a special role in laser physics. As indicated in Ref. 4, a linear array of semiconductor diodes with structural elements of a certain size can also be regarded as an analog of a Bragg structure with respect to the lateral propagation of radiation.

In ordinary distributed-feedback devices, which reflect radiation incident on them along the normal, a large reflectance builds up over a distance of many periods of the structure, $N \gg 1$. This is because the modulation depth of the permittivity is usually small, $|\delta\epsilon| \ll \epsilon_0$, where ϵ_0 is the mean value and $\delta\epsilon$ is the modulation amplitude (in general, complex) of the permittivity. $|\delta\epsilon|$ is also small in a linear diode array, but it should be comparable with the small angle of incidence of the radiation on the layer, since the wave vector on the average is directed along the axis of the diode lasers, i.e., parallel to the faces of the layers. The typical structure of a linear diode array is shown schematically in Fig. 1.

The small fraction of the radiation reflected from one distributed-feedback period makes it possible to derive the differential equations widely used in the literature for coupled waves.[5] Similar equations can be derived for linear diode arrays, which constitute a resonant structure for lateral waves.[6] The reason that differential equations can be derived is also that the reflectance of a lateral wave from one cell of the periodic structure is small. However, this smallness is ensured by a resonant transmission effect and accordingly occurs for waves whose lateral wave vector differs little from the resonance value. It is precisely such structures and waves that have turned out to be extremely promising from the viewpoint of obtaining powerful coherent radiation from a linear diode array.[7] From a more general viewpoint, one is dealing with periodic structures in which the band gap is close to zero because of the corresponding ϵ profile (in quantum mechanics, the potential) over a period.

This paper develops the approach described in Ref. 8, for which the general conditions needed to derive equations for a smooth envelope in a layered medium are formulated on the basis of the theory of second-order coupling matrices. The nonsteady-state equation that describes radiation propagation, allowing for nonlinearity of the medium, is derived for a resonant linear diode array. The stability with respect to the self-focusing of radiation is analyzed in terms of the resulting equation, and steady-state solutions that describe an analog of the self-focusing channel are constructed.

2. GENERAL CONSIDERATIONS

A linear array of diodes is a rather complex structure (see Fig. 1), in which field propagation should be described by the vector Maxwell equations. However, in most cases, the scalar Helmholtz equation for a coherent field can be used as a first approximation. Since the size of the active zone is extremely small, the field structure in this direction, called the transverse direction, stabilizes quickly. This gives a basis for averaging the Helmholtz equation along the transverse direction, as a result of which the averaged equation includes a form factor that allows for the overlap of the field with the active zone and an effective permittivity produced by averaging the product of the two-dimensional field distribution and the local ϵ:[7]

$$\frac{d^2\mathscr{U}(x)}{dx^2} + \mathscr{U}(x)(k^2\epsilon(x) - \beta^2) = 0, \tag{1}$$

where $\mathscr{U}(x)$ is the wave amplitude, x is the lateral coordinate, k is the wave vector, β is the propagation constant, and $\epsilon(x)$ is the effective permittivity of the medium, which is a periodic function of x. An equation in the form of Eq. (1)

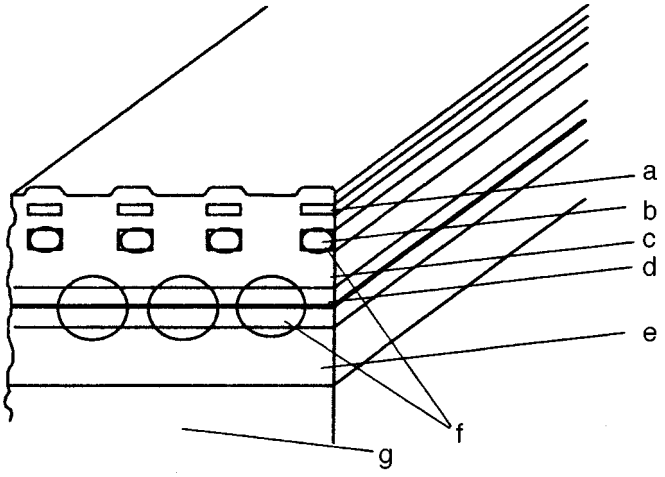

FIG. 1. Layout of a linear array of semiconductor lasers: a—p-GaAs is a supplementary absorber, b—p-$Al_{0.3}Ga_{0.7}As$ is a waveguide, c—p-$Al_{0.6}Ga_{0.4}As$, d—GRIN–SCH–SQW is an active zone, e—n-$Al_{0.6}Ga_{0.4}As$, f—radiation field, g—n^+-GaAs is the substrate.

also arises in solid-state theory and quantum mechanics. Interest in its properties has been maintained by studies on nanotechnology.

One method of solving Eq. (1) with a periodic $\epsilon(x)$ is the T-matrix method. Introducing two basis functions $f(x_m)$ and $g(x_m)$, which are linearly independent solutions of Eq. (1) on the period with number m (i.e., x_m is defined inside one cell of the periodic structure), we represent the general solution of Eq. (1) in the form

$$\mathscr{U}=a_m f(x_m)+b_m g(x_m). \tag{2}$$

The values of the coefficients a and b in the mth element are associated with the values in the $(m+1)$st element by the coupling matrix:

$$\begin{pmatrix} a_{m+1} \\ b_{m+1} \end{pmatrix} = T \begin{pmatrix} a_m \\ b_m \end{pmatrix}.$$

The elements of matrix T can be expressed in general in terms of the values of functions $f(x_m)$ and $g(x_m)$ at the cell boundaries.[10] In this case, the determinant of T equals unity.

It is well known[11] that any matrix identically satisfies its own secular equation:

$$T^2-\mathrm{Tr}\,T\cdot T+E=0, \tag{3}$$

where $\mathrm{Tr}\,T=T_{11}+T_{22}$ is the sum of the diagonal elements of the matrix. Acting with the left-hand side of the secular equation on the column of field amplitudes $\binom{a_{m-1}}{b_{m-1}}$, we get the equation that connects the amplitudes of the f waves in three adjacent cells:

$$a_{m+1}-\mathrm{Tr}\,T\cdot a_m+a_{m-1}=0. \tag{4}$$

Precisely the same equation is also valid for the g wave, and hence also for the total field amplitude $\mathscr{U}$. The necessary condition that the field amplitude changes slowly as one goes to an adjacent cell directly follows from Eq. (4). It can be written in the form

$$|\mathrm{Tr}\,T\pm 2|\ll 1. \tag{5}$$

When $\mathrm{Tr}\,T=2$ holds, Eq. (4) implies that the discrete second-order derivative vanishes. Such an equation evidently has two independent solutions $a_m=c_0$ and $a_m=c_1 m$. The second solution can obviously be regarded as smoothly varying if one is dealing with a large number of cells in the linear array, $N\gg 1$. We shall assume in what follows that this condition is satisfied. If $\mathrm{Tr}\,T=-2$ holds, the problem can be reduced to the preceding one by the replacement $\tilde{a}_m=(-1)^m a_m$. The condition given by Eq. (5) thus generalizes the evident criterion ordinarily used in the theory of distributed feedback:

$$\|T-E\|\ll 1,$$

where $\|\ \|$ signifies the norm of the matrix.

In an alternative approach to the analysis of Eq. (1), based on Bloch functions, it is well known that, in the spectrum of the waves propagating in a periodic lattice, band gaps appear whose center corresponds to the Bragg conditions. The eigenvalue λ that is found by substituting a power solution into Eq. (4) is associated with the Bloch vector Q by the relationship

$$\lambda=\exp(\pm iQ\Lambda),$$

where Λ is the period of the structure. When $\mathrm{Tr}\,T=2$, the eigenvalue $\lambda=1$ is doubly degenerate. This corresponds to the disappearance of the band gap. When inequality (5) is satisfied, the band gap is accordingly small. The conditions of ordinary distributed-feedback theory correspond to the consideration of high-lying bands whose width is exponentially small. Thus, in the language of band theory, the condition of smoothness can be formulated as the requirement that the band gaps are narrow.

Since a simple replacement reduces the case $\mathrm{Tr}\,T=-2$ to the case $\mathrm{Tr}\,T=2$, we restrict ourselves to the latter. Introducing the quantity $\Delta=\mathrm{Tr}\,T-2$, we replace the discrete equation (3) by a differential equation, which is allowable for smoothly varying functions; as noted above, this equation is valid for $\mathscr{U}(x,x_m)$, where $x=m\Lambda$ is a smooth variable:

$$\Lambda^2\frac{d^2\mathscr{U}(x,x_m)}{dx^2}=\Delta\cdot\mathscr{U}(x,x_m). \tag{6}$$

The value of Δ depends on $\omega=kc$ (c is the velocity of light) and β as on parameters. If one distinguishes values of the carrier frequency ω_0 of the radiation and the propagation constant β_0 close to their values for a pure resonance, when $\mathrm{Tr}\,T=\pm 2$, $\Omega=\omega-\omega_0$ and $\delta\beta=\beta-\beta_0$ will essentially be Fourier variables when the total field is expanded in plane waves. Expanding Δ in powers of Ω and β, we have

$$\Delta=ic_4\Omega+ic_5\delta\beta+c_2\Omega^2+c_3\Omega\cdot\delta\beta+c_1(\delta\beta)^2+c_6. \tag{7}$$

Assuming that the field variations in time and along the propagation axis z are fairly smooth for the solutions of interest to us, we can easily obtain the following equation from Eqs. (6) and (7):

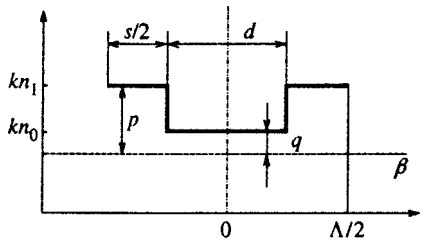

FIG. 2. Layout of one element of a linear array of semiconductor lasers in the effective-permittivity approximation.

$$\Lambda^2\frac{\partial^2\mathscr{U}}{\partial x^2}+c_1\frac{\partial^2\mathscr{U}}{\partial z^2}+c_2\frac{\partial^2\mathscr{U}}{\partial t^2}+c_3\frac{\partial^2\mathscr{U}}{\partial z\,\partial t}+c_4\frac{\partial\mathscr{U}}{\partial t}+c_5\frac{\partial\mathscr{U}}{\partial z}+c_6\mathscr{U}=0. \tag{8}$$

To find the coefficients $c_1,\ldots,c_6$, it is necessary to specify the $\epsilon(x)$ dependence in Eq. (1).

Forms of potential wells at which there is no scattering are well known from quantum mechanics.[9] In these cases, it should be expected that an anomalously narrow band gap also arises in a periodic structure constructed from layers with a corresponding profile on a period. Below we consider the simplest version of such a structure in which theoretical calculations can be completed.

3. THE EQUATION FOR A LINEAR DIODE ARRAY IN THE APPROXIMATION OF A STEPPED $\epsilon(x)$

In the theory of linear diode arrays $\epsilon(x)$ is usually chosen to be a step function with two values, ϵ_0 in the active element and ϵ_1 in the passive gap. We restrict ourselves below to a linear array (see Fig. 2) in which amplification is localized in an antiguide core of width d, whereas absorption can be introduced in a zone with a high refractive index and width s (zone a in Fig. 1). Moreover, we include the nonlinear part of the complex permittivity, which we write, following Ref. 12, in the form

$$\epsilon_n(x)=\Gamma(x)\left.\frac{d\epsilon}{dN}\right|_{N=N_0}(N(x)-N_0),$$

where $N(x)$ is the carrier concentration, N_0 is the same in the absence of a radiation field, and $\Gamma(x)$ is a factor that allows for the overlap of the wave field with the active region in the transverse direction.

The choice of basis functions f and g and the derivation of expressions for the elements of matrix T that satisfy natural conditions on the side faces of the semiconductor are described in Ref. 10. A cell of the periodic structure is shown in Fig. 2. The function $f(x_m)$ corresponds to a wave transmitted through the antiguide core barrier when it is incident from the left, whereas $g(x_m)$ corresponds to a wave incident on the antiguide core from the right. In a resonant structure in which $(l+1)$ half-waves fit inside the antiguide core and m half-waves fit in the waveguide gap (l and m are whole numbers), in the absence of amplification and absorption the waves f and g pass through the cell with no losses, being only multiplied by the number $(-1)^\sigma$, where $\sigma=l+m+1$. In this case, there is no wave coupling. In actuality, such a situation occurs only for a wave with a strictly determined frequency and propagation constant $\beta=\beta_0$. The waves rescatter into each other both when the cell parameters deviate from the resonant values (in this case, the waves transmitted without scattering in general disappear), and for a resonant structure for waves that differ from the resonance values. If it is recalled that there is always absorption or amplification in actual structures, it can be understood that pure resonance is virtually never observed. However, as shown numerically,[13] and later analytically,[10] structures and modes that can be called near-resonant present the greatest interest for generation in a linear array. In this case, the trace of matrix T can be represented approximately in the form[10]

$$\mathrm{Tr}\,T=(-1)^\sigma(2-F^2), \tag{9}$$

where

$$F^2=\left(p_1s+\frac{p_0}{q_0}q_1d\right)\left(p_1s+\frac{q_0}{p_0}q_1d\right). \tag{10}$$

Here q_1 and p_1 are defined as the differences of the complex-valued lateral wave vectors within the antiguide core and the interelement gap from their resonant values, respectively, i.e.,

$$q_1=q-q_0,\quad p_1=p-p_0;$$

$$q_0=\pi(l+1)/d,\quad p_0=m\pi/s.$$

The magnitudes of the vectors p and q are determined in this case from

$$q^2=k^2\epsilon_0-\beta^2,\quad p^2=k^2\epsilon_1-\beta^2$$

(see also Fig. 2). For typical parameters of a linear laser diode array,

$$|\epsilon_1-\epsilon_0|\ll\mathrm{Re}\,\epsilon_0;$$

i.e., the permittivity step is much less than the magnitude of the permittivity. Since the conditions $\lambda\ll s$ and $\lambda\ll d$ are satisfied when the wavelength of the radiation in vacuum is λ, the lateral angles at which the radiation propagates are small. Consequently, $\beta\approx kn_0$ holds to within small corrections (here n_0 is the mean refractive index). These corrections play a decisive role in calculating the lateral wave vectors.

For even σ, it follows from Eq. (9) that the quantity introduced above is $\Delta=-F^2$. In order to find the coefficients in Eqs. (7) and (8), p_1 and q_1 must be expressed in terms of $\delta\beta$ and Ω. This is easy to do, using the explicit expressions given above for p and q:

$$\begin{cases}p_1p_0=k\Omega\epsilon_1/c+k^2\delta\epsilon_s-\beta\delta\beta,\\ q_1q_0=k\Omega\epsilon_0/c+k^2\delta\epsilon_d-\beta\delta\beta,\end{cases} \tag{11}$$

where $2\,\delta\epsilon_s$ and $2\,\delta\epsilon_d$ are the detunings of the permittivity in the waveguide and antiguide core zones, respectively. Note that these deviations in general consist of two parts each: one occurs because the actual parameters of the structure deviate from the strictly resonant values, and the second part is caused by the variation of the nonlinear component of the

permittivity. The possibility of operating with a piecewise-constant nonlinear addition to ϵ is not evident in general and requires special consideration. Recalling that the original one-dimensional model is an approximation, as well as the role of carrier diffusion in semiconductors, it can be assumed that replacing the actual distribution of the nonlinear component of ϵ, averaged, as noted above, over the transverse coordinate, with a stepwise value can satisfactorily describe the nonlinear effects in a linear array in our case. In this case, the total field intensity, averaged over the active (antiguide core) and passive zones, respectively, appears in the theory in a natural way. Using Eqs. (10) and (11), we write $\Delta=-F^2$ in the form of Eq. (7), whence for the coefficients $c_1,\ldots,c_6$ we get

$$c_1=-\beta^2L^4,\quad c_2=-k^2\epsilon^2L^4/c^2,$$
$$c_3=-2\beta k\epsilon L^4/c,\quad c_4=ik^3\epsilon[\delta\epsilon' L^4+L_\epsilon^4]/c,$$
$$c_5=i\beta k^2[\delta\epsilon' L^4+L_\epsilon^4],\quad c_6=k^4\delta\epsilon' L_\epsilon^4, \tag{12}$$

where we have introduced the mean permittivity

$$\epsilon=(\epsilon_1 s+\epsilon_0 d)/\Lambda,$$

the mean detuning

$$\delta\epsilon'=(s\,\delta\epsilon_s+d\,\delta\epsilon_d)/\Lambda,$$

and the characteristic lengths

$$L^4=\Lambda(s/p_0^2+d/q_0^2),\quad L_\epsilon^4=\Lambda(s\,\delta\epsilon_s/p_0^2+d\,\delta\epsilon_d/q_0^2).$$

These expressions for the coefficients of Eq. (8) are some of the most important results of this paper; to illustrate the possibilities of using them, we will solve specific problems in what follows. Note that in Eq. (6) the coordinate x_m inside the cell is a free parameter. The dependence on the internal coordinate enters into the total field via functions $f(x_m)$ and $g(x_m)$, which also, strictly speaking, depend on the Fourier variables over which the expansion in Eq. (7) is carried out. However, a more careful analysis shows that the distortion of the basis functions in the space–time variations of the field can be neglected in calculating the total field, since their contribution is smaller by a factor of N than for the coefficients a_m and b_m. Analogously, provided one remains close to resonance, as given by inequality (5), the variations of the refraction indices should also be taken into account only when solving Eq. (8).

We should point out a number of limiting cases in which Eq. (8) is easy to interpret. In particular, for a resonant structure in which amplification, absorption, and nonlinearity are neglected, $\delta\epsilon_s=\delta\epsilon_d=0$, we have

$$c_4=c_5=c_6=0.$$

Then, for steady-state propagation of radiation along z, we get a hyperbolic equation, describing propagation along straight lines at an angle of

$$\frac{\Lambda}{\sqrt{-c_1}}=\frac{\lambda}{2n_0\Lambda}\left[\frac{d^3}{\Lambda^3(l+1)^2}+\frac{s^3}{\Lambda^3m^2}\right]^{-1/2}.$$

This unexpected outcome is confirmed by experiment[14] from the amplification of the radiation in an array of twenty-one antiguide core elements. The figures in Ref. 14 show clearly that the boundary of the beam propagating inside the set is a straight line. An estimate gives an angle close to that written above. Note that when radiation initially fills part of the cross section of the cavity homogeneously along the z axis the entire cross section also becomes filled by a wave in the lateral direction with constant velocity, determined by the formula $\Lambda/\sqrt{-c_2}$. This also strongly differs from the diffusion-like spreading usual for a plane-parallel cavity.

Finally, when the structure differs appreciably from a resonant structure, so that a special role is played in the propagation of radiation by the first-order derivatives with respect to z or t, Eq. (8) approaches in its properties the ordinary parabolic equation of optics.

Since the actual boundary or initial conditions are often such that, close by (or initially), the field cannot be represented in the form of Eq. (3) with a_m and b_m smoothly varying from element to element, the natural evolution of the field in space and in time has the effect in this case that, after some distance (time), Eq. (8) becomes usable for the subsequent description. However, the question of the boundary or initial conditions requires special treatment in this case.

Below we shall consider problems for an infinite set of diodes, which corresponds to the approximation of small scales of structures that in our case embrace a rather large number of elements, although less than the total number N of elements.

4. SELF-FOCUSING OF RADIATION WITH A PLANE WAVEFRONT

Let us consider the steady-state propagation of radiation in an infinite array of semiconductor lasers. In this case we have $c_2=c_3=c_4=0$. A solution in the form of a plane wavefront for the envelope of the field $\mathscr{E}(x,z)$,

$$\mathscr{E}=E_0\exp(i\,\delta\beta z), \tag{13}$$

corresponds to $\Delta=0$ (or $F=0$), whence we get two allowable values for $\delta\beta$:

$$\delta\beta_1=\frac{k\,\delta\epsilon'}{n_0}, \tag{14}$$

$$\delta\beta_2=\frac{k}{n_0}\frac{L_\epsilon^4}{L^4}, \tag{15}$$

each of which makes one of the parentheses in Eq. (10) go to zero, and here the propagation constant is taken to be approximately $\beta\approx kn_0$. The intensity dependence in Eqs. (14) and (15) implicitly enters the detunings $\delta\epsilon_s$ and $\delta\epsilon_d$ via the carrier concentration. It is easy to verify by direct substitution that $\delta\beta_1$ corresponds to a field distribution of

$$\mathscr{E}_1=f(x_m)+g(x_m),$$

whereas the solution with propagation constant $\delta\beta_2$ has the form

$$\mathscr{E}_2=f(x_m)-g(x_m).$$

The field structure for each of the indicated waves is illustrated in Fig. 3 for a grating with parameters $l=0$ and

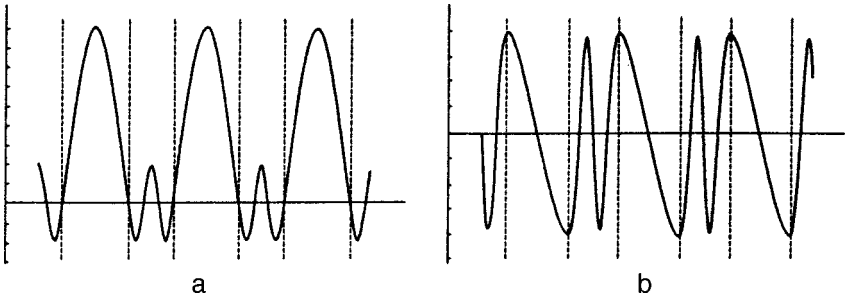

FIG. 3. Mode structure for $l=0$, $m=3$: a—mode $f+g$, corresponding to $\delta\beta_1$, b—mode $f-g$, corresponding to $\delta\beta_2$.

$m=3$. Note that these solutions were obtained earlier in Ref. 4 for a purely resonance grating and in the absence of nonlinearity. The second of the solutions, $\mathscr{U}_2$, is evidently unsuitable from the viewpoint of laser action, since the field is determined in equal degree by the active and passive zones. Moreover, for a finite number of elements, losses to the radiation through the edge will be large for field propagation of type $\mathscr{U}_2$.

Let us apply the standard analysis to the linear stability. We introduce the field perturbation $\tilde{E}(x,z)$, so that

$$\mathscr{U}=(E_0+\tilde{E})\exp(i\delta\beta z).$$

The linearized equation for $\tilde{E}$ is written in the form

$$\Lambda^2\frac{\partial^2\tilde{E}}{\partial x^2}+c_1\left(\frac{\partial^2\tilde{E}}{\partial z^2}+2i\delta\beta\frac{\partial\tilde{E}}{\partial z}\right)+c_5\frac{\partial\tilde{E}}{\partial z}+(\tilde{E}+\tilde{E}^*)\times\left(i\delta\beta\frac{\partial c_5}{\partial I}+\frac{\partial c_6}{\partial I}\right)E_0^2=0, \tag{16}$$

where the derivatives of the coefficients of Eq. (8) and $\delta\beta$ are taken for the mean intensity $I_0=|E_0|^2$ in an element. Expanding the perturbation in plane waves,

$$\tilde{E}=\tilde{a}(\kappa,\gamma)\exp(i\kappa x+\gamma z)+\tilde{b}(\kappa,\gamma)\exp(-i\kappa x+\gamma^* z),$$

we get the dispersion relation

$$\gamma^2|A|^2+2\gamma(\mathrm{Re}(AB^*)-Q^2\mathrm{Re}\,A)+Q^2(Q^2-2\,\mathrm{Re}\,B)=0, \tag{17}$$

where

$$A=c_5+2i\delta\beta c_1,\quad B=i\delta\beta I_0\partial c_5/\partial I+I_0\partial c_6/\partial I,$$
$$Q^2=\kappa^2\Lambda^2-c_1\gamma^2.$$

The values of the coefficients A and B for which, for real values of κ, values of γ exist in which the real part is positive correspond to a region of instability of the propagation of a plane wavefront (self-focusing). In accordance with the Rouse–Hurwits criterion, Eq. (17) has no roots with a positive real part (i.e., the solution is stable) if the following system of inequalities is satisfied:

$$\mathrm{Re}\,A<0, \tag{18}$$

$$|A|^2-2c_1\kappa^2\Lambda^2+2c_1\mathrm{Re}\,B>0, \tag{19}$$

$$\mathrm{Re}\,AB^*-\kappa^2\Lambda^2\mathrm{Re}\,A>0, \tag{20}$$

$$\kappa^2(\kappa^2\Lambda^2-2\mathrm{Re}\,B)>0, \tag{21}$$

$$\kappa^2\Lambda^2|A|^2(\mathrm{Re}\,A)^2+c_1(\mathrm{Re}\,AB^*)^2>\mathrm{Re}\,AB^*\mathrm{Re}\,A(|A|^2+2c_1\mathrm{Re}\,B). \tag{22}$$

It follows from inequality (18) that a solution with $\mathrm{Re}\,A>0$ is unstable for any perturbations, whereas, for $\mathrm{Re}\,A<0$, the perturbations increase only for certain values of κ, for which at least one of the inequalities (19)–(22) breaks down. Moreover, when inequality (21) is satisfied, recalling that $c_1<0$, inequality (19) follows, so that it is possible to restrict the discussion to the inequalities (20)–(22).

In practice, one tries to localize the field mainly in the antiguide core zone; also, to reduce the divergence of the radiation in the far field, it is desirable that the size of the radiative zone be close to the period of the structure. Both these requirements are satisfied in the limit $|p_0/q_0|\gg 1$. The coefficient $\mathrm{Re}\,A$ can be written in this case as

$$\mathrm{Re}\,A=\mp\frac{\beta k^2 sd}{q_0^2}(\mathrm{Im}\,\delta\epsilon_d-\mathrm{Im}\,\delta\epsilon_s),$$

where the upper sign refers to $\delta\beta$ determined from Eq. (14), and the lower sign to that from Eq. (15). Since the amplification is concentrated in the antiguide cores in the type of linear diode array considered here, while losses are sometimes specially introduced into the waveguide gap, we find that the solution with $\delta\beta=\delta\beta_2$ is unstable, and therefore it is assumed in the subsequent analysis that $\delta\beta=\delta\beta_1$.

If detuning $\delta\epsilon=\delta\epsilon_d-\delta\epsilon_s$ of the permittivity step is introduced at the boundary of the antiguide core, the coefficient $\mathrm{Re}\,AB^*$ can be written as

$$\mathrm{Re}\,AB^*=-\frac{k^6\beta|\delta\epsilon|^2 s^2 d^2 I_0}{q_0^4}\mathrm{Im}\left(\frac{d\,\delta\epsilon'}{dI}\right),$$

so that if $\mathrm{Re}\,AB^*>0$ holds, as in the case that is important in practice, when the sign on the derivative is determined by the decrease of the amplification as the intensity increases the inequality (20) is satisfied for any κ. Then self-focusing occurs when at least one of the following conditions is satisfied:

$$\kappa^2<K_0 \tag{23}$$

or

$$\kappa^2<(K_0-K_1)Gd/s, \tag{24}$$

where the following notation has been introduced:

$$K_0=\frac{2k^4sdI_0}{q_0^2\Lambda^2}\,\mathrm{Re}\left(\delta\epsilon\frac{d\delta\epsilon'}{dI}\right),$$

$$K_1=\frac{k^4sd^2|\delta\epsilon|^2}{q^2\Lambda^3}\left(G+\frac{s}{d}\right),$$

$$G=\Lambda I_0\,\mathrm{Im}\left(\frac{d\delta\epsilon'}{dI}\right)/d\,\mathrm{Im}\,\delta\epsilon.$$

For further simplification, it is necessary to specify the form of the intensity dependences of $\delta\epsilon_s$ and $\delta\epsilon_d$. Since the overlap of the radiation field with the amplification zone is weak in the interelement space, we neglect the amplification in this zone. The permittivity detuning in the interelement gap can also be considered independent of intensity, assuming $\delta\epsilon_s=\Delta+i\alpha_T$, where Δ is the real detuning from the resonance value, and α_T is the additional interelement loss. The permittivity detuning in the antiguide core is determined as[12]

$$\delta\epsilon_d=\epsilon_{20}\mathscr{P}(\alpha-i)/(1+I/I_s),$$

where

$$\mathscr{P}=(J\tau_c/ed-N_0)/N_0$$

is a dimensionless parameter that allows for amplification and the difference of $N(x)$ from N_0 as a consequence of pumping, J is the injection current density, τ_c is the spontaneous lifetime, e is the charge of the electron, α is a linewidth enhancement factor, I_s is the saturation intensity, and $\epsilon_{20}=\Gamma\sigma_g n_0/k$ (σ_g is the amplification cross section, and Γ is a coefficient that allows for the overlap of the field of the mode with the active zone in the transverse direction).

As follows from inequalities (23) and (24), self-focusing is possible when at least one of the conditions $K_0>K_1$ or $K_0>0$ is satisfied.

Under the assumptions made here, the first condition in the absence of losses in the interelement gap ($\alpha_T=0$) reduces to a quadratic inequality for $\tau=\Delta(1+I_0/I_S)/(\epsilon_{20}\mathscr{P})-\alpha$:

$$\tau^2(H+1)-2\tau+H-1<0,$$

where

$$H=(1+I_S/I_0)s/d.$$

This inequality has a nonempty set of values only when $H<\sqrt{2}$ holds, i.e., for fairly large intensities:

$$I_0>I_S/(\sqrt{2}d/s-1).$$

For smaller intensities, only the inequality (23) remains.

The condition $K_0>0$ can be reduced to the form $\mathrm{Re}[(\alpha-i)\delta\epsilon]<0$ or, in more detail,

$$\alpha\Delta+\alpha_T-\frac{\alpha^2-1}{1+I_0/I_s}\epsilon_{20}\mathscr{P}>0. \tag{25}$$

Thus, when there is no modulation of the losses and the radiation intensity is not too large, self-focusing occurs when the detuning of the permittivity step is above the resonance value ($\mathrm{Re}\,\delta\epsilon>0$), whereas self-defocusing occurs when it is below that value. This agrees with the results of the numerical calculations of Ref. 12.

For the condition $|\gamma|\ll\delta\beta$ [which corresponds to predominance of the parabolic term in the original Eq. (8)], we can set $Q=\kappa\Lambda$ in Eq. (17), reducing it to square in γ. For large positive detunings, $\Delta\gg|\epsilon_2|$, Eq. (17) can be rewritten as

$$\tilde{\gamma}^2+2\tilde{\gamma}\left(Q_0^2+Q^2\frac{\epsilon_{20}\mathscr{P}}{\Delta(1+I/I_s)}\right)+Q^2(Q^2-2\alpha Q_0^2)=0, \tag{26}$$

where we have introduced the notation

$$\tilde{\gamma}=\gamma|A|,\quad Q_0^2=k^4sd^2\epsilon_{20}\Delta\mathscr{P}/q^2I_s\Lambda(1+I_0/I_s)^2.$$

We find from Eq. (26) that $\tilde{\gamma}$ reaches its maximum value when $Q^2\approx\alpha Q_0^2$, so that the maximum growth rate of the perturbations,

$$\gamma_{\max}=\frac{k^2dI_0}{\beta\Lambda I_s}\frac{\epsilon_{20}\mathscr{P}}{(1+I_0/I_s)^2}((\alpha^2+1)^{1/2}-1),$$

is reached when the characteristic transverse size of a perturbation is

$$L_{\mathrm{char}}=\frac{\Lambda\lambda^2}{2d^2}\frac{1+I_0/I_s}{(\alpha\epsilon_{20}\Delta\mathscr{P}sI_0/\Lambda I_s)^{1/2}}.$$

The condition that makes it possible to reduce Eq. (17) to a square can be written as

$$\Delta\gg\frac{((\alpha^2+1)^{1/2}-1)\epsilon_{20}\mathscr{P}I_0}{(1+I_0/I_s)^2I_s}, \tag{27}$$

while the condition that the field vary little over one period, $|Q|\ll1$, implies the inequality

$$\frac{(\alpha\epsilon_{20}\Delta\mathscr{P}sI_0/\Lambda I_s)^{1/2}}{1+I_0/I_s}\ll\frac{\lambda^2}{2d^2}. \tag{28}$$

When the pump power is close to the lasing threshold ($\mathscr{P}\sim1$), inequalities (27) and (28) are simultaneously satisfied for the entire range of Δ values that are small in comparison with the permittivity step at the boundary of the antiguide core.

5. SOLITON SOLUTIONS IN A LINEAR ARRAY OF DIODE LASERS

Reference 15 used numerical calculations to show that, besides solutions in the form of a plane wavefront, solutions of the isolated wave or soliton type are also possible in a linear array of diode lasers. Such a soliton is small by comparison with the size of the entire linear array, but must be large by comparison with the size of one element. Below we construct such a solution in the model of a smooth envelope of the field for the condition that the pump power weakly exceeds the threshold.

We shall use a model of a laser cavity in the form of a medium homogeneous in the longitudinal direction, with distributed losses responsible for the output of radiation from

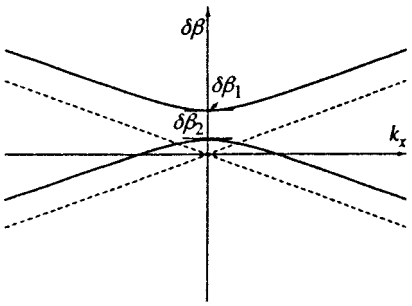

FIG. 4. Formation of a band when a linear array of lasers is detuned from resonance: the dashed lines are the dispersion curves in strict resonance, and the solid curves are the dispersion curves for positive detuning of the permittivity step from resonance.

the cavity.[16] Moreover, the longitudinal coordinate and time appear in Eq. (8) in a symmetric way, so that Eq. (8) in general reduces to an equation in two variables. Specifically, if a field-independent replacement of the coordinates

$$\tilde{x}=L^2k^2x/\Lambda \quad \text{and} \quad \tilde{t}=kct/\epsilon+kz/n_0$$

is made, Eq. (8) can be written as

$$\frac{\partial^2\mathscr{U}}{\partial\tilde{x}^2}-\frac{\partial^2\mathscr{U}}{\partial\tilde{t}^2}+i\left(\delta\epsilon'+\frac{L_\epsilon^4}{L^4}\right)\frac{\partial\mathscr{U}}{\partial\tilde{t}}+\frac{L_\epsilon^4}{L^4}\delta\epsilon'\,\mathscr{U}=0. \tag{29}$$

The dispersion equation corresponding to the plane waves of the envelope, as can easily be seen from Eq. (29), is a second-order algebraic equation. Neglecting the imaginary parts of the permittivity, the dispersion equation can be illustrated on plane wave vectors $\delta\beta$ and k_x (see Fig. 4). The values of $\delta\beta_{1,2}$ in Fig. 4 are determined by Eqs. (14) and (15). The dashed lines show the dispersion curves for strict resonance, where (0,0) is the point of degeneracy. For nonzero detuning (which can be caused by nonlinear field dependence of $\delta\epsilon_s$ and $\delta\epsilon_d$), the pair of straight lines transforms into two branches of a hyperbola. A band gap arises in this case. Thus, for a linear array with parameters different from the resonance values, there is an interval of constant propagation $\delta\beta_1<\delta\beta<\delta\beta_2$ (or $\delta\beta_2<\delta\beta<\delta\beta_1$), where plane waves are forbidden. If we now construct a solution, taking $\delta\beta$ from the indicated interval for the intensity of the light to be equal to zero, the parameters can be chosen so that, as the intensity increases, the original dispersion curve is transformed so that $\delta\beta$ goes beyond the limits of the band gap (we recall that $\delta\beta_1$ and $\delta\beta_2$ depend on the intensity of the light). In such a situation, Eq. (29) obviously must allow a soliton solution.

The translationally invariant solution of Eq. (29) is written as

$$\mathscr{U}=\exp(-i\theta\tilde{t})\,\hat{\mathscr{U}}(\tilde{x}-v\tilde{t}),$$

where θ is a free parameter, and v is the velocity of the structure (θ and v are real). Equation (29) can then be rewritten as

$$\frac{\partial^2\hat{\mathscr{U}}}{\partial\xi^2}(1-v^2)+\left(\delta\epsilon'+\frac{L_\epsilon^4}{L^4}-2\theta\right)iv\frac{\partial\hat{\mathscr{U}}}{\partial\xi}+\left(\frac{L_\epsilon^4}{L^4}-\theta\right)\times(\delta\epsilon'-\theta)\hat{\mathscr{U}}=0, \tag{30}$$

where $\xi=\tilde{x}-v\tilde{t}$. Equation (30), with intensity-dependent coefficients for the first derivative and the free term, was investigated as a consequence of the generalized Ginzburg–Landau equation in Ref. 17. It was also shown there that it allows as solutions both plane wavefronts [which correspond to stationary points in the phase space of Eq. (29)] and spatially limited structures (which correspond to the separatrices that connect the stationary points; see, for example, Ref. 18 for an application to optical systems), such as wavefronts, solitons, sources, and sinks. At the same time, it is very difficult to analyze Eq. (30) even for the simplest versions of the dependence of the amplification on the field intensity, because all the quantities in the equation are complex.

With pumping close to threshold, the imaginary part of the permittivity detuning in the antiguide core is small by comparison with the real part, and the intensity is small by comparison with the saturation intensity. This makes it possible to represent the detuning in the antiguide core as

$$\delta\epsilon_d=\delta\epsilon_0-\epsilon_2|\mathscr{U}|^2/I_S,$$

where $\delta\epsilon_0$ and ϵ_2 are real parameters and I_S is the saturation intensity. Then, in the approximation of strong confinement of the field in the antiguide core, $|p_0|\gg|q_0|$, and in the absence of modulation of the losses, the field-dependent coefficients in Eq. (30) can be written as

$$\delta\epsilon'-\theta=\frac{s}{\Lambda}\delta\epsilon-\theta_1-\epsilon_2\frac{d}{\Lambda}\frac{|\mathscr{U}|^2}{I_S},$$

$$\frac{L_\epsilon^4}{L^4}-\theta=-\theta_1-\epsilon_2\frac{|\mathscr{U}|^2}{I_S},$$

where $\theta_1=\theta-\delta\epsilon_0$ is a renormalized free parameter and $\delta\epsilon$ is the detuning of the permittivity step between the waveguide and the antiguide core zones. In this case, a solution exists in the form of a stationary ($v=0$) soliton having the form

$$\hat{\mathscr{U}}=\left(\frac{\theta_1\delta\epsilon I_S}{\epsilon_2[(d+\Lambda)\theta_1/s-\delta\epsilon]}\right)^{1/2}\cosh^{-1}\left(\sqrt{\theta_1\delta\epsilon\frac{s}{\Lambda}}\xi\right). \tag{31}$$

It is obvious from Eq. (31) that the conditions for the existence of such a soliton are $\theta_1\delta\epsilon>0$, which means that there is no radiation propagation when $I=0$, and $\epsilon_2((d+\Lambda)\theta_1/s-\delta\epsilon)>0$, which means that transmission is no longer forbidden when the intensity is greater than a certain threshold value. This confirms the qualitative considerations expressed above.

Figure 5 illustrates the overall structure of the soliton. There are small discontinuities at the boundaries of the elements of the periodic structure because the envelope of the soliton actually covers about three elements. This means that the transition to a continuous variable that we used in deriv-

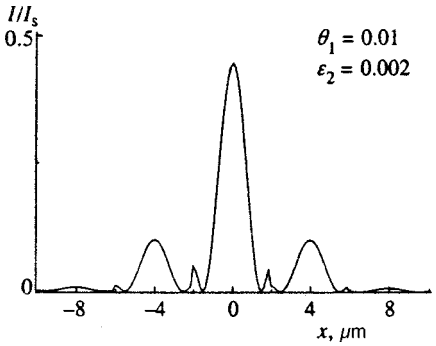

FIG. 5. Intensity profile of a fundamental soliton in a linear array of diode lasers, normalized to the saturation intensity, when the permittivity step is detuned from resonance by $\delta\epsilon = 0.003$.

ing the main Eq. (8) introduces an appreciable error. On the whole, however, it is evident that Fig. 5 correctly reflects the shape of the soliton.

6. CONCLUSION

An equation has thus been derived in this paper for the smooth envelope of the wave field in a plane stratified medium characterized by weak scattering over one period, caused by resonance effects. It has been found that, starting from the elliptical type of Helmholtz equation in the variables (x,z), the equation for the envelopes can be transformed into a hyperbolic equation that describes the field propagation in a stratified medium at some angle. The conditions have been specified under which the concept of a smooth envelope is valid and under which this equation can be used. For a linear array in which the permittivity has a stepwise dependence on the coordinates, the conditions for stable propagation of radiation with a plane wavefront have been analyzed, and an explicit nonlinear isolated soliton solution has been found.

This work was carried out with the financial support of the Russian Fund for Fundamental Research (Grant No. 96-02-19203).

*)E-mail: dima@fly.triniti.troitsk.ru

[1] L. A. Brekhovskikh, *Waves in Layered Media*, Academic Press, New York (1980) [Russian original, Nauka, Moscow (1973)].
[2] P. Yeh, *Optical Waves in Layered Media*, Wiley, New York (1988).
[3] N. N. Il'ichev, L. A. Malyutin, P. P. Pashinin *et al.*, Pis'ma Zh. Tekh. Fiz. **8**, 460 (1982) [Sov. Tech. Phys. Lett. **8**, 200 (1982)].
[4] C. A. Zmudzinski, D. Botez, and L. J. Mawst, Appl. Phys. Lett. **60**, 1049 (1992).
[5] H. Kogelnik and C. V. Shank, J. Appl. Phys. **43**, 2327 (1972).
[6] A. P. Napartovich and D. Botez, Proc. SPIE **2994**, 600 (1997).
[7] D. Botez, in *Diode Laser Arrays*, D. Botez and D. R. Scifres (Eds.), Cambridge University Press (1994).
[8] V. V. Likhanskii and A. P. Napartovich, Proc. SPIE **1840**, 256 (1991).
[9] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics–Non-Relativistic Theory*, vol. 3 of *Course of Theoretical Physics*, Pergamon Press, New York (1977); Fizmatgiz, Moscow (1963).
[10] D. Botez and A. P. Napartovich, IEEE J. Quantum Electron. **30**, 975 (1994).
[11] F. R. Gantmakher, *The Theory of Matrices*, Chelsea, New York (1959) [Russian original Gl. Red. Fiz. Mat. Lit., Moscow (1975)].
[12] R. F. Nabiev and D. Botez, IEEE J. Sel. Top. Quantum Electron. **1**, 138 (1995).
[13] D. Botez, L. J. Mawst, G. L. Peterson, and T. J. Roth, IEEE J. Quantum Electron. **26**, 482 (1990).
[14] C. Zmudzinski, D. Botez, L. J. Mawst *et al.*, IEEE J. Sel. Top. Quantum Electron. **1**, 129 (1995).
[15] R. F. Nabiev, P. Yeh, and D. Botez, Opt. Lett. **18**, 1612 (1993).
[16] Ya. I. Khanin, *The Dynamics of Quantum Generators*, Sov. Radio, Moscow (1966).
[17] W. van Saarlos and P. C. Hohenberg, Physica D **56**, 303 (1992).
[18] N. N. Rozanov, *Optical Bistability and Hysteresis in Distributed Nonlinear Systems*, Nauka, Moscow (1997).

Translated by W. J. Manthey

Quasi-collinear and partially degenerate four-wave mixing: an alternative explanation of the phase-conjugation property of backward stimulated scattering

Dun Liu

Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, Michigan 48109, USA

Guang S. He*)

Photonics Research Laboratory, State University of New York at Buffalo, Buffalo, New York 14260-3000, USA

(Submitted 9 September 1997)
Zh. Éksp. Teor. Fiz. **115**, 431–448 (February 1999)

A quasi-collinear and partially degenerate four-wave mixing model is proposed to explain the optical phase-conjugation property of various types of stimulated backscattering. According to this model, after passing through a phase-disturbed medium or an aberration plate, the input pump beam can be resolved into two portions: a stronger undisturbed regular portion and a weaker phase-disturbed irregular portion. These two portions interfere with each other and create a volume holographic grating in the pumped region of the scattering medium. Only the stronger undisturbed portion of the pump field can excite an initial backward stimulated scattering beam with a regular wavefront. When the latter (as a reading beam) passes through the induced holographic grating, a diffracted wave will be created and then amplified together with the reading beam. A rigorous mathematical analysis shows that under certain conditions the combination of these two portions (the reading wave and the diffracted wave) of the backward stimulated scattering can be an approximate phase-conjugate field of the input pump field. The major theoretical conclusions are basically supported by the experimental results based on a specially designed two-beam interference setup. © *1999 American Institute of Physics.* [S1063-7761(99)00502-8]

1. INTRODUCTION

Optical phase-conjugation is one of the most interesting research subjects in nonlinear optics.[1,2] So far, there are two major technical approaches to generate optical phase-conjugate waves: one is based on degenerate or partially degenerate four-wave mixing,[3–8] the other is based on backward stimulated Brillouin, Raman, or Rayleigh-wing scattering.[9–14] In addition, for certain arrangements, special three-wave mixing,[15,16] the photon echo technique,[17,18] and the self-pumped photorefractive effect can also be used to produce phase-conjugate waves.[19–21]

Now backward stimulated scattering (BSS) is one of the most sophisticated techniques to generate optical phase-conjugate waves. A considerable number of theoretical papers on this specific issue have been published since the late 1970s.[22–31] Most of those theoretical studies have been based on a particular assumption that there is gain discrimination between the phase-conjugate portion and non-phase-conjugate portions of the backward stimulated scattering; only the former can obtain the maximum gain and be effectively amplified. However, for a long time, a clear theoretical model or physical explanation to support this assumption was lacking. Therefore, a better physical understanding of this effect is still necessary, as indicated by Ref. 32. Another feature of most theoretical papers cited here is that both the pump field and the BSS field were represented by a function expressed as an infinite series.[9,22,24–31] In those cases, it was extremely difficult to obtain a rigorous analytical solution of the wave equation.

In this paper we intend to propose an alternative physical model as well as a novel mathematical approach to explain the phase-conjugation property of BSS. The suggested model is based on a quasi-collinear and partially-degenerate four-wave mixing process, or equivalently, a quasi-collinear holographic wavefront-reconstruction process. The proposed mathematical approach is based on the assumption that both the pump field and the BSS field can be viewed as composed of two portions: a portion of the regular-wave and a portion of the aberration-wave. The advantage of the suggested approach is that an explicit analytic solution of the wave equations can be obtained.

2. QUASI-COLLINEAR HOLOGRAPHIC INTERACTION MODEL

Before starting our theoretical discussions, it is helpful to describe briefly the typical experimental setup for demonstrating the phase-conjugation property of the BSS from a given medium. Figure 1a shows the typical experimental setup without using an aberration plate. In this case, a quasi-plane pump beam is focused into the center of a scattering medium. Most experimental observations show that the BSS exhibits nearly the same beam size and beam divergence as

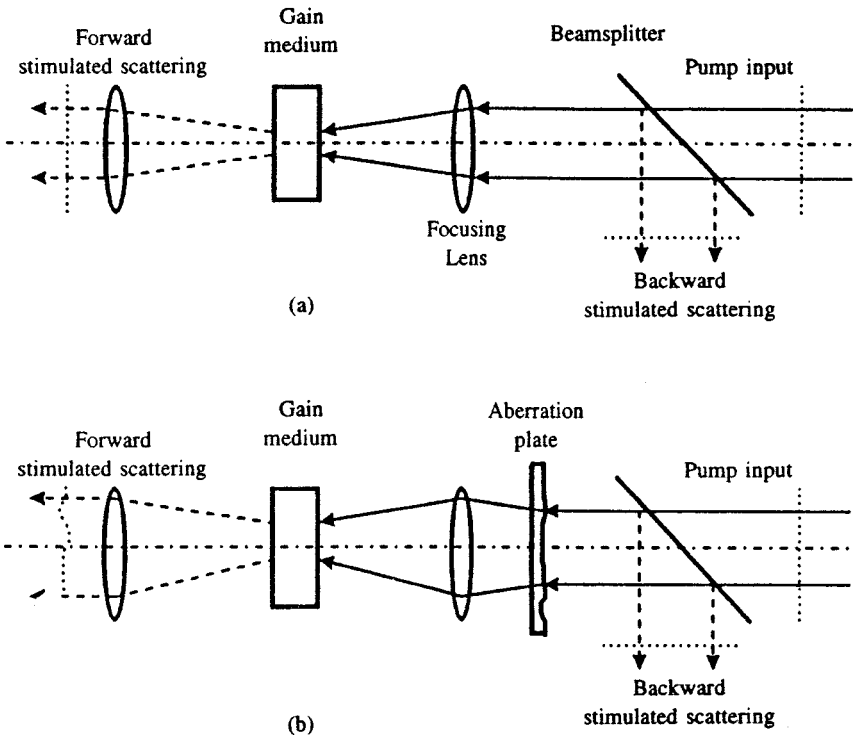

FIG. 1. Experimental setups for observation of a phase-conjugate backward stimulated scattering: (a) without using an aberration plate and (b) with an aberration plate.

does the pump beam. These facts can be explained by the following two considerations. First, only those portions of the initial BSS which are propagating within the solid aperture angle (measured from the sample center to the focusing lens) of the pump beam can get the maximum gain length. Second, the divergence angle of the collimated BSS is determined by its spot size in the pumped region of the gain medium, which is limited by the spot size of the pump beam. We can assume that near the focal point region the focused pump beam exhibits a Gaussian transverse intensity distribution. Considering the threshold requirement of the burst of stimulated scattering, the minimum spot size (beam waist) of the stimulated scattering should be slightly smaller or quite closer to that of the pump beam. Therefore, after passing back through the focusing lens the BSS output manifests a slightly smaller or nearly the same beam divergence angle as the pump beam.

Figure 1b shows the same experimental setup except that a transparent aberration plate is placed between the beamsplitter and the focusing lens. In this case, after passing back through the aberration plate the BSS may show a significantly reduced aberration influence. That is the typical experimental evidence of the phase-conjugation property of the BSS.[10–14]

To describe how can we employ the quasi-collinear holographic interaction model to explain the basic experimental results mentioned above, it is worth returning to the original idea of Gabor's holography principle. In that case, a coherent light wave passing through a transparent object (phase object), the object is assumed to be such that a considerable part of the wave penetrates undisturbed through it, and a hologram is formed by the interference between the secondary wave arising from the presence of the object and the strong background wave, as clearly described in Ref. 33. According to this principle, after passing through a phase object the total optical field can be expressed as a superposition of two portions:[33]

$$U = U^{(i)} + U^{(s)} = A^{(i)} \exp(i\varphi_i) + A^{(s)} \exp(i\varphi_s)$$
$$= \exp(i\varphi_i)\{A^{(i)} + A^{(s)} \exp[i(\varphi_s - \varphi_i)]\}. \quad (1)$$

Here $U^{(i)}$ is the undisturbed part of the transmitted field, $U^{(s)}$ the secondary wave arising from the presence of the object; $A^{(i)}$ and $A^{(s)}$ are their amplitude functions; and φ_i and φ_s are the corresponding phase functions, respectively.

The Gabor principle described above is applicable to most phase-conjugation experiments based on BSS. In this case, as shown schematically in Fig. 2, $E(\omega)$ is a quasiplane pump wave; after passing through an aberration plate, the pump field appears as a superposition of two portions: a stronger undisturbed wave $E_1(\omega)$ and a weaker distorted wave $E_2(\omega)$. After passing through a focusing lens, these two portions interfere with each other in the focal region inside a gain medium and create an induced volume holographic grating that is due to the intensity-dependent refractive index change of the gain medium. Only the $E_1(\omega)$ wave is strong enough to fulfill the threshold requirement and to generate an initial BSS wave $E_3(\omega')$, the latter exhibits the same regular wavefront as does the former. When the $E_3(\omega')$ wave passes backward through the induced holographic grating region, a diffracted wave $E_4(\omega')$ is created. Here we see a typical holographic wavefront-reconstruction process: the induced grating is formed by the regular $E_1(\omega)$ wave (reference beam) and the irregular $E_2(\omega)$ wave (signal beam); the initial backward stimulated scattering $E_3(\omega')$ wave is a reading beam with a regular wavefront like the $E_1(\omega)$; therefore, the diffracted wave $E_4(\omega')$, as a recon-

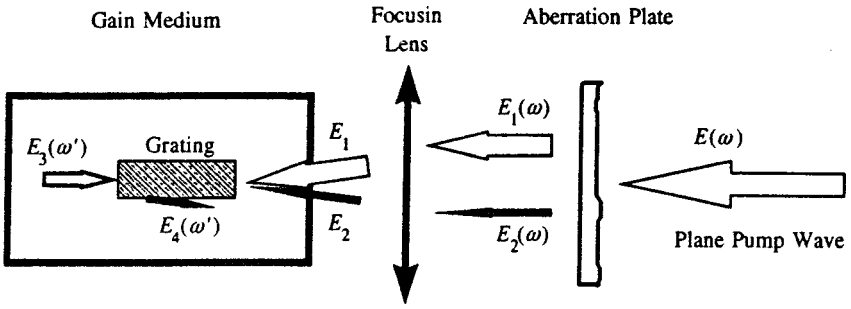

FIG. 2. Schematic illustration of Gabor's holographic model for the phase-conjugation formation of backward stimulated scattering.

structive beam, will be the phase-conjugate replica of the $E_2(\omega)$ wave.

Furthermore, the $E_4(\omega')$ wave will experience a further amplification together with the $E_3(\omega')$ wave because both waves have the same signal frequency. In the case of stimulated Brillouin scattering, $\omega \approx \omega'$, it is a nearly degenerate quasi-collinear four-wave mixing (FWM) process in the sense of phase-conjugate formation. In the case of stimulated Raman scattering, $\omega > \omega'$, there is a partially degenerate and frequency down-converted FWM process. In the case of anti-Stokes stimulated scattering, $\omega < \omega'$, there is a partially degenerate and frequency up-converted FWM process. Based on the explanations described above, one can see that there is a common mechanism (pump field-induced holographic grating) playing the same key role for phase-conjugation formation in both FWM and BSS. This common mechanism is applicable to any types of backward stimulated scattering processes including stimulated Brillouin, Raman, Rayleigh-wing,[34,35] and Kerr scattering,[36–38] even though the specific scattering mechanisms are totally different among them.

3. PUMP FIELD-INDUCED HOLOGRAPHIC GRATING AND STIMULATED SCATTERING GAIN

Let us consider an isotropic scattering medium. We assume that ω and ω' are the frequencies of pump wave and stimulated scattering wave, both of which are linearly polarized along the x axis. The induced refractive index change experienced by the stimulated scattering wave can be expressed as

$$\Delta n(\omega') = \frac{1}{2n_0(\omega')} \chi_e^{(3)}(-\omega';\omega',-\omega,\omega)|E(\omega)|^2, \quad (2)$$

where $n_0(\omega')$ is the linear refractive index, $\chi_e^{(3)} = \chi_{xxxx}^{(3)}$ is a real matrix element of the third-order nonlinear susceptibility tensor, and $E(\omega)$ is the electric field function of the incident pump beam. In the focal region inside the gain medium, the values of local intensity of the pump field, $I(x,y,z,\omega) \propto |E(x,y,z,\omega)|^2$, can be quite high with a spatial intensity fluctuation that is due to the interference between the two portions (disturbed and undisturbed) of the pump beam. As a result, an intensity-dependent holographic grating can be formed via the mechanism described by Eq. (2).

If the local pump intensity is high enough, stimulated scattering and subsequent stimulated amplification may occur in the focal region of the pump beam inside the medium. As in the case of one-photon pumped lasing, in steady-state and small-signal approximation the growth of an initial stimulated scattering signal can be described as

$$I(l,\omega') = I(0,\omega')e^{(G-\alpha)l}, \quad (3)$$

where ω' is the frequency of the stimulated Stokes scattering, $I(l=0,\omega')$ is the initial intensity, α is the linear attenuation coefficient, G is the exponential gain coefficient (in units of cm^{-1}), and l is the effective gain length (in units of cm) of the medium. The threshold requirement for the burst of an observable stimulated scattering can be written as

$$e^{(G-\alpha)l} \gg 1. \quad (4)$$

The exponential gain coefficient G is assumed to be proportional to the local intensity of the pump field, i.e.,

$$G(z) = gI_0(z,\omega). \quad (5)$$

Here the pump intensity $I_0(z,\omega)$ is in units of MW/cm^2, and g is the exponential gain factor in units of cm/MW.

4. GENERALIZED DEFINITIONS OF AN OPTICAL PHASE-CONJUGATE WAVE

The term ''optical phase-conjugation'' is specially used to described the wavefront reversal property of a backward propagating optical wave with respect to a forward propagating optical wave. Suppose there is an input quasi-monochromatic wave with a certain phase-distortion departing from an ideal plane wavefront, i.e.,

$$E(z,x,y,\omega) = E(z,x,y)e^{-i\omega t} = A_0(z,x,y)e^{i[kz+\varphi(z,x,y)]}e^{-i\omega t}. \quad (6)$$

Here, z is the longitudinal variable along the propagation direction, x and y are the transverse variables along the bean section, ω is the frequency of the field, $k = 2\pi n_0/\lambda$ is the magnitude of the wave vector; $E(z,x,y)$ is the complex amplitude function, $A_0(z,x,y)$ is the real amplitude function, and, $\varphi(z,x,y,k)$ is a phase-distortion function. If there is a backward propagating wave which can be expressed as

$$E'(z,x,y,\omega) = aE^*(z,x,y)e^{-i\omega t} = aA_0(z,x,y)e^{-i[kz+\varphi(z,x,y)]}e^{-i\omega t}, \quad (7)$$

where a is any real constant, the field $E'(z,x,y,\omega)$ is called the frequency-degenerate phase-conjugate wave of the input field $E(z,x,y,\omega)$. This type of optical phase-conjugate wave can be experimentally generated by using the well-known degenerate four-wave mixing technique.[3–7]

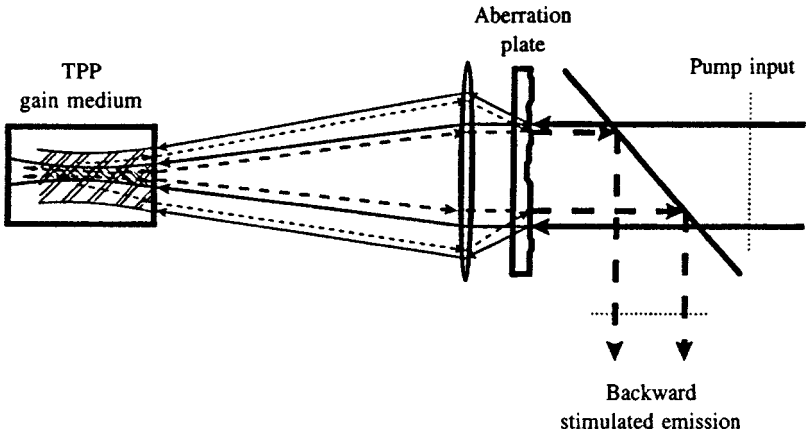

FIG. 3. Schematic illustration of the optical-path geometry for the formation of a phase-conjugate backward stimulated scattering.

In a more general case, if there is a backward propagating optical field with a different frequency ω', which can be written as

$$E''(z,x,y,\omega')=aA_0(z,x,y)e^{-i[k'z+\varphi(z,x,y)]}e^{-i\omega' t}, \quad (8)$$

then $E''(z,x,y,\omega')$ can be termed the frequency-nondegenerate phase-conjugate wave of the same original field $E(z,x,y,\omega)$. This type of phase-conjugate wave can be experimentally generated by using (i) various BSS with a considerable frequency-shift,[12,13] or (ii) a partially degenerate FWM.[8,39–42]

5. DESCRIPTION OF A FOCUSED INPUT PUMP FIELD

A detailed schematic illustration for the beam-path geometry of the BSS in a scattering medium is shown in Fig. 3. In this case, a quasi-parallel pump beam (thick lines) passes through an aberration plate and is focused into the center of the medium. After passing through the aberration plate, the pump beam can be imagined as composed of two portions. One is a relatively strong undisturbed quasi-plane wave (medium-thick lines); the other is a relatively weak distorted wave (thin lines). These two portions of the input pump beam can interfere with each other in the overlap region and create a volume holographic grating. For the undistorted portion of the input pump beam, the optical field in the focal region can be approximately treated as the fundamental mode of a Hermite-Gaussian beam, i.e.,[25,43]

$$E_1(z,x,y,\omega)=C_1\frac{w_0}{w(z)}\exp\left\{-(x^2+y^2)\times\left[\frac{1}{w^2(z)}-\frac{ik}{2R(z)}\right]-i\tan^{-1}\left(\frac{z}{\delta z}\right)\right\}\times\exp[i(kz-\omega t)], \quad (9)$$

where C_1 is a real amplitude constant, and the parameters w, R and δz are determined by

$$w^2(z)=w_0^2\left[1+\left(\frac{z}{\delta z}\right)^2\right],\quad R(z)=z\left[1+\left(\frac{z}{\delta z}\right)^2\right],\quad \delta z=\frac{kw_0^2}{2}. \quad (10)$$

Here, $w(z)$ represents the change in the pump-beam size along the z axis near the focal point region, w_0 is the minimum spot size of the beam at the focal plane, δz is the focal depth of the focused beam, and $R(z)$ describes the change of the radius of curvature as a function of z. In writing Eq. (9), we simply neglect the amplitude depletion of the pump field within a distance comparable to δz.

According to Gabor's principle and Eq. (1), the total pump field near the focal point region can be expressed as a combination of two portions:

$$\begin{aligned}E(z,x,y,\omega)&=E_1(z,x,y,\omega)+E_2(z,x,y,\omega)\\&=[A_1(z,x,y)+A_2(z,x,y)]\exp[i(kz-\omega t)]\\&=\left\{C_1\frac{w_0}{w(z)}\exp\left(-(x^2+y^2)\left[\frac{1}{w^2(z)}-\frac{ik}{2R(z)}\right]-i\tan^{-1}\frac{z}{\delta z}\right)+C_2\exp[i\theta(z,x,y)]\right\}\\&\quad\times\exp[i(kz-\omega t)]. \quad (11)\end{aligned}$$

Here C_1 and C_2 are real amplitude constants for normalization, and $\theta(z,x,y)$ is an unknown function depending on the aberration plate. The first term with C_1 in the braces of Eq. (11) represents the undistorted portion of the pump field; the second term with C_2 represents the distorted portion of the pump field. It is assumed that the latter amplitude is distributed nearly uniformly near the focal point with a cross section considerably larger than w_0. As shown in Fig. 3, this assumption is based on the fact that the aberration plate causes a wide and diffuse halo in the focal plane compared to the small focal spot of the undistorted wave. Nevertheless, apart from the focal plane these two portions overlap considerably with each other and can induce a holographic grating.

6. FORMATION OF PHASE-CONJUGATE WAVE BY BACKWARD STIMULATED SCATTERING

As shown in Fig. 3, the two portions of the input pump beam can interfere with each other in the overlapping region and create a volume holographic grating. However, on the other hand, since the focused undisturbed portion of the pump beam has a much smaller focal spot size and much higher local intensity than that of the distorted portion, an

initial BSS (medium-thick dashed lines) with a regular wavefront can be excited only by the stronger undisturbed pump field. In propagating backward this initial BSS experiences further amplification, and creates a diffracted wave through the induced holographic grating. This diffracted portion (thin dashed lines) can be viewed as a secondary seeded signal and therefore will undergo further amplification together with the initial BSS signal. The remaining issue is whether the combination of these two portions of the backward emission can be a phase-conjugate replica of the combination of the two portions of the input pump field.

6.1. Description of the BSS field: the $E_1'(\omega')$ wave

According to the proposed model, it is assumed that the backward stimulated emission field also consists of two portions and can be expressed as

$$\begin{aligned} E'(z,x,y,\omega') &= E_1'(z,x,y,\omega') + E_2'(z,x,y,\omega') \\ &= [A_1'(z,x,y) + A_2'(z,x,y)] \\ &\quad \times \exp[i(-k'z - \omega' t)], \end{aligned} \tag{12}$$

where $A_1'(z,x,y)$ is the complex amplitude function of the initial BSS wave and $A_2'(z,x,y)$ is the complex amplitude function of the diffracted wave created by the former through the induced holographic grating.

According to the physical model described in Sec. 2, the A_1' field is generated by the A_1 field through the BSS process. Therefore, the electric polarization component corresponding to A_1' field can be formally written as (in SI units)

$$\begin{aligned} P_1'(\omega') &= i\varepsilon_0 \chi_e' |A_1|^2 A_1' \exp[i(-k'z - \omega' t)] \\ &= i\varepsilon_0 \chi_e' \frac{C_1^2 w_0^2}{w^2(z)} \exp\left[-\frac{2(x^2+y^2)}{w^2(z)} \right] A_1' \\ &\quad \times \exp[i(-k'z - \omega' t)], \end{aligned} \tag{13}$$

where ε_0 is the permittivity of vacuum and χ_e' is a phenomenologically introduced effective third-order susceptibility coefficient (a real coefficient) that is employed to describe the gain behavior of the BSS process. The above expression is based on the assumption that the initial BSS experiences an exponential gain, and the exponential gain coefficient is proportional to the intensity of the A_1 field [see Eq. (5)]. The wave equation of E_1' can be written as

$$\nabla^2 E_1' - \frac{n_0^2(\omega')}{c^2} \frac{\partial^2 E_1'}{\partial t^2} = \mu_0 \frac{\partial^2 P_1'(\omega')}{\partial t^2}, \tag{14}$$

where μ_0 is the permeability in vacuum, c the speed of light in vacuum, and $n_0(\omega')$ the linear refractive index at the frequency of ω'. In the slowly-varying-amplitude approximation, the solution of Eq. (14) obeys the following equation:

$$\begin{aligned} \frac{\partial A_1'}{\partial z} + \frac{i}{2k'} \nabla_\perp^2 A_1' &= \frac{g' C_1^2}{2[1 + (z/\delta z)^2]} \\ &\quad \times \exp\left[-\frac{2(x^2+y^2)}{w^2(z)} \right] A_1', \end{aligned} \tag{15}$$

where $g' = k' \chi_e'$ is the exponential gain factor of the stimulated emission for a given medium. In order to solve this equation, we could solve for the Fourier transform of A_1', then obtain the solution of A_1' through the reverse Fourier transform.[25] Omitting the intermediate mathematical procedure, the final solution of A_1' is given by

$$\begin{aligned} A_1' = C_1' \frac{w_0'}{w'(z)} \exp\left[\frac{1}{2} g' C_1^2 (z - z_0) \right] \exp\Bigg\{ &-(x^2+y^2) \\ &\times \left[\frac{1}{w'^2(z)} + \frac{ik'}{2R'(z)} \right] + i \tan^{-1} \frac{z}{\delta z'} \Bigg\}. \end{aligned} \tag{16}$$

Here C_1' is a real amplitude constant, the first exponential term represents the amplitude gain effect, the second exponential term describes the behavior of the transverse intensity and wavefront curvature, and z_0 is the starting position of the initial backward stimulated emission. The other new parameters in the above equation are defined as

$$\begin{aligned} &\delta z' = \eta \delta z, \quad w_0'^2 = (k/k')\, \eta w_0^2, \\ &w'^2(z) = w_0'^2 \left[1 + \left(\frac{z}{\delta z'} \right)^2 \right], \\ &R'(z) = z \left[1 + \left(\frac{\delta z'}{z} \right)^2 \right], \end{aligned} \tag{17}$$

and

$$\eta = \frac{1}{3} g' C_1^2 \delta z \left[\left(\frac{z}{\delta z} \right)^3 - \left(\frac{z_0}{\delta z} \right)^3 \right]. \tag{18}$$

In the condition

$$\eta = \frac{1}{3} g' C_1^2 \delta z \left[\left(\frac{z}{\delta z} \right)^3 - \left(\frac{z_0}{\delta z} \right)^3 \right] \approx 1, \tag{19}$$

we have $\delta z' = \delta z$, $w_0'^2 \approx w_0^2$, $w'^2(z) \approx w^2(z)$ and $R'(z) = R(z)$. Comparing Eq. (16) to Eq. (11), one can see that only in that case can the E_1' field be approximately the phase-conjugate replica of the input E_1 field. Later, we will discuss the real experimental conditions that fulfill the above requirement and will also consider the influence of the difference between k' and k.

Now let us consider the gain behavior of the backward stimulated emission described by the first exponential term in Eq. (16). We have already chosen $z = 0$ at the focal point position in the sample center, and we assume that the initial backward stimulated emission starts roughly from the position $-z_0 \approx -\delta z/2$. If the optical path length of the gain medium is much longer than δz, the effective single-path amplitude gain can be written as

$$\Delta(d) \approx \exp\left(\frac{1}{2} g' C_1^2 \delta z \right). \tag{20}$$

If the sample thickness is much shorter than the focal depth, i.e., $d \ll \delta z$, then we have

$$\Delta(d) \approx \exp\left(\frac{1}{2} g' C_1^2 d \right). \tag{21}$$

This indicates that the BSS grows exponentially along the whole optical path length within the medium.

6.2. Description of the BSS field: the E_2' wave

Next we shall consider the diffracted E_2' wave created by the E_1' wave through the induced holographic grating near the focal region. Like a partly degenerate FWM process, the nonlinear polarization component P_2', which corresponds to the E_2' wave, can be written as[44,45]

$$P_2'(\omega')=i\varepsilon_0\chi_e''A_1A_2^*A_1'\exp[i(-k'z-\omega' t)], \tag{22}$$

where χ_e'' is a nominally introduced effective third-order susceptibility (a real coefficient) that is used to characterize the gain behavior of FWM or a grating diffraction process. Assuming that the condition of Eq. (19) is satisfied and the E_1' field is phase-conjugated with the E_1 field, then, based on Eqs. (11), (16), and (19), we can obtain an explicit expression for P_2':

$$\begin{aligned}P_2'(\omega')=i\varepsilon_0\chi_e''\frac{C_1C_2C_1'w_0^2}{w^2}\exp\left[-\frac{2(x^2+y^2)}{w^2(z)}\right]\\ \times\exp(-i\theta)\exp\left[\frac{1}{2}g'C_1^2(z-z_0)\right]\\ \times\exp[i(-k'z-\omega' t)].\end{aligned} \tag{23}$$

Substituting Eq. (23) into a nonlinear wave equation like Eq. (14), we find that the complex A_2' function is governed by the following equation:

$$\begin{aligned}\frac{\partial A_2'}{\partial z}+\frac{i}{2k'}\nabla_\perp^2A_2'=\frac{g''}{2}\frac{C_1C_2C_1'}{1+(z/\delta z)^2}\\ \times\exp\left[-\frac{2(x^2+y^2)}{w^2(z)}\right]\exp(-i\theta)\\ \times\exp\left[\frac{1}{2}g'C_1^2(z-z_0)\right],\end{aligned} \tag{24}$$

where $g''=k'\chi_e''$. Since A_2' does not appear on the right-hand side of Eq. (24), we can give a trial solution of A_2' as

$$A_2'=\frac{C_2C_1'}{C_1}\exp\left[-\frac{2(x^2+y^2)}{w^2(z)}\right]A_2'', \tag{25}$$

where A_2'' is an unknown function to be solved. Substituting Eq. (25) into Eq. (24) leads to

$$\begin{aligned}&\frac{\partial A_2''}{\partial z}-i\frac{2}{\delta z[1+(z/\delta z)^2]}\left(x\frac{\partial A_2''}{\partial z}+y\frac{\partial A_2''}{\partial y}\right)\\ &\quad+\frac{i}{2k'}\left(\frac{\partial^2A_2''}{\partial x^2}+\frac{\partial^2A_2''}{\partial y^2}\right)-\frac{i}{\delta z[1+(z/\delta z)^2]}\\ &\quad\times\left[1-\frac{4(x^2+y^2)}{w^2(z)}\right]A_2''+\frac{4(x^2+y^2)}{\delta zw^2(z)}A_2''\\ &=\frac{1}{2}g''C_1^2\frac{1}{1+(z/\delta z)^2}\exp\left[\frac{1}{2}g'C_1^2(z-z_0)\right]\exp(-i\theta).\end{aligned} \tag{26}$$

This is a rather complicated partial differential equation; however, it can be greatly simplified with the following considerations. First, since there is an exponential term on the right-hand side of Eq. (26), we may expect that the solution of A_2'' should also involve a corresponding exponential gain term. As a result, the condition $\partial A_2''/\partial z\gg A_2''/\delta z$ should be fulfilled; therefore, compared to the first term, the fourth and the fifth terms on the left-hand side of Eq. (26) can be neglected. Second, the focal depth is much larger that the beam size, i.e., $\delta z\gg x,y$; the longitudinal variation of the beam intensity is faster than the transverse variation, i.e., $\partial A_2''/\partial z>\partial A_2''/\partial x$, $\partial A_2''/\partial z>\partial A_2''/\partial y$. Therefore, the second term on the left-hand side of Eq. (26) can also be neglected. Then Eq. (26) can be finally simplified as

$$\begin{aligned}\frac{\partial A_2''}{\partial z}+\frac{i}{2k'}\nabla_\perp^2A_2''=\frac{g''C_1^2}{2[1+(z/\delta z)^2]}\\ \times\exp\left[\frac{1}{2}g'C_1^2(z-z_0)\right]\exp(-i\theta).\end{aligned} \tag{27}$$

To solve this equation we can further assume a trial solution,

$$A_2''(z,x,y)=A_2'''(z)\exp[-i\theta'(z,x,y)]. \tag{28}$$

Here the $A_2'''(z)$ term represents the real amplitude as a function of z, and the exponential term represents the phase front as a function of z, x, and y. Substituting Eq. (28) into Eq. (27), we obtain the following pair of equations:

$$\begin{aligned}&\frac{\partial A_2'''}{\partial z}+\frac{A_2'''}{2k'}\nabla_\perp^2\theta'=\frac{g''C_1^2}{2[1+(z/\delta z)^2]}\\ &\qquad\times\exp\left[\frac{1}{2}g'C_1^2(z-z_0)\right]\cos\delta\theta,\\ &\frac{\partial\theta'}{\partial z}+\frac{1}{2k'}\left[\left(\frac{\partial\theta'}{\partial x}\right)^2+\left(\frac{\partial\theta'}{\partial y}\right)^2\right]\\ &\quad=\frac{g''C_1^2}{2[1+(z/\delta z)^2]}\exp\left[\frac{1}{2}g'C_1^2(z-z_0)\right]\frac{\sin\delta\theta}{A_2'''},\end{aligned} \tag{29}$$

where

$$\delta\theta=\theta-\theta'.$$

In the small aberration approximation, the second-order spatial derivative or the square of the first-order spatial derivative of the function θ' can be neglected, then Eqs. (29) can be simplified as

$$\begin{aligned}\frac{\partial A_2'''}{\partial z}=\frac{g''C_1^2}{2[1+(z/\delta z)^2]}\exp\left[\frac{1}{2}g'C_1^2(z-z_0)\right]\cos\delta\theta,\\ \frac{\partial\theta'}{\partial z}=\frac{g''C_1^2}{2[1+(z/\delta z)^2]}\exp\left[\frac{1}{2}g'C_1^2(z-z_0)\right]\frac{\sin\delta\theta}{A_2'''}.\end{aligned} \tag{30}$$

As mentioned in Sec. 5, the distorted portion of the input pump field has a uniform amplitude distribution near the focal point region [see Eq. (11)]. This assumption is based on the fact that the beam waist of this portion is considerably larger than that of the undistorted portion of the pump field (see Fig. 3). In Fig. 3 one can also see that for the distorted portion of the pump field, the variation of the shape of the

beam waist along the z axis is negligible within the focal depth range. Hence we can further assume that the longitudinal variation of the wavefront of the distorted portion of the pump field within the focal depth range can be neglected, i.e.,

$$\partial\theta/\partial z \approx 0. \tag{31}$$

Subtracting the second equation of Eq. (30) from Eq. (31), we obtain a new pair of coupled equations as

$$\frac{\partial A_2'''}{\partial z} = \frac{g'' C_1^2}{2[1+(z/\delta z)^2]} \exp\left[\frac{1}{2} g' C_1^2 (z-z_0)\right] \cos\delta\theta,$$

$$\frac{\partial(\delta\theta)}{\partial z} = -\frac{g'' C_1^2}{2[1+(z/\delta z)^2]} \exp\left[\frac{1}{2} g' C_1^2 (z-z_0)\right] \frac{\sin\delta\theta}{A_2'''}. \tag{32}$$

Dividing the first equation by the second in Eqs. (32), we find

$$\frac{\partial A_2'''}{\partial(\delta\theta)} = -A_2''' \frac{\cos\delta\theta}{\sin\delta\theta}. \tag{33}$$

This implies that

$$A_2'''(z)\sin\delta\theta(z) = A_2'''(-z_0)\sin\delta\theta(-z_0) = B, \tag{34}$$

where B is a constant considerably smaller than A_2'''. Substituting $\cos\delta\theta = \sqrt{(A_2''')^2 - B^2}/A_2'''$ into the first equation of Eqs. (32), we have

$$\frac{A_2'''}{\sqrt{(A_2''')^2 - B^2}} \frac{\partial A_2'''}{\partial z} = \frac{g'' C_1^2}{2[1+(z/\delta z)^2]} \times \exp\left[\frac{1}{2} g' C_1^2 (z-z_0)\right]. \tag{35}$$

Since $B^2 \ll (A_2''')^2$, Eq. (35) can be approximately rewritten as

$$\frac{\partial A_2'''}{\partial z} = \frac{g'' C_1^2}{2[1+(z/\delta z)^2]} \exp\left[\frac{1}{2} g' C_1^2 (z-z_0)\right]. \tag{36}$$

Now let us consider the physical meaning of the factors $g' C_1^2$ and $g'' C_1^2$. The former is used to describe the gain of the A_1' field due to stimulated scattering amplification, the latter is nominally employed to describe the growth of the A_2' field through the holographic grating diffraction (or the equivalent four-wave mixing). It is important to point out that the initial A_2' field can be viewed as a secondary seed signal, which will experience an additional gain from the stimulated scattering amplification like that of the A_1' field. Under most BSS experimental conditions, the subsequent gain of the A_2' field comes mainly from the stimulated amplification rather than the equivalent FWM or the grating diffraction. Hence, in Eq. (36) we can replace the $g'' C_1^2$ term by the $g' C_1^2$ term, and rewrite Eq. (36) as

$$\frac{\partial A_2'''}{\partial z} = \frac{g' C_1^2}{2[1+(z/\delta z)^2]} \exp\left[\frac{1}{2} g' C_1^2 (z-z_0)\right]. \tag{37}$$

The final solution of this equation is

$$A_2''' = \exp\left[\frac{1}{2} g' C_1^2 (z-z_0)\right]. \tag{38}$$

To obtain this solution, we have assumed that $1/\sqrt{1+(z/\delta z)^2} \approx 1$, which is valid for $z < \delta z$. From Eq. (38) one can find the initial value of A_2''' at $z = -z_0$

$$A_2'''(-z_0) = 1. \tag{39}$$

Next, the remaining issue is to consider the phase function $\theta'(z,x,y)$ of the A_2'' field expressed by Eq. (28). From Eq. (34) we have

$$\sin\delta\theta(z) = \sin\delta\theta(-z_0) \frac{A_2'''(-z_0)}{A_2'''(z)}. \tag{40}$$

Substituting Eqs. (38) and (39) into Eq. (40) leads to

$$\sin\delta\theta(z) = \sin[\theta(z) - \theta'(z)] = \sin\delta\theta(-z_0) \times \exp\left[-\frac{1}{2} g' C_1^2 (z-z_0)\right]. \tag{41}$$

From Eq. (41) we see that during backward propagation the phase front of the A_2'' field is getting closer and closer to the phase front of the A_2 field. Specifically, if the stimulated gain is high enough, i.e.,

$$\frac{1}{2} g' C_1^2 (z-z_0) \gg 1, \tag{42}$$

we find

$$\sin[\theta(z) - \theta'(z)] \to 0, \quad \theta'(z) \to \theta(z). \tag{43}$$

Based on Eqs. (25), (28), (38) and (43), the diffracted portion of the backward stimulated scattering field can be finally obtained as

$$A_2' = \frac{C_2 C_1'}{C_1} \exp\left[\frac{1}{2} g' C_1^2 (z-z_0)\right] \times \exp\left[-\frac{2(x^2+y^2)}{w^2(z)}\right] \exp(-i\theta). \tag{44}$$

6.3. The total BSS field: the $E_1' + E_2'$ wave

Based on Eqs. (12), (16) and (44), the total BSS field can finally be written as

$$\begin{aligned} E'(z,x,y,\omega) &= [A_1'(z,x,y) + A_2'(z,x,y)] \exp[i(-k'z - \omega' t)] \\ &= \frac{C_1'}{C_1} \exp\left[\frac{1}{2} g' C_1^2 (z-z_0)\right] \left\{ C_1 \frac{w_0}{w(z)} \exp\left(-(x^2 + y^2)\left[\frac{1}{w^2(z)} + \frac{ik'}{2R(z)}\right] + i \tan^{-1}\frac{z}{\delta z}\right) \right. \\ &\quad \left. + C_2 \exp\left[-\frac{2(x^2+y^2)}{w^2(z)}\right] \exp(-i\theta) \right\} \\ &\quad \times \exp[i(-k'z - \omega' t)]. \end{aligned} \tag{45}$$

The above expression is obtained assumming that the requirements described by Eqs. (19) and (42) are fulfilled. Here we can combine these two requirements as

$$\eta = \frac{1}{3} g' C_1^2 \delta z \left[\left(\frac{z}{\delta z}\right)^3 - \left(\frac{z_0}{\delta z}\right)^3\right] \approx 1,$$

$$\frac{1}{2}g'C_1^2(z-z_0)\gg 1. \tag{46}$$

On the other hand, the total input pump field is given by Eq. (11) and can be rewritten as

$$\begin{aligned} E(z,x,y,\omega)&=[A_1(z,x,y)+A_2(z,x,y)]\exp[i(kz-\omega t)]\\ &=\left\{C_1\frac{w_0}{w(z)}\exp\left(-(x^2+y^2)\left[\frac{1}{w^2(z)}\right.\right.\right.\\ &\quad\left.\left.\left.-\frac{ik}{2R(z)}\right]-i\tan^{-1}\frac{z}{\delta z}\right)\right.\\ &\quad\left.+C_2\exp(i\theta)\right\}\exp[i(kz-\omega t)]. \end{aligned} \tag{47}$$

Comparing Eq. (45) to Eq. (47), if we neglect the difference between k and k' and only consider the central portion of the field within the region of

$$[2(x^2+y^2)/w^2(z)]<1, \tag{48}$$

we have

$$[A_1'(z,x,y)+A_2'(z,x,y)]\propto[A_1(z,x,y)+A_2(z,x,y)]^*. \tag{49}$$

Based on the above relationship, one can conclude that the total BSS field can be approximately phase-conjugate to the total input pump field provided that certain conditions hold.

7. DISCUSSIONS

The conclusion that a BSS field can be an approximate phase-conjugate of the input pump field holds under certain conditions. First, the gain requirements expressed by Eq. (46) should be fulfilled. Assuming that the effective gain length of the nonlinear medium is nearly determined by the focal depth of the focused pump beam, e.g., $z=(0.4–0.5)\delta z$ and $z_0=-(0.4–0.5)\delta z$, according to the first requirement expressed by Eq. (46)

$$g'C_1^2\delta z\approx 12–23, \tag{50}$$

the required gain of the intensity of a small BSS signal should be

$$\begin{aligned} \exp[g'C_1^2(z-z_0)]&\approx\exp(g'C_1^2\delta z)\approx\exp(12–23)\\ &\approx 1\times10^5–1\times10^{10}. \end{aligned} \tag{51}$$

In this case, the second requirement of Eq. (46) is automatically fulfilled. The high gain requirement can be readily fulfilled in most experimental conditions for observing the backward stimulated scattering without using any optical feedback devices. For example, the values of the exponential gain factor g_B of stimulated Brillouin scattering for the common transparent solvents (such as acetone, benzene, and toluene) are estimated to be 0.01–0.03 cm/MW.[32,44] Assuming the typical pump intensity is $I_0\simeq 500\,\mathrm{MW/cm^2}$ and the effective gain length (focal depth) is $\delta z\simeq 1.5\,\mathrm{cm}$, the gain should be

$$g_BI_0\delta z\approx 8–22. \tag{52}$$

Compared Eq. (52) to Eq. (51) we see that the requirements for observing the phase-conjugate property of a BSS wave can be basically fulfilled by common experimental conditions.

The phase-conjugation property of the BSS from a gain medium, in general, is not perfect for the following reasons. First, all the mathematical derivations described above are based on the small-aberration approximation, so that a substantial part of the distorted pump wave may overlap the undisturbed pump wave in the focal region to generate a holographic grating. Second, only the major (central) part of the BSS may exhibit high-fidelity phase-conjugation, as indicated by the requirement of Eq. (48). One can expect that under a larger aberration influence, the fidelity of phase-conjugation behavior should become poorer.

So far the difference between k and k' has been ignored, which has a certain influence on the fidelity of the phase-conjugation of a BSS wave. This influence should be essentially the same as when we create a hologram by using two beams of wavelength λ and then read this hologram by using another beam of wavelength λ'. Nevertheless, comparing Eq. (45) to Eq. (47) one can see that the difference between k and k' has no influence on the relationship $\theta'(z)=\theta(z)$, which is the most essential requirement for distortion compensation. But the difference between k and k' does affect the radius of curvature of the undistorted part of a BSS field and causes an apparent displacement of the focal point of the BSS wave.

It should also be noted that the holographic model employed in this work is qualitatively compatible with the existing theoretical explanation for phase-conjugate formation of backward stimulated Brillouin scattering, which is based on the assumption that there is a highly nonuniform pump intensity distribution, i.e., a volume speckle pattern in the focal region. Only the phase-conjugate portion of the BSS field whose intensity distribution best matches the nonuniform gain distribution, experiences the maximum exponential gain coefficient, twice that of the non-phase-conjugate portion.[24,44] In terms of the holographic model the volume speckle is a result of interference between the undistorted and distorted parts of the pump field.

Finally, it is important to indicate that the Gabor holographic interaction model and the theoretical treatment presented in this paper are suitable not only for various BSS processes but also for other kinds of backward coherent emission processes, provided that there is a high exponential gain mechanism for small initial backward coherent signal as well as an effective holographic grating induced by the pump field.[46] Recently, we have observed a nearly perfect phase-conjugation property in the backward frequency-upconverted stimulated emission from a two-photon pumped lasing medium.[47] This observation may suggest a new technical approach to generate optical phase-conjugate waves and can be explained very well based on the same theoretical model described here.[48–50]

8. SOME RECENT EXPERIMENTAL RESULTS

So far, most experimental studies of phase-conjugation fidelity of BSS have been based on measurements of the

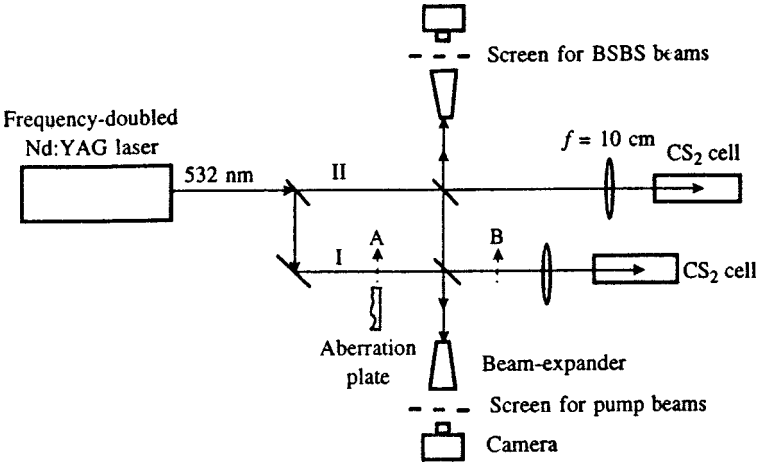

FIG. 4. Experimental setup for measuring the phase-conjugation property of the backward stimulated Brillouin scattering with two-beam interference method.

near-field and far-field distributions. We present here some recent experimental results on the phase-conjugation property of backward stimulated Brillouin scattering (BSBS), based on measuring the fidelity of wavefront reconstruction by using the two-beam interference technique. The advantage of this method is that it can provide both near-field and the wavefront information of the tested beams.

The experimental setup is shown schematically in Fig. 4. A 532-nm master pump laser beam was provided by a *Q*-switched and frequency-doubled pulsed Nd:YAG laser source; the pulse duration, beam size and divergence angle of this beam were 10 ns, 4 mm, and 1 mrad, respectively. After passing through a beamsplitter and a reflecting mirror, that master beam was divided into two beams, which were finally focused into two 10-cm-long CS_2 liquid cells through two $f=10\,\mathrm{cm}$ focusing lenses. The intensities of these two pump beams could be adjusted separately and were high enough to generate BSBS in both liquid cells. Furthermore, by means of two-edge beamsplitters and a 10× beam expander, the interference pattern of the two incident pump beams could be observed on a screen and recorded by a camera. In the same manner, the interference pattern of the two BSBS beams from these two liquid cells could be also observed in another screen. In order to test the wavefront-reconstruction ability, a hydrofluoric acid-etched glass slide was used as an aberration plate, which could introduce an aberration of 10–15 mrad on the pump beam (I). This aberration plate can be placed either at position A or position B.

Figure 5a shows the photograph of the pump beam II: here we see a relatively uniform intensity distribution in the beam section. Figure 5b shows the photograph of the pump beam I after passing through an aberration plate placed at position A; here we see a random transverse intensity fluctuation caused by the aberration plate. Figure 5c shows the

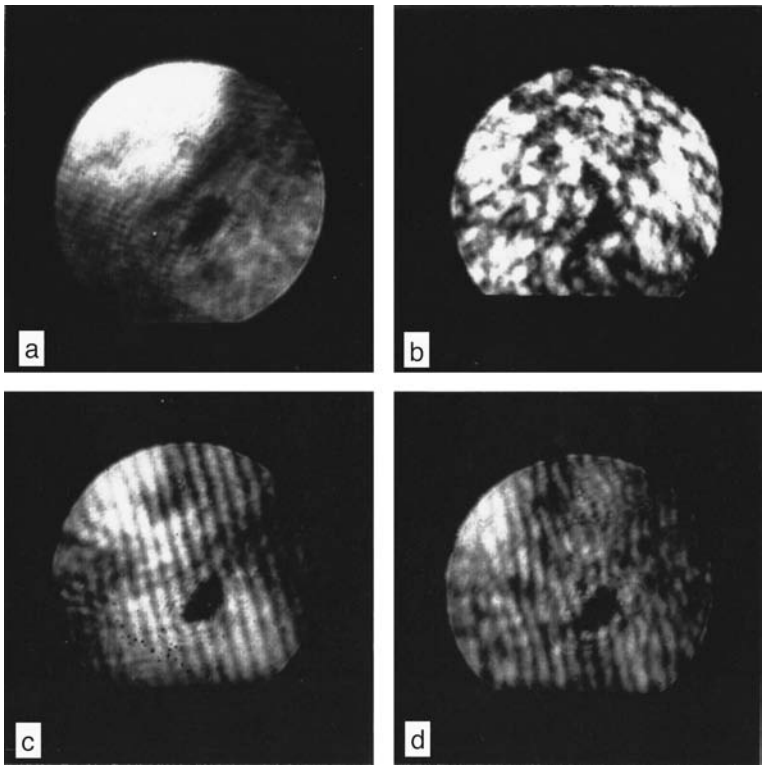

FIG. 5. Photographs of (a) the pump beam II, (b) the pump beam I after passing through an aberration plate, (c) the interference pattern of the two pump beams without no aberration plate, and (d) the interference pattern of the two pump beams with aberration plate in position A shown in Fig. 4.

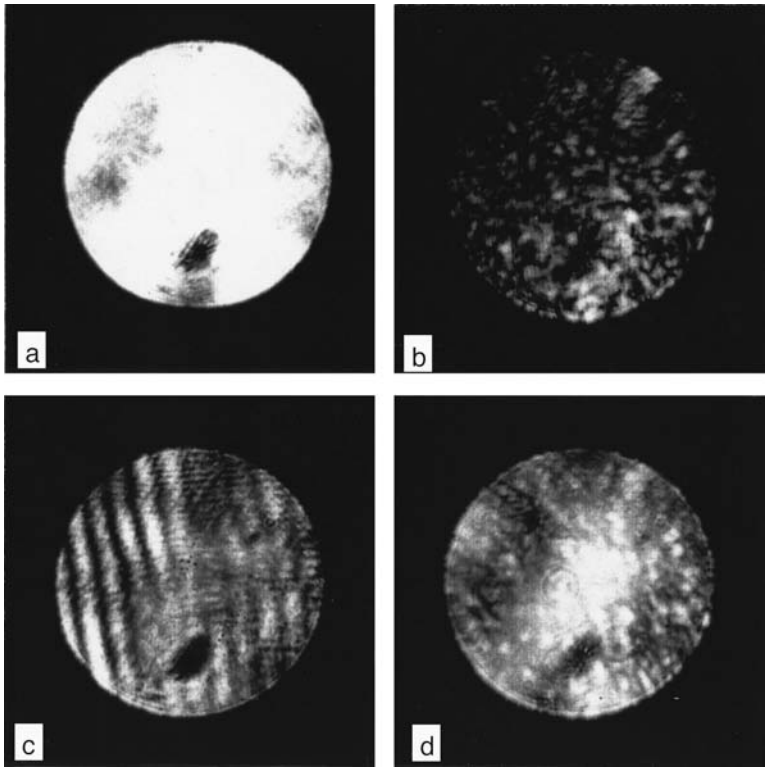

FIG. 6. Photographs of (a) the BSBS beam excited by the pump beam II, (b) the BSBS beam excited by the pump beam I passing through an aberration plate, (c) the interference pattern of the two BSBS beams with no aberration plate, and (d) the interference pattern of the two BSBS beams with an aberration plate in position B shown in Fig. 4.

photograph of the interference pattern formed by the two incident pump beams with no aberration plate; here we can see the regular and straight fringes that indicate a nearly ideal plane wavefront for both beams. Finally, Fig. 5d shows a photograph of the interference pattern when the aberration plate was placed at position A; here we can no longer see any clear and regular fringes because of the severe aberration influence on one beam. To obtain the photographs shown in Fig. 5c and 5d, the intensities of the two interfering beams were kept nearly the same.

Under the same conditions, a set of photographs can be obtained for the BSBS beams from the two CS_2 liquid cells. Figure 6a shows the photograph of the BSBS beam excited by the pump beam II: here we see a relatively uniform intensity distribution with no discrete-spot structure. Figure 6b shows a photograph of the BSBS beam excited by the pump beam I with the aberration plate at position A. Here we see a randomly fluctuating intensity distribution, analogous to that shown in Fig. 5b. Figure 6c shows a photograph of the regular interference pattern formed by two BSBS beams with no aberration plate. Compared Fig. 6c to Fig. 5c, one can find that the two stimulated backscatter beams exhibit a nearly ideal plane wavefront, with no aberration plate. Finally, Fig. 6d shows a photograph of the interference pattern formed by the two BSBS beams with the aberration plate at position B. In this case we can still see the clear fringes, although there is a small irregularity among them in comparison with Fig. 6c.

The results shown in Fig. 6c and 6d can be readily explained in terms of the two basic theoretical conclusions given in Sec. 7. The first conclusion is that if the stimulated scattering gain is high enough and the aberration influence is small, one can expect a nearly perfect wavefront reconstruction. The result shown in Fig. 6c basically supports this conclusion. The second conclusion is that under a large aberration influence, the wavefront reconstruction will not be perfect. The result shown in Fig. 6d basically supports that conclusion. In the latter case, when the BSBS beam from the second liquid cell passed through the aberration plate at position B, the original wavefront of the pump beam I can be essentially, but not perfectly, reconstructed.

*)E-mail: gshe@acsu.buffalo.edu

[1] *Optical Phase Conjugation*, ed. by R. A. Fisher, Academic Press, New York (1983).
[2] A. Yariv, IEEE J. Quantum Electron. **14**, 650 (1978).
[3] R. W. Hellwarth, J. Opt. Soc. Am. **67**, 1 (1977).
[4] D. M. Bloom and G. C. Bjorklund, Appl. Phys. Lett. **31**, 592 (1977).
[5] A. Yariv and D. M. Pepper, Opt. Lett. **1**, 16 (1977).
[6] D. G. Steel, R. C. Lind, J. F. Lam, and C. R. Giuliano, Appl. Phys. Lett. **35**, 376 (1979).
[7] R. L. Abrams and R. C. Lind, Opt. Lett. **2**, 94 (1978); **3**, 205 (1978).
[8] G. Martin and R. W. Hellwarth, Appl. Phys. Lett. **34**, 371 (1979).
[9] B. Ya. Zel'dovich, V. I. Popovichev, V. V. Ragul'skii, and F. S. Faizullov, JETP Lett. **15**, 109 (1972).
[10] O. Yu. Nosach, V. I. Popovichev, V. V. Ragul'skii, and F. S. Faizullov, JETP Lett. **16**, 435 (1972).
[11] V. Wang and C. R. Giuliano, Opt. Lett. **2**, 4 (1978).
[12] A. I. Sokolovskaya, G. L. Brekhovskikh, and A. D. Kudryavtseva, Opt. Commun. **24**, 74 (1978).
[13] A. D. Kudriavtseva, A. I. Sokolovskaia, J. Gazengel, N. P. Xuan, and G. Rivoire, Opt. Commun. **26**, 446 (1978).
[14] E. J. Miller, M. S. Malcuit, and R. W. Boyd, Opt. Lett. **15**, 1188 (1990).
[15] A. Yariv, Opt. Commun. **21**, 49 (1977).
[16] P. V. Avizonis, F. A. Hopf, W. D. Bamberger, S. F. Jacobs, A. Tomita, and K. H. Womack, Appl. Phys. Lett. **31**, 435 (1977).
[17] N. C. Griffen and C. V. Heer, Appl. Phys. Lett. **33**, 865 (1978).
[18] M. Fujita, H. Nakasuka, H. Nakanishi, and M. Matsuoka, Phys. Rev. Lett. **42**, 974 (1979).

[19] J. O. White, M. Cronin-Golomb, B. Fischer, and A. Yariv, Appl. Phys. Lett. **40**, 450 (1982).
[20] J. Feinberg, Opt. Lett. **7**, 486 (1982).
[21] M. Cronin-Golomb, B. Fischer, J. O. White, and A. Yariv, Appl. Phys. Lett. **42**, 919 (1983).
[22] R. W. Hellwarth, J. Opt. Soc. Am. **68**, 1050 (1978).
[23] H. Hsu, Appl. Phys. Lett. **34**, 855 (1979).
[24] B. Ya. Zel'dovich, N. F. Pilipetsky, and V. V. Shkunov, *Principles of Phase Conjugation*, Springer-Verlag, Berlin (1985).
[25] G. G. Kochemasov and V. D. Nikolaev, Kvant. Elektron. (Moscow) **4**, 115 (1977) [Sov. J. Quantum Electron. **7**, 60 (1977)].
[26] I. M. Bel'dyugin, M. G. Galushkin, E. M. Zemskov, and V. I. Mandrosov, Kvant. Elektron. (Moscow) **3**, 734 (1976) [Sov. J. Quantum Electron. **6**, 1349 (1976)].
[27] B. Ya. Zel'dovich and V. V. Shkunov, Kvant. Elektron. (Moscow) **4**, 1090 (1977) [Sov. J. Quantum Electron. **7**, 610 (1977)].
[28] N. B. Baranova, B. Ya. Zel'dovich, and V. V. Shkunov, Kvant. Elektron. (Moscow) **5**, 973 (1978) [Sov. J. Quantum Electron. **8**, 559 (1978)].
[29] N. B. Baranova and B. Ya. Zel'dovich, Kvant. Elektron. (Moscow) **7**, 973 (1980) [Sov. J. Quantum Electron. **10**, 555 (1980)].
[30] V. I. Bespalov, V. G. Manishin, and G. A. Pasmanik, Zh. Éksp. Teor. Fiz. **77**, 1756 (1979) [Sov. Phys. JETP **50**, 879 (1979)].
[31] V. G. Sidorovich and V. V. Shkunov, Zh. Tekh. Fiz. **49**, 816 (1979) [Sov. Phys. Tech. Phys. **24**, 472 (1979)].
[32] Y. R. Shen, *The Principles of Nonlinear Optics*, Wiley, New York (1984), pp. 189, 254.
[33] M. Born and E. Wolf, *Principles of Optics*, 6-th ed., Pergamon, Oxford (1983), p. 453.
[34] D. I. Mash, V. V. Morozov, V. S. Starunov, and I. L. Fabelinskii, JETP Lett. **2**, 25 (1965).
[35] G. I. Zaitsev, Y. I. Kyzylasov, V. S. Serarunov, and I. L. Fabelinskii, JETP Lett. **6**, 35 (1967).
[36] G. S. He and P. N. Prasad, Phys. Rev. A **41**, 2687 (1990).
[37] G. S. He, R. Burzynski, and P. N. Prasad, J. Chem. Phys. **93**, 7647 (1990).
[38] G. S. He and G. C. Xu, IEEE J. Quantum Electron. **28**, 323 (1992).
[39] J. Nilsen and A. Yariv, Opt. Commun. **39**, 199 (1981).
[40] V. M. Arutunyan, G. G. Adonts, A. R. Aramyan, S. P. Ishkhanyan, E. G. Kanetsyan, T. A. Papasyan, and S. M. Sarkisyan, Opt. Appl. **13**, 347 (1983).
[41] H. A. Mackenzie, D. J. Hagan, and H. A. Al-Attar, IEEE J. Quantum Electron. **QE-22**, 1328 (1986).
[42] M. T. De Araujo, S. S. Vianna, and G. Grynberg, Opt. Commun. **80**, 79 (1990).
[43] A. Yariv, *Optical Electronics*, 3-rd ed., Holt, Rinehart and Winston, New York (1985), pp. 28–34.
[44] R. W. Boyd, *Nonlinear Optics*, Academic, Boston (1992), pp. 245–252, 337, 343–347.
[45] P. Yeh, *Introduction to Photorefractive Nonlinear Optics*, Willey, New York (1993), pp. 211–220.
[46] V. G. Koptev, A. M. Lazaruk, I. P. Petrovich, and A. S. Rubanov, JETP Lett. **28**, 434 (1978).
[47] G. S. He, Y. Cui, M. Yoshida, and P. N. Prasad, Opt. Lett. **22**, 10 (1997).
[48] G. S. He and P. N. Prasad, J. Opt. Soc. Am. B **15**, 1078 (1998).
[49] G. S. He, N. Cheng, P. N. Prasad, D. Liu, and S. H. Liu, J. Opt. Soc. Am. B **15**, 1086 (1998).
[50] G. S. He and P. N. Prasad, IEEE J. Quantum Electron. **34**, 473 (1998).

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

Atom cooling by VSCPT: accumulation plus filtering

Fam Le Kien

Institute for Laser Science, University of Electro-Communications, Tokyo 182, Japan; Department of Physics, University of Hanoi, Hanoi, Vietnam

V. I. Balykin*)

Institute for Laser Science, University of Electro-Communications, Tokyo 192 Japan; Institute of Spectroscopy, Russian Academy of Sciences, 142092 Troitzk, Moscow Region, Russia
(Submitted 2 December 1997; resubmitted 7 September 1998)
Zh. Éksp. Teor. Fiz. **115**, 449–462 (February 1999)

We study laser cooling by velocity-selective coherent population trapping (VSCPT) in a double-Λ scheme with decay beyond the working levels. We show that this additional decay channel filters diffused atoms from trapped ones and provides an ultrasharp atomic momentum distribution. © *1999 American Institute of Physics.* [S1063-7761(99)00602-2]

1. INTRODUCTION

Coherent population trapping was observed for the first time by Alzetta *et al.*[1] as a decrease in fluorescent emission in a laser optical pumping experiment involving a three-level atomic system with two ground levels and one excited level. This effect results from coherent superposition of the ground states which is stable against absorption from the radiation field. Various theoretical and experimental aspects of coherent population trapping have been reviewed in several papers.[2–6] This phenomenon has been exploited in very different applications: metrology, optical bistability, high-resolution spectroscopy, laser multiphoton ionization, four-wave mixing, laser-induced structures in the continuum, laser manipulation of atoms, adiabatic transfer, lasing without inversion, and matched pulse propagation.

The application of velocity-selective coherent population trapping (VSCPT) to laser manipulation of atoms has been studied intensively.[7–18] The basic idea of VSCPT is to pump atoms into a noncoupled state that has a well-defined momentum, where the atoms do not interact with the laser radiation. Accumulation of atoms in this special velocity-selective trapping state is due to spontaneous emission. In VSCPT experiments[7–10] with ^{4}He metastable atoms, very narrow final distributions of atomic momenta are observed. The form and width of the momentum distribution has been described theoretically for various schemes and in various regimes.[11–18] It is common to associate either the peak width or the dark-state population of an atomic momentum distribution with an effective temperature. Based on this effective temperature, the authors of Refs. 7–18 conclude that laser cooling below the recoil limit has been achieved. However, spontaneous emission produces, as shown in these references, not only the accumulation of atoms in the trapping state but also diffusion of some of those atoms toward high values of momentum. Unlike the velocity-selective trapping phenomenon, the diffusion of atoms in the momentum space tends to increase the temperature. Due to such a random process, the wings of the momentum distribution are not Gaussian. In the meantime, the wing shape and the fraction of the atoms that have diffused toward the wings are not reflected in the peak width and dark-state population. Consequently, the increase in the peak value and decrease in the width of the final atomic momentum distribution, which is far from Gaussian, do not mean the cooling; they merely indicate the accumulation of atoms in the dark state. A simple analytical description of trapping versus diffusion was recently given in Ref. 19. A nice way to confine atomic velocities during the VSCPT phase has been proposed by Marte *et al.*[20] This method is based on the coexistence of VSCPT and polarization-gradient cooling.

The purpose of this paper is to show that one can use VSCPT to perform not only accumulation but also filtering of atoms in momentum space. Based on the theoretical Λ model of Ref. 11, we introduce a double-Λ model, where an additional upper level with possible decay outside the working configuration is added. We find that this new decay channel can separate the atoms in the wings from atoms near the peaks of the momentum distribution, and hence laser cooling of atoms below the recoil limit can be achieved.

The paper is organized as follows. In Sec. 2 we formulate the model and present the basic equations for the atomic density-matrix elements. In Sec. 3 we perform a numerical analysis. Finally, Sec. 4 contains conclusions.

2. MODEL AND BASIC EQUATIONS

We consider an ensemble of atoms of mass M moving in the z direction. The atoms have two degenerate ground levels, g_- and g_+, and two nondegenerate excited levels, e and e'. We denote the energy of the atomic level j ($j=e,e',g_-,g_+$) by $\hbar\omega_j$. The ground sublevels, g_- and g_+, can be coupled to the level e by two counterpropagating laser beams with the same frequency ω_L, the opposite wave vectors k and $-k$ being aligned along the z direction, and the strengths being characterized by the Rabi frequencies Ω_+ and Ω_-, see Fig. 1. Similarly, the level e' can also be coupled to the levels g_- and g_+ by two counterpropagating laser beams with frequency ω_L' and strengths characterized

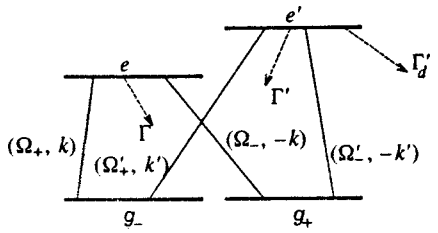

FIG. 1. Energy levels and optical transitions in the double-Λ configuration.

by the Rabi frequencies Ω'_+ and Ω'_-. The second pair of laser beams is aligned in the z' direction, which may be different from the z direction, but should be very close. We denote the projections of the wave vectors of these laser beams onto the z axis by k' and $-k'$. In what follows we are interested only in the atomic center-of-mass motion along the z axis. We take into account the spontaneous emission of atoms from the upper levels e and e' to the ground sublevels g_- and g_+. In addition, we assume that the level e' can decay into another level, which is not shown in the figure. Furthermore, the fields can be switched on and off at different times, that is, the Rabi frequencies Ω_+, Ω_-, Ω'_+, and Ω'_- are generally functions of time.

We introduce the state $|j,p\rangle$, which represents an atom in the internal state j with linear momentum p along the z axis. Because of momentum conservation, the interaction of the atom with the fields can couple $|e,p\rangle$ only with $|g_-,p-\hbar k'\rangle$ and $|g_+,p+\hbar k'\rangle$, and $|e',p\rangle$ only with $|g_-,p-\hbar k'\rangle$ and $|g_+,p+\hbar k'\rangle$. The Hamiltonian corresponding to the unitary evolution of the system is of the form

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

where

$$H_A=\frac{P^2}{2M}+\hbar\omega_e|e\rangle\langle e|+\hbar\omega_{e'}|e'\rangle\langle e'| \tag{2}$$

describes the translational and internal degrees of freedom of the atom and

$$\begin{aligned}H_{\rm int}=\sum_p\ &(\hbar\Omega_+|e,p\rangle\langle g_-,p-\hbar k|+\hbar\Omega_-|e,p\rangle\langle g_+,p\\&+\hbar k|)e^{-i\omega_L t}+\sum_p\ (\hbar\Omega'_+|e',p\rangle\langle g_-,p-\hbar k'|\\&+\hbar\Omega'_-|e',p\rangle\langle g_+,p+\hbar k'|)e^{-i\omega'_L t}+\text{H.c.}\end{aligned} \tag{3}$$

describes the interaction of the atom with the laser fields, which are taken to be classical.

We assume that an atom in the upper level e can decay into the lower levels g_- and g_+, emitting a fluorescence photon in any direction. This spontaneous emission leads to the damping of the population of e and the associated coherences and to the feeding of g_- and g_+. We assume that the atomic decay rate Γ and the normalized probability $H(u)$ of emitting a photon with momentum u along the z axis do not depend on the center-of-mass motion and are the same for transitions from e to g_- and g_+. Analogously, we assume that an atom in the upper level e' can decay into each of the lower levels g_- and g_+ with the rate Γ'. The corresponding probability of emitting a photon with momentum u along the z axis is denoted by $H'(u)$. In addition, we assume that an atom in the upper level e' can decay with the rate Γ'_d into a fifth level, which is not shown in the figure. This dissipative irreversible decay will remove the untrapped diffused atoms from the working configuration.

Such a filtering process enables us to separate heated atoms from cooled ones, that is, to get a cooled system. The cooling efficiency is determined by the accumulation efficiency from one side and the separation efficiency from the other side. Since the atomic separation efficiency is proportional to the rate Γ'_d of the decay into the outside of the working configuration, we expect that for a subsequent application of the pairs of the laser beams, a higher decay rate Γ'_d will enable us to obtain a cooler atomic system.

We derive in Appendix the generalized optical Bloch equations

$$\begin{aligned}
\frac{d}{dt}\rho_{ee}(p_1,p_2)&=\left[\frac{d}{dt}\rho_{ee}(p_1,p_2)\right]_{\rm Ham}+\left[\frac{d}{dt}\rho_{ee}(p_1,p_2)\right]_{\Gamma},\\
\frac{d}{dt}\rho_{e'e'}(p_1,p_2)&=\left[\frac{d}{dt}\rho_{e'e'}(p_1,p_2)\right]_{\rm Ham}+\left[\frac{d}{dt}\rho_{e'e'}(p_1,p_2)\right]_{\Gamma'}\\
&\quad+\left[\frac{d}{dt}\rho_{e'e'}(p_1,p_2)\right]_{\Gamma'_d},\\
\frac{d}{dt}\rho_{\pm\pm}(p_1,p_2)&=\left[\frac{d}{dt}\rho_{\pm\pm}(p_1,p_2)\right]_{\rm Ham}+\left[\frac{d}{dt}\rho_{\pm\pm}(p_1,p_2)\right]_{\Gamma}\\
&\quad+\left[\frac{d}{dt}\rho_{\pm\pm}(p_1,p_2)\right]_{\Gamma'},\\
\frac{d}{dt}\rho_{e\pm}(p_1,p_2)&=\left[\frac{d}{dt}\rho_{e\pm}(p_1,p_2)\right]_{\rm Ham}+\left[\frac{d}{dt}\rho_{e\pm}(p_1,p_2)\right]_{\Gamma},\\
\frac{d}{dt}\rho_{e'\pm}(p_1,p_2)&=\left[\frac{d}{dt}\rho_{e'\pm}(p_1,p_2)\right]_{\rm Ham}+\left[\frac{d}{dt}\rho_{e'\pm}(p_1,p_2)\right]_{\Gamma'}\\
&\quad+\left[\frac{d}{dt}\rho_{e'\pm}(p_1,p_2)\right]_{\Gamma'_d},\\
\frac{d}{dt}\rho_{-+}(p_1,p_2)&=\left[\frac{d}{dt}\rho_{-+}(p_1,p_2)\right]_{\rm Ham},\\
\frac{d}{dt}\rho_{ee'}(p_1,p_2)&=\left[\frac{d}{dt}\rho_{ee'}(p_1,p_2)\right]_{\rm Ham}+\left[\frac{d}{dt}\rho_{ee'}(p_1,p_2)\right]_{\Gamma}\\
&\quad+\left[\frac{d}{dt}\rho_{ee'}(p_1,p_2)\right]_{\Gamma'}+\left[\frac{d}{dt}\rho_{ee'}(p_1,p_2)\right]_{\Gamma'_d}.
\end{aligned} \tag{4}$$

These equations govern the evolution of the density-matrix elements of the atom with the internal and external degrees of freedom, and will be solved numerically in the next section. Using this numerical solution, we will calculate and study the atomic momentum distribution. We will then

calculate the mean deviation of the atomic momentum from the nearest peak, which characterizes the effective temperature of the atomic subsystem.

Note that the above model is very simple, but it can reveal the underlying physics of actual situations, in which atomic level configurations are usually much more complicated. A specific example that is nearest to our model is neon in the metastable $1s_5$, level, with $1s_5-2p_9$ and $1s_5-2p_2$ transitions and decay channel from $2p_2$ to $1s_3$. This atomic level configuration is used in the laser cooling experiment reported by Shimizu *et al.*[21] A scheme to achieve velocity-selective coherent population trapping in a multilevel system under two-frequency laser excitation is proposed in Ref. 18. One difference between our model and the model of Ref. 18 is that a trapping state in our model is a superposition of two ground levels, and is created by two laser beams of the same frequency, while a trapping state in the other model is a time-dependent superposition of three ground levels and is created by two-color fields. Moreover, dissipative irreversible decay out of the working configuration plays a significant role in our model, while the authors of Ref. 18 considered only decay from the excited levels to the ground levels.

3. NUMERICAL ANALYSIS

We now study numerically the generalized optical Bloch equations (4) for the case in which the laser detunings $\delta_L=\omega_L-\omega_e$ and $\delta'_L=\omega'_L-\omega_{e'}$ are zero and the spontaneous-emission rates from the upper levels to the ground sublevels are equal ($\Gamma=\Gamma'$). We choose, for our example, an atomic mass M and the wave number k such that the recoil frequency $\omega_{\rm rec}\equiv\hbar k^2/2M$ is $\omega_{\rm rec}=0.027\Gamma$, which corresponds to the experiment[7] on He atoms. The decay rate Γ'_d of the atoms from the level e' into the outside of the double-Λ configuration is chosen to be $\Gamma'_d=10\Gamma$. Since the atomic separation efficiency is proportional to Γ'_d, the value chosen for this parameter is good enough to demonstrate filtering, and consequently, cooling the atoms below sub-recoil energy for reasonable interaction times. For a significantly higher or lower value of Γ'_d, the cooling efficiency, which depends on the rate of separation of heated and cooled atoms, is, in the case of subsequent application of the pairs laser beams, expected to be higher or lower, respectively. The temporal evolution is obtained by incrementation starting from various initial conditions. The time increment is $0.01\Gamma^{-1}$, small enough to avoid artificial instabilities introduced by the incremental approach. Since the exact shapes of the kernels $H(u)$ and $H'(u)$, characterizing the spontaneous-radiation patterns, are not important,[11] we take the constant forms $H(u)=1/2\hbar k$ and $H'(u)=1/2\hbar k'$.

For the initial atomic state, we take a statistical mixture of the two ground sublevels, g_- and g_+, with momentum distribution

$$W_0(p)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{p^2}{2\sigma^2}\right), \tag{5}$$

which is a normalized Gaussian function with a peak at $p=0$ and a standard width σ. The initial density-matrix elements $\rho_{j_1j_2}(p_1,p_2)|_{t=0}$ are thus vanish, except for

$$\rho_{--}(p,p)|_{t=0}=\frac{1}{2}W_0(p),$$

$$\rho_{++}(p,p)|_{t=0}=\frac{1}{2}W_0(p). \tag{6}$$

For the standard width of the initial momentum distribution we choose the value $\sigma=3\hbar k$. The variable p is discretized in steps of $\epsilon=\hbar k/30$, from $-p_{\rm max}$ to $p_{\rm max}$ where $p_{\rm max}=30\hbar k$. The chosen value of ϵ is small enough compared to the narrowest structure that emerges in the solution of Eqs. (4). The chosen value of $p_{\rm max}$ is large enough that the interesting part of the solution (near $p=0$) for the largest value of t considered here ($t=600\Gamma^{-1}$) is not affected by the truncation of the p range. For such values of $p_{\rm max}$ and t, the momentum diffusion from p values larger than $p_{\rm max}$ to $p=0$ is negligible.

We are interested in the momentum distribution of those atoms that have not decayed from the double-Λ configuration at the end of the interaction with the laser fields. The probability density of finding such an atom with linear momentum p along the z axis is

$$W(p)=\rho_{ee}(p,p)+\rho_{e'e'}(p,p)+\rho_{--}(p,p)+\rho_{++}(p,p). \tag{7}$$

Due to the decay of the atoms from the double-Λ configuration, the function $W(p)$ is not normalized with respect to the variable p. When we integrate this function over p, we obtain the probability

$$W_{\rm remain}=\int_{-\infty}^{\infty}W(p)dp, \tag{8}$$

that an atom remains in one of the working levels. The normalized function

$$W_{\rm norm}(p)=\frac{1}{W_{\rm remain}}W(p) \tag{9}$$

is the momentum distribution corresponding to the subensemble of atoms that remain in the working levels after the interaction with the fields.

In order to characterize the effective temperature in the cooling process, we introduce the quantity

$$V\equiv\left\{\int_{-\infty}^{\infty}[p-p_{\rm max}(p)]^2W_{\rm norm}(p)dp\right\}^{1/2}, \tag{10}$$

which is the mean deviation of the atomic momentum p from the nearest peak $p_{\rm max}(p)$. The effective temperature of the atomic system is defined as

$$\theta_{\rm eff}\equiv V^2/2M. \tag{11}$$

The sub-recoil cooling occurs when $\theta_{\rm eff}<\theta_{\rm rec}$ where $\theta_{\rm rec}\equiv(\hbar k)^2/2M$ is the recoil energy.

The deviation V characterizes the statistical spread of the atomic ensemble in momentum space. However, the geometrical peak width and the dark-state population used to characterize VSCPT[7,11] correspond only to a part of the en-

semble. This is the principal difference between our effective temperature and the effective temperature used in VSCPT.[7,11]

When the momentum distribution $W_{\rm norm}(p)$ has only one peak $p_{\max}$ which is, due to the symmetry of the initial conditions and the evolution equations, positioned at the mean momentum $\overline{p}=0$, the quantity V coincides with the momentum standard deviation,

$$V^2=\int_{-\infty}^{\infty}p^2W_{\rm norm}(p)dp=\int_{-\infty}^{\infty}(p-\overline{p})^2W_{\rm norm}(p)dp, \tag{12}$$

and characterizes the spread of the atoms around the peak as well as the spread of the whole momentum distribution.

When the central peak at $p=0$ splits into two symmetrical side peaks positioned at $\pm p_0(p_0>0)$, we can consider the system of the atoms that remain in the working level configuration after the interaction with the fields as a two-component system, one component with $p\geqslant 0$ and the other component with $p\leqslant 0$. Note that the normalized momentum distributions of these two components are $2W_{\rm norm}(p\geqslant 0)$ and $2W_{\rm norm}(p\leqslant 0)$, respectively. Hence, we see from the expression

$$V^2=2\int_0^{\infty}(p-p_0)^2W_{\rm norm}(p)dp \tag{13}$$

that V characterizes the spread of the atoms around each peak, as well as the spread of the momentum distribution of each component.

In what follows, we show and discuss numerical results for the case in which the two pairs of laser beams are applied in succession. The detailed sequence is the following. The first pair of laser beams resonant with the transitions $e\leftrightarrow g_-$ and $e\leftrightarrow g_+$ is turned on for the time T and then shut down. The second pair of laser beams resonant with the transitions $e'\leftrightarrow g_-$ and $e'\leftrightarrow g_+$ is then turned on for the time T'. The time-dependent Rabi frequencies corresponding to the laser pulses in each pair are taken to be the same, and have the rectangular forms

$$\Omega_-=\Omega_+=\Omega_0[\theta(t)-\theta(t-T)],$$
$$\Omega'_-=\Omega'_+=\Omega'_0[\theta(t-T)-\theta(t-T-T')]. \tag{14}$$

Here, Ω_0 and Ω'_0 are the maximal values of the Rabi frequencies, and $\theta(t)$ is the Heaviside step function.

We have solved Eqs. (4) for the parameters $\Omega_0=\Omega'_0=0.3\,\Gamma$, $T=150\,\Gamma^{-1}$, and $T'=450\,\Gamma^{-1}$ for two different cases: $k'=k$ and $k'=1.1\,k$. According to Aspect *et al.*,[11] the peak width of the momentum distribution is of order $M\Omega_0/k\sqrt{\Gamma T}$ for the first step, and $M\Omega'_0/k'\sqrt{(\Gamma'+\Gamma_d)T'}$ for the second. We therefore expect that the chosen interaction times $T=150\,\Gamma^{-1}$ and $T'=450\,\Gamma^{-1}$ are large enough to show two resolved peaks at the end of the first step and two very narrow peaks at the end of the second. For larger values of T and T', the effect becomes more dramatic.

3.1. The case of $k'=k$

We first present results for equal wavenumbers, that is, the case when $k'=k$. We show by the dashed and the solid curves in Fig. 2 the probability density $W(p)$ for an atom to have momentum p at times $t=T$ and $t=T+T'$, respectively, while remaining in the double-Λ level configuration. Since the decay out of the double-Λ configuration occurs only when the second pair of laser beams is turned on, the probability density $W(p)$, which corresponds to the original ensemble, and the normalized momentum distribution $W_{\rm norm}(p)$, which corresponds to the sub-ensemble of atoms inside the working configuration, are identical at $t=T$ but different at $t=T+T'$. Therefore, we additionally plot the function $W_{\rm norm}(p)$ obtained at $t=T+T'$ by the dashed-dotted curve. For comparison, the initial momentum distribution $W_0(p)$ is denoted by a dotted curve.

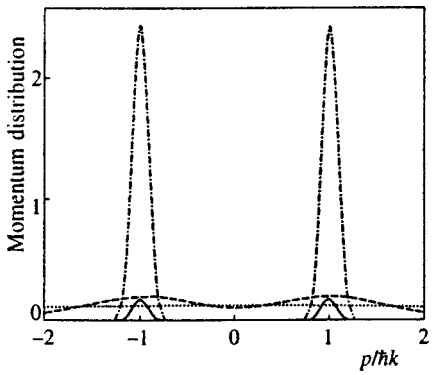

FIG. 2. Atomic momentum distributions produced by the successive application of two pairs of laser beams with equal wavenumbers $k'=k$. The dashed curve illustrates $W(p)$ at the end of the operation of the first pair of laser beams. The solid curve and the dashed-dotted curve correspond to $W(p)$ and $W_{\rm norm}(p)$, respectively, at the end of the operation of the second pair of the laser beams. The dotted curve corresponds to the initial momentum distribution. Here we have chosen the parameters $\sigma=3\hbar k$, $\Gamma'=\Gamma$, $\Gamma'_d=10\,\Gamma$, $\omega_{\rm rec}\equiv\hbar k^2/2M=0.027\,\Gamma$, $\Omega'_0=\Omega_0=0.3\,\Gamma$, $T=150\,\Gamma^{-1}$, and $T'=450\,\Gamma^{-1}$.

The dashed curve in Fig. 2 shows that the atomic momentum distribution obtained at the end of the interaction with the first pair of the laser beams exhibits two resolved peaks emerging at $\pm\hbar k$ above the initial distribution.[11] Such a structure results from the accumulation of atoms in the state

$$|\Psi_0\rangle=\frac{1}{\sqrt{2}}(|g_-,-\hbar k\rangle-|g_+,\hbar k\rangle), \tag{15}$$

which is a velocity-selective coherent trapping state with respect to the first pair of laser beams. The mechanism for accumulating atoms in this trapping state is momentum redistribution resulting from the spontaneous emission from the upper level e to the ground sublevels g_- and g_+. Besides the double narrow-peak structure, one sees that some of the atoms have diffused toward higher momentum values. All the above features of the momentum distribution depicted by the dashed curve in Fig. 2 have been studied in details in Ref. 11.

The solid curve in Fig. 2 shows that the second pair of the laser beams take off those atoms whose momenta are not near to $\pm\hbar k$. The spread of the distribution function $W(p)$ with respect to the peaks is greatly reduced, while the peak heights, which have the meaning of the probability densities

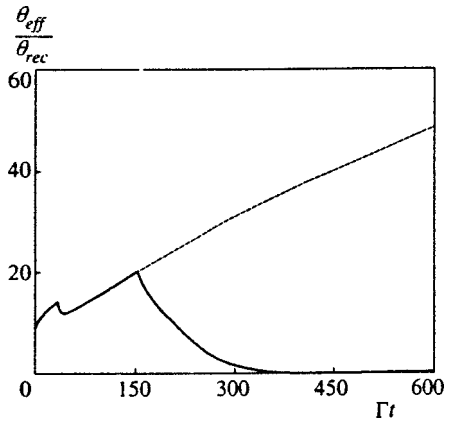

FIG. 3. Temporal evolution of the effective temperature $\theta_{\rm eff}$ in units of the recoil energy $\theta_{\rm rec}$ for the period of successive application of two pairs of laser beams with equal wavenumbers $k'=k$ (solid curve). All parameters are the same as for Fig. 2. For comparison, the dashed curve represents corresponding values of $\theta_{\rm eff}/\theta_{\rm rec}$ for the case of a single pair of laser beams that operate on an upper level with no decay out of the working configuration.

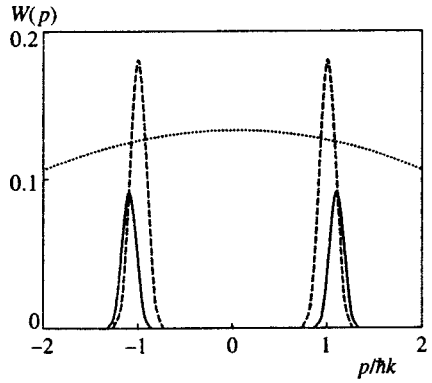

FIG. 4. Final atomic momentum probability density $W(p)$ in the case of unequal wavenumbers $k'=1.1\,k\neq k$ (solid curve). All parameters are the same as for Fig. 2. For comparison, we show the initial momentum distribution and the final momentum probability density $W(p)$ in the case of equal wavenumbers $k'=k$ (dotted curve and dashed curve, respectively).

for an atom from the original ensemble to remain within the working level configuration with linear momentum $\hbar k$ or $-\hbar k$, do not change appreciably. The reason is that when $k=k'$, the state (15) is also a velocity-selective coherent trapping state with respect to the second pair of laser beams, and those atoms which are not in this trapping state at the end of the first stage must undergo decay into the outside of the working level configuration in the second stage. The corresponding probability for an atom to remain in one of the working levels is $W_{\rm remain}\simeq 0.074$.

The dash-dot curve in Fig. 2 shows that the peaks of the normalized momentum distribution $W_{\rm norm}(p)$ created at $t=T+T'$ become not only much narrower but also much higher than the peaks of the dashed curve created at $t=T$. Thus, we observe a decrease in the momentum deviation around the peaks, or in other words cooling of the subensemble of atoms that continue to stay in the double-Λ configuration after the interaction with the two pairs of laser beams. This cooling is due to the accumulation of atoms in the trapping state in combination with filtering of the atoms in momentum space.

In Fig. 3, we plot by the solid line the temporal development of the effective temperature $\theta_{\rm eff}$ in units of the recoil energy $\theta_{\rm rec}$ for the whole time period of subsequent operation of the two pairs of the laser beams. For comparison, we depict by the dashed line the corresponding values of $\theta_{\rm eff}/\theta_{\rm rec}$ for the case in which the first pair of laser beams continues to operate without the intervention of the second pair for the entire time period.[7,11] It is worth noting that during the operation of the first pair of laser beams, that is, for the part at $t\leqslant 150\,\Gamma^{-1}$ for the solid line and for the entire dashed line, the effective temperature $\theta_{\rm eff}$ is in general increasing, except for a short time during which a sharp drop, like a phase transition, suddenly occurs. This increase in $\theta_{\rm eff}$ is due to atomic momentum diffusion and clearly shows that the atomic system at this stage is not cooled but heated, despite the accumulation of atoms in the dark state. This result is quite different from statements of Refs. 7 and 11, where a very similar system was considered but the atomic momentum diffusion was not taken into account in the definition of the effective temperature. The sudden decrease in $\theta_{\rm eff}$ occurring for a short time period is associated with the splitting of the central peak into the side peaks, whose locations very quickly move from 0 to $\pm\hbar k$. In other words, this phase-transition-like behavior is a result of the splitting of the atomic system from one component, with the momentum peak at 0, into two components, with momentum peaks at $\pm\hbar k$. In contrast to the first pair of laser beams, the second pair, operating from $t=150\,\Gamma^{-1}$ to $t=600\,\Gamma^{-1}$, causes a monotonic decrease in the effective temperature $\theta_{\rm eff}$. By the end of the interaction, $\theta_{\rm eff}$ is approximately $0.0085\,\theta_{\rm rec}$—a value which is small in comparison with the initial value $9\,\theta_{\rm rec}$ as well as with the recoil energy $\theta_{\rm rec}$. This decrease in the effective temperature is a signature of laser cooling below the one-photon recoil energy. It should be emphasized here that the underlying physics of the cooling obtained here involves velocity-selective coherent population trapping, on the one hand, and the filtering of the atoms in momentum space, on the other.

3.2. The case of $k'\neq k$

We now present numerical results for unequal wavenumbers k' and k. All conditions are the same as in the previous case except that $k'=1.1\,k$.

In Fig. 4, the solid curve is the probability density $W(p)$ for an atom to have momentum p at time $t=T+T'$, remaining inside the double-Λ configuration of levels. The total probability for an atom to remain inside this configuration is $W_{\rm remain}\simeq 0.035$. The dotted curve represents the initial momentum distribution. The dashed curve represents the probability density $W(p)$ for an atom to have momentum p at time $t=T+T'$ in the case when $k'=k$. As in the case of equal wavenumbers (dashed curve) the probability density function $W(p)$ in the case of unequal wavenumbers (solid curve) is very narrow compared to the initial momentum distribution (dotted curve) and photon momenta $\hbar k$ and $\hbar k'$, and is free from diffusion wings. However, the two peaks resulting from interaction with the second pair of laser beams

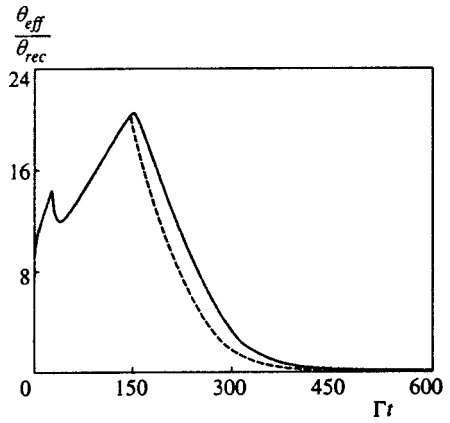

FIG. 5. Temporal evolution of effective temperature $\theta_{\rm eff}$ in units of the recoil energy $\theta_{\rm rec}$ during successive application of two pairs of laser beams with unequal wavenumbers $k'=1.1\,k\neq k$ (solid curve). All other parameters are the same as for Fig. 2. For comparison, we show $\theta_{\rm eff}/\theta_{\rm rec}$ for equal wavenumbers (dashed curve).

are now positioned at $p=\hbar k'=1.1\,\hbar k$. Moreover, the peaks of $W(p)$ in the case of unequal wavenumbers (solid curve) are lower than the corresponding peaks in the case of equal wavenumbers (the dashed curve).

The reason is that: when $k'\neq k$, the velocity-selective trapping state

$$|\Psi_0'\rangle=\frac{1}{\sqrt{2}}(|g_-,\hbar k'\rangle-|g_+,\hbar k'\rangle) \tag{16}$$

of the atoms interacting with the second pair of laser beams is different from the trapping state $|\Psi_0\rangle$, Eq. (15), of the atoms with respect to the first pair of the laser beams. After interaction with the first pair of laser beams, the number of atoms in the state $|\Psi_0\rangle$, which corresponds to momenta $\pm\hbar k$, is greater than the number of atoms in the state $|\Psi_0'\rangle$, which corresponds to momenta $\pm\hbar k'$. However, the second pair of laser beams takes away the atoms in $|\Psi_0\rangle$ while it does not affect the atoms in $|\Psi_0'\rangle$. This explains the shift in the peak positions from $\pm\hbar k$ to $\pm\hbar k'$, as well as the decrease in the peak heights.

In Fig. 5, we depict by the solid line the time development of the effective temperature $\theta_{\rm eff}$ in units of recoil energy $\theta_{\rm rec}$ for unequal wavenumbers $k'=1.1\,k\neq k$. For comparison, we replot (dashed line) the corresponding values of $\theta_{\rm eff}/\theta_{\rm rec}$ for equal wavenumbers $k'=k$. The figure shows that the second pair of laser fields in the case of unequal wavenumbers can also reduce the effective temperature $\theta_{\rm eff}$ to a much lower value than the recoil energy $\theta_{\rm rec}$, which is indicative of sub-recoil cooling. During the operation of the second pair, i.e., for $t>150\,\Gamma^{-1}$, the cooling effect for $k'\neq k$ (solid line) is not as strong as the cooling effect in the case of $k'=k$ (the dashed line). The reason is the difference between the velocity-selective trapping states $|\Psi_0'\rangle$ and $|\Psi_0\rangle$ with $k'\neq k$. This difference leads to the shift in the peaks of the atomic momentum distribution from locations $\pm\hbar k$, created by the first pair of the laser beams, to the locations $\pm\hbar k'$, created by the second pair. Such a process is not favorable for the dark-state population accumulation as well as the cooling of the atoms. Furthermore, we note that the two curves in Fig. 5 tend to merge with each other when $t>450\,\Gamma^{-1}$. The reason is that after interaction with the second pair of laser beams for a long enough time, the new peak structure of the atomic momentum distribution is well established and most of the atoms having momenta in the diffusion wings are removed. The effective temperature $\theta_{\rm eff}$ is then proportional to the peak width, which does not depend on the initial conditions in this limit.[11] This is why the difference in the effective temperature $\theta_{\rm eff}$ between the case of $k'\neq k$ and the case of $k'=k$ becomes small when the time of the interaction with the second pair of laser beams is long enough.

4. CONCLUSIONS

We have studied the application of velocity-selective coherent population trapping to laser manipulation of atomic center-of-mass motion in the framework of a simple double-Λ model.

We have introduced a variance of the momentum distribution with respect to the peaks, which can be used to characterize the effective temperature of atoms in the presence atomic momentum diffusion.

We have found that during the operation of the first pair of laser beams, operating in an upper level with no atomic decay out of the working configuration, the effective temperature is in general increasing, except for a short time during which a sharp drop, like a phase transition, suddenly occurs. Such an increase in the effective temperature is due to atomic momentum diffusion, and indicates heating of the atomic system at this stage in despite of the dark-state population accumulation. The sudden decrease in the effective temperature occurring for a short time is associated with the central peak splitting into two side peaks.

The subsequent application of the second pair of laser beams, operating on the other upper level with possible atomic decay into out of the working configuration, can filter atoms in the wings from the atoms near the peaks of the momentum distribution, cause a monotonic decrease in the effective temperature, and thereby lead to the laser cooling below the one-photon recoil energy.

The difference between the wave numbers of the two pairs of laser beams results in a shift of the peak positions, a decrease in the peak heights, and a decrease in the cooling rate.

Our model is very simple, and has been used just to show the underlying physics of real situations, where atomic level configurations are usually much more complicated. From this model we have seen clearly that (a) the known standard VSCPT scheme does not cool, but, in fact, it heats the whole ensemble of atoms, since the trapped atoms are mixed with the diffused ones, (b) an additional dissipative irreversible decay channel can filter atoms in momentum space, and (c) the use of VSCPT in a combination of accumulation and filtering steps can cool atoms below the recoil energy. To study the cooling effect in a specific real medium, further work will be required.

F.L.K. gratefully acknowledges the support of the Nishina Memorial Foundation and the Vietnamese Basic Re-

search Program in Natural Sciences, and appreciates the hospitality of Prof. K. Shimizu during his stay at the Institute for Laser Science. V.I.B. would like to express his gratitude for the hospitality during his stay at the Institute for Laser Science.

APPENDIX A

Generalized optical Bloch equations

We use the simplified notation

$$\begin{aligned}
&\rho_{ee}(p_1,p_2)=\langle e,p_1|\rho|e,p_2\rangle,\\
&\rho_{e'e'}(p_1,p_2)=\langle e',p_1|\rho|e',p_2\rangle,\\
&\rho_{\pm\pm}(p_1,p_2)=\langle g_\pm,p_1|\rho|g_\pm,p_2\rangle,\\
&\rho_{e\pm}(p_1,p_2)=\langle e,p_1|\rho|g_\pm,p_2\rangle e^{i\omega_L t},\\
&\rho_{e'\pm}(p_1,p_2)=\langle e',p_1|\rho|g_\pm,p_2\rangle e^{i\omega'_L t},\\
&\rho_{-+}(p_1,p_2)=\langle g_-,p_1|\rho|g_+,p_2\rangle,\\
&\rho_{ee'}(p_1,p_2)=\langle e,p_1|\rho|e',p_2\rangle e^{i(\omega_L-\omega'_L)t},\\
&\rho_{j_2j_1}(p_2,p_1)=\rho^*_{j_1j_2}(p_1,p_2),
\end{aligned} \tag{A1}$$

where $j_1,j_2=e,e',g_-,g_+$. By using the Schrödinger equation

$$i\hbar\left[\frac{d}{dt}\rho\right]_{\rm Ham}=[H,\rho], \tag{A2}$$

the equations of unitary evolution for the matrix elements of the density operator ρ are found to be

$$\begin{aligned}
\left[\frac{d}{dt}\rho_{ee}(p_1,p_2)\right]_{\rm Ham}&=i\frac{p_2^2-p_1^2}{2M\hbar}\rho_{ee}(p_1,p_2)-i\Omega_+\rho_{-e}(p_1\\
&\quad-\hbar k,p_2)+i\Omega^*_+\rho_{e-}(p_1,p_2-\hbar k)\\
&\quad-i\Omega_-\rho_{+e}(p_1+\hbar k,p_2)\\
&\quad+i\Omega^*_-\rho_{e+}(p_1,p_2+\hbar k),\\
\left[\frac{d}{dt}\rho_{e'e'}(p_1,p_2)\right]_{\rm Ham}&=i\frac{p_2^2-p_1^2}{2M\hbar}\rho_{e'e'}(p_1p_2)\\
&\quad-i\Omega'_+\rho_{-e'}(p_1-\hbar k',p_2)\\
&\quad+i\Omega'^*_+\rho_{e'-}(p_1,p_2-\hbar k')\\
&\quad-i\Omega'_-\rho_{+e'}(p_1+\hbar k',p_2)\\
&\quad+i\Omega'^*_-\rho_{e'+}(p_1,p_2+\hbar k'),\\
\left[\frac{d}{dt}\rho_{\pm\pm}(p_1,p_2)\right]_{\rm Ham}&=i\frac{p_2^2-p_1^2}{2M\hbar}\rho_{\pm\pm}(p_1,p_2)\\
&\quad+i\Omega_\mp\rho_{\pm e}(p_1,p_2\mp\hbar k)\\
&\quad+i\Omega'_\mp\rho_{\pm e'}(p_1,p_2\mp\hbar k')\\
&\quad-i\Omega^*_\mp\rho_{e\pm}(p_1\mp\hbar k,p_2)\\
&\quad-i\Omega'^*_\mp\rho_{e'\pm}(p_1\mp\hbar k',p_2),
\end{aligned}$$

$$\begin{aligned}
\left[\frac{d}{dt}\rho_{e\pm}(p_1,p_2)\right]_{\rm Ham}&=i\left(\delta_L+\frac{p_2^2-p_1^2}{2M\hbar}\right)\rho_{e\pm}(p_1,p_2)\\
&\quad-i\Omega_\mp[\rho_{\pm\pm}(p_1\pm\hbar k,p_2)\\
&\quad-\rho_{ee}(p_1,p_2\mp\hbar k)]\\
&\quad-i\Omega_\pm\rho_{\mp\pm}(p_1\mp\hbar k,p_2)\\
&\quad+i\Omega'_\mp\rho_{ee'}(p_1,p_2\mp\hbar k'),
\end{aligned} \tag{A3}$$

$$\begin{aligned}
\left[\frac{d}{dt}\rho_{e'\pm}(p_1,p_2)\right]_{\rm Ham}&=i\left(\delta'_L+\frac{p_2^2-p_1^2}{2M\hbar}\right)\rho_{e'\pm}(p_1,p_2)\\
&\quad-i\Omega'_\mp[\rho_{\pm\pm}(p_1\pm\hbar k',p_2)\\
&\quad-\rho_{e'e'}(p_1,p_2\mp\hbar k')]\\
&\quad-i\Omega'_\pm\rho_{\mp\pm}(p_1\mp\hbar k',p_2)\\
&\quad+i\Omega_\mp\rho_{e'e}(p_1p_2\mp\hbar k),\\
\left[\frac{d}{dt}\rho_{-+}(p_1,p_2)\right]_{\rm Ham}&=i\frac{p_2^2-p_1^2}{2M\hbar}\rho_{-+}(p_1,p_2)\\
&\quad+i\Omega_-\rho_{-e}(p_1,p_2-\hbar k)\\
&\quad-i\Omega^*_+\rho_{e+}(p_1+\hbar k,p_2)\\
&\quad+i\Omega'_-\rho_{-e'}(p_1,p_2-\hbar k')\\
&\quad-i\Omega'^*_+\rho_{e'+}(p_1+\hbar k',p_2),\\
\left[\frac{d}{dt}\rho_{ee'}(p_1,p_2)\right]_{\rm Ham}&=i\left(\delta_L-\delta'_L+\frac{p_2^2-p_1^2}{2M\hbar}\right)\rho_{ee'}(p_1,p_2)\\
&\quad-i\Omega_+\rho_{-e'}(p_1-\hbar k,p_2)\\
&\quad-i\Omega_-\rho_{+e'}(p_1+\hbar k,p_2)\\
&\quad+i\Omega'^*_+\rho_{e-}(p_1,p_2-\hbar k')\\
&\quad+i\Omega'^*_-\rho_{e+}(p_1,p_2+\hbar k').
\end{aligned}$$

The terms describing the spontaneous emission from the upper level e to the lower levels g_- and g_+ are

$$\begin{aligned}
\left[\frac{d}{dt}\rho_{ee}(p_1,p_2)\right]_\Gamma&=-\Gamma\rho_{ee}(p_1,p_2),\\
\left[\frac{d}{dt}\rho_{e\pm}(p_1,p_2)\right]_\Gamma&=-\frac{\Gamma}{2}\rho_{e\pm}(p_1,p_2),\\
\left[\frac{d}{dt}\rho_{ee'}(p_1,p_2)\right]_\Gamma&=-\frac{\Gamma}{2}\rho_{ee'}(p_1,p_2),\\
\left[\frac{d}{dt}\rho_{--}(p_1,p_2)\right]_\Gamma&=\left[\frac{d}{dt}\rho_{++}(p_1,p_2)\right]_\Gamma\\
&=\frac{\Gamma}{2}\int_{-\hbar k}^{\hbar k}duH(u)\rho_{ee}(p_1+u,p_2+u).
\end{aligned} \tag{A4}$$

Analogously, the terms describing the spontaneous emission from the upper level e' to the lower levels g_- and g_+ are

$$\left[\frac{d}{dt}\rho_{e'e'}(p_1,p_2)\right]_{\Gamma'} = -\Gamma'\rho_{e'e'}(p_1,p_2),$$

$$\left[\frac{d}{dt}\rho_{e'\pm}(p_1,p_2)\right]_{\Gamma'} = -\frac{\Gamma'}{2}\rho_{e'\pm}(p_1,p_2),$$

$$\left[\frac{d}{dt}\rho_{ee'}(p_1,p_2)\right]_{\Gamma'} = -\frac{\Gamma'}{2}\rho_{e'}(p_1,p_2), \tag{A5}$$

$$\left[\frac{d}{dt}\rho_{--}(p_1,p_2)\right]_{\Gamma'} = \left[\frac{d}{dt}\rho_{++}(p_1,p_2)\right]_{\Gamma'}$$

$$= \frac{\Gamma'}{2}\int_{-\hbar k'}^{\hbar k'} du H'(u)\rho_{e'e'} \times (p_1+u,p_2+u).$$

The terms corresponding to the decay from the upper level e' into the outside of the working configuration are

$$\left[\frac{d}{dt}\rho_{e'e'}(p_1,p_2)\right]_{\Gamma'_d} = -\Gamma'_d\rho_{e'e'}(p_1,p_2),$$

$$\left[\frac{d}{dt}\rho_{e'\pm}(p_1,p_2)\right]_{\Gamma'_d} = -\frac{\Gamma'_d}{2}\rho_{e'\pm}(p_1,p_2), \tag{A6}$$

$$\left[\frac{d}{dt}\rho_{ee'}(p_1,p_2)\right]_{\Gamma'_d} = -\frac{\Gamma'_d}{2}\rho_{ee'}(p_1,p_2).$$

The generalized optical Bloch equations are obtained by adding the terms of Eqs. (A3)–(A6).

*)E-mail: balykin@isan.troitsk.ru

[1] G. Alzetta, A. Gozzini, L. Moi, and G. Orriols, Nuovo Cimento B **36**, 5 (1976).
[2] B. J. Dalton and P. L. Knight, in *Laser Physics, Lecture Notes in Physics*, ed. by J. D. Harvey and D. F. Walls, Springer, Berlin (1983), Vol. 182, p. 213.
[3] H. I. Yoo and J. H. Eberly, Phys. Rep. **118**, 239 (1985).
[4] E. Arimondo, in *Interaction of Radiation with Matter, Volume in Honor of Adriano Gozzini*, ed. by G. Alzetta, F. Bassani, and L. Radicati, Scuola Normale Superiore, Pisa. (1987), p. 343.
[5] B. D. Agap'ev, M. B. Gornyi, and B. G. Matiso, Physics-Uspekhi **36**, 763 (1993).
[6] E. Arimondo, in *Progress in Optics*, ed. by E. Wolf, Elsevier, Amsterdam (1996), XXXV, p. 257.
[7] A. Aspect, E. Arimondo, R. Kaiser, N. Vansteenkiste, and C. Cohen-Tannoudji, Phys. Rev. Lett. **61**, 826 (1988).
[8] F. Bardou, B. Saubamea, J. Lawall, K. Shimizu, O. Émile, C. Westbrook, A. Aspect, and C. Cohen-Tannoudji, C. R. Acad. Sci. Paris, Série II **318**, 877 (1994).
[9] J. Lawall, F. Bardou, B. Saubamea, K. Shimizu, M. Leduc, A. Aspect, and C. Cohen-Tannoudji, Phys. Rev. Lett. **73**, 1915 (1994).
[10] M. R. Doery, M. T. Widmer, M. J. Bellanca, W. F. Buell, T. H. Bergeman, H. Metcalf, and E. J. D. Vredenbregt, Phys. Rev. A **52**, 2295 (1995).
[11] A. Aspect, E. Arimondo, R. Kaiser, N. Vansteenkiste, and C. Cohen-Tannoudji, J. Opt. Soc. Am. B **6**, 2112 (1989).
[12] Y. Castin, H. Wallis, and J. Dalibard, J. Opt. Soc. Am. B **6**, 2046 (1989).
[13] V. Z. Alekseev and D. D. Krylova, JETP Lett. **55**, 321 (1992).
[14] V. Z. Alekseev and D. D. Krylova, Laser Phys. **2**, 781 (1992).
[15] V. Z. Alekseev and D. D. Krylova, Opt. Commun. **95**, 319 (1993).
[16] E. A. Korsunsky, D. V. Kosachiov, B. G. Matisov, Yu. V. Rozhdestvensky, L. Windholz, and C. Neureiter, Phys. Rev. A **48**, 1419 (1993).
[17] F. Bardou, J. P. Bouchaud, O. Émile, A. Aspect, and C. Cohen-Tannoudji, Phys. Rev. Lett. **72**, 203 (1994).
[18] M. R. Doery, R. Gupta, T. Bergeman, H. Metcalf, and E. I. D. Vredenbregt, Phys. Rev. A **51**, 2334 (1995).
[19] S. Schaufler and V. P. Yakovlev, Laser Phys. **6**, 414 (1996).
[20] P. Marte, R. Dum, R. Taïeb, P. Zoller, M. S. Shahriar, and M. Prentiss, Phys. Rev. A **49**, 4826 (1994).
[21] F. Shimizu, K. Shimizu, and H. Takuma, Jpn. J. Appl. Phys., Part 2 **26**, L1847 (1987).

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

Correlation effects in electron–ion collisions in a strong laser field

G. M. Fraĭman,*) and V. A. Mironov

Institute of Applied Physics, Russian Academy of Sciences, 603600 Nizhniĭ Novgorod, Russia

A. A. Balakin

Higher School of General and Applied Physics, Nizhniĭ Novgorod State University, 603600 Nizhniĭ Novgorod, Russia

(Submitted 21 May 1998)
Zh. Éksp. Teor. Fiz. **115**, 463–478 (February 1999)

We examine elastic and inelastic scattering of electrons by ions in intense laser light. A method of numerical investigation of the scattering characteristics based on regularizing the Coulomb singularity is proposed. We show that over a broad range of parameter values the transport scattering cross section is weakly dependent on the intensity of the high-frequency field. We detect a significant modification of the dependence of the effective inelastic scattering cross section. We also show that the energy exchange with the field is determined by a fairly small group of electrons, called the representative electrons. Finally, we propose a qualitative model that explains our results by the fact that the leading contribution is provided by inelastic collisions of electrons with relatively small impact parameters traversing the region important for the interaction at large angles. © *1999 American Institute of Physics.* [S1063-7761(99)00702-7]

1. INTRODUCTION

The effect of a strong electromagnetic field on electron–ion collisions in a plasma has long been under intensive investigations.[1–6] This is because the problem is not only fundamental for plasma physics in general but also important for applications. In the last decade this became especially evident after new powerful terawatt lasers were built.[7]

Starting with the pioneering work of Dawson and Oberman[1] and Silin,[2] it was noticed that this effect may be very large. The greatest progress was achieved with the model of small-angle scattering[8–10] (in this model the unperturbed electron path is a straight line). Another model, the low-frequency approximation,[3–5] describes strong collisions with large scattering angles. It is assumed that an external (and fairly strong) electric field accelerates the electron before and after the collision, while in the scattering process proper (which is instantaneous) only the static field of the nearest ion is important. The region within which the small-angle approximation can be used is naturally that of large impact parameters and high drift velocities (compared to the oscillator velocity). In the low-frequency approximation it is assumed (often implicitly) that each passage of an electron near an ion (each act of ''instantaneous'' scattering) is totally independent of the previous one, i.e.,the squares of the scattering angles are always added rather than the angles themselves.

The main idea of the present work is to consistently analyze the correlation effects. It is possible to divide these effects into two classes. First, a rapidly oscillating electron moves past an ion at a low drift velocity and returns many times to that ion as it travels through the interaction region. In the course of these multiple oscillations the electron drift velocity slowly changes (only slightly during each field period). Hence the scattering angles acquired in each oscillation or, what is the same, the increments of the drift velocity, are simply added. That is, while traveling past an ion, the oscillating electrons are deflected (in the drift velocity) in the same direction—toward the ion (in the region with a stronger field). Here the scattering in the total particle velocity is small-angle. As a result, the angle of scattering in the drift velocity, which is actually responsible for variation in the transverse energy and the transport scattering cross section, may be much larger than the angle obtained in the hypothesis of independent collisions.

Another effect, as we will shortly see, arises because the probability of effective inelastic collision (cross section) grows due to the same multiple oscillations. It is found that in addition to the majority of electrons, which experience small-angle scattering and hence exchange little energy with the field, there is a relatively small group of electrons (the ''representative'' electrons) that undergo strong inelastic scattering accompanied by a large change in their drift energy. Although the number of such electrons is relatively small, their contribution to the effective scattering cross section is dominant. Such scattering events are strongly correlated with the field. This is seen in the fact that owing to the focusing properties of the Coulomb potential the oscillating electrons with large impact parameters are attracted to an ion (without changing the electron average energy) and, if the phase of the field is suitable when an electron lands in the region of substantial energy exchange, the electrons are effectively scattered with a substantial change in their average energy.

Correlation effects were examined in Refs. 8–10 using straight electron paths (in relation to the drift velocity). How-

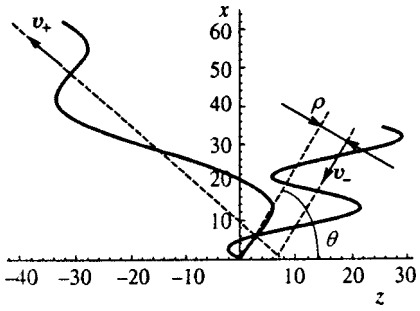

FIG. 1. Typical electron path (solid curve), drift path (dashed straight line), and the parameters of the scattered electron (v is the velocity, ρ is the impact parameter, and θ is the angle velocity and field **E**).

ever, in view of the adopted approximations, these effects cannot explain the results of numerical investigations (by the cell method[8]), which show that the effective scattering cross section increases at low drift velocities (compared to the oscillator velocity).

In this paper we use the classical (nonrelativistic) approach to analyze the scattering of monoenergetic (in the drift velocity) beams of electrons, either isotropic or unidirectional, by a single ion in a high-frequency electric field. As a result we obtain and analyze expressions for the cross sections of such collisions. We find that under conditions where the electron drift velocity is low compared to the oscillator velocity (the ''thermal energy'' is low compared to the oscillator energy), the effective inelastic scattering cross section and the effective collision frequency may be many times the standard values obtained by estimating the Rutherford scattering cross section using the oscillator velocity.

The plan of this paper is as follows. In Sec. 2 we discuss the formal statement of the problem and introduce expressions that are convenient for numerically finding the quantities of interest (the transport and total scattering cross sections and the effective collision frequency). There we also discuss the numerical model. In Sec. 3 we analyze inelastic collisions, which are responsible for exchange of energy between particle and field. Section 4 is devoted to elastic effects, which are responsible for the rate with which the electron distribution function becomes isotropic in the course of electron–ion collisions. Finally, in Sec. 5 we perform several analytical interpolations of our results and briefly discuss the future for such investigations.

2. BASIC RELATIONS

Let us examine the problem of the scattering by an ion of a single electron (Fig. 1) with charge e moving with a drift velocity **v** in a strong uniform electric field **E** polarized along the z axis:

$$\mathbf{E}(R,t)=\mathbf{E}\sin\omega t. \tag{1}$$

We assume that the radius of oscillations is small compared to the particle separation. Following a many-particle plasma approach we begin with the following concepts, which are quite common in the model of pair collisions. Suppose that we are dealing with many scattering ions. The distance between the ions is larger than the size of the region where there is substantial energy exchange between the electron and the high-frequency field near an ion (for details see Sec. 5).[1)] Then each oscillating electron can pass the ion many times before it leaves the scattering region. Moreover, far from an ion, at each moment in time the ensemble of electrons may be considered uniformly distributed over the impact parameters and the drift velocities. This means that instead of analyzing a many-body plasma problem we can consider the scattering by a single ion of many electrons with a uniform impact-parameter distribution. In a real plasma, Debye scattering is equivalent to replacing the Coulomb potential by a Yukawa potential.

Analysis of the electron paths was done by using the classical Newton equation

$$m\ddot{\mathbf{R}}=-\frac{Ze^2}{R^3}\mathbf{R}+e\mathbf{E}\sin\omega t, \tag{2}$$

which describes the motion of an electron in the field of an ion with charge Ze and in a uniform electric field **E** varying according to the harmonic law with a frequency ω. Here the radius of electron oscillations in the laser field is assumed small compared to the wavelength.

We introduce new dimensionless variables $\mathbf{R}_n=\mathbf{R}/r_E$ and $t_n=\omega_E t$ and reduce Eq. (2) to dimensionless form. To describe the strong variations in the Keplerian path that the field initiates, it is convenient to select the characteristic scales in the form

$$r_E=\sqrt{\frac{eZ}{E}},\quad \omega_E=\sqrt[4]{\frac{eE^3}{m^2Z}},\quad v_E=\sqrt[4]{\frac{Ze^3E}{m^2}}, \tag{3}$$

where r_E is the distance from the Coulomb center at which the field ion eZ/r_E^2 is equal to E, and ω_E is the characteristic frequency for the motion along a ''Keplerian'' orbit of radius r_E. As a result (dropping the subscript n in the formulas that follow) we get

$$\ddot{\mathbf{R}}=-\frac{\mathbf{R}}{R^3}+\mathbf{n}\sin\Omega t, \tag{4}$$

where **n** is the unit vector pointing in the same direction as field **E** ($\mathbf{E}=E\mathbf{n}$), and

$$\Omega=\omega\left(\frac{m^2Z}{eE^3}\right)^{1/4}=\frac{\omega}{\omega_E} \tag{5}$$

is the dimensionless frequency. Thus, dimensional analysis shows that one parameter, Ω, suffices to describe our problem. This parameter contains the frequency and the field strength in the combination $\omega/E^{3/4}$, which means that the limit of an ultrahigh field ($E\to\infty$) corresponds to a quasistatic field ($\Omega\to 0$).

We introduce the concept of the drift coordinate $\mathbf{r}(t)$ of the electron by assuming that the total electron coordinate $\mathbf{R}(t)$ can be written

$$\mathbf{R}(t)=\mathbf{r}(t)-\mathbf{r}_{\sim}\sin\Omega t,\quad \mathbf{r}_{\sim}=\frac{1}{\Omega^2}\mathbf{z}_0. \tag{6}$$

Thus, prior to a collision the electron had a velocity

$$\mathbf{V}_-(t)=\mathbf{v}_-+\mathbf{v}_\sim\cos\Omega t,\quad \mathbf{v}_\sim=-\frac{1}{\Omega}\mathbf{z}_0, \tag{7}$$

and after the collision the velocity is

$$\mathbf{V}_+(t)=\mathbf{v}_+(t)+\mathbf{v}_\sim\cos\Omega t. \tag{8}$$

We chose the region where the drift velocity ceases to change for our computer simulations. This means that the size L of this region meets the following condition:

$$L\gg r_\sim. \tag{9}$$

Of course, we must formally pass to the limit $L\to\infty$.

In these terms the effective differential cross section responsible for the deviation of particles in their drift velocity from the initial direction is

$$d\sigma_{\rm eff}(\mathbf{v}_-,\mathbf{v}_\sim,\boldsymbol{\rho})=\frac{\langle v_+^2-v_-^2\rangle}{v_\sim^2/2}d^2\rho; \tag{10}$$

it characterizes the extent to which the scattering is inelastic. Here $d^2\rho$ is the area element in the plane perpendicular to the direction of the incident beam around the impact parameter vector $\boldsymbol{\rho}$ (Fig. 1). Similarly, the transport cross section

$$d\sigma_{\rm tr}(\mathbf{v}_-,\mathbf{v}_\sim)=\frac{\langle(v_+)_\perp^2\rangle}{v_-^2}d^2\rho=\frac{dw_{\rm tr}}{v_-^2} \tag{11}$$

is responsible for the deviation of particles in their drift velocity from the initial direction, since $\mathbf{v}_{+\perp}$ is the component of the total velocity at the exit that is perpendicular to the direction of the incident beam. The angle brackets in (10) and (11) indicate averages over the period of the field. The denominator in (10) corresponds to normalization to the time-average of the energy of the scattered particles and the denominator in (11) corresponds to normalization to the square of the mean velocity.

Let us examine more thoroughly the physical meaning of the cross section (10) and (11) just introduced. First, we recall that to characterize scattering in static fields one usually introduces the concept of differential (effective) scattering cross section

$$d\sigma=\frac{dN}{N} \tag{12}$$

as the ratio of the flux of electrons scattered into the solid angle $d\Omega$ to the initial number of particles passing in unit time through unit area of the beam cross section (see p. 49 in Ref. 11). For static scattering, where everything is determined only by the drift velocity of the incident electrons, such a definition is indeed convenient, since the scattering angle is independent of the time when a particle entered the scattering region. But if the oscillation velocity is high, i.e., the scattering strongly depends on the phase of the field as the particles enter the scattering region,[12] the quantity $\langle d\sigma\rangle_{t_0}$, the differential scattering cross section averaged over the incident phases, reflects the situation more correctly. What we are interested in, however, is not the scattering cross section but its ''energy'' manifestations. The quantities (10) and (11) characterize these effects.

To clarify the meaning of the formula (10) for $\sigma_{\rm eff}$, we reason in the following standard way. First we note that, by its very definition, a variation in the energy W of the external field in a physically small volume U is related to a variation in the electron energy in the same volume as follows:

$$\frac{dW}{dt}U=-\frac{dW_u}{dt}U\equiv-\nu_{\rm eff}n\left\langle\frac{mv_\sim^2}{2}\right\rangle U, \tag{13}$$

where n is the electron number density. This relationship is actually the definition of the effective collision frequency $\nu_{\rm eff}$, which in turn is related to the effective scattering cross section $\sigma_{\rm eff}$:

$$\nu_{\rm eff}=\sigma_{\rm eff}Nv_-, \tag{14}$$

where N is the scattering-center (ion) density. Since particle energy variations in the volume are due only to energy exchange with the field in the electron–ion scattering process, Eq. (13) yields

$$\frac{dW}{dt}U\equiv\nu_{\rm eff}n\left\langle\frac{mv_\sim^2}{2}\right\rangle U=UNnv_-m\int\frac{\langle v_+^2-v_-^2\rangle}{2}d^2\rho. \tag{15}$$

Combining (14) and (15), we get (10). As expected, this quantity is a characteristic of only pair interactions and does not depend on the electron and ion densities. Of course, both $\nu_{\rm eff}$ and $\sigma_{\rm eff}$ coincide with the expressions used in Refs. 8 and 9. What makes this expression convenient is that one can directly find this quantity by means of computer modeling.

In our computer experiment we found that the motion of an electron at small distances from the Coulomb singularity is characterized by a sharp change of path, which may lead to errors in calculations. Since the requirement with respect to precision in this section of the path is found to be the determining factor in studying the long-term electron dynamics, the problem consisted in seeking a coordinate and time transformation that would make the equation describing the motion again regular (i.e., the equation was to have no singularity at the point occupied by the ion). Here it proved convenient to introduce, by analogy with Levi–Civita coordinates,[13] a new time s that coincides with the ''normal'' time t at large distances ($R\gg r_E$) and with the ''rapidly passing'' time as the electron approaches the Coulomb singularity ($R\leqslant r_E$). As an example, here is the replacement of time used in our calculations:

$$\frac{dt}{ds}=\frac{R^2}{1+R^2}. \tag{16}$$

The final form of the system of equations is

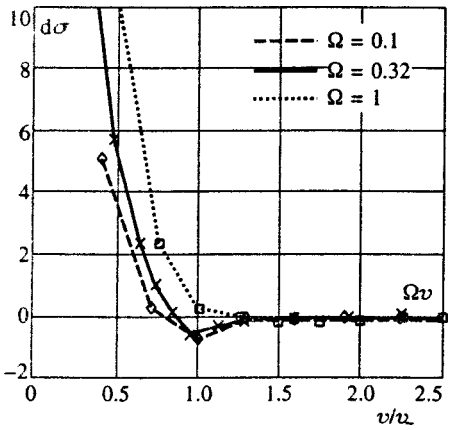

FIG. 2. Effective cross section of longitudinal ($\mathbf{v}_-\|\mathbf{E}$) inelastic collisions as a function of the velocity $v/v_\sim$ for three values of frequency.

$$\begin{cases} t' = \dfrac{R^2}{1+R^2}, \\ x' = \dfrac{p_x R^2}{1+R^2}, \\ z' = \dfrac{p_z R^2}{1+R^2}, \\ p_x' = -\dfrac{x}{r(1+R^2)}, \\ p_z' = \dfrac{R^2}{1+R^2}\cos\Omega t - \dfrac{z}{R(1+R^2)}. \end{cases} \tag{17}$$

Three values of frequency were used in the computer experiment: $\Omega = 0.1$, 0.32, and 1. We studied beams of electrons with drift velocities in the $10^{-2}-10^4$ range and impact parameters ρ (in the drift velocity) in the $10^{-1}-10^4$ range that were uniformly distributed over all incident phases. The initial distance from the Coulomb singularity to the leading center of the particles was chosen to be $10r_\sim(10/\Omega^2)$; a particle was assumed to have left the interaction region when the distance from the leading center to the Coulomb singularity was at least $10r_\sim$. Special attention was paid to selecting a large enough number of particles in one field period. For instance, in calculating inelastic interactions, the number of particles in one field period reached 2000 in some calculations.

3. INELASTIC COLLISIONS

To estimate the extent to which a collision is inelastic, we calculated the effective scattering cross section $d\sigma_{\text{eff}}(\theta,v,\Omega)$ defined in (10) and responsible, as noted in Sec. 2, for the energy exchange between particles and field. Recall that θ is the angle between the velocity $\mathbf{v}$ of the incident particles and the field $\mathbf{E}$ (Fig. 1). The main calculations were done for longitudinal ($\mathbf{v}_-\|\mathbf{E}$) and transverse ($\mathbf{v}_-\perp\mathbf{E}$) collisions for several values of θ: $\pi/8$, $\pi/4$, and $3\pi/8$.

Let us analyze the dependence of the effective scattering cross section for longitudinal collisions on Ωv (Fig. 2).

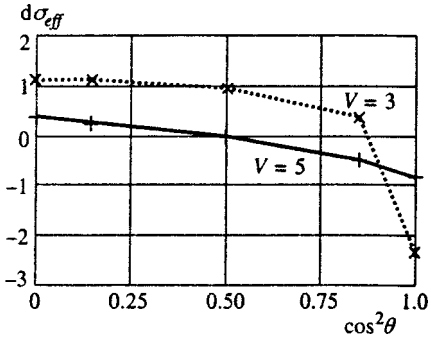

FIG. 3. Dependence of $d\sigma_{\text{eff}}$ on $\cos^2\theta$ at $\Omega = 0.32$ for $v = 3$ and 5.

Below we list the most important features of the behavior of the effective scattering cross section as a function of velocity.

1. In the limit $v \ll 1$ the cross section increases according to the law $d\sigma_{\text{eff}} \sim \Omega^2/v^2$. What is important is that such an approximation is true for all scattering angles θ.

2. For $\Omega \leqslant 1$ the cross section changes sign near $v \sim v_\sim$.

3. In the limit $v \gg v_\sim$ the absolute value of the scattering cross section decreases. In this range a good approximation of the curve at all angles is $d\sigma_{\text{eff}} \sim 1/v^4$.

To determine how θ affects the sign of the scattering cross section, we did a detailed study at $\Omega = 0.32$ for two values of the drift velocity of the scattered beam, $v = 3$ and $v = 5$ (Fig. 3).

The figure shows that the region of solid angles with negative values of $d\sigma_{\text{eff}}$ increases as a function of the drift-to-oscillator velocity ratio. The asymptotic behavior for $v \gg v_\sim$ probably corresponds to the Marcuse effect $\sigma_{\text{eff}} \sim 1 - 3\cos^3\theta$ (see Ref. 6). At the same time, the absolute value of the maximum of the negative value, which is reached at $\theta = 0$, decreases with increasing θ. Within our accuracy for $\Omega = 0.32$ and $\Omega = 1$ we see that in isotropic scattering the total effective scattering cross section remains positive.[2)]

Note that on the whole only a small fraction of the electrons is responsible for the energy exchange. We call these electrons ''representative'' electrons. The effect is illustrated by Fig. 4, in which the regions of substantial energy exchange, more than half the oscillator energy, for longitudinal ($\mathbf{v}_-\|\mathbf{E}$) particle incidence are marked in the (ρ,φ) plane (impact parameter vs. incident phase). The fact that there are vertical lines is due to the discreteness of calculations in impact parameters. Of course, the true ''branches'' are continuous rather than discrete. Estimates show that these particles are responsible for more than 70% of the total energy exchange. The fraction of these particles is small, of order $\Omega^2 \ll 1$. Note that although the representative electrons have different incident phases, all collisions take place at strongly correlated field phases (Fig. 5), which depend on the value of the initial drift velocity. Here the collision correlation is prepared by the ''elastic'' drift of particles in the Coulomb field of the ion. The figure depicts the phase of the field when the electrons are nearest to the ion as a function of the initial phase. All electrons begin their motion along the field with an impact parameter $\rho = 4.6$, which corresponds to represen-

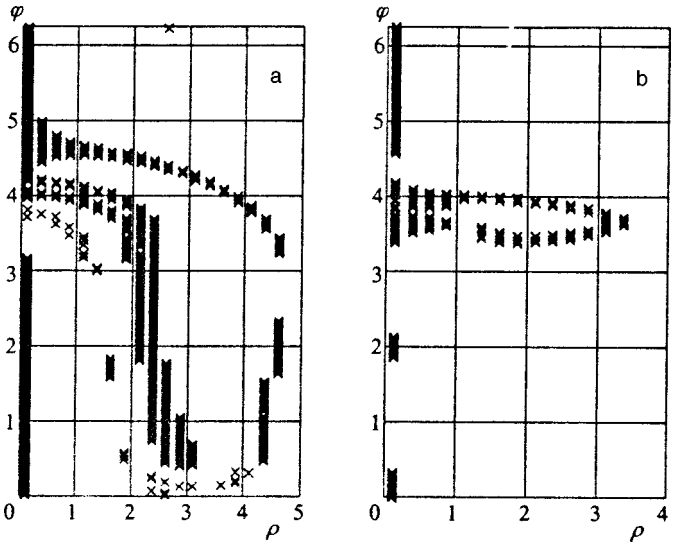

FIG. 4. Regions of substantial energy exchange (points) in the (ρ,φ) plane (impact parameter vs. incident phase) in longitudinal scattering ($\mathbf{v}_{\sim}\|\mathbf{E}$) at (a) $v=1$ and (b) $v=3$.

tative electrons (the vertical bar in Fig. 4). Clearly, the actual energy exchange with the field takes place within a narrow band of field phases. The ''steps'' in this dependence correspond to a shift in the moments of collision by one field period. Figure 6 shows that the dependence of the energy variation as a function of the impact parameter in longitudinal scattering, averaged over the incident phases, is essentially nonmonotonic. The same conclusion can be drawn from Fig. 4. The number of side maxima depends on the ratio of the drift velocity to the oscillator velocity (Ωv) and increases with the decreasing parameter. When $v \gg v_{\sim}$, there is only the central peak. Its width can estimated by the size of the Rutherford region determined from the total incident electron velocity. The contribution of the side maxima is found to be dominant, so that the effective scattering cross section for drift velocities lower than $v_{\sim}$ increases substantially and is found to be of order Ω^2/v^2, the geometric mean of the Rutherford cross sections estimated by the oscillator velocity, $\sigma_{\sim}\sim\Omega^4$, and by the drift velocity, $\sigma_{\rm dr}\sim 1/v^4$.

On the other hand, for high drift velocities (in comparison to the oscillator velocity) the result agrees, as expected, with the small-angle approximation (Fig. 2), so that the scattering cross section is of order $1/v^4$. In this case the dependence on the incident phase is unimportant, and hence the low energy exchange approximation is valid.[1,2,8–10]

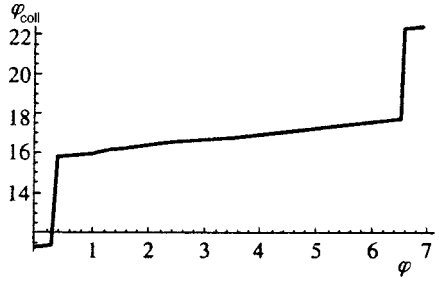

FIG. 5. Dependence of the phase of the field in a ''collision'' on the initial phase of the field.

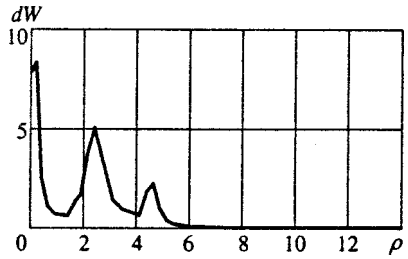

FIG. 6. Dependence of the energy variation on the impact parameter in longitudinal scattering, averaged over the incident phases, at $\Omega=0.32$ and $v=1$.

Calculations show that there are two types of representative electrons. There are electrons whose energy-exchange times are much shorter than the period of the field (''fast'' collisions; Fig. 7a). There are also electrons whose energy-exchange times are comparable to the period of the field (''slow'' collisions; Fig. 7b). As noted earlier, both representative electron types are present not only in head-on collisions but also for significant impact parameters, when it takes one or several field periods to deliver an electron to the inelastic-collision region.

Note that slow collisions occur at velocities of order of the oscillator velocity. Their effectiveness rapidly decreases at low and high velocities. We found that slow collisions mainly lead to a decrease in the electron energy (drift energy) and rapidly disappear as the angle between the direction of the initial drift velocity and the direction of the field increases.

Figure 8 depicts the final dependence of the energy acquired by the electrons on the initial electron drift velocity in transverse scattering. What sets this type of scattering apart from longitudinal scattering is that $\sigma_{\rm eff}$ is positive definite for all velocities of the electrons being scattered. For convenience we use the log–log scale. We see that, as in longitudinal scattering, this dependence is approximated fairly well by functions of the form $\sigma\propto 1/v^2$ at low energies and $\sigma\propto 1/v^4$ at high energies.

Note that this results is not accidental. Dimensional analysis of the scattering problem for an arbitrary incidence angle shows that for all drift velocities of the scattered beam the variation of energy as a function of the dimensionless parameters of the problem obeys the relationship

$$\sigma_{\rm eff}=\frac{\Omega^2}{v^2}f_1(\Omega v,\cos^2\theta)+\Omega^2 f_2(\Omega v,\cos^2\theta). \qquad (18)$$

Here, as computer simulation shows, the first function experiences a sharp drop near $\Omega v=1$ and decreases in proportion to $1/(\Omega v)^2$ at high velocities, while the second function decreases at both high and low velocities. Furthermore, we can expect the second function to always be negative.

4. ELASTIC COLLISIONS

Figure 9 depicts characteristic paths of particles undergoing strong elastic scattering for parameter values Ω

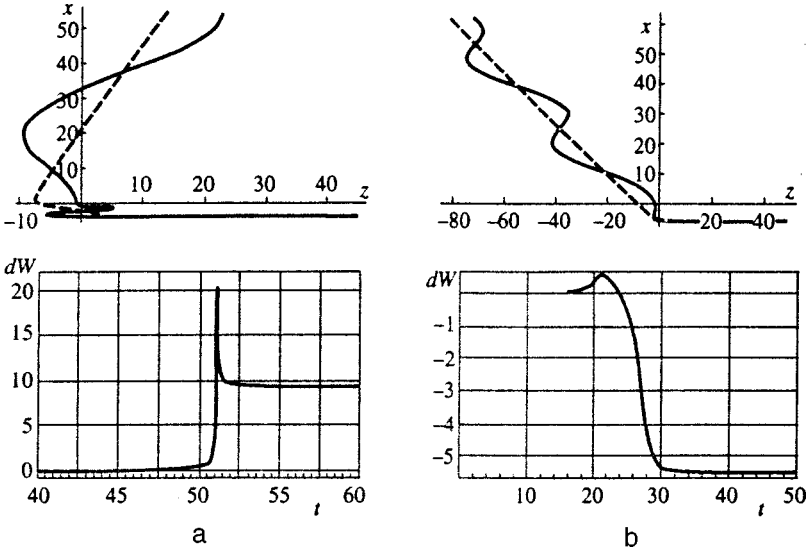

FIG. 7. (a) Fast energy exchange at $\Omega=0.32$ and $v=1$; (b) slow collision at $\Omega=0.32$ and $v=3$. The upper half of the figure depicts the paths of the particle (solid curves) and of the drift center (dashed curves), while the lower half depicts the time dependence of the variation of the electron energy.

$=0.32$ and $v=1$. For the same parameter values but for different incident phases there are electrons undergoing strong inelastic scattering.

Typical curves representing the impact-parameter dependence of the beam transverse energy averaged over the incident phases, $dW_{\rm tr}$ [Eq. (11)], are depicted in Figure 10. Note that these curves are not always monotonic at low drift velocities of the particles, and they suddenly break off at large impact parameters. In particular, when $\rho \gg r_{\sim}$, their asymptotic behavior is clearly that of $1/\rho^2$, which corresponds to scattering in the static field of the ion. Thus, as expected, the contribution of the high-frequency field to the transport scattering cross section decreases at large distances in an integrable manner. When the scattering is not longitudinal, the dependence of the averaged energy on the impact parameter is anisotropic (see Fig. 11).

Qualitatively, such behavior is quite understandable if instead of the problem of elastic scattering in a strong high-frequency field we consider the problem of scattering in the drift velocity by an averaged potential whose analytical expression is[14]

$$V(\mathbf{r})=-\frac{2}{\pi}\frac{1}{\sqrt{r_+ r_-}}K\left(\sqrt{\frac{1}{2}\left(1-\frac{\mathbf{r}_+\mathbf{r}_-}{r_+ r_-}\right)}\right), \tag{19}$$

where $\mathbf{r}_{\pm}=\mathbf{r}\pm\mathbf{r}_{\sim}$, and $K(m)$ is the complete elliptic integral of the first kind. Here the side maxima can be related to the presence of two scattering centers. In particular, as the energy of the scattered beam increases, the side maxima disappear, since for high-energy beams the spherically symmetric part of the scattering potential, the Coulomb potential, is the most significant part of the potential. The constant-averaged-potential surfaces are axisymmetric (the symmetry axis is parallel to the external field) and are shaped as ''dumbbells'' with a logarithmic divergence $\ln(1/\rho)$ at the center and a square-root singularity $1/\sqrt{r_{\pm}r_{\sim}}$ at $z\sim\pm r_{\sim}$.

Studies of the dependence of the transport scattering cross section $d\sigma_{\rm tr}$ on the angle θ between the electric field and the drift velocity have shown that even at fairly low velocities (higher than v_E in dimensional variables) there is a weak dependence on θ. Substantial anisotropy is observed only at low velocities. In this case it is found that the average-potential model is insufficient, i.e., inelastic effects are important in such collisions. To support this statement we have drawn the lines of constant $d\sigma_{\rm tr}$ for two frequency values (Fig. 12). Clearly, appreciable anisotropy is observed only at fairly low velocities.

The results of the calculations have been used to construct curves representing the dependence of the total transport scattering cross section $\sigma_{\rm tr}$ (i.e., the differential cross section (11) was integrated over the solid angles) on the initial particle velocity (see Fig. 13). For comparison we also depict the corresponding dependence of the Rutherford scattering cross section determined from the drift velocity, $\sigma_R=4\pi/v^4$ (dotted curve). Qualitatively, the plot of $\sigma_{\rm tr}$ vs. v can be divided into three parts.

1. $v \gg v_{\sim}$. Small-angle elastic scattering provides the main contribution.[9,10]

2. $\Omega \leqslant v \leqslant v_{\sim}$. In addition to small-angle elastic scattering, inelastic processes (see Sec. 3), responsible for the nonmonotonic dependence of $\Delta W_{\rm tr}$ on ρ, begin to contribute (cf.

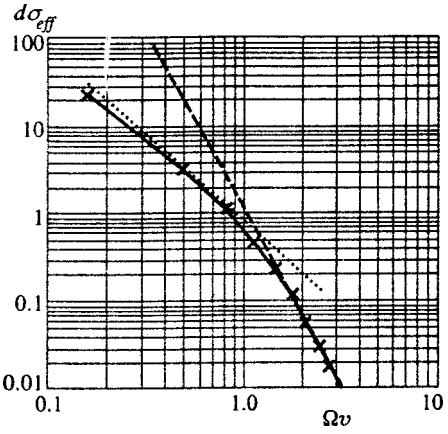

FIG. 8. Velocity dependence of the effective inelastic scattering cross section in transverse scattering (on the log–log scale) at $\Omega=0.32$; the dotted and dashed curves represent the $d\sigma_{\rm eff}$ vs. $1/v^2$ and $d\sigma_{\rm eff}$ vs. $1/v^4$ plots.

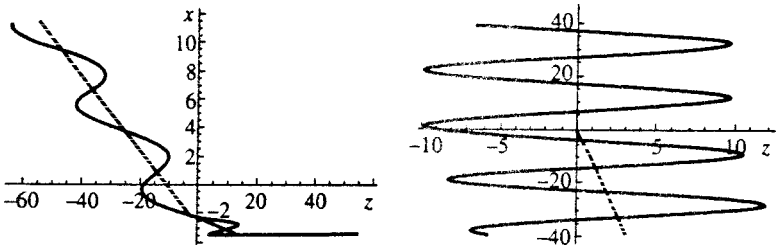

FIG. 9. Characteristic paths of a particle scattered by a Coulomb center for different values of the impact parameter ($\Omega=0.32$ and $v=1$).

Figs. 10 and 6). However, since $\sigma_{\text{eff}}\sim\Delta W\sim 1/v^2$ [see Eq. (18)], the contribution of inelastic collisions is of the same order as that of elastic (small-angle) collisions.

3. $v\leqslant\Omega$. Quasicapture (trapping) of particles near a Coulomb center for a time much longer than one field period probably become important. Because of the complexity of the calculations, we have not studied this range of parameters thoroughly enough, since the requirements with respect to precision become more stringent due to the need to analyze the long-term evolution of each electron. These results demand further investigations.

The most interesting feature of these curves is the remarkable coincidence (embracing the huge drop in σ and v) with the Rutherford scattering cross section, calculated from the drift velocity. This result is even more remarkable if we account for the fact that the impact-parameter dependence of the average transverse energy (Fig. 10) is highly nonmonotonic for velocities lower than the oscillator velocity.

5. CONCLUSION

The main idea of the present investigation is that inelastic collisions of electrons with ions in ultrahigh fields involve, in a coherent manner, only the representative electrons, which land in the r_E-neighborhood of the ion with the same values of the phase of the field (more precisely, with almost the same values). Bearing this in mind, below we formulate the main results of our work.

1. Elastic collisions (the transport scattering cross section) are unaware, so to speak, of the external field (Fig. 13). This cross section is close to the Rutherford cross section, calculated from the drift velocity. Formally, this result can be obtained even in the small-angle approximation.[8–10] However, a detailed investigation has shown (Sec. 4) that to a certain extent this coincidence is not an obvious fact, since a comparable contribution to the transport scattering cross section is provided by particles inelastically scattered to large angles (relatively small impact parameters). The dependence on the scattering angle at velocities much lower than the oscillator velocity is substantially anisotropic. Nevertheless, down to the very low velocities at which quasicapture processes become significant, integration over the angles, i.e., the transition to the total transport scattering cross section, leads to the Rutherford formula. Our calculations have shown that above these velocities the total scattering cross section is approximated fairly well by the expression

$$\sigma_{\text{tr}}=\frac{4\pi}{v^2(v^2+\Delta(\Omega,v))},$$

where $\Delta(\Omega,v)$ is a correction that is essential when $v\leqslant\Omega$ (the correction requires further refinement).

2. Inelastic collisions at drift velocities lower than the oscillator velocity are chiefly determined by a small fraction of the electron, the representative electrons. Energy exchange between these electrons and the field occurs at correlated moments in time corresponding to the arrival of all these electrons with a predetermined field phase in a small neighborhood of an ion. Outside this neighborhood there is adiabatic drift of the oscillating electrons due to the focusing properties of the Coulomb potential.

At high velocities (in comparison to the oscillator velocity), our results are in good agreement with those of the small-angle approximation in the Dawson–Oberman–Silin (DOS) model.[1,2,10] What makes them different is the absence of a logarithmic factor, whose presence in the small-angle approximation (the DOS model) is due to small impact parameters. In terms of dimensionless variables, the result of the DOS model is $\sigma_{\text{eff}}\propto 1/v^4$. At low velocities, representative electrons provide the leading contribution to the effec-

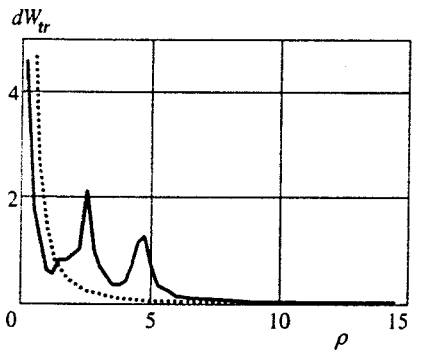

FIG. 10. The averaged transverse energy as a function of the impact parameter ρ (longitudinal incidence) with $\Omega=0.32$ and $v=1$. The solid curve represents the case of longitudinal incidence and the dotted line, the case of a zero high-frequency field.

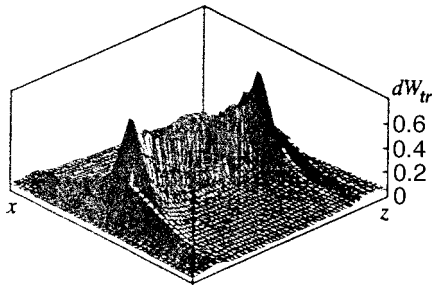

FIG. 11. The averaged transverse energy as a function of the impact parameter vector $\boldsymbol{\rho}$ (longitudinal incidence) with $\Omega=0.32$ and $v=1$.

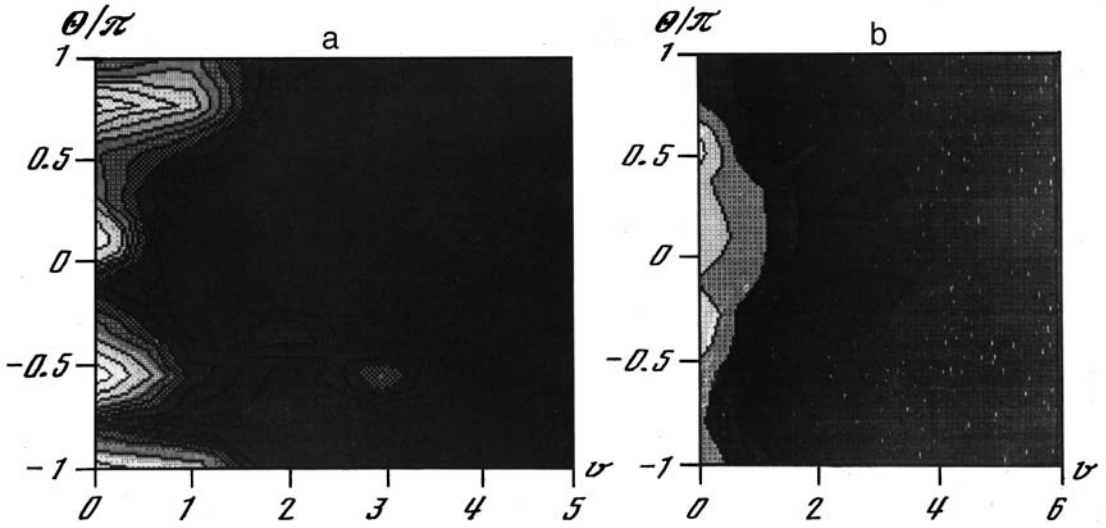

FIG. 12. The transport scattering cross section as a function of the incidence angle (along the vertical axis) and velocity (along the horizontal axis) with $\Omega=0.1$ (to the left) and $\Omega=0.32$ (to the right).

tive scattering cross section (see Sec. 3). Among representative electrons there are slow electrons, with energy-exchange times of order field period, and fast electrons, with energy-exchange times much shorter than one field period. Slow collisions are important only at drift velocities comparable to the oscillator velocity.

At lower velocities the leading contribution is provided by fast collisions, which on the face of it are described fairly well by the low-frequency approximation.[3,4] This model leads, in terms of dimensionless variables, to values $\sigma_{\text{eff}}\sim\Omega^4$, i.e., to the Rutherford scattering cross section, estimated from the oscillator velocity. In the same range we arrived at a much larger quantity, $\sim\Omega^2/v^2$, the geometric mean of the scattering cross sections obtained in the low-frequency and small-angle approximations. These differences probably arise because the low-frequency approximation does not allow for ''coherent aiming,'' where the electrons do not change their drift energy and are attracted to an ion in the course of one or several field periods, and then are rapidly scattered, changing their drift energy in the process. These effects lead to a significant increase in the effective scattering cross section compared to the results provided by standard models.

Formally our results are valid only in an extremely low-density plasma, $N\ll(\lambda v/c)^{-3}$, where λ is the radiation wavelength. It can be expected, however, that similar effects occurs in a dense plasma. Indeed, the one thing that is important is that in the volume defining the effecting scattering region, $r_\sim\sigma_{\text{eff}}$, there was no more than one particle. If we take into account our numerical results for the effective inelastic scattering cross section, this statement is equivalent to

$$r_e\ll r_D \text{ for } v_T\ll v_\sim, \tag{20}$$

where r_D is the Debye radius. Under these conditions, we can make an estimate of the effective collision frequency in dimensional variables:

$$\nu_{\text{eff}}\sim\frac{\omega_p}{nr_D^3}\frac{v_T^2}{v_\sim^2}. \tag{21}$$

To within a logarithmic factor, this relationship arises because the effective scattering cross section found in the course of our investigation is equal to the geometric mean of the effective scattering cross sections estimated by the thermal velocity and by the oscillator velocity.

In conclusion we note that the conventional ideas undergo a marked transformation only for very strong fields, i.e., for $r_\sim\gg r_E$ $(\omega_E\gg\omega)$, or in dimensional variables,

$$\omega\ll\omega_E\sim Z^{1/4}\times 2\times 10^{10}\left(P\left[\frac{\text{W}}{\text{cm}^2}\right]\right)^{3/8}. \tag{22}$$

One can expect that in such fields the plasma is a non-

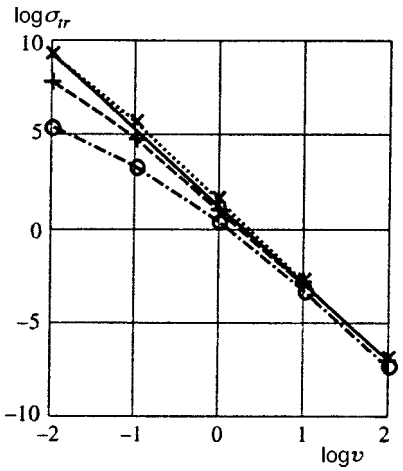

FIG. 13. The transport scattering cross section as a function of the initial particle velocity for three values of the frequency Ω: 1 (○), 0.32 (+), and 0.1 (×).

equilibrium medium (σ_{eff} may change sign in relation to the process of generation of harmonics of the incident radiation); the spectrum of the bremsstrahlung harmonics becomes much richer, stimulated recombination of electrons and ions may occur, etc. We believe that these and other obvious problems, such as an analytical model of elastic and inelastic collisions in ultrahigh fields in a disperse plasma, collisions in a dense plasma, and generation of harmonics of ultrastrong radiation in a dense plasma, will be the most promising topics of future investigations.

The authors are grateful to the staff of the Department of Plasma Physics of the Institute of Applied Physics, Russian Academy of Sciences and especially to M. A. Miller, V. B. Gil'denburg, A. G. Litvak, and V. V. Semenov for fruitful discussions. The authors would also like to express their gratitude to N. J. Fisch (PPL, Princeton, USA) for stimulating discussions. The work was made possible by grants from the Russian Fund for Fundamental Research (Grants Nos. 96-02-019482 and 98-02-17205).

*)E-mail: fraiman@appl.sci-nnov.ru

1)This is not a stringent restriction. Debye screening leads to similar effects.

2)The final answer to the question of whether negative values of the total cross sections σ at low frequencies and an isotropic monoenergetic distribution function for scattered electrons can exist has yet to be found. We find no reason in principle why this is impossible for smaller values of Ω.

[1] J. Dawson and C. Oberman, Phys. Fluids **5**, 517 (1962).
[2] V. P. Silin, Zh. Éksp. Teor. Fiz. **47**, 2254 (1964) [Sov. Phys. JETP **20**, 1510 (1965)].
[3] N. M. Kroll and K. M. Watson, Phys. Rev. A **8**, 804 (1973).
[4] F. V. Bunkin, A. E. Kazakov, and M. V. Fedorov, Usp. Fiz. Nauk **107**, 559 (1972) [Sov. Phys. Usp. **15**, 416 (1972)]; F. V. Bunkin and M. V. Fedorov, Zh. Éksp. Teor. Fiz. **49**, 1215 (1965) [Sov. Phys. JETP **22**, 844 (1966)].
[5] M. H. Mittleman, *Introduction to the Theory of Laser–Atom Interactions*, Plenum Press, New York (1993).
[6] M. V. Fedorov, *Interaction of Intense Laser Light with Free Electrons*, Harwood Academic Publishers, Chur (1991) [Russian original, Nauka, Moscow (1991)].
[7] S. A. Akhmanov, V. A. Vysloukh, and A. S. Chirkin, *Optics of Femtosecond Laser Pulses*, AIP Press, New York (1992) [Russian original, Nauka, Moscow (1988).
[8] C. D. Decker, W. B. Mori, J. M. Dawson, and T. Katsouleas, Phys. Plasmas **1**, 4043 (1994).
[9] G. J. Pert, J. Phys. B **12**, 2756 (1979).
[10] G. Shvets and N. J. Fisch, Phys. Plasmas **4**, 428 (1997).
[11] L. D. Landau and E. M. Lifshitz, *Mechanics*, 3rd ed., Pergamon Press, Oxford (1976).
[12] L. Wiesenfeld, Phys. Lett. A **144**, 467 (1990).
[13] E. L. Steifel and G. Scheifele, *Linear and Regular Celestial Mechanics*, Springer, Berlin (1971).
[14] M. Pont, N. R. Walet, and M. Gavrila, Phys. Rev. A **41**, 477 (1990).

Translated by Eugene Yankovsky

Transformation of strong picosecond pulses in radiation with an extended quasirotational spectrum during self-focusing in high-pressure hydrogen

V. B. Morozov, A. N. Olenin, and V. G. Tunkin*)

International Science-Educational Laser Center, M. V. Lomonosov Moscow State University, 119899 Moscow, Russia

(Submitted 25 June 1999)
Zh. Éksp. Teor. Fiz. **115**, 479–493 (February 1999)

Radiation with a spectrum representing a discrete analog of the extended spectrum observed in the generation of a supercontinuum in gases is generated in the self-focusing of 30-ps pulses with a wavelength of 1.06 μm in hydrogen at pressures up to 120 atm. The spectrum contains lines with similar intensities, an average frequency spacing approximately equal to the rotational transition frequency in hydrogen (587 cm^{-1}), and a smooth spatial profile. The lines consist of several vibrational–rotational components. As the pressure is increased, the spectral lines are transformed so that at a pressure above 60 atm each line in the spectrum contains one or two components formed as a result of the smaller number of cascade (rotational and vibrational) processes. Self-focusing is manifested in the occurrence of a radiating channel up to 12 cm in length. The formation of a channel of this length is associated mainly with the variation of the refractive index in vibrational excitation of the hydrogen molecules by electrons heated in the pump field.
[S1063-7761(99)00802-1]

1. INTRODUCTION

The generation of coherent optical radiation with a spectrum consisting of equidistant spectral lines by means of vibrational, rotational, or combined vibrational–rotational stimulated Raman scattering (SRS) and resonance parametric processes, specifically in gaseous hydrogen, has been reported in a multitude of papers.[1–18] Cascade vibrational lines in the Stokes and anti-Stokes regions have been observed in the presence of a linearly polarized pump, because the vibrational Raman gain is a maximum in this case.[1–4] For a circularly polarized pump isotropic scattering is eliminated, the vibrational Raman gain decreases, and rotational SRS becomes dominant. Cascade rotational lines have been observed using a circularly polarized pump.[5] Several authors have reported the detection of cascade vibrational lines with rotational lines on both sides of the vibrational lines[6–9] as well as cascade rotational lines[10,11] after the pumping of hydrogen by an elliptically polarized wave[6,10] or with the application of a polarized biharmonic pump having a difference frequency equal to the frequency (587 cm^{-1}) of the hydrogen rotational transition $S_0(1)$ (Refs. 7, 8, and 11). Kawasaki *et al.*[7] have used radiation from two dye layers for the same purpose. Others[8,11] have generated frequency-shifted radiation in an auxiliary cell, in which pure rotational SRS was induced by circular polarization of the pump wave. The resulting SRS was steady-state in some of the cited papers and time-dependent in others. The first and second Stokes vibrational components have been efficiently generated[12] in connection with time-dependent SRS in hydrogen at pressures up to 40 atm pumped by pulses of duration 30 ps, wavelength 1.06 μm, and energy 90 mJ; to avoid optical breakdown, lenses having focal lengths of at least 1 m were used to focus the pump beam into the hydrogen-filled cell. The maximum pressure of the hydrogen in Refs. 1–12 was 40 atm.

If the series of vibrational–rotational lines, which extend all the way from the infrared to the ultraviolet in many cases, do not contain any parts having a distinctive intensity, these series can be regarded as a discrete analog of the extended spectrum observed in the generation of a supercontinuum in gases. The generation of a supercontinuum in gases has been studied in detail in several papers,[13–18] where it has been shown to be associated with self-focusing of the pump beam. In some papers on supercontinuum generation[13–15] the significant broadening of the frequency spectrum has been accompanied by only a slight variation of the spatial spectrum, the divergence of the beam after the gas-filled cell being almost equal to its convergence entering the cell.

We have previously reported that when linearly polarized picosecond pulses with wavelength 1.06 μm and energy up to 40 mJ are self-focused in hydrogen at high pressure (up to 120 atm), radiation is generated[19] with an extended spectrum consisting of a series of single-component or two-component vibrational–rotational lines having similar intensities within the visible range, an average frequency interval between lines (593 cm^{-1}) approximately equal to the rotational transition frequency in hydrogen (587 cm^{-1}), and a smooth spatial profile.

In this paper we give the results of an experimental study of the distinctive features of the structural formation of the frequency and spatial spectra of radiation generated by rotational and vibrational SRS and by parametric cascade processes in the presence of self-focusing of strong picosecond pulses in hydrogen at pressures up to 120 atm.

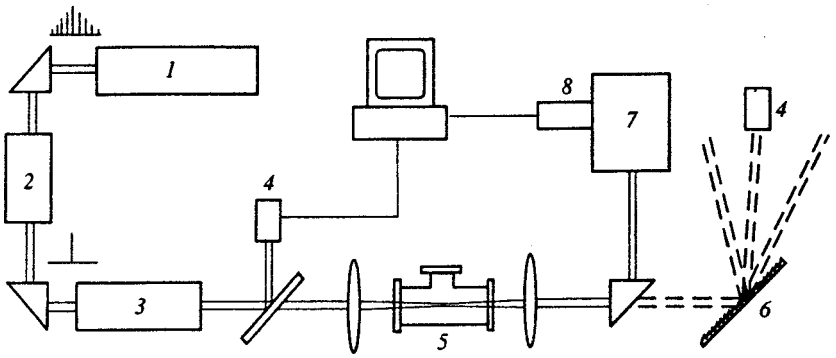

FIG. 1. Experimental setup: (*1*) passive mode-locked picosecond Nd:YAG laser oscillator with negative feedback ($\lambda_p = 1.06\ \mu$m, $\tau_p = 30$ ps); (*2*) electrooptical modulator; (*3*) amplifiers; (*4*) pyroelectric detector; (*5*) hydrogen-filled cell; (*6*) diffraction grating; (*7*) diffraction monochromator; (*8*) multichannel optical analyzer.

2. EXPERIMENTAL SETUP AND OPTICAL SYSTEM

The experimental setup is shown schematically in Fig. 1. A passively mode-locked Nd:YAG driving laser with negative feedback generates a train of 10–12 pulses of duration 30 ps. A solitary pulse is isolated from the middle of the train. Two $LiTaO_3$ electrooptical modulators are used to enhance the contrast of the isolated pulse against the background. The solitary pulse is amplified in a system of amplifiers to an energy $\leqslant 40$ mJ. The contrast of the amplified solitary pulse relative to the background is at least 2×10^3 in intensity and 10^2 in energy. This linearly polarized pulse is focused into a cell of length 25 cm containing compressed hydrogen. The diameter of the pump beam at the focusing lens is 4.4 mm. The hydrogen pressure is varied from 0 to 120 atm. The radiation emitted from the cell is resolved into its spectrum by diffraction gratings, and the energy distribution among the spectral lines is measured by means of a pyroelectric detector. The detailed structure of the spectrum is is investigated by means of a diffraction monochromator and (depending on the spectral range) a multichannel optical analyzer by photomultiplier. The cell has a side window, through which the radiating channel created by self-focusing of the pump pulses can be observed. The lateral luminescence spectrum of this channel is recorded by using a multichannel optical analyzer by projecting the image of the channel onto the slit of a diffraction monochromator.

3. EXPERIMENTAL RESULTS

At a sufficiently high hydrogen pressure p and pump energy W_p the radiation emitted from the cell is an almost pure white beam with a $1.5-2°$ divergence, which exceeds the divergence of the pump beam after the cell (0.9°). This radiation is expanded into its spectrum by a 600 lines/mm grating for recording in the IR region or by a 1200 lines/mm grating for recording in the visible and UV regions. The spectral dependence obtained by means of the pyroelectric detector is shown in Fig. 2, in which the orders i of the lines are read from the pump. At a pump energy of 30 mJ the energy of each line in the green region is approximately 10 μJ. The graphs of the spectra obtained by means of the 600 lines/mm and 1200 lines/mm gratings are matched at line $i=10$ on the assumption that the reflection coefficients of the two gratings are identical in this spectral region. The spectrum shown in Fig. 2 is similar in the width and shape of its envelope to the supercontinuum emission spectrum generated in gases[13–18] and is a discrete analog of it. It will be shown below that for a hydrogen pressure >60 atm and a sufficiently high pump energy each line in the investigated spectrum consists of one or two components with an average line spacing approximately equal to the rotational transition frequency in hydrogen ($587\ \text{cm}^{-1}$). Vibrational lines cannot be distinguished by their energies. As the pump energy and the hydrogen pressure are increased, the energies of the lines increase across the board without any change in their ratio to one another. We call this series of spectral lines a quasirotational spectrum.

The far-field beam profiles and, hence, their divergences are determined by means of the multichannel optical analyzer. The beam profiles of the various lines are smooth (Fig. 3). The divergence of the pump beam for the evacuated cell is 0.75°, and the divergences of the pump beam and the red-line ($i=10$, $\lambda=0.65\ \mu$m) and green-line ($i=16$, $\lambda=0.53\ \mu$m) beams at $p=100$ atm are equal to 0.9°, 1.5°, and 1.7°, respectively.

A quasirotational spectrum is observed when the pump beam is focused into the hydrogen-filled cell by lenses having focal lengths $F=25-35$ cm. When lenses with $F\geqslant 50$ cm are used, the output radiation undergoes a radical change of character, acquiring the shape of rings corresponding to the generation of a series of vibrational lines. To avoid optical breakdown of the cell windows during the experiments using

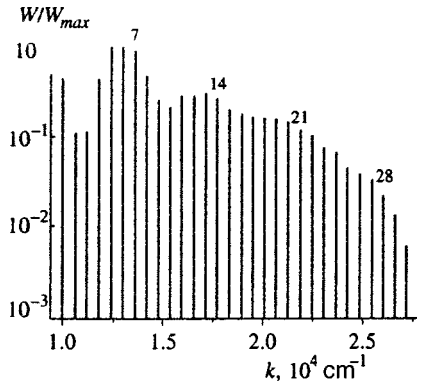

FIG. 2. Energy of the lines of the quasirotational spectrum at a hydrogen pressure of 100 atm and a pump energy of 30 mJ. The lines are numbered from the pump. Lines 1–10 were recorded with a 600 lines/mm grating, and lines 10–30 were recorded with a 1200 lines/mm grating. The energy of each line in the green region is ≈ 10 μJ ($W_{\max}$ is the energy of the strongest line).

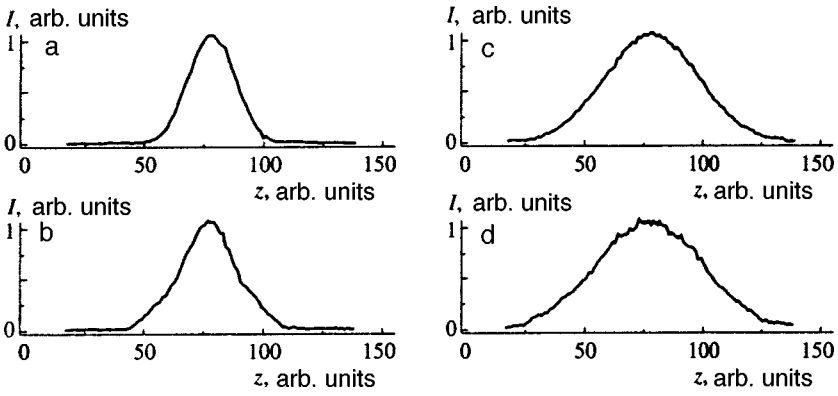

FIG. 3. Spatial beam profiles for various hydrogen pressures, measured with the projection of far-field images on matte glass onto the multichannel optical analyzer: (a) 1.06-μm pump, $p=0$ atm; (b) the same, $p=100$ atm; (c) red line ($i=10$, $\lambda=0.65$ μm) at $p=100$ atm; (d) green line ($i=16$, $\lambda=0.53$ μm) at 100 atm (z is the transverse coordinate).

lenses with $F\geqslant 50$ cm, the cell is lengthened by means of special adapters.

A quasirotational spectrum is not observed when the quality of the spatial profile of the pump radiation deteriorates as a result of poor adjustment of the experimental apparatus, breakdown of the surface of the active element, or other factors. The experimental setup has been designed with considerable attention to the attainment of a good spatial profile on the part of the pump radiation; in particular, spatial filters are used in the amplification channel with this objective in mind.

Cascade multiplication of the Stokes components of the rotational and vibrational SRS by virtue of four-photon parametric processes plays a significant role in the formation of the series of vibrationa;–rotational lines. A quantum diagram of the processes of generation of the first Stokes rotational and vibrational SRS components and several components in the anti-Stokes region relative to the pump is shown in Fig. 4.

The initial stage in the formation of the total spectrum is the generation of the first Stokes rotational and vibrational components. The experimentally measured dependences of their energy on the hydrogen pressure for pump energies are shown in Figs. 5 and 6, respectively. The energy reaches its maximum value at a certain pressure $p_{\rm sat}$ (in these figures the energies of the components are normalized to their saturation values). For rotational SRS the pressure $p_{\rm sat}$ decreases as W_p increases, the product $p_{\rm sat}W_p$ remaining almost constant (Fig. 7).

A different situation is encountered for the first Stokes vibrational component (Fig. 6). For $W_p=14$ mJ its energy reaches saturation at hydrogen pressures attainable in our cell, specifically at $p_{\rm sat}=110$ atm. For this component $p_{\rm sat}$ remains unchanged with any further increase in W_p.

When the hydrogen pressure is higher than $p_{\rm sat}$ for rotational SRS, cascade rotational components are observed, their intensity decreasing with distance from the pump line and their positions extending to the yellow region of the spectrum. As the pressure is further increased, vibrational components emerge with cascade rotational components on both sides. Various vibrational–rotational components are grouped into lines. In analyzing the detailed structure of the spectrum of these lines over a wide range of pressures, we

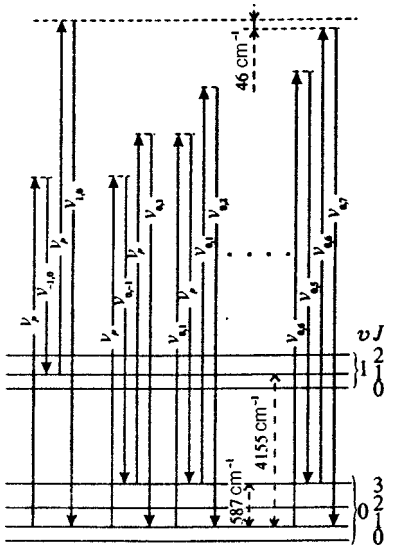

FIG. 4. Quantum diagram of the processes of generation of the first Stokes vibrational $\nu_{-1,0}$ and rotational $\nu_{0,-1}$ SRS components and the components $\nu_{1,0}$, $\nu_{0,1}$, $\nu_{0,2}$, ..., $\nu_{0,7}$ as a result of four-photon parametric processes.

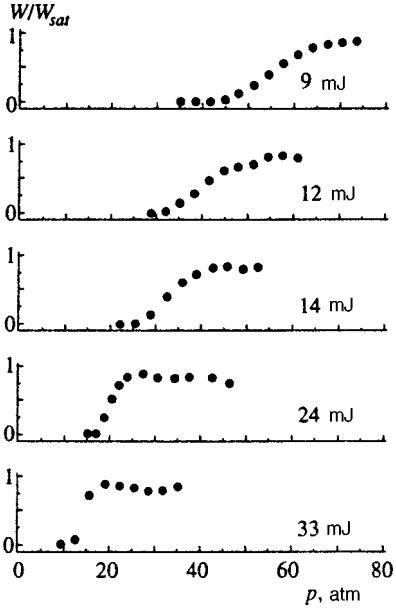

FIG. 5. Dependence of the energy of the first Stokes purely rotational component ($\lambda=1.13$ μm) on the hydrogen pressure at various pump energies. For each graph the energy of the Stokes rotational component is normalized to its maximum value ($W_{\rm sat}$ is the energy of the component at $p_{\rm sat}$).

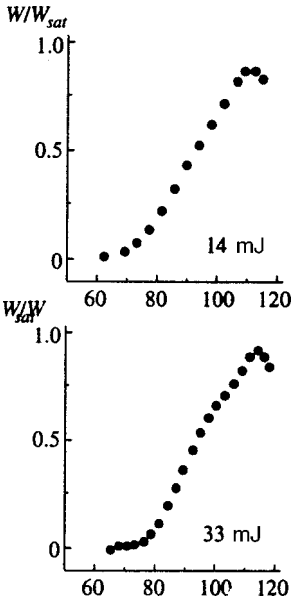

FIG. 6. The same as Fig. 5 for the first Stokes vibrational component ($\lambda = 1.91$ μm).

have found that the frequencies of all the observed components are described by the relation

$$\nu_{n,m} = 9397\ \mathrm{cm}^{-1} + n \cdot 4155\ \mathrm{cm}^{-1} + m \cdot 587\ \mathrm{cm}^{-1}, \qquad (1)$$

where $4155\ \mathrm{cm}^{-1}$ and $587\ \mathrm{cm}^{-1}$ are the vibrational and rotational transition frequencies in hydrogen, respectively, and n and m are the orders of the vibrational and rotational cascade processes (they can be either positive or negative). We label the various components (n, m) to indicate the orders of the corresponding cascade processes.

The hydrogen molecule is distinguished by the property that the vibrational transition frequency ($4155\ \mathrm{cm}^{-1}$) is $46\ \mathrm{cm}^{-1}$ higher than seven times the rotational transition frequency ($587\ \mathrm{cm}^{-1}$). As the pressure is varied, the component constituency of the lines is transformed, but the distance between consecutive components is always $46\ \mathrm{cm}^{-1}$. The ith line contains components (n, m), where i, n, and m are bound by the relation $i = 7n + m$. Figure 8a shows the transformation of the spectral line $i = 13$, Fig. 8b shows the same for line $i = 14$, and Fig. 8c corresponds to $i = 17$. All the spectra in these figures are normalized to their values at the maximum. In Fig. 8a the $(0, 13)$ component is the 13th purely rotational cascade component. Initially, as the pressure is increased, this component becomes stronger than the others, and then, with a further increase in the pressure, it and the $(1, 6)$ and $(3, -8)$ components are suppressed, leaving only the $(2, -1)$ component. The $(0, 14)$ component in Fig. 8b is the 14th purely rotational cascade component. As the pressure is increased, the structure of the line is transformed so as to suppress the $(0, 14)$, $(1, 7)$, and $(3, -7)$ components and leave only the $(2, 0)$ component. The situation with lines $i = 13, 14, 17$ is typical: At a pressure above 60 atm for not too low a pump energy the transformation of the structure of the spectra ceases, and the spectral lines become essentially single-component or two-component lines (a quasirotational spectrum is formed). In the pressure range above 60 atm the energies of the various lines in the visible and UV regions of the spectrum increase with increasing pressure in the same way as the energy of the first Stokes vibrational component.

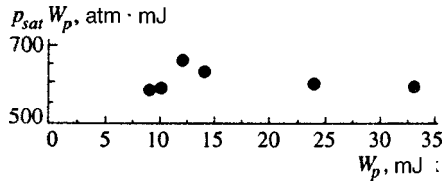

FIG. 7. Dependence of $p_{\mathrm{sat}} W_p$ on the pump energy W_p.

The quasirotational spectrum is formed in the event of self-focusing of the pump, which results in the formation of a radiating self-focusing channel, and the latter is observed through the side window of the cell. The spectrum of the lateral emission of this channel in the spectral sensitivity range of the analyzer mimics the quasirotational spectrum recorded at the cell output. The lateral emission of the channel is perceived as white light. We are therefore looking at the scattering of radiation generated in the self-focusing channel. This kind of scattering was investigated in an early paper on the laser spark,[20] where it was established that the scattering is attributable to Fresnel reflections form the interfaces of regions with different electron densities. The higher the energy of the quasirotational spectrum at the cell output, the brighter is the emission of the channel. As a rule, the middle of the channel is the brightest, the brightness diminishing toward the beginning or end.

As the pressure and the pump energy are increased, the end of the channel scarcely moves at all and occupies the same position as the linear focus. The beginning of the channel, on the other hand, shifts toward the focusing lens as the pump energy and the pressure are increased. For $W_p = 25$ mJ and $p = 100$ atm the length of the channel is ≈ 12 cm, and the focusing lens must therefore be shifted along the axis of the beam so that the beginning of the channel can be determined.

The diameter of the channel is measured by projecting its lens-magnified image onto the multichannel optical analyzer so that the image is perpendicular to the light-sensitive rule. For $W_p = 25$ mJ and $p = 100$ atm the diameter of the channel at its widest middle section is approximately 200 μm. The diameter of the channel decreases with increasing distance from the middle, becoming equal to 80 μm near the linear focus.

4. DISCUSSION OF THE RESULTS

The initial stage in the formation of the quasirotational spectrum is stimulated Raman scattering by rotational ($v = 0$, $J = 1 \rightarrow v = 0$, $J = 3$) and vibrational ($v = 0$, $J = 1 \rightarrow v = 1$, $J = 1$) transitions of hydrogen. The transformation to the quasirotational spectrum takes place inside the channel formed in self-focusing of the pump beam. The excitation of hydrogen molecules by electrons heated in a strong laser field has a significant influence on the self-focusing conditions.

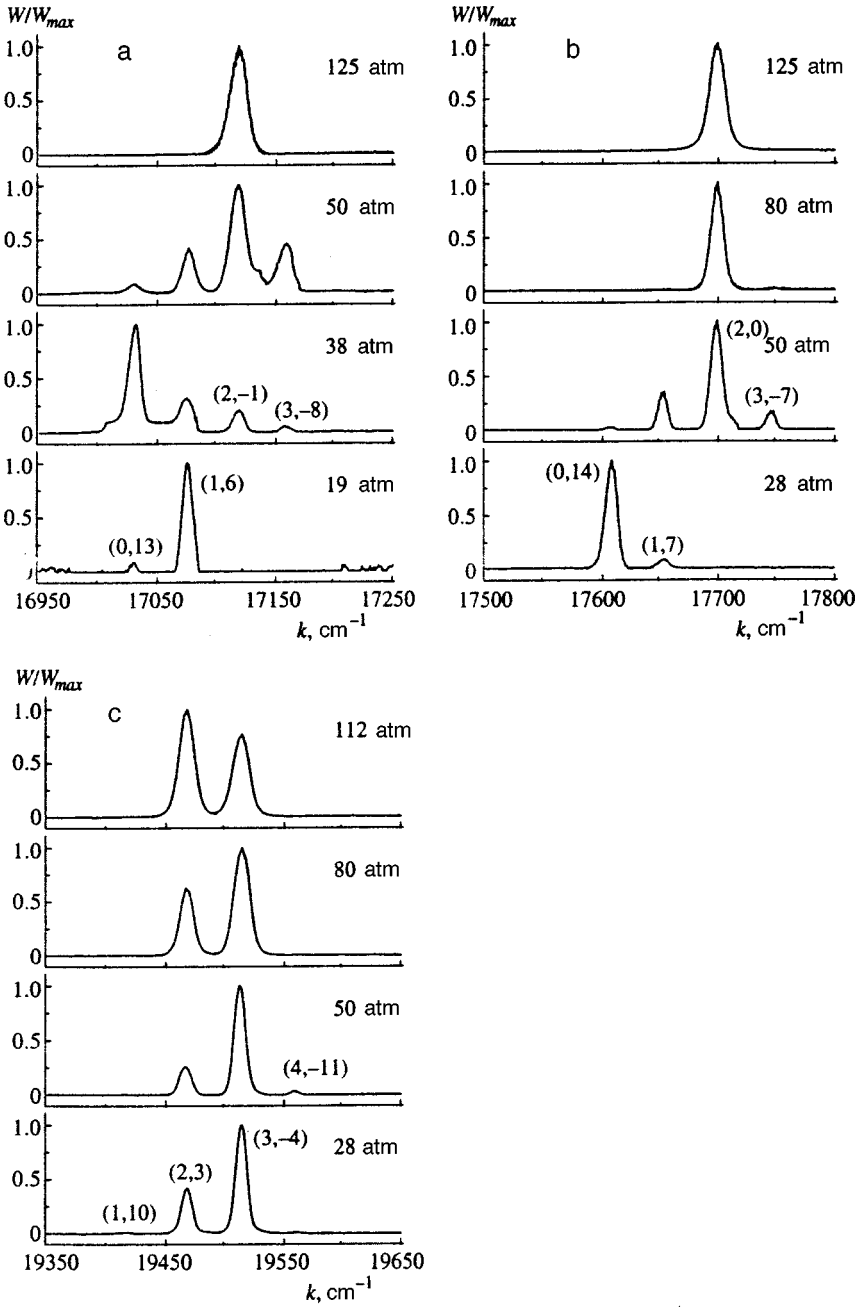

FIG. 8. Transformation of the structure of the spectrum of quasirotational lines as the hydrogen pressure is increased at $W_p=40$ mJ: (a) $i=13$; (b) $i=14$; (c) $i=17$. Each spectrum is normalized to its maximum value.

4.1. Characteristics of Rotational and Vibrational Stimulated Raman Scattering

The SRS observed in the present study is time-dependent. First, even at $p=100$ atm the pump pulse is shorter than the dephasing time T_2, which at this pressure is equal to 32 ps for the rotational transition and 62 ps for the vibrational transition.[21,22] Second, in the ensuing discussion we address the fact that the formation of the self-focusing channel is caused by a moving focus, in the presence of which the time of interaction of the molecules with the field becomes shorter than the pulse duration. For time-dependent SRS the peak intensity $I_s^{\max}$ of the Stokes component at the exit from a nonlinear medium of length L is given by the expression[23,24]

$$I_s^{\max}=I_0\exp\left[\left(A\omega_s|M_{ij}|^2LN\int I_p dt\right)^{1/2}\right], \tag{2}$$

where A is a constant, ω_s is the frequency of the Stokes wave, M_{ij} is the matrix element for Raman scattering by the $i\rightarrow j$ transition, N is the density of hydrogen molecules, and I_p is the pump intensity. It follows from Eq. (2) that the peak intensity $I_s^{\max}$ is given by the product

$$N\int I_p dt=N\overline{I_p}\tau_p\sim NW_p\,.$$

If we assume that as the pressure or the pump energy increases, $I_s^{\max}$ must increase from the initial level I_0 (e.g., the noise level) to a fixed value dictated by saturation, we infer

from Eq. (2) that the product of the pressure p_{sat} at which I_s^{max} saturates and the pump energy W_p must be independent of W_p, as is indeed th case for rotational SRS (Fig. 7).

The saturation of rotational scattering at a high pump energy is obviously attributable to transfer of a large fraction of the molecules in the interaction zone into coherent rotation. The strong emissions of the pump and the Stokes wave are capable of coherently exciting in the interaction zone a large portion of the molecules situated at the first rotational level. Following Ref. 25, we write the expression for the Rabi frequency of two-photon resonance nutations:

$$\Omega_R^2=\frac{16\pi^2c^2}{\hbar^2\omega_s^4}I_pI_s\frac{d\sigma}{do}. \tag{3}$$

Here $d\sigma/do$ is the Raman scattering cross section of the transition. Based on the cross section of the $1\rightarrow3$ rotational transition of hydrogen at a laser wavelength of 488 nm (Ref. 26), for our case we obtain $d\sigma/do=6.1\times10^{-29}$ cm^2/sr. Let the pump intensity with allowance for ''smearing'' along the moving focus be $I_p=10^{10}$ W/cm^2, and let the intensity of the rotational Stokes wave be $I_s=10^9$ W/cm^2; then for the nutation period we obtain $T_R=1$ ps. This estimate is crude, but it characterizes the efficiency of rotational excitation of molecules in the interaction zone.

The energy of the vibrational Stokes component saturates in the pressure range where the rotational scattering energy reaches saturation and self-focusing, together with the increase in the length of the self-focusing channel, is at a maximum. In the presence of a moving focus each group of molecules in the region of the self-focusing channel interacts only with part of the pump pulse, not with the entire pulse. Consequently, as the pump energy is increased, the interaction time τ_{int} decreases relative to the pulse duration τ_p, so that the effective energy density of the pump $I_p\tau_{int}$ depends weakly on its total energy W_p. Saturation of the vibrational SRS therefore sets in approximately at the same pressures, irrespective of the pump energy.

Even though the pump is linearly polarized, the energy and pressure threshold for rotational SRS is much lower than for vibrational SRS. This disparity stems from the fact that in pumping by pulses having a wavelength of 1.06 μm the Stokes frequency for the rotational transition is 1.7 times the same frequency for the vibrational transition; according to Eq. (2), the gain is proportional to the square root of the Stokes frequency. On the other hand, if the wavelength of the pump pulses is 0.53 μm, the Stokes frequency for the rotational transition is only 1.25 times the same frequency for the vibrational transition, and it has been shown experimentally that vibrational SRS is generated predominantly for a linearly polarized pump in this case.

4.2. Self-Focusing Conditions

The spatial structure of the radiation at the cell output, unlike its spectral composition, does not undergo any radical changes. The divergence of the pump beam after the cell at a hydrogen pressure of 100 atm and a pump energy of 24 mJ (0.9°) is only slightly greater than the convergence of the pump beam entering the cell (0.75°). This situation is very similar to that reported by Corkum *et al.*[13–15] where the divergence of the beam is almost equal to its convergence in a gas-filled cell. The authors regarded this condition as an extremely dramatic and important observation. It can be attributed to the fact that self-focusing was not accompanied by constriction of the beam to dimensions significantly smaller than the diameter of the waist in the case of a linear focus. In our experiments this result is further corroborated by direct observation of the diameter of the self-focusing channel, which is greater than or close to the diameter of the linear waist (110 μm).

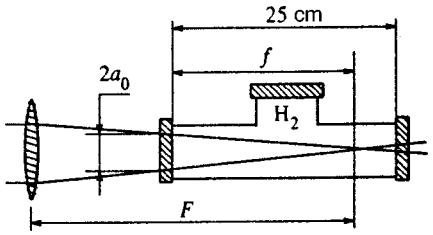

FIG. 9. Optical diagram of the focusing of the pump beam into the hydrogen-filled cell for locating the beginning of the self-focusing channel.

Is the experimentally observed length of the channel consistent with the focal shift that must occur in self-focusing as a result of the optical Kerr effect? In our case the focusing lens is situated outside the nonlinear medium (Fig. 9), and the position f' of the focus with allowance for self-focusing can be determined from the relations[27]

$$\frac{1}{f'}=\frac{1}{f}+\frac{1}{z_f},\quad z_f=\frac{ka_0^2}{\sqrt{2(P/P_{cr}-1)}},$$
$$P_{cr}=1.2\cdot10^{-2}\lambda^2c/n_2, \tag{4}$$

where f is the position of the linear focus (f' and f are read from the beginning of the linear medium), z_f is the length at which self-focusing of a parallel beam of radius a_0 at the entrance to the nonlinear medium takes place, P is the power of the pump beam, P_{cr} is the critical power for self-focusing, and n_2 is the nonlinear refractive index. The radius of the beam at the focusing lens has been measured and is equal to 2.2 mm. The value of n_2 for hydrogen cannot be found in the literature; we have estimated it from the values of n_2 for CO_2 (Ref. 28) and N_2 (Ref. 29): $n_2=(2-2.5)\times10^{-17}$ cm$^3\cdot$erg$^{-1}\cdot$atm^{-1}. Hence, at $p=100$ atm we obtain $P_{cr}=200$ MW, and the focal shift at $P=1$ GW should only be 0.9 cm. Even if we assume that our estimate of n_2 is too low and we increase this value of n_2 sevenfold, the focal shift is still only 2.6 cm.

The significant disparity between the observed channel length and the above-calculated focal shift leaves no choice but to assume that the effective nonlinear refractive index is considerably higher, as can be the case if excitation of the medium is taken into account.[30] The shift of the focus is governed by distortion of the pump wave front by self-focusing, which begins to take effect right at the entrance to the nonlinear medium. In the region from the entrance to the beginning of the radiating channel the only possible source of excitation of the hydrogen molecules is collisions with electrons heated in the pump field. Vibrational excitation of

the hydrogen molecules sets in when the electron energy attains 0.5 eV. The cross section of this process is a maximum at an energy ≈ 2 eV, where the effective frequency of the collisions responsible for vibrational excitation of the hydrogen molecules attains $\nu_v = 10^{13}\,\mathrm{s}^{-1}$ ($p = 100$ atm).[31] In the energy interval from 0.5 eV to 8.7 eV, i.e., up to the energy of the first excited electron level of the hydrogen molecule, an electron can expend its accumulated energy mainly in the vibrational excitation of molecules. For estimates we can assume here that the above-indicated effective collision frequency is constant in this energy interval. The electron begins to efficiently excite the electron levels when the time to build up an energy of 8.7 eV is comparable with the effective time constant of vibrational excitation $1/\nu_v = 100$ fs. The average time between collisions of the electron with hydrogen molecules is $1/\nu_m = 2$ fs at $p = 100$ atm (Ref. 31). Thus, if an energy of 8.7 eV is attained after 50 collisions, the probability of electronic excitation will be close to the probability of vibrational excitation, i.e., if the energy $\Delta\varepsilon_E$ acquired by the electron in one collision is 0.18 eV. According to Ref. 32, in a field of power density I and frequency $\nu = c/\lambda$ we have

$$\Delta\varepsilon_E[\mathrm{eV}] = \frac{6.3\cdot 10^{17} I[\mathrm{W/cm^2}]}{(4\pi^2\nu^2 + \nu_m^2)},$$

so that for $\lambda = 1.06\,\mu$m and $\nu \gg \nu_m$ we obtain $\Delta\varepsilon_E[\mathrm{eV}] = 2\times 10^{-13} I[\mathrm{W/cm^2}]$. The required condition is satisfied for $I = 9\times 10^{11}$ W/cm^2, which is attained at a distance ≈ 3 cm from the middle of the linear waist for a pump power of 1 GW, a lens with $F = 35$ cm, and a beam radius of 2.2 mm at the lens. In our case, therefore, vibrational excitation prevails over electronic excitation in the region from the entrance to the nonlinear medium almost to the linear focus.

But can vibrational excitation significantly alter the refractive index? A certain refinement is needed here. If we assume that the number of vibrationally excited molecules increases by an integral law within the pulse duration (30 ps), the time dependence of the refractive index can be written as follows in our case:

$$n(t) = n_0 + \Delta n_1\left(\int_{-\infty}^{t} E^2(\tau)\,d\tau\right) + n_2\left(\int_{-\infty}^{t} E^2(\tau)\,d\tau\right)E^2(t). \qquad (5)$$

The second term in Eq. (5) corresponds to time-dependent self-focusing in which the trailing edge of the pulse is focused closest to the lens. The third term corresponds to the optical Kerr effect with allowance for excitation of the medium; here, as in the case of quasisteady self-focusing, the leading and trailing edges of the pulse are focused near the linear focus.

The calculation of the second term in Eq. (4) is not too difficult. The variations of the polarizability for several molecules in the presence of vibrational excitation and the optical Kerr effect are given in Ref. 33: The hydrogen molecule holds a special place in this scheme. The optical Kerr effect is very small for this molecule by virtue of its small anisotropy, and yet it exhibits the greatest variation of the polarizability (19%) as a result of vibrational excitation. The refractive index of hydrogen at $p = 100$ atm is equal to 1.014; accordingly, for a relative density of excited molecules N^*/N the variation of the refractive index is $\Delta n_1 = 3\times 10^{-3} N^*/N$ (Ref. 33).

In the initial part of the nonlinear medium, where the vibrational excitation of hydrogen molecules by electrons is prevalent, the only possible source of electrons is the ionization of easily ionized impurities. Let us calculate the relation between the relative density of excited molecules N^*/N and the relative electron density N_e/N. An energy of 1.5 eV, at which the efficient vibrational excitation of hydrogen molecules already takes place, is built up by an electron after $1.5/\Delta\varepsilon_E[\mathrm{eV}]$ collisions, i.e., at a pressure of 100 atm in a time of $2\cdot 1.5/\Delta\varepsilon_E[\mathrm{eV}]$ fs. The time of interaction of molecules with the field is approximately equal to the pulse duration of 30 ps in this region of the nonlinear medium, and within this time one electron excites $10^4\Delta\varepsilon_E[\mathrm{eV}] = 2\times 10^{-9} I[\mathrm{W/cm^2}]$ molecules. Consequently, $N^*/N = 2\times 10^{-9} I[\mathrm{W/cm^2}] N_e/N$, and $\Delta n_1 = 6\times 10^{-12} I[\mathrm{W/cm^2}] N_e/N$. At a pressure of 100 atm the variation of the refractive index due to the optical Kerr effect in the case of unexcited molecules is $\Delta n = n_2 I/c = 0.8\times 10^{-18} I[\mathrm{W/cm^2}] N_e/N$, and if $N_e/N > 1.3\times 10^{-7}$, then Δn_1 is greater than Δn. This means that for $N_e/N > 1.3\times 10^{-7}$ the focal power of the lens created by the variation of the refractive index as a result of the vibrational excitation of hydrogen molecules is greater (and indeed substantially greater if the electron density is sufficiently large) than the focal power of the lens induced by the optical Kerr effect in the presence of unexcited molecules.

The efficient ionization of molecules takes place directly in the focal region.[34] The generated plasma has a stabilizing influence on the diameter of the beam waist, owing to the defocusing effect of the electrons. The corresponding increment in the refractive index is $\Delta n_{\mathrm{el}} = -\omega_{\mathrm{pl}}^2/\omega^2$. We have not attempted any estimates of the plasma frequency ω_{pl}, having found it impossible to estimate the electron density. However, the proportionality of this increment to the square of the frequency can account for the increase in the divergence of the beams as the frequency of the lines of the quasirotational spectrum increases. As the frequency increases, the increment Δn_{el} decreases, so that the defocusing effect of the electronic plasma diminishes accordingly, and high-frequency radiation beams constrict to a smaller diameter and undergo greater divergence in the far field.

The fixed position of the end of the channel is an argument in favor of the conclusion that self-channeling does not take place, and the formation of the self-focusing channel is attributable to the moving focus effect.

4.3. Structure of the Spectral Lines

The detailed structure acquired by the spectral lines at a pressure above 60 atm in the anti-Stokes region is shown schematically in Fig. 10 for the example of lines $i = 7-14$. Lines $i = 7, 14$, like all lines $i = 7n$, have one component each, and their frequencies are given by Eq. (1) with $m = 0$,

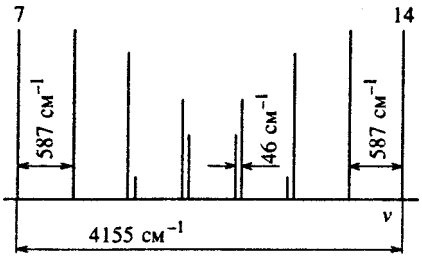

FIG. 10. Structure of the quasirotational spectrum in the interval of lines $i=7-14$. The spectrum exhibits a similar structure in the intervals $i=14-21$ and $i=21-28$.

i.e., are equal to the frequency of the nth vibrational anti-Stokes component. For this reason, the average distance between the lines is determined by 1/7 of the vibrational transition frequency, i.e., is approximately $593\,\mathrm{cm}^{-1}$. Lines 8 and 13 are also essentially single-component lines. Lines $i=7+m$ $(2\leqslant m\leqslant 5)$ consist of components $(1,m)$ and $(2,m-7)$. In Fig. 9 the intensities of the components are normalized so that the sum of the intensities of the components forming one line will be independent of m. This situation corresponds to a weak dependence of the total intensity of the lines in the visible region of the spectrum on their order number (see Fig. 2).

If we define the total order of a cascade process as the sum $|n|+|m|$, the following law emerges: one component, or two components of the lowest total order, become dominant. In the pressure range where the quasirotational spectrum has already formed (>60 atm) we do not perceive any of the phase synchronization effects typical of parametric processes: As the pressure increases, the quasirotational spectrum behaves as a unit whole, and the energies of the lines increase identically. If phase synchronization is not significant, this law is perfectly natural: the components formed as a result of a smaller number of cascade processes become dominant.

5. CONCLUSION

High-intensity pump radiation in high-pressure hydrogen is transformed into radiation whose spectrum consists of a series of lines having close intensities and extending from the IR to the UV. The lines of the spectrum have a smooth spatial profile and are separated by spectral intervals approximately equal to the rotational transition frequency. Several factors are operative in the focal region and determine the ray behavior in this region: self-focusing due to vibrational and electronic excitations of the hydrogen molecules and defocusing due to the electronic plasma. Akhmanov, Sukhorukov, and Khokhlov have remarked[35] that even if a self-channeling regime does not set in, the actual behavior of rays in a focused laser beam differs significantly from that calculated in the linear approximation. In this case it is meaningful to consider the waveguide propagation of the pump and emitted radiation and to attribute this phenomenon to the generation of quasirotational lines with a smooth spatial profile. It is conceivable that when focusing lenses with $F\geqslant 50$ cm are used, this propagation regime breaks down, whereupon the spectrum and spatial profile of the output radiation change radically.

The waveguide propagation of light beams along an ionized channel created, for example, by the focusing of another pulse in a gaseous medium by a conical lens has been demonstrated in several papers.[36–38] Il'inskiĭ and Mikheev[39] have investigated so-called waveguide SRS, where Stokes radiation propagates along the waveguide created by the increase in the refractive index during the vibrational excitation of hydrogen molecules. The presence of a conical component along with the axial component in the Stokes radiation was explained by the mode structure of the radiation propagating along the waveguide.

The authors are grateful to N. I. Koroteev, A. V. Andreev, V. T. Platonenko, K. N. Drabovich, V. P. Kandidov, A. M. Zheltikov, O. G. Kosareva, and S. Yu. Nikitin for helpful discussions. This work was supported by the Russian Fund for Fundamental Research (Grant No. 95-02-05259-a) and the State Scientific-Technical Program ''Fundamental Spectroscopy'' (020/2, Project 4.4).

*)E-mail: tunkin@sr.phys.msu.su

[1] V. Wilke and W. Schmidt, Appl. Phys. **18**, 177 (1979).
[2] G. B. Jarvis, S. Mathew, and J. E. Kenny, Appl. Opt. **33**, 4938 (1994).
[3] K. G. H. Baldwin, J. P. Harangos, and D. D. Burgess, Opt. Commun. **52**, 351 (1985).
[4] A. Z. Grasyuk, I. G. Zubarev, A. V. Kotov, S. I. Mikhaĭlov, and V. G. Smirnov, Kvantovaya Élektron. **3**, 1062 (1976) [Sov. J. Quantum Electron. **6**, 568 (1976)].
[5] N. V. Kravtsov and N. N. Naumkin, Vestn. Mosk. Univ., Fiz., Astron. **36**, 84 (1995).
[6] Y. Irie and T. Imasaka, Opt. Lett. **20**, 2072 (1995).
[7] S. Kawasaki, T. Imasaka, and N. Ishibashi, J. Opt. Soc. Am. B **8**, 1461 (1991).
[8] S. Ohtake, S. Yoshikava, and T. Imasaka, Appl. Opt. **34**, 4337 (1995).
[9] V. G. Bespalov and D. I. Stasel'ko, Opt. Spektrosk. **65**, 1061 (1988) [Opt. Spectrosc. **65**, 625 (1988)].
[10] C. W. Wilkerson Jr., E. Sekreta, and J. P. Reilly, Appl. Opt. **30**, 3855 (1991).
[11] L. L. Losev, A. P. Lutsenko, and S. N. Sazonov, Kvantovaya Élektron. **17**, 960 (1990) [Sov. J. Quantum Electron. **20**, 878 (1990)].
[12] A. Z. Grasyuk, L. L. Losev, D. N. Nikogosyan, and A. A. Oraevskiĭ, Kvantovaya Élektron. **11**, 1872 (1984) [Sov. J. Quantum Electron. **14**, 1257 (1984)].
[13] P. B. Corkum, C. Rolland, and T. Srinivasan-Rao, Phys. Rev. Lett. **57**, 2268 (1986).
[14] P. B. Corkum and C. Rolland, IEEE J. Quantum Electron. **QE-25**, 2634 (1989).
[15] D. Strickland and P. B. Corkum, J. Opt. Soc. Am. B **11**, 492 (1994).
[16] T. R. Gosnell, A. J. Taylor, and D. P. Greene, Opt. Lett. **15**, 130 (1990).
[17] F. A. Ilkov, L. Sh. Ilkova, and S. L. Chin, Opt. Lett. **18**, 681 (1993).
[18] K. Ueda, H. Nishioka, W. Odajima, and H. Takuma, Laser Phys. **6**, 260 (1996).
[19] V. B. Morozov, A. N. Olenin, and V. G. Tunkin, Kvantovaya Élektron. **25**, 293 (1998).
[20] R. G. Tomlinson, IEEE J. Quantum Electron. **QE-5**, 591 (1969).
[21] W. K. Bischel and M. J. Dyer, Phys. Rev. A **33**, 3113 (1986).
[22] R. A. J. Keijser, J. R. Lombardi, K. D. Van den Hout, B. C. Sanctuary, and H. F. P. Knaap, Physica (Amsterdam) **76**, 585 (1974).
[23] S. A. Akhmanov, K. N. Drabovich, A. P. Sukhorukov, and A. S. Chirkin, Zh. Éksp. Teor. Fiz. **59**, 485 (1970) [Sov. Phys. JETP **32**, 266 (1971)].
[24] R. Carman, F. Shimizu, C. Wang, and N. Bloembergen, Phys. Rev. A **2**, 60 (1970).
[25] S. A. Akhmanov and N. I. Koroteev, *Methods of Nonlinear Optics in Light Scattering Spectroscopy* [in Russian], Nauka, Moscow (1981), 232 pp.

[26] H. A. Hyatt, J. M. Cherlow, W. R. Fenner, and S. P. S. Porto, J. Opt. Soc. Am. **63**, 73 (1973).
[27] S. A. Akhmanov, A. S. Sukhorukov, and R. V. Khokhlov, Usp. Fiz. Nauk **93**, 19 (1967) [Sov. Phys. Usp. **10**, (1967)].
[28] R. H. Lehmberg, C. J. Pawley, A. V. Deniz, M. Klapisch, and Y. Leng, Opt. Commun. **121**, 78 (1995).
[29] A. J. Alcock, C. DeMichelis, and M. C. Richardson, IEEE J. Quantum Electron. **QE-6**, 622 (1970).
[30] G. A. Askar'yan, JETP Lett. **4**, 270 [400] (1966).
[31] E. W. McDaniel, *Collision Phenomena in Ionized Gases*, Wiley, New York (1964), 262 pp.
[32] Yu. P. Raĭzer, *Gas Discharge Physics*, Springer-Verlag, New York (1991) [Russian original, Nauka, Moscow (1992), 526 pp.].
[33] B. Vil'gel'mi and É. Goĭman, Zh. Prikl. Spektrosk. **19**, 550 (1973).
[34] Ya. B. Zel'dovich and Yu. P. Raĭzer, Zh. Éksp. Teor. Fiz. **47**, 1150 (1964) [Sov. Phys. JETP **20**, 772 (1965)].
[35] S. A. Akhmanov, A. V. Sukhorukov, and R. V. Khokhlov, Zh. Éksp. Teor. Fiz. **50**, 1537 (1966) [Sov. Phys. JETP **23**, 1025 (1966)].
[36] C. G. Durfee and H. M. Milchberg, Phys. Rev. Lett. **71**, 2409 (1993).
[37] C. G. Durfee, J. Lynch, and H. M. Milchberg, Opt. Lett. **19**, 1937 (1994).
[38] T. R. Clark and H. M. Milchberg, Phys. Rev. Lett. **78**, 2373 (1997).
[39] Yu. A. Il'inskiĭ and G. M. Mikheev, Zh. Éksp. Teor. Fiz. **101**, 1445 (1991) [Sov. Phys. JETP **74**, 772 (1991)].

Translated by James S. Wood

Fine-structure effects in relativistic calculations of the static polarizability of the helium atom

A. Derevianko and W. R. Johnson

Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556, USA

V. D. Ovsyannikov

Voronezh State University, 394693 Voronezh, Russia

V. G. Pal'chikov*)

National Research Institute of Physicotechnical and Radio-Engineering Measurements (State-Owned), 141570 Mendeleevo, Moscow Region, Russia

D. R. Plante

Department of Mathematics & Computer Science, Stetson University, De Land, Florida 32720, USA

G. von Oppen

Institute für Atomare und Analytische Physik, Technische Universität, D-10623 Berlin, Germany

(Submitted 20 July 1998)
Zh. Éksp. Teor. Fiz. **115**, 494–504 (February 1999)

We use the relativistic configuration-interaction method and the model potential method to calculate the scalar and tensor components of the dipole polarizabilities for the excited states $1s3p\,^3P_0$ and $1s3p\,^3P_2$ of the helium atom. The calculations of the reduced matrix elements for the resonant terms in the spectral expansion of the polarizabilities are derived using two-electron basis functions of the relativistic Hamiltonian of the atom, a Hamiltonian that incorporates the Coulomb and Breit electron–electron interactions. We formulate a new approach to determining the parameters of the Fuss model potential. Finally, we show that the polarizability values are sensitive to the choice of the wave functions used in the calculations.
 [S1063-7761(99)00902-6]

1. INTRODUCTION

The anticrossing of atomic levels in an external field constitutes an effective method for precise measurements of the fine- and hyperfine-structure intervals and other spectroscopic constants, such as the exchange energy and the diagonal and off-diagonal matrix elements of the spin–orbit coupling operators. In a recent paper, Schumann *et al.*[1] studied the $0^-\times 0^-$ anticrossing of the $1s3p\,^3P_J$ $(J=0,2)$ levels of helium by methods of high-resolution laser spectroscopy. At the anticrossing point, the error in measuring the fine-structure interval $\delta=E_{3\,^3P_0}-E_{3\,^3P_2}$ amounted to ± 5 MHz, and the use of microwave devices makes it possible to reduce this value by a factor of at least 100. Nevertheless, the degree of accuracy already achieved makes it possible to draw important conclusions concerning the optimum choice of theoretical approaches describing the effect of $100-200\,\mathrm{kV\,cm^{-1}}$ electric fields on the spectrum of the helium atom.

In particular, the widely used semiempirical approach, which makes it possible to analyze the observed spectrum in terms of averaged values of the atomic Hamiltonian with allowance for relativistic corrections (either spin-dependent or spin-independent) of order α^2 Ry (Refs. 2 and 3), with α the fine-structure constant, does not require allowing for higher-order perturbation effects, which are needed for a meaningful interpretation of the results of measurements. For this reason, Schumann *et al.*[1] justified the need to employ methods of quantum mechanics and quantum electrodynamics to analyze the results of measurements of the δ-to-$\delta^{(0)}$ ratio ($\delta^{(0)}=8772.517(16)$ MHz is the fine-structure interval in a zero field[4]) at the point of anticrossing of the $1s3p\,^3P_0$ and $1s3p\,^3P_2$ levels of helium. Another interesting result of that paper was the possibility of studying the effects of spin–spin mixing of helium levels with different values of orbital angular momentum L but the same parity.[5,6]

The aim of the present paper is to analyze theoretically the contribution of relativistic effects in calculations of the scalar and tensor components of dipole polarizabilities, which determine the shift and splitting of the $1s3p\,^3P_J$ $(J=0,2)$ levels of helium. We employ two alternative approaches based on the relativistic configuration-interaction method[7] and on the Fuss model potential method.[8]

2. ALLOWING FOR RELATIVISTIC EFFECTS IN CALCULATIONS OF DIPOLE POLARIZABILITIES OF THE HELIUM ATOM

The shift and splitting of a level $|nJLM\rangle$ in a uniform field F is described by the formula

$$\Delta E_{nJLM} = -\frac{1}{2}\,\alpha_{nJLM} F^2, \tag{1}$$

where the polarizability α_{nJLM} contains a scalar component (α^s_{nJL}) and a tensor component (α^t_{nJL}), i.e.,

$$\alpha_{nJLM} = \alpha^s_{nJL} + \alpha^t_{nJL}\frac{3M^2 - J(J+1)}{J(2J-1)}. \tag{2}$$

As the field strength F increases, the splitting of the level may reach values comparable to the distance between adjacent levels of the same parity (the components of the fine structure of an atomic multiplet). Hence in the case under investigation, i.e., the $1s3p\ ^3P_0$ and $1s3p\ ^3P_2$ levels and a field strength F of several hundred kilovolts per centimeter, the level shift $\Delta E_{nJLM} = E - E_{nJLM}$ can be found by solving the secular equation

$$\det \| \Delta E_{nJLM}\delta_{JJ'} - V_{JJ'} \| = 0. \tag{3}$$

Here the finite off-diagonal matrix elements $V_{JJ'}$ correspond to dipole transitions between the fine-structure components in second-order perturbation theory in the external field F.

To derive a formula describing the dependence of δ on F, we use the solutions of Eq. (3) and the definition of the scalar and tensor components of the polarizability (2). The result is

$$\begin{aligned}\delta &= \sqrt{\left[\delta^{(0)} - \frac{1}{2}F^2(\alpha^s_{3\,^3P_0} - \alpha^s_{3\,^3P_2} + \alpha^t_{3\,^3P_2})\right]^2 + 2F^4(\alpha^t_{3\,^3P_2})^2} \\ &\simeq \delta^{(0)} - \frac{1}{2}F^2(\alpha^s_{3\,^3P_0} - \alpha^s_{3\,^3P_2} + \alpha^t_{3\,^3P_2}) \\ &\quad + \frac{F^4}{\delta^{(0)}}(\alpha^t_{3\,^3P_2})^2 + \cdots.\end{aligned} \tag{4}$$

This expression allows for the principal, or resonant, part of the hyperpolarizability (fourth-order corrections in the external field) of the interacting sublevels of the multiplet $3\,^3P_J$ with a zero projection of total angular momentum, $M = 0$. The contribution of the nonresonant part to the hyperpolarizability is at most a few percent.[9] In deriving (4) we allowed for the fact that the matrix element $V_{JJ'}$ is finite at $J' = J \pm 1$, $J \pm 2$ and contains only a tensor part, which depends on the projection M of the total angular momentum J. If we ignore the multiplet splitting $E^{(0)}_{nJL} - E^{(0)}_{nJ'L}$ in comparison to the energy difference $E^{(0)}_{nJL} - E^{(0)}_{n'J'L'}$ between different multiplets with $n' \neq n$, the matrix element $V_{JJ'}$ can be expressed in terms of the tensor polarizability of the $1s3p\ ^3P_2$ state. In this approximation the matrix element V_{02} is given by the formula[1)]

$$V_{02} = -\frac{F^2}{\sqrt{2}}\alpha^t_{3\,^3P_2}. \tag{5}$$

The difference of the scalar polarizabilities in Eq. (4), $\alpha^s_{3\,^3P_0} - \alpha^s_{3\,^3P_2}$, is determined by the contribution of relativistic effects, which means it is a small quantity of order α^2. Note that at $M = 0$ the matrix element $V_{J\,J\pm1} \equiv 0$ (see, e.g., Ref. 10), so that the state $3\,^3P_1$ remains isolated, i.e., does not mix with states with $J = 0$ and $J = 2$.

Equation (4) also implies that the minimum value for δ in an electric field (anticrossing of the fine-structure sublevels) is attained at

$$F = \overline{F} = \sqrt{\frac{2\,\delta^{(0)}(\alpha^s_{3\,^3P_0} - \alpha^s_{3\,^3P_2} + \alpha^t_{3\,^3P_2})}{(\alpha^s_{3\,^3P_0} - \alpha^s_{3\,^3P_2} + \alpha^t_{3\,^3P_2})^2 + 8(\alpha^t_{3\,^3P_2})^2}}. \tag{6}$$

The problem of exact *ab initio* relativistic calculations of the quantities $\alpha^s_{3\,^3P_0} - \alpha^s_{3\,^3P_2}$ and $\alpha^t_{3\,^3P_2}$ in Eqs. (4) and (6) is extremely difficult and involves calculating spectral sums over the complete set of unperturbed states. The need for such summation (irrespective of the general approach) emerges in the process of determining the perturbed wave functions or energy shifts of atomic levels in the perturbation-theory setting. In addition to direct summation over the discrete spectrum of intermediate states and integration over the continuous spectrum of intermediate states, which are extremely involved processes in the relativistic case, we basically used two methods to effectively calculate such spectral sums (composite matrix elements), a method for integrating inhomogeneous differential equations and a method that uses the formalism of Green's functions.

In the first approach, the polarizability of the state $|0\rangle$ is given by the formula

$$\alpha_{|0\rangle} = -2\langle\Psi|D|\psi_0\rangle, \tag{7}$$

in which the perturbed wave function $|\Psi\rangle$ satisfies the inhomogeneous equation

$$(\hat{H} - E_0)\Psi = -D\Psi_0. \tag{8}$$

In (7) and (8), D is the dipole moment operator, and $\hat{H}$ is the relativistic Hamiltonian.

An important advantage of this method is the possibility of using different expressions for the atomic potential in the numerical integration of Eq. (8), and the calculations can be done not only for a purely Coulomb interaction but in the multiconfiguration interaction approximation,[7] the Hartree–Fock–Dirac approximation,[11] and the relativistic random phase approximation with exchange.[12] The most exact relativistic calculations were done by Johnson and Cheng[13] for the polarizability of the ground state of a heliumlike atom with $2 \leqslant Z \leqslant 30$, but at present there are no similar results for excited states of helium with $L \neq 0$.

The effectiveness of the method of Green's functions is largely determined by the existence of appropriate representations of these functions. Since in the relativistic case the expressions for the Green's functions are known only for the Coulomb field, the use of this approach is restricted to problems in which the difference of the potential and the Coulomb potential is insignificant or can be taken into account by perturbation-theory techniques.[14,15]

To allow for the contribution of relativistic corrections in calculations of the scalar and tensor components of polarizabilities, we used the resonance approximation for the second-order composite matrix elements, i.e., in the spec-

trum of intermediate states we isolated the terms that correspond to transitions in which the principal quantum number does not change.

For instance, for the scalar polarizability of the triplet $|nJLM\rangle$ state,

$$\alpha^s_{nJL} = -\frac{2}{3(2J+1)} \sum_{n'J'L'} \frac{|\langle nJL\|\mathbf{r}\|n'J'L'\rangle|^2}{E_{nJL}-E_{n'J'L'}}, \tag{9}$$

the resonant term $\alpha^{s(r)}_{nJL}$ in (9) has the form

$$\alpha^{s(r)}_{nJL} = -\frac{2\beta^2_{nJL}}{3}\left\{ \sum_{J'} \frac{L(2J'+1)}{E_{nJL}-E_{nJ'\,L-1}}\left[\beta_{nJ'\,L-1} R^{nJ'\,L-1}_{nJL} \times \begin{Bmatrix} L & J & 1 \\ J' & L-1 & 1 \end{Bmatrix}\right]^2 + \sum_{J'} \frac{(L+1)(2J'+1)}{E_{nJL}-E_{nJ'\,L+1}} \times \left[\beta_{nJ'\,L+1} R^{nJ'\,L+1}_{nJL} \begin{Bmatrix} L & J & 1 \\ J' & L+1 & 1 \end{Bmatrix}\right]^2 \right\}. \tag{10}$$

Here the $\{\begin{smallmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{smallmatrix}\}$ are Wigner $6j$-symbols,[16] and the $R^{n'J'L'}_{nJL}$ are radial matrix elements. In Eq. (10), the effects of singlet–triplet mixing of levels are taken into account by the parameter

$$\beta_{nJL} = \begin{cases} \cos\theta_{nL}, & J=L, \\ 1, & J\neq L, \end{cases} \tag{11}$$

where θ_{nL} is the singlet–triplet mixing angle. Chang[17] noted that θ_{nL} is almost entirely independent of n, and the typical values of θ_{nL} at $L=1,2,3$ are 0.02°, 0.5°, and 30°, respectively. The tensor part of the polarizability has a similar structure and differs from (10) only in angular coefficients; for the sake of brevity we will not write it here.

To calculate $\alpha^{s(r)}_{nJL}$ and $\alpha^{t(r)}_{nJL}$ we used exact relativistic results for the radial integrals, while for the energy denominators we used precise experimental data.[18] Calculations of the other terms in (9) with $n'\neq n$ were done with nonrelativistic values for the radial matrix elements of the dipole moment operator. To this end we used exact numerical data for the oscillator strengths of the s–p and p–d transitions in helium calculated with multiparameter variational wave functions.[19,20]

This approach is applicable primarily because the contribution of states with $n'=n$ in (9) is numerically predominant and amounts to roughly 95% in the case of excited $1s3p\ ^3P_J$ $(J=0,2)$ levels of helium (see Tables II and III below).

3. SELECTION OF THE BASIS WAVE FUNCTIONS AND DISCUSSION OF THE RESULTS OF CALCULATIONS

In recent years the configuration-interaction method has been successfully used to obtain precise wave functions and matrix elements for atoms with a small number of electrons, e.g., heliumlike systems. In the present paper we use the technique developed by Johnson *et al.*[7,13] to calculate the reduced matrix elements corresponding to the resonant terms in the expansion (10).

The wave function of the initial (I) and final (F) states can be written

$$\Psi_{I(F)} = \sum_{k\geqslant l} C^{I(F)}_{kl} \Phi_{kl}. \tag{12}$$

Here the Φ_{kl} are two-particle basis functions with fixed values of total angular momentum J, its projection M, and parity. We found the weighting factors $C^{I(F)}_{kl}$ from the variational principle by using the relativistic no-pair Hamiltonian, which incorporates the Coulomb and Breit electron–electron interaction operators,[21,22] averaged over the functions (12). To exclude the contribution of negative-energy states (the positron spectrum), the two-particle operators in the relativistic Hamiltonian were multiplied by products of single-particle operators projecting on the subspaces of solutions of the positive-energy Dirac equation.[7]

The wave functions (12) are normalized by the condition

$$\sum_{k\geqslant l} |C^{I(F)}_{kl}|^2 = 1. \tag{13}$$

The single-particle basis orbitals employed in the configuration-interaction method incorporates the s, p, d, f, and g partial waves, with a spline approximation used for each wave. Estimates of the convergence rate of the method (for calculations of given accuracy) can be found in Refs. 13 and 22.

The results of relativistic calculations for the reduced matrix elements (without allowance for retardation effects for the dipole moment operator) are listed in Table I.

TABLE I. Reduced matrix elements of the dipole moment operator.

Transition	Matrix element
$1s3p\ ^3P_0 \rightarrow 1s3s\ ^3S_1$	−6.4797
$1s3p\ ^3P_2 \rightarrow 1s3s\ ^3S_1$	−14.489
$1s3p\ ^3P_0 \rightarrow 1s3d\ ^3D_1$	8.2923
$1s3p\ ^3P_2 \rightarrow 1s3d\ ^3D_1$	1.8542
$1s3p\ ^3P_2 \rightarrow 1s3d\ ^3D_2$	7.1805
$1s3p\ ^3P_2 \rightarrow 1s3d\ ^3D_3$	16.994

In Tables II and III we list the results for the contributions of the intermediate S and D states in calculations of the scalar polarizabilities of the $1s3p\ ^3P_J$ $(J=0,2)$ states of helium. The difference between the tensor part of the polarizability of the $1s3p\ ^3P_2$ level and the scalar part is that the contributions of the intermediate states $1sn's\ ^3S_1$ and $1sn'd\ ^3D_1$ have opposite signs and that the total contribution of the $1sn'd\ ^3D_3$ states contains an additional numerical factor, $-2/7$.

Table IV summarizes the results of numerical calculations of the scalar and tensor components of helium polarizabilities. The dependence of δ on F near the anticrossing point is plotted in Fig. 1.

Note that the difference $\Delta\alpha$ remains almost the same if we use nonrelativistic variational values for the resonant matrix elements in (10): $\alpha^s_{3\,^3P_0}=17\,207$, $\alpha^s_{3\,^3P_2}=17\,198$, and $\Delta\alpha=9$. The explanation is that the principal contribution to $\Delta\alpha$ is provided by the relativistic corrections for the fine structure of the levels in the energy denominators of (12) rather than the relativistic corrections for the matrix elements of the dipole moment operator (see Table I).

TABLE II. Contributions of the S and D states of the intermediate spectrum to the polarizability of the $1s3p\ ^3P_0$ state of helium.

n'	$1sn's\ ^3S_1$ ($\lambda_0=0.698$)	$1sn's\ ^3S_1$ ($\lambda_0=-0.302$)	$1sn's\ ^3S_1$	$1sn'd\ ^3D_1$
2	−0.34	−1.67	−1.57	
3	−2377.93	−2682.56	−2638.45	18 742.21
4	451.87	319.37	312.17	623.53
5	25.38	18.27	17.92	86.01
6	6.36	4.60	4.72	27.16
7	2.59	1.88	1.85	12.29
8	1.34	0.97	0.96	6.72
9	0.79	0.57	0.57	4.13
10	0.51	0.37	0.34	2.59
Total	−1888.6	−2337.7	−2301.50	19 540.63

To test the above results, we did alternative polarizability calculations using the method of the Green's functions of an optical electron to sum over the complete intermediate-states spectrum in (9).

Note that the correct selection of the initial analytical representation of the Green's functions $G_E(\mathbf{r}_1,\mathbf{r}_2)$ plays an important role in specific polarizability calculations, since it makes it possible to obtain the result in the form most rational and convenient for further applications. In this paper we have taken the Green's function for the Fuss model potential from Ref. 14. The angular part of $G_E(\mathbf{r}_1,\mathbf{r}_2)$ is simply the product of spherical harmonics, while for the radial part $g_l(E;r_1,r_2)$ we have taken an expansion in Sturm functions, which have only a discrete spectrum:[15]

$$G_E(\mathbf{r}_1,\mathbf{r}_2)=\sum_{lm} g_l(E;r_1,r_2)Y_{lm}(\mathbf{n}_1)Y^*_{lm}(\mathbf{n}_2), \tag{14}$$

$$g_l(E;r_1,r_2)=\frac{4Z}{\nu}\sum_{k=0}^{\infty}\frac{U_{kl}(2Zr_1/\nu)U_{kl}(2Zr_2/\nu)}{k+\lambda_l+1-\nu}, \tag{15}$$

where $\nu=Z/\sqrt{-2E}$, and

$$U_{kl}(x)=\sqrt{\frac{k!}{\Gamma(k+2+2\lambda_l)}}x^{\lambda_l}\exp\left(-\frac{x}{2}\right)L_k^{2\lambda_l+1}(x). \tag{16}$$

The radial wave functions $R_{nl}(r)$ are obtained from the residues of the Green's function at the poles of $g_l(E;r_1,r_2)$: $\nu=\nu_{nl}=n_r+\lambda_l+1$, with $n_r=0,1,2,\ldots$ the radial quantum number, λ_l the effective orbital angular momentum, and ν_{nl} the effective principal quantum number in the formula for the energy of the atomic state $|nl\rangle$,

$$E_{nl}=-\frac{Z^2}{2\nu_{nl}^2}. \tag{17}$$

Here and in (15), Z is the charge of the residual ion. The explicit expression for $R_{nl}(r)$ coincides in form with hydrogenlike wave functions:[23]

$$R_{nl}(r)=\frac{2Z^{3/2}}{\nu_{nl}^2}U_{n_rl}\left(\frac{2Zr}{\nu_{nl}}\right). \tag{18}$$

The parameter λ_l (l may represent a specific set of spin–orbit quantum number, in addition to representing a specific angular momentum) can be found by comparing (17) with experimental values of the lowest state of a valence electron with a given l (see Ref. 23). The radial quantum number n_r of this state is assumed to be zero. As shown by Simons,[8] λ_l represents the entire experimental spectrum of the atom fairly well (in most cases the weak dependence of λ_l on the position of the energy level can be ignored).

With such a definition of λ_l for atomic series whose lowest states are the ground and the metastable, the error in calculating the radial matrix elements $\langle nl|r^L|n'l\rangle L\geqslant 1$ with wave functions (18) may reach 50%. In view of this we formulated a modified approach to the definition of n_r and λ_l in these series,[24] which allowed us to significantly refine the calculations of the spectroscopic characteristics of atoms in ground and excited states. Here the radial quantum number of the lowest state (ground or metastable) of the series is assumed to be unity, so that the effective orbital angular

TABLE III. Contributions of the S and D states of the intermediate spectrum to the polarizability of the $1s3p\ ^3P_2$ state of helium.

n'	$1sn's\ ^3S_1$	$1sn'd\ ^3D_1$	$1sn'd\ ^3D_2$	$1sn'd\ ^3D_3$
2	−1.57			
3	−2638.76	187.32	2809.37	15 735.95
4	312.15	6.23	93.52	523.74
5	17.92	0.86	12.90	72.25
6	4.72	0.27	4.07	22.81
7	1.85	0.12	1.84	10.32
8	0.96	0.07	1.01	5.65
9	0.57	0.04	0.62	3.47
10	0.34	0.03	0.39	302.17
Total	−2301.83	194.94	2923.72	16 376.36

TABLE IV. Scalar and tensor polarizabilities of the helium atom.

Quantity	Numerical value
$\alpha^s_{3\,^3P_0}$	17 203
$\alpha^s_{3\,^3P_2}$	17 191
$\Delta\alpha=\alpha^s_{3\,^3P_0}-\alpha^s_{3\,^3P_2}$	10
$\alpha^t_{3\,^3P_2}$	351.65
$\overline{F}$	$0.29\,343\times10^{-4}$ $=150.99$ kV cm^{-1}

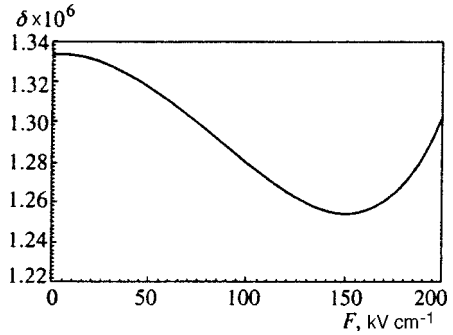

FIG. 1. Dependence of δ (in atomic units) on the electric field strength F in the vicinity of anticrossing $\overline{F}$.

momentum λ_l is equal to $\nu_g - 2$, with ν_g the effective principal quantum number of the lowest state. Note that the wave function of the lowest level at $n_r = 1$ coincides precisely with the wave function obtained by the quantum-defect method in atoms.[25]

Thus, the sets of states of an atom with spin–orbit quantum numbers of the ground and metastable levels are, strictly speaking, incomplete since they do not contain states with $n_r = 0$. Hence the radial Green's functions in the subspaces of the series in question contain additional ''imaginary'' terms with $n_r = 0$ and an effective principal quantum number $\nu_{\text{im}} = \nu_g - 1$. The binding energy of the ''imaginary'' state, defined in (17), is almost ten times higher than the excitation energy of any level in a series, so that its contribution to the optical-transition amplitude can be ignored.

Note also that the wave function of a valence electron in state $|nl\rangle$ has the correct sign in the asymptotic region (which provides the principal contribution to multiple matrix elements) only if we multiply it by an addition phase factor $(-1)^k$, where $k = n - n_r - l - 1$.

The second and third columns of Table II contain the contributions of the intermediate $1sn's\,^3S_1$ states to the scalar polarizability of the $1s3p\,^3P_0$ state calculated by the traditional and modified approaches, which yield values of λ_0 equal to 0.698 and -0.302, respectively. A comparison with the data of precise variational calculations (the fourth column in Table II) suggests that the discrepancy of the final results is less than 2% if the value $\lambda_0 = -0.302$ is used, while calculations with $\lambda_0 = 0.698$ yield an error exceeding 20%.

The results of polarizability calculations based on the use of Green's functions in the method of the Fuss model potential are listed in Table V. The numerical discrepancies of the data of Tables IV and V are due, on the one hand, to the semiempirical approximation of the model potential method and, on the other, to the allowance for the contribution of the continuous spectrum in the Green's functions method. Note that the results for the difference of scalar polarizabilities and for the tensor polarizability of the $1s3p\,^3P_2$ state differ from those in Table IV only by 10%. To obtain numerical values of the difference of scalar polarizabilities with an accuracy of about 1%, the calculations of the radial integral should be done with an accuracy of five or six figures, since, as Table V implies, the first three significant figures cancel out. This loss of accuracy can be avoided if we expand the polarizabilities as functions of the energy of an atomic level in a Taylor series. The calculation then reduces to finding the energy derivatives of the polarizabilities, and these can be expressed in terms of third-order dipole matrix elements with two Green's functions, which, in particular, enter into the expression for the hyperpolarizability of the atomic states.[26]

TABLE V. Scalar and tensor polarizabilities of the helium atom calculated by the model potential method.

Quantity	Numerical value
$\alpha^s_{3\,^3P_0}$	17 266
$\alpha^s_{3\,^3P_2}$	17 255
$\Delta\alpha = \alpha^s_{3\,^3P_0} - \alpha^s_{3\,^3P_2}$	11
$\alpha^t_{3\,^3P_2}$	374.16
$\overline{F}$	$0.28\,458 \times 10^{-4}$
	$= 146.45$ kV cm^{-1}

It is no accident that the numerical results in Tables IV and V arer close, since the model potential method yields a correct dependence in the higher-order matrix elements for the energy of the atomic levels,[23] provided that we use the exact (experimental) values for the energies of the fine-structure sublevels. This condition makes it possible to take into account the contributions of relativistic and correlation effects in the model potential method. Furthermore, using this method, one can easily show that the states belonging to a complete set and not taken into account in relativistic calculations (the states belong to the continuum and to the discrete spectrum with $n' > 10$) contribute no more than 0.2% to the numerical values of the quantities considered here.

The present work was made possible by the financial support the Russian Fund for Fundamental Research [Grant 97-02-16407and an international grant with the German Research Society 96-02-00257, 436 RUS 113/164/O(R,S)] and the U.S. National Science Foundation (Grant PHY 95-13179).

*)E-mail: pal@kalium.physik.tu-berlin.de
bpt97@ftri.extech.msk.su

[1)]Here and in the text that follows we use the atomic system of units.

[1] R. Schumann, M. Dammasch, U. Eichmann, Y. Kriescher, G. Ritter, and G. von Oppen, J. Phys. B **30**, 2581 (1997).
[2] D. R. Cok and S. R. Lundeen, Phys. Rev. A **19**, 1830 (1979).
[3] D. R. Cok and S. R. Lundeen, Phys. Rev. A **24**, 3283 (1981).
[4] H. D. Yang, P. McNicholl, and H. Metcalf, Phys. Rev. A **53**, 1725 (1996).
[5] V. G. Pal'chikov and G. von Oppen, Phys. Scr. **52**, 366 (1995).
[6] V. G. Pal'chikov and G. von Oppen, Zh. Éksp. Teor. Fiz. **110**, 876 (1996) [JETP **83**, 481 (1996)].
[7] W. R. Johnson, D. R. Plante, and J. Sapirstein, Adv. At., Mol., Opt. Phys. **35**, 255 (1995).
[8] G. Simons, J. Chem. Phys. **55**, 756 (1971).
[9] A. Derevianko, W. R. Johnson, V. D. Ovsiannikov, V. G. Pal'chikov, D. R. Plante, and G. von Oppen, in *Proc. Int. Conf. At. Phys. (ICAP-98, Windsor)* (1998), p. 168.
[10] N. L. Manakov and V. D. Ovsiannikov, J. Phys. B **10**, 569 (1977).
[11] W. R. Johnson and C. D. Lin, Phys. Rev. A **9**, 1486 (1974).
[12] W. R. Johnson, D. Kolb, and K. N. Huang, At. Data Nucl. Data Tables **28**, 333 (1983).
[13] W. R. Johnson and K. T. Cheng, Phys. Rev. A **53**, 1375 (1996).
[14] S. A. Zapryagaev, N. L. Manakov, and V. G. Pal'chikov, *Theory of One-*

and Two-Electron Multiply Charged Ions [in Russian], Énergoatomizdat, Moscow (1985).
[15] N. L. Manakov, V. D. Ovsiannikov, and L. P. Rapoport, Phys. Rep. **141**, 319 (1986).
[16] D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskiĭ, *Quantum Theory of Angular Momentum*, World Scientific, Singapore (1987).
[17] E. S. Chang, Phys. Rev. A **35**, 2777 (1987).
[18] W. C. Martin, Phys. Rev. A **36**, 3575 (1987).
[19] A. Kono and S. Hattori, Phys. Rev. A **29**, 2981 (1984).
[20] M. K. Chen, J. Phys. B **27**, 865 (1994).
[21] M. H. Chen, K. T. Cheng, and W. R. Johnson, Phys. Rev. A **47**, 3692 (1993).
[22] K. T. Cheng, M. H. Chen, W. R. Johnson, and J. Sapirstein, Phys. Rev. A **50**, 247 (1994).
[23] N. L. Manakov, V. D. Ovsyannikov, and L. P. Rapoport, Opt. Spektrosk. **38**, 206 (1975) [Opt. Spectrosc. **38**, 115 (1975)].
[24] V. D. Ovsiannikov and V. G. Pal'chikov, in *Proc. Int. Conf. At. Phys. (ICAP-98, Windsor)* (1998), p. 166.
[25] M. J. Seaton, Rep. Prog. Phys. **46**, 167 (1983).
[26] V. A. Davydkin and V. D. Ovsiannikov, J. Phys. B **19**, 2071 (1986).

Translated by Eugene Yankovsky

Inversionless superradiance of an ensemble of three-level atoms in a high-Q cavity

A. I. Zaĭtsev, I. V. Ryzhov, and E. D. Trifonov

A. I. Gerzen Russian State Teachers' Training University, 191186 St. Petersburg, Russia

V. A. Malyshev*)

National Center ''S. I. Vavilov State Optical Institute,'' 199034 St. Petersburg, Russia
(Submitted 21 May 1998)
Zh. Éksp. Teor. Fiz. **115**, 505–520 (February 1999)

We analyze the possibility of superradiance in an ensemble of three-level atoms in the absence of population inversion. We show that in the case of a Λ configuration of the active transitions this effect can occur for an initially coherent superposition of the states of the lower doublet. We also study how splitting of the lower levels influences this effect and discuss ways of creating low-frequency coherence. © *1999 American Institute of Physics.* [S1063-7761(99)01002-1]

1. INTRODUCTION

The theory of collective spontaneous emission of radiation (superradiance) was first developed by Dicke[1] for an ensemble of two-level atoms. Further studies of this effect[2–9] were done primarily with this model. Stepping outside the scope of the two-level approximation (say, by assuming that the ground state is a doublet) leads to new effects in superradiance, effects produced by competition between the transitions. Among such effects we note quantum beats and polarization features in the superradiance emission signal.[6] Elyutin *et al.*[10] discussed the possibility of superradiance and photon echo in a magnetodipole transition when coherent superposition of all (three) states is produced by a strong optical resonant pulse.

It is well known that a necessary condition for inducing superradiance in the two-level model is the presence of initial population inversion of the levels involved in the active transition. Recently the problem of inversionless amplification has been widely discussed (see, e.g., Refs. 11–15). This effect is possible, for instance, if there is an additional level that is close to the ground level (the Λ configuration of levels). If the initial state of the lower doublet is prepared as a coherent superposition to which a transition from the upper level is forbidden, the orthogonal superposition, to which a transition is allowed, is found to be unpopulated. Hence a resonant pulse passing through a medium prepared in this way will be amplified.

The main goal of the present investigation is to show that in addition to inversionless amplification there can be inversionless superradiance (a preliminary discussion of the problem can be found in be found in Ref. 16). In Sec. 2 we formulate the model and present the basic equations of the theory. In Sec. 3 we study the case of a degenerate doublet in the ground state (this case allows for an analytical investigation). The effect of splitting of the lower level is examined in Sec. 4. Section 5 is devoted to an analysis of the possible ways of forming the low-frequency coherence, which is necessary for inversionless superradiance. Finally, in Sec. 6 we solve the general equations of the model, incorporating into the scheme an external field that produces the low-frequency coherence.

2. MODEL AND BASIC EQUATIONS

Let us examine an ensemble of three-level atoms with a Λ configuration of the active transitions (see Fig. 1). We assume that the splitting frequency ω_{21} for the lower levels is much smaller than the frequencies ω_{31} and ω_{32} of the transitions between the upper state 3 and the states of the lower doublet, 1 and 2. To describe the interaction of such a system and an electromagnetic field in quasiresonance with the high-frequency transitions we use the one-dimensional version of the semiclassical approach, in which we assume that the quantities describing the state of the atomic system and the field vary in only one direction while all the vectors are directed in the same way along the perpendicular direction. Then the evolution of the system obeys the following system of Maxwell–Bloch equations:

$$\dot{\rho}_{31}=-i\omega_{31}\rho_{31}-i\frac{d_{31}\mathscr{E}}{\hbar}(\rho_{33}-\rho_{11})+i\frac{d_{32}\mathscr{E}}{\hbar}\rho_{21}, \tag{1}$$

$$\dot{\rho}_{32}=-i\omega_{32}\rho_{32}-i\frac{d_{32}\mathscr{E}}{\hbar}(\rho_{33}-\rho_{22})+i\frac{d_{31}\mathscr{E}}{\hbar}\rho_{12}, \tag{2}$$

$$\dot{\rho}_{21}=-i\omega_{21}\rho_{21}-i\frac{d_{31}\mathscr{E}}{\hbar}\rho_{23}+i\frac{d_{32}\mathscr{E}}{\hbar}\rho_{31}, \tag{3}$$

$$\dot{\rho}_{11}=i\frac{d_{31}\mathscr{E}}{\hbar}(\rho_{31}-\rho_{13}), \tag{4}$$

$$\dot{\rho}_{22}=i\frac{d_{32}\mathscr{E}}{\hbar}(\rho_{32}-\rho_{23}), \tag{5}$$

$$\dot{\rho}_{33}=-i\frac{d_{31}\mathscr{E}}{\hbar}(\rho_{31}-\rho_{13})-i\frac{d_{32}\mathscr{E}}{\hbar}(\rho_{32}-\rho_{23}), \tag{6}$$

$$\left(\frac{\partial^2}{\partial x^2}-\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\mathscr{E}=\frac{4\pi}{c^2}\frac{\partial^2\mathscr{P}}{\partial t^2}, \tag{7}$$

where the $\rho_{\alpha\beta}$ are the elements of the density matrix of the three-level atom at the point with coordinate x at time

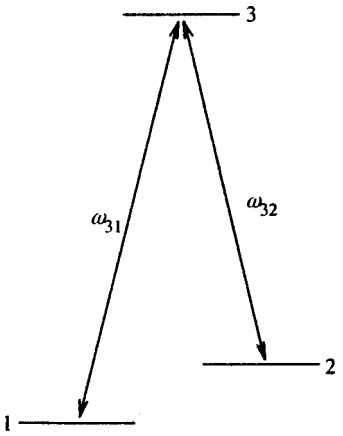

FIG. 1. Diagram of energy levels and optical transitions (designated by arrows) of a three-level atom.

t ($\alpha,\beta=1,2,3$), d_{31} and d_{32} are the dipole moments of the $3\leftrightarrow1$ and $3\leftrightarrow2$ transitions (these moments are assumed to be real and positive), $\mathscr{P}=N_0(d_{31}\rho_{31}+d_{32}\rho_{32})+$c.c. is the polarization of the medium, N_0 is the the atom number density, and $\mathscr{E}$ is the electric field strength. Population relaxation and polarization relaxation (either homogeneous or related to inhomogeneous broadening) are not taken into account, since the superradiance process is assumed to be faster.

Let us suppose that the atoms are uniformly distributed over the vole of a high-Q ring cavity (field damping due to cavity losses is ignored) and that the transitions with frequencies ω_{31} and ω_{32} are in quasiresonance with one of the cavity modes (frequency ω_c). We also assume that the superradiance spectrum and the value of the double splitting ω_{21} do not overlap the gap between cavity modes, i.e., we an limit ourselves to the single-mode approximation.

We seek the solution of the system of equations (1)–(7) in the form

$$\mathscr{E}=E\exp[-i(\omega_c t-k_c x)]+\text{c.c.}, \tag{8}$$

$$\rho_{31}=R_{31}\exp[-i(\omega_c t-k_c x)], \tag{9}$$

$$\rho_{32}=R_{32}\exp[-i(\omega_c t-k_c x)], \tag{10}$$

where $k_c=\omega_c/c$ (c is the speed of light in vacuum); E, R_{31}, and R_{32} are the amplitudes of the field and of the off-diagonal elements of the density matrix (the last two quantities are known as coherences), which are assumed to vary slowly on the scale of the optical period $2\pi/\omega_{31}$. Note that the analogous approximation on the $2\pi/\omega_{21}$-scale, i.e., for the low-frequency coherence ρ_{21}, is not used.

To simplify the problem still more, we assume that the state of the cavity changes little during the time light traverses the cavity (the mean-field approximation). This makes it possible to ignore the spatial dependence of the amplitude and thus exclude effects of propagation and amplification of light in the active medium. Such simplifications are not very important for the effect considered in this paper, i.e., inversionless superradiance. The study of inversionless superradiance in an open extended system with allowance for the factors ignored by the adopted model constitutes a separate problem.

If we now pass in the usual way from the system of equations (1)–(7) to a similar system for the amplitudes, we get

$$\dot{R}_{31}=-i\Delta_{31}R_{31}+\varepsilon[\mu_{31}(\rho_{33}-\rho_{11})-\mu_{32}\rho_{21}], \tag{11}$$

$$\dot{R}_{32}=-i\Delta_{32}R_{32}+\varepsilon[\mu_{32}(\rho_{33}-\rho_{22})-\mu_{31}\rho_{12}], \tag{12}$$

$$\dot{\rho}_{21}=-i\omega_{21}\rho_{21}+\mu_{31}\varepsilon R_{32}^*+\mu_{32}\varepsilon^* R_{31}, \tag{13}$$

$$\dot{\rho}_{11}=\mu_{31}(\varepsilon R_{31}^*+\varepsilon^* R_{31}), \tag{14}$$

$$\dot{\rho}_{22}=\mu_{32}(\varepsilon R_{32}^*+\varepsilon^* R_{32}), \tag{15}$$

$$\dot{\rho}_{33}=-\mu_{31}(\varepsilon R_{31}^*+\varepsilon^* R_{31})-\mu_{32}(\varepsilon R_{32}^*+\varepsilon^* R_{32}), \tag{16}$$

$$\dot{\varepsilon}=\Omega^2(\mu_{31}R_{31}+\mu_{32}R_{32}), \tag{17}$$

where $\Omega=\sqrt{2\pi\omega_{31}d^2N_0/\hbar}$, $\varepsilon=-idE/\hbar$ is the amplitude of the electric field in frequency units, $\Delta_{31}=\omega_{31}-\omega_c$, $\Delta_{32}=\omega_{32}-\omega_c$, $\mu_{31}=d_{31}/d$, and $\mu_{32}=d_{32}/d$, with $d=\sqrt{(d_{31}^2+d_{32}^2)/2}$. Note that this system of equations has the following constants of motion:

$$\rho_{11}+\rho_{22}+\rho_{33}=1, \tag{18}$$

$$\rho_{11}^2+\rho_{22}^2+\rho_{33}^2+2(|\rho_{21}|^2+|R_{31}|^2+|R_{32}|^2)=\text{const}, \tag{19}$$

$$|\varepsilon|^2+\rho_{33}=\text{const}. \tag{20}$$

3. A DEGENERATE DOUBLET

In the two-level scheme of inversionless superradiance we would have only attenuation of the initial fluctuations of R_{31} and R_{32}; they certainly do not become stronger. The situation changes dramatically, however, when the ground state is a doublet. Note that Eqs. (11) and (12) for the high-frequency coherences contain terms proportional to the low-frequency coherence ρ_{21}. To be specific, let us assume $\rho_{33}(0)-\rho_{11}(0)=\rho_{33}(0)-\rho_{22}(0)=0$. Now, if $\rho_{21}(0)\neq0$, the behavior (weakening or strengthening) of the initial fluctuations of R_{31} and R_{32} depends on the phase of $\rho_{21}(0)$: for positive values of $\rho_{21}(0)$ these fluctuations still become weaker, but for negative values the strength of these fluctuation increases in a snowballing manner, thus initiating superradiance. Note that this becomes possible without population inversion in the high-frequency channels $3\leftrightarrow1$ and $3\leftrightarrow2$ and is ensured by conversion of the low-frequency coherence (ρ_{21}) to the high-frequency coherences (R_{31} and R_{32}). The latter effect manifests itself explicitly in the constant of motion (19).

The mechanism of this conversion and development of inversionless superradiance allows a simple interpretation when the low level is two-fold degenerate and the frequency of the atomic transition coincides precisely with the cavity frequency ($\Delta_{31}=\Delta_{32}=\omega_{21}=0$). Then, passing to a new basis of atomic states, $|+\rangle=(\mu_{31}|1\rangle+\mu_{32}|2\rangle)/\sqrt{2}$ and $|-\rangle=(\mu_{31}|1\rangle-\mu_{32}|2\rangle)/\sqrt{2}$, we see that the state $|+\rangle$ is active in the superradiant transition (the dipole moment of the $|3\rangle\rightarrow|+\rangle$ transition is finite and equal to $\langle3|\hat{d}|+\rangle=\sqrt{2}\,d$). At the same time, the superposition $|-\rangle$, which is orthogonal to $|+\rangle$, does not interact with the upper state $|3\rangle$ ($\langle3|\hat{d}|-\rangle$

$=0$) and in this sense is passive in superradiance. Thus, in the new basis the problem is equivalent to the two-level one, and the system of equations (11)–(17) reduces to

$$\dot{R}_{3+}=\frac{1}{2}\,\epsilon(\rho_{33}-\rho_{++}), \tag{21}$$

$$\dot{\rho}_{--}=0, \tag{22}$$

$$\dot{\rho}_{++}=\epsilon R_{3+}, \quad \dot{\rho}_{33}=-\epsilon R_{3+}, \tag{23}$$

$$\dot{\epsilon}=4\Omega^2 R_{3+}, \tag{24}$$

where ρ_{++} and ρ_{--} are, respectively, the populations of the active and passive states, $\epsilon=2\sqrt{2}\,\varepsilon$, and R_{3+} is the active-channel coherence:

$$\rho_{++}=\frac{1}{2}(\mu_{31}^2\rho_{11}+\mu_{32}^2\rho_{22}+2\mu_{31}\mu_{32}\,\mathrm{Re}\,\rho_{21}), \tag{25}$$

$$\rho_{--}=\frac{1}{2}(\mu_{32}^2\rho_{11}+\mu_{31}^2\rho_{22}-2\mu_{31}\mu_{32}\,\mathrm{Re}\,\rho_{21}), \tag{26}$$

$$R_{3+}=\frac{1}{\sqrt{2}}(\mu_{31}R_{31}+\mu_{32}R_{32}). \tag{27}$$

Equations (26) and (25) imply that ρ_{--} and $\rho_{33}+\rho_{++}$ are constants of motion. By introducing the half-difference of populations in the active channel, $W=(\rho_{33}-\rho_{++})/2$, we reduce (25)–(27) to

$$\dot{R}_{3+}=\epsilon W, \quad \dot{W}=-\epsilon R_{3+}, \tag{28}$$

$$\dot{\epsilon}=4\Omega^2 R_{3+}. \tag{29}$$

This system of equations describes the cooperative emission in an ensemble of two-level atoms in the mean-field approximation.[2–4,6,8,9] Its solution demonstrates superradiant behavior in the presence of initial inversion in the active channel ($W(0)>0$), i.e., at time zero the following condition must be met:

$$\rho_{33}(0)>\rho_{++}(0)=\frac{1}{2}[\mu_{31}^2\rho_{11}(0)+\mu_{32}^2\rho_{22}(0)+2\mu_{31}\mu_{32}\mathrm{Re}\rho_{21}(0)]. \tag{30}$$

If we prepare the system of atoms in such a way that the active states are not populated ($\rho_{++}(0)=0$),

$$\rho_{11}(0)=\frac{1}{2}\mu_{32}^2[1-\rho_{33}(0)], \quad \rho_{22}(0)=\frac{1}{2}\mu_{31}^2[1-\rho_{33}(0)], \tag{31}$$

$$\rho_{21}(0)=-\frac{1}{2}\mu_{31}\mu_{32}[1-\rho_{33}(0)], \tag{32}$$

where $1-\rho_{33}(0)$ is the total population of the lower level, then inversion between the upper and active states is achieved for all populations of the upper state. This situation can also be achieved in the case $\rho_{33}(0)<\rho_{11}(0)+\rho_{22}(0)$, i.e., without population inversion in the large. At the same time, if initially the states of the lower doublet are populated in an incoherent manner, i.e., $\rho_{21}(0)=0$, the fact that there is population inversion for one of the active transitions is not sufficient for superradiance to occur (superradiance is also a coherent effect).

If we substitute $R_{3+}=(\Omega_0/2\Omega)^2\sin\theta$, $W=(\Omega_0/2\Omega)^2\cos\theta$, and $\epsilon=\dot{\theta}$, where $\Omega_0=2\Omega[R_{3+}^2(0)+W^2(0)]^{1/4}$ in the system of first-order differential equations (28) and (29), we obtain the second-order equation

$$\ddot{\theta}-\Omega_0^2\sin\theta=0, \tag{33}$$

which describes nonlinear oscillations of a simple pendulum. The oscillation period T_0 depends on the initial deflection of the pendulum from the equilibrium position, $\theta(0)$. When the initial polarization R_{3+} is small and $W(0)$ is positive (conditions characteristic of superradiance), the angle $\theta(0)$ is close to zero. Here the oscillation period T_0 is approximately $4\Omega_0\ln[8/\theta(0)]$ (see Ref. 9).

4. A NONDEGENERATE DOUBLET

The kinetics of superradiance in the presence of splitting of the lower doublet ($\omega_{21}\neq 0$) was studied by solving the system of equations (11)–(17)numerically. The natural frequency of the cavity, ω_c, was selected as the arithmetic mean of the frequencies ω_{31} and ω_{32}, i.e., $\omega_c=(\omega_{31}+\omega_{32})/2$. For simplicity the dipole moments of the $3\leftrightarrow 1$ and $3\leftrightarrow 2$ transitions were taken to be equal: $\mu_{31}=\mu_{32}=1$.

The calculations were done with the following initial data: $\rho_{11}(0)=\rho_{22}(0)=0.35$, $\rho_{33}(0)=0.3$, $\rho_{21}(0)=\pm 0.35$, $R_{32}(0)=R_{31}(0)=10^{-8}$, and $\varepsilon(0)=0$. The positive (negative) sign of $\rho_{21}(0)$ means that the system has been prepared in the active (passive) state. Here there is no population inversion in the $3\rightarrow 1$ and $3\rightarrow 2$ channels: $\rho_{33}(0)-\rho_{11}(0)=\rho_{33}(0)-\rho_{22}(0)=-0.05$. At the same time, population inversion between the upper and active states is 0.3. Finite values of $R_{31}(0)$ and $R_{32}(0)$ are needed for initiating superradiance. Here we are interested in fluctuations of superradiance, so that the initial values $R_{31}(0)$ and $R_{32}(0)$ are specified as deterministic parameters, which corresponds to the conditions for stimulated superradiance.[17,18]

The upper half of Fig. 2 represents the results of calculating the kinetics of deexcitation of the system obtained for negative values of $\rho_{21}(0)$ (initially the active state is unpopulated). In the absence of splitting of the lower levels, we have the periodic superradiance regime, which is described by the nonlinear pendulum model (33). Splitting of the lower levels ($\omega_{21}\neq 0$) results in modulation of superradiance signals (slow modulation on the scale of the cooperative frequency Ω for $\omega_{21}<\Omega$ and rapid modulation for $\omega_{21}>\Omega$). The reason is that when the splitting of the lower doublet is finite, the $|+\rangle$ and $|-\rangle$ states are not stationary, so that with the passage of time the active state periodically becomes the passive state.

The lower half of Fig. 2 present similar results obtained, however, for a positive value of the initial low-frequency coherence $\rho_{21}(0)$ (initially, the active state is populated, and population inversion between this state and the upper state needed for superradiance to develop is absent). In complete agreement with the results of Sec. 3, no superradiance signal is detected at zero splitting. For small values of ω_{21} (smaller

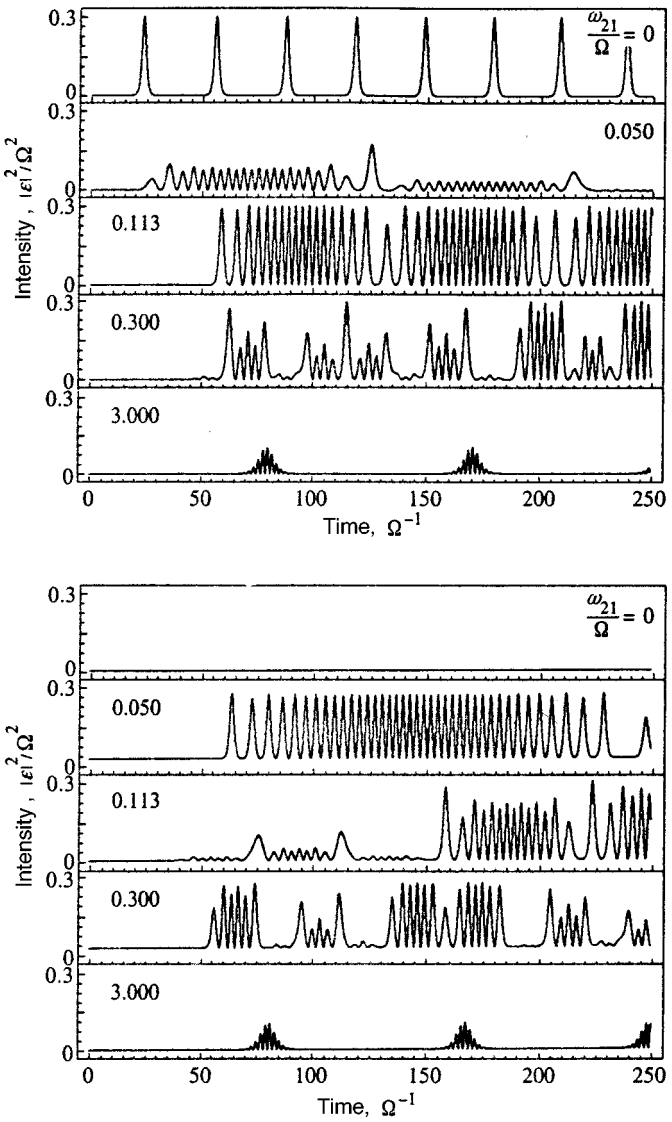

FIG. 2. Kinetics of inversionless superradiance for different values of the doublet splitting ω_{21} (indicated in the figure). The cavity frequency is tuned exactly to the midpoint between the levels 1 and 2, $\omega_c=(\omega_{31}+\omega_{32})/2$. The results presented in the upper half of the figure were obtained for a negative value of the low-frequency coherence $[\rho_{21}(0)=-0.35$; initially the antisymmetric superposition of the states of the doublet is populated], while the results presented in the lower half were obtained for a positive value of the low-frequency coherence $[\rho_{21}(0)=0.35$; initially the symmetric superposition of the states of the doublet is populated]. The other parameters have been chosen as follows: $\rho_{11}(0)=\rho_{22}(0)=0.35$, $\rho_{33}(0)=0.3$, $R_{32}(0)=R_{31}(0)=10^{-8}$, and $\varepsilon(0)=0$.

than Ω) the signal is delayed by a time interval that is exactly equal to the half-period of the low-frequency coherence oscillations, π/ω_{21}, i.e., a time interval after which the sign of the signal changes.

At first it seems strange that these signals do not exhibit the expected exact periodicity in time (signal period $2\pi/\omega_{21}$). We relate this behavior to the population trapping effect[19] (a fraction of the population of the lower state is trapped in a passive state not related to the upper state). In the cases under investigation such an effect is possible, since for a small value of the splitting ($\omega_{21}<\Omega$) the superradiance spectrum overlaps the low-frequency doublet, which is the condition for population trapping.

As the doublet splitting increases, superradiance signals in the two cases in question resemble each other more and more irrespective of the sign of the initial coherence $\rho_{21}(0)$ (cf. the upper and lower halves of Fig. 2). Hence, when the lower level is split, the initial phase $\rho_{21}(0)$ plays no significant role. As ω_{21} increases, the superradiance intensity decreases due to the increasing detuning of the transition frequency ω_{31} and ω_{32} from the cavity frequency ω_c.

5. FORMATION OF LOW-FREQUENCY COHERENCE BY AN EXTERNAL FIELD

The above treatment of inversionless superradiance presupposes the existence of initial coherence of the states $|1\rangle$ and $|2\rangle$. In this section we study the possibility of creating the necessary coherence by introducing an external pulsed field interacting with the $1\leftrightarrow 2$ transition. This interaction may occur either in the electrodipole approximation if the quantum system has no inversion center or in the magnetodipole approximation if the doublet in question is magnetically active.

Let us examine a situation in which the length of the pulse forming the coherence in the low-frequency channel is smaller than the time by which the next superradiance pulse is delayed. Then it is obvious that the evolution of the low-frequency channel $2\leftrightarrow 1$ in the external field and the development of superradiance are separated in time. Below we demonstrate the possibility of such separation directly by a numerical solution of the Maxwell–Bloch equations that incorporate an external field into the low-frequency channel.

In the notation adopted here, the evolution of the doublet is described by the following system of equations for a two-level atom in an external field:

$$\dot{\rho}_{21}=-i\omega_{21}\rho_{21}-2i\mathscr{F}_i(t)Z, \tag{34}$$

$$\dot{Z}=i\mathscr{F}_i(t)(\rho_{21}-\rho_{12}), \tag{35}$$

where $Z=(\rho_{22}-\rho_{11})/2$, and $\mathscr{F}_i(t)$ is the external field measured in frequency units.

5.1. Wideband excitation

If the pulse length T_p is such that $T_p<2\pi/\omega_{21}$ and, in addition, $|\mathscr{F}_i|>\omega_{21}$, the term with ω_{21} in (34) can be dropped. Then the solution of the system of equations (34) and (35) has the form

$$Z=B\cos(\mathscr{A}+\phi_1), \quad \mathrm{Im}\,\rho_{21}=-B\sin(\mathscr{A}+\phi_1), \tag{36}$$

where $\mathscr{A}=2\int_0^t\mathscr{F}_i(t)\,dt$ is the field pulse area, $B=\sqrt{Z^2(0)+[\mathrm{Im}\rho_{21}(0)]^2}$, and $\sin\phi_1=\mathrm{Im}\rho_{21}(0)/B$. Accordingly, for the density-matrix elements we have

$$\rho_{21}=\mathrm{Re}\,\rho_{21}(0)-iB\sin(\mathscr{A}+\phi_1), \tag{37}$$

$$\rho_{11}=\frac{1}{2}[\rho_{11}(0)+\rho_{22}(0)]-B\cos(\mathscr{A}+\phi_1), \tag{38}$$

$$\rho_{22}=\frac{1}{2}[\rho_{11}(0)+\rho_{22}(0)]+B\cos(\mathscr{A}+\phi_1). \tag{39}$$

If initially the system is in an incoherent state, i.e., $\rho_{21}(0)=0$, Eqs. (37)–(39) become

$$\rho_{21}=-\frac{1}{2}i[\rho_{22}(0)-\rho_{11}(0)]\sin\mathscr{A}, \tag{40}$$

$$\rho_{11}=\frac{1}{2}[\rho_{22}(0)+\rho_{11}(0)]-\frac{1}{2}[\rho_{22}(0)-\rho_{11}(0)]\cos\mathscr{A}, \tag{41}$$

$$\rho_{22}=\frac{1}{2}[\rho_{22}(0)+\rho_{11}(0)]+\frac{1}{2}[\rho_{22}(0)-\rho_{11}(0)]\cos\mathscr{A}. \tag{42}$$

As can be seen from (40)–(42), maximum coherence (in absolute value) is achieved when only one level, 1 or 2, is populated initially and the field pulse area is $\pi/2$.

Attaining maximum coherence corresponds to the optimum condition for inversionless superradiance only if the values of the transition dipole moments μ_{31} and μ_{32} are the same. When μ_{31} differs from μ_{32}, the pulse area must be selected on the basis of (31) and (32). Then if level 1 is initially populated we have

$$\cos\mathscr{A}=\frac{1}{2}(\mu_{32}^2-\mu_{31}^2), \quad \sin\mathscr{A}=\mu_{31}\mu_{32}. \tag{43}$$

Note that the value of the pulse area determines only the magnitude of ρ_{21}, although the phase is also important for realizing the optimum condition for inversionless superradiance [see (32)]. As (37) implies, in this case an external pulse generates only the imaginary part of the low-frequency coherence, which does not directly initiate superradiant emission. The optimum phase of coherence emerges in the process of the time evolution of $\rho_{21}\propto\exp(-i\omega_{21}t)$ after the pulse has already acted.

5.2. A resonant low-frequency pulse

Now we study the situation with $\omega_i=\omega_{21}$, $T_p\gg 2\pi/\omega_{21}$, and $|\mathscr{F}_i|\ll\omega_{21}$. Under these conditions we can use the expansions $\mathscr{F}_i(t)=f_i(t)\exp(-i\omega_{21}t)+\text{c.c.}$ and $\rho_{21}(t)=R_{21}(t)\exp(-i\omega_{21}t)$, where $f_i(t)$ and $R_{21}(t)$ are the amplitudes that slowly vary on the scale of the period $2\pi/\omega_{21}$, and pass in the usual way from Eqs. (34) and (35) to the equations for the slow variables, whose solution is

$$R_{21}=C\sin(\mathscr{A}+\phi_2)+i\,\mathrm{Im}\,R_{21}(0), \tag{44}$$

$$Z=C\cos(\mathscr{A}+\phi_2), \tag{45}$$

where $\mathscr{A}=2\int_0^t f_i(t)\,dt$ is the pulse area, $C=\sqrt{Z^2(0)+[\mathrm{Re}R_{21}(0)]^2}$, and $\tan\phi_2=\mathrm{Re}R_{21}(0)/Z(0)$. In particular, for a state that is initially incoherent ($R_{21}(0)=0$) we have

$$R_{21}=Z(0)\sin\mathscr{A}, \quad Z=Z(0)\cos\mathscr{A}. \tag{46}$$

This means that the maximum value of the absolute value of the coherence, $|R_{21}|=|\rho_{22}(0)-\rho_{11}(0)|/2$, is attained for a pulse area $\mathscr{A}=\pi/2$ (similar to the case discussed in Sec. 5.1). Here the magnitude $|R_{21}|$ of the coherence is largest (for a given total population of the doublet) if only one level is populated (which is also completely analogous to Sec. 5.1). However, its optimum value for distinct μ_{31} and μ_{32} is still given by (43). Since in the limit in question ($T_p\gg 2\pi/\omega_{21}$), ρ_{21} rapidly oscillates with ω_{21}, the phase of R_{21} after the external pulse has travelled through the system is unimportant, in contrast to the previous case.

6. INITIATING INVERSIONLESS SUPERRADIANCE BY AN EXTERNAL LOW-FREQUENCY FIELD

The system of equations describing the behavior of the three-level medium in question and allowing for the action of an external field in the low-frequency channel $2\leftrightarrow 1$ has the form

$$\dot{R}_{31}=-i\Delta_{31}R_{31}+\varepsilon[\mu_{31}(\rho_{33}-\rho_{11})-\mu_{32}\rho_{21}]-i\mathscr{F}_iR_{32}, \tag{47}$$

$$\dot{R}_{32}=-i\Delta_{32}R_{32}+\varepsilon[\mu_{32}(\rho_{33}-\rho_{22})-\mu_{31}\rho_{12}]-i\mathscr{F}_iR_{31}, \tag{48}$$

$$\dot{\rho}_{21}=-i\omega_{21}\rho_{21}+\mu_{31}\varepsilon R_{32}^*+\mu_{32}\varepsilon^*R_{31}-i\mathscr{F}_i(\rho_{22}-\rho_{11}), \tag{49}$$

$$\dot{\rho}_{11}=\mu_{31}(\varepsilon R_{31}^*+\varepsilon^*R_{31})+i\mathscr{F}_i(\rho_{21}-\rho_{12}), \tag{50}$$

$$\dot{\rho}_{22}=\mu_{32}(\varepsilon R_{32}^*+\varepsilon^*R_{32})-i\mathscr{F}_i(\rho_{21}-\rho_{12}), \tag{51}$$

$$\dot{\rho}_{33}=-\mu_{31}(\varepsilon R_{31}^*+\varepsilon^*R_{31})-\mu_{32}(\varepsilon R_{32}^*+\varepsilon^*R_{32}), \tag{52}$$

$$\dot{\varepsilon}=\Omega^2(\mu_{31}R_{31}+\mu_{32}R_{32}). \tag{53}$$

Figure 3 depicts the results of numerical calculations of superradiance kinetics corresponding to the excitation of the low-frequency coherence by the scheme discussed in Sec. 5.1. The calculations were done with initial data $\rho_{11}(0)=0.7$, $\rho_{22}(0)=0$, $\rho_{21}(0)=0$, $\rho_{33}(0)=0.3$, $R_{32}(0)=R_{31}(0)=10^{-8}$, and $\varepsilon(0)=0$. The external field was a rectangular $\pi/2$ pulse of height Ω (in frequency units). Note that if there were no external field in the low-frequency channel, the system would not radiate.

When analyzing the results, it is convenient to bear in mind the following four time scales that are natural to the problem in question: the length T_p of the external-field pulse (the shortest time), the period $T_{21}=2\pi/\omega_{21}$ of variation of the low-frequency coherence, the superradiance time $T_R\approx\Omega^{-1}$, and the time delay of the superradiance pulse, T_D, due to the smallness of the initial values of R_{31} and R_{32}.

The external pulse generates the purely imaginary off-diagonal matrix element ρ_{21}, which evolves in time and initiates superradiance. As Fig. 3 shows, a small variation in the value of the T_D-to-T_{21} ratio (for $T_{21}>T_R$, which is actually the case) has a strong effect on the intensity of the radiated pulses. The explanation is that there is a time interval ΔT (shorter than $T_{21}/4$) that separates the moments of onset of the threshold ($\rho_{33}=\rho_{++}$) and optimum ($\rho_{++}=0$) conditions for inversionless superradiance. For $T_D<\Delta T$ the radiating process begins prematurely, in the sense that the optimum condition is not attained (Fig. 3a). At $T_D\approx\Delta T$ the radiating process begins exactly in optimum conditions (Fig. 3b). As the doublet splitting increases and hence T_{21} decreases, so that $T_D>\Delta T$, superradiance is delayed (Fig. 3c). But if T_D is much longer than ΔT, the onset of superradiance

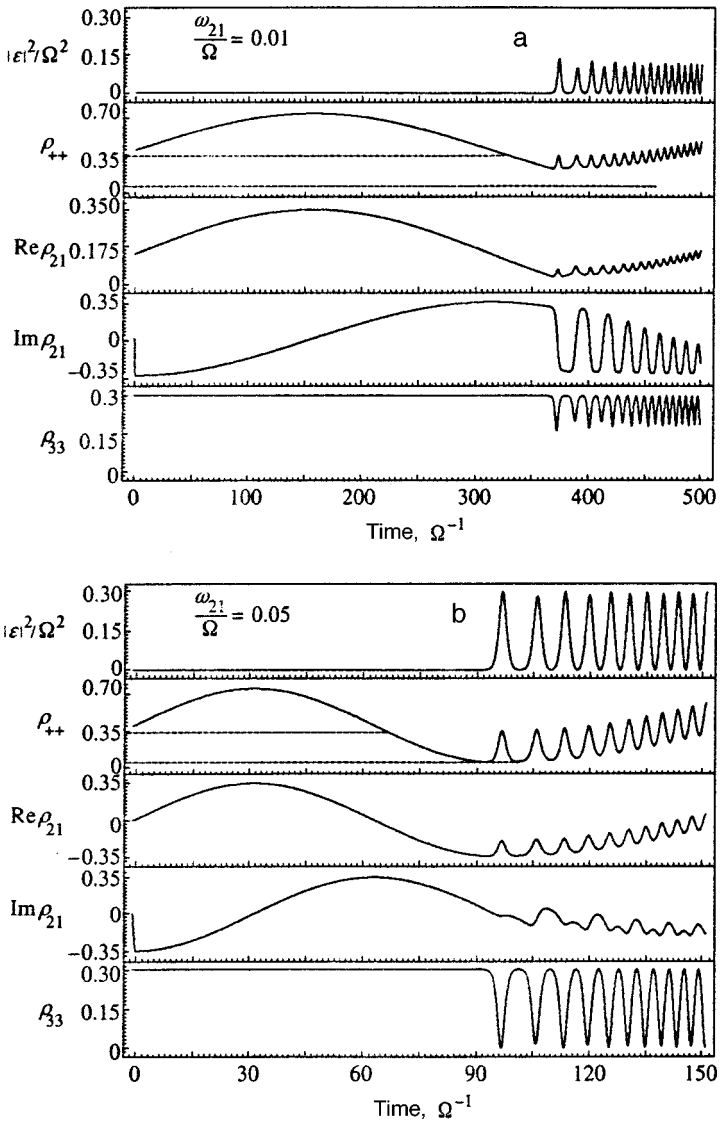

FIG. 3. Kinetics of inversionless superradiance in the formation of the low-frequency coherence by a an external field in the form of a rectangular $\pi/2$-pulse of height Ω (in frequency units) for different values of the doublet splitting ω_{21} (indicated in the figure); $\omega_c = (\omega_{31}+\omega_{32})/2$. The initial data are $\rho_{11}(0)=0.7$, $\rho_{22}(0)=0$, $\rho_{21}(0)=0$, $\rho_{33}(0)=0.3$, $R_{32}(0)=R_{31}(0)=10^{-8}$, and $\varepsilon(0)=0$.

may be postponed, so to say, to the next favorable cycle (Fig. 3d). Superradiance intensity is determined by the difference in the populations of the upper and active states at the time the radiating process begins and remains, for all practical purposes, the same.

The kinetics of inversionless superradiance calculated under the condition that the low-frequency coherence is excited by a resonant $\pi/2$ pulse (the scheme developed in Sec. 5.2) is depicted in Fig. 4. The initial data for the calculation were take from the previous case. The low-frequency coherence was formed by a $\pi/2$ pulse of frequency $\omega_i=\omega_{21}=3\Omega$ and a rectangular envelope of length $T_p=80\Omega^{-1}$. As a result of the action of the pulse, the subsystem of states of the doublet is prepared in the superposition state $|t\rangle = (1/\sqrt{2})[|1\rangle+\exp(-i\omega_{21}t)|2\rangle]$, which is periodically converted (with the passage of time) to the passive combination $|-\rangle=(1/\sqrt{2})(|1\rangle-|2\rangle)$, leaving the active state $|+\rangle = (1/\sqrt{2})(|1\rangle+|2\rangle)$ unpopulated for this period and thus creating the conditions for inversion between states $|3\rangle$ and $|+\rangle$. After a certain delay, needed for the cooperative dipole moment to develop, the system emits a superradiance pulse representing beats of two signals of the same amplitude with the difference frequency ω_{21}. Note that the superradiance kinetics in this case coincides almost perfectly with that calculated in conditions where the low-frequency coherence is part of the initial data (see the curves in Fig. 2 obtained for $\omega_{21}=3\Omega$).

We conclude this section by discussing the possibility of detecting inversionless superradiance in atomic sodium vapor, which Kocharovskaya and Mandel[20] considered a suitable object for realizing inversionless amplification (related to inversionless superradiance). It is well known that the ground state of a sodium atom $(3S)$ is the spin–orbit doublet $3\,^2S_{3/2}$,$3^2\,S_{1/2}$ with the transition frequency 1.77 GHz, which in combination with an appropriate excited state of opposite parity (say, $3P$) may serve as a model of a three-level Λ-system. Estimates of the characteristic superradiance time scale (T_R) for a vapor density of roughly $10^{11}\,\mathrm{cm}^{-3}$ with a 10% occupancy of the excited state yield ~ 0.1 ns, which is shorter by a factor of ten than the relaxation time $T_2^*\sim 1$ ns due to Doppler broadening.[9] Here the superradiance delay time $T_D\sim 10T_R\sim 1\ \mathrm{ns}\sim T_2^*$, which makes it possible (at least in principle) to realize the deexcitation regime in question. Thus, creating a coherent state of the doublet in the

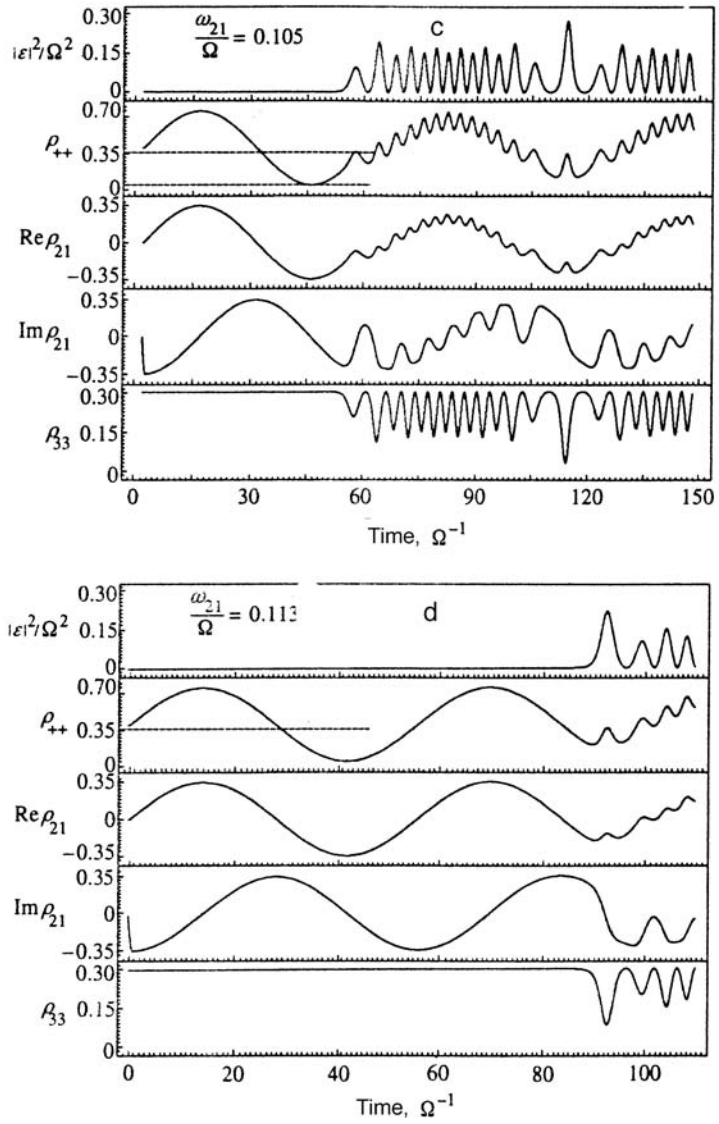

FIG. 3 (*Continued.*)

ground state requires a gigahertz electromagnetic $\pi/2$ pulse shorter than 10 ns (i.e., shorter than the lifetime of the excited state), values that can be attained by the spin echo technique.[21] Another possible effective way of creating the low-frequency coherence is to use an external optical nanosecond pulse that is in resonance with another (nonsuperradiant) transition, as suggested in Ref. 10.

Among other possible objects in which the inversionless superradiance effect can appear are crystals activated by rare-earth ions. Auzel *et al.*[22] reported detecting superradiance in $LiYF_4$ with Er^{3+} ions that involved the $^4I_{11/2} \rightarrow {}^4I_{13/2}$ transition. The lifetime of the excited state $^4I_{11/2}$ is of order 0.1 s (see Ref. 22; this transition is dipole-forbidden in a single Er^{3+} ion). This means that it is easy to create the low-frequency coherence involving the Zeeman sublevels of Er^{3+} in the ground state by applying a microwave pulsed field. For a density of excited Er^{3+} ions in the $^4I_{11/2}$ state of order $5\times10^{11}\,\mathrm{cm}^{-3}$ (see Ref. 22), the characteristic superradiance time scale T_R is in the nanosecond range (~ 10 ns).

7. CONCLUSION

In multiatomic systems consisting of three-level atoms with a doublet structure of the ground state (the Λ configuration of transitions), superradiance without population inversion in the large is possible. This effect requires the presence of initial coherence between the states of the doublet, which can be created either by a wideband pulse or a resonant low-frequency coherent pulse with a certain area.

When the low-frequency coherence is produced by a pulsed field, the superradiance intensity strongly depends on the relationship between the splitting of the low-frequency doublet and delay time of the superradiant pulse. As is known, the latter can be controlled by using the stimulated-superradiance approach,[17,18] in which the process is initiated by a small-area ultrashort pulse in quasiresonance with the $3\leftrightarrow1$ and $3\leftrightarrow2$ transitions. Such a pulse generates initial polarization in the superradiance channels $3\leftrightarrow1$ and $3\leftrightarrow2$ that exceeds the level of the spontaneous fluctuations of polarization and thus reduces the superradiance delay time.

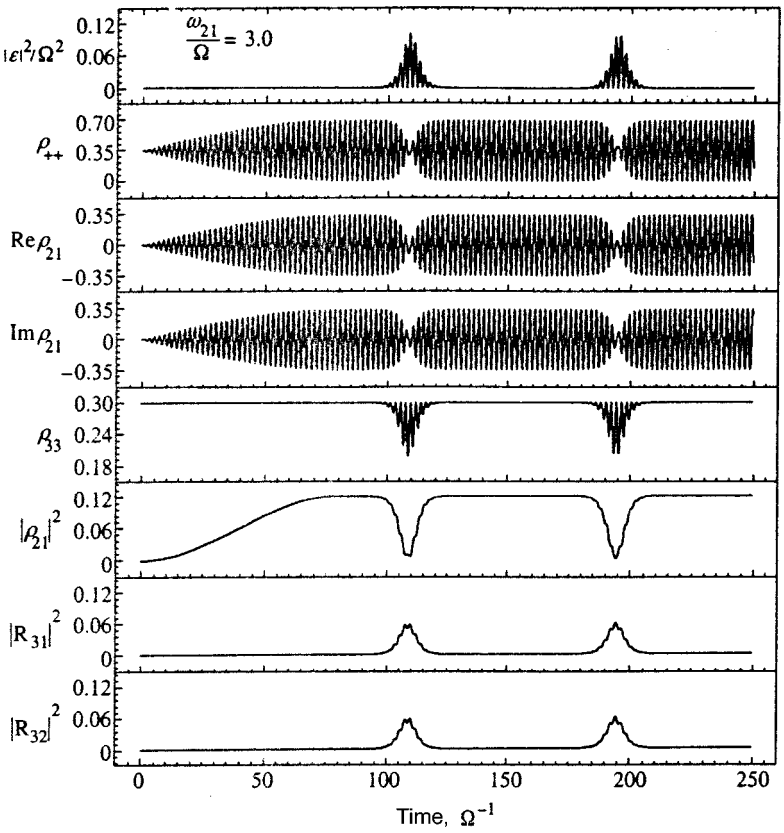

FIG. 4. As in Fig. 3, but the low-frequency coherence is formed by an external field in the form of a resonant $\pi/2$ pulse with a rectangular envelope of length $T_p = 80\Omega^{-1}$.

Controlling stimulated superradiance by varying the area of the ultrashort pulse, we can optimize the intensity of inversionless superradiance. Note that if the dipole moments of the active transitions are not collinear, the effect weakens due to the weakening of channel coupling and may disappear completely when their polarizations are orthogonal.

ACKNOWLEDGMENTS

The authors would like to thank the participants of the St. Petersburg Workshop in quantum optics for discussing the results of the present research. The work was supported by the Russian Fund for Fundamental Research (Project 97-02-18027).

*)E-mail: vicmal@spain.stoic.spb.su

[1] R. H. Dicke, Phys. Rev. **93**, 99 (1954).
[2] R. Bonifacio, P. Schwendimann, and F. Haake, Phys. Rev. A **4**, 302 (1971); **4**, 854 (1971).
[3] R. Bonifacio and L. A. Lugiato, Phys. Rev. A **11**, 1507 (1975); **12**, 587 (1975).
[4] M. S. Feld and J. C. MacGillivrey, in *Coherent Nonlinear Optics*, M. S. Feld and V. S. Letokhov (Eds.), Springer, Berlin (1980), p. 7.
[5] A. V. Andreev, V. I. Emel'yanov, and Yu. A. Il'inskiĭ, Usp. Fiz. Nauk **131**, 653 (1980) [Sov. Phys. Usp. **23**, 439 (1980)].
[6] M. Gross and S. Haroche, Phys. Rep. **93**, 301 (1982).
[7] V. V. Zheleznyakov, V. V. Kocharovskiĭ, and Vl. V. Kocharovskiĭ, Usp. Fiz. Nauk **159**, 193 (1989) [Sov. Phys. Usp. **32**, 835 (1989)].
[8] A. V. Andreev, V. I. Emel'yanov, and Yu. A. Il'inskiĭ, *Cooperative Effects in Optics*, IOPP, Bristol (1993).
[9] M. G. Benedict, A. M. Ermolaev, V. A. Malyshev, I. V. Sokolov, and E. D. Trifonov, *Super-radiance: Multiatomic Coherent Emission*, IOPP, Bristol (1996).
[10] S. O. Elyutin, S. M. Zakharov, and É. A. Manykin, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **22**, 1213 (1979).
[11] O. A. Kocharovskaya and Ya. I. Khanin, JETP Lett. **48**, 630 (1988).
[12] S. E. Harris, Phys. Rev. Lett. **62**, 1033 (1989).
[13] Ya. I. Khanin and O. A. Kocharovskaya, J. Opt. Soc. Am. B **7**, 2016 (1990).
[14] O. A. Kocharovskaya, Phys. Rep. **219**, 175 (1992).
[15] Quantum Optics **6**(8) (1994).
[16] V. A. Malyshev, I. V. Ryzhov, E. D. Trifonov, and A. I. Zaitsev, Laser Phys. **8**, 494 (1998).
[17] N. W. Carlson, D. J. Jackson, A. L. Schawlow, M. Gross, and S. Haroche, Opt. Commun. **32**, 350 (1980).
[18] R. F. Malikov and E. D. Trifonov, Opt. Commun. **52**, 74 (1984).
[19] B. D. Agap'ev, M. B. Gornyĭ, B. D. Matisov, and Yu. V. Rozhdestvenskiĭ, Usp. Fiz. Nauk **163**, No. 9, 1 (1993) [Phys. Usp. **36**, 763, 1086 (1993)].
[20] O. Kocharovskaya and P. Mandel, Phys. Rev. A **42**, 523 (1990).
[21] K. M. Salikhov, A. G. Semenov, and Yu. D. Tsvetkov, *Spin Echo and Its Application* [in Russian], Nauka, Novosibirsk (1976).
[22] F. Auzel, S. Hubert, and D. Meichenin, Europhys. Lett. **7**, 459 (1988).

Translated by Eugene Yankovsky

Asymptotic coefficients for atoms and ions

V. D. Mur and B. M. Karnakov*)

Moscow Engineering Physics Institute (Technical University), 115409 Moscow, Russia

V. S. Popov

Institute of Theoretical and Experimental Physics, 117218 Moscow, Russia

(Submitted 14 July 1998)
Zh. Éksp. Teor. Fiz. **115**, 521–541 (February 1999)

We derive convenient analytical formulas in the effective-range approximation for the asymptotic coefficient C_κ of the radial wave function at infinity and for the average radius of the system. A comparison with the results of numerical calculations (done by the Hartree–Fock method) for multi-electron atoms and ions reveals that this approximation has good accuracy for valence *s*-electrons in all atoms from hydrogen to uranium. We calculate the values of the scattering lengths and the effective ranges for electron–atom and electron–ion scattering. We also examine the quasiclassical approximation for C_κ. Finally, we discuss the logarithmic increase in the effective ranges of ns states as $n\to\infty$. © *1999 American Institute of Physics.* [S1063-7761(99)01102-6]

1. The wave function of the *s* state has the asymptotic behavior

$$\psi_\kappa(r)=C_\kappa\sqrt{\frac{\kappa^3}{\pi}}e^{-\kappa r}(\kappa r)^{\nu-1}\left[1-\frac{\nu^2-\nu}{2\kappa r}+\cdots\right],$$

$$r\gg r_0,\quad \frac{|\nu-1|}{\kappa^2 a_B}, \tag{1}$$

where $\int\psi_\kappa^2 d^3r=1$, $\lambda=\sqrt{-2E}$, E is the level energy, $\nu=Z/\kappa$, with Z the charge of the atomic core, r_0 is the radius of the core, $a_B=1/Z$ is the Bohr radius, and atomic units ($\hbar=m=e=1$) are used throughout

The asymptotic coefficients C_κ are important physical parameters of bound states and are continually encountered in quantum mechanics,[1] atomic and nuclear physics,[2–8] in the inverse problem of quantum scattering theory,[9,10] etc.[1)] Their exact calculation constitutes a fairly involved problem and, e.g, for heavy atoms requires numerical calculation of the Hartree–Fock equations for multi-electron systems, with the calculation error sometimes reaching 10–30% or even larger values.[11] In this paper we specify simple analytical approximations for the coefficients C_κ, approximations that describe the results of numerical calculations fairly well. These formulas can also be used for other Coulomb systems with short-range interaction (e.g., in the theory of hadronic atoms[12]).

The plan of the paper is as follows. In Sec. 2 we introduce the effective range expansion for the coefficients C_κ, compare it with the results of numerical calculations, and give the results of calculations of the scattering lengths and effective ranges for some atoms and ions. In Sec. 3 we illustrate the expansion (3) using model potentials that allow an exact solution (in particular, we give the explicit expressions for the *s*-scattering length a_s and the effective range r_s for the case of the Hulthén potential). In Sec. 4 we consider the quasiclassical approximation for the coefficients C_κ. In Sec. 5 we discuss the case $l\neq 0$ and in Sec. 6, the logarithmic increase in the effective ranges of ns states as $n\to\infty$. The Appendix deals with the derivation of the expansion (3), the details of calculations, and some lengthy formulas.

2. The effective range expansion. We assume that the potential is given by the formula

$$V(r)=-Zr^{-1}+U(r),$$

where $U(r)$ is its short-range part with a characteristic range r_0 of the forces. The normalized wave function of the shallow *s* level has the form

$$\chi_\kappa(r)=\sqrt{4\pi}\,r\psi_\kappa(r)$$
$$=2^{1-\nu}\kappa^{1/2}C_\kappa\begin{cases}\varphi(r)/\Gamma(1-\nu), & r\lesssim r_0,\\ W_{\nu,1/2}(2\kappa r), & r\gg r_0,\end{cases} \tag{2}$$

where $l=0$, $\kappa r_0\ll 1$, $\varphi(r)$ is the wave function corresponding to zero energy, $W_{\nu,1/2}$ is the Whittaker function, and ν is the Sommerfeld parameter (for the discrete spectrum). The region where the expressions in (2) are matched is $r_0\ll r\ll\min(\kappa^{-1},a_B)$, and the buildup of the normalization integral occurs primarily at distances $r>r_0$ and can be calculated analytically. The correction due to small distances ($r\lesssim r_0$) depends on the shape of the potential $U(r)$ and can be expressed in terms of the nuclear Coulomb effective range r_{cs} (see Refs. 13–15). As a result we arrive at the expansion (see Appendix A)

$$C_\kappa=C_\kappa^{(0)}\{1-c_1\kappa r_{cs}+O((\kappa r_0)^3)\}^{-1/2}, \tag{3}$$

where $C_\kappa^{(0)}$ and c_1 are universal functions of the parameter ν (see Figs. 1 and 2):

$$C_\kappa^{(0)}=\frac{2^{\nu-1}}{\Gamma(1+\nu)}F(\nu),\quad c_1=\frac{1}{2}\left[\frac{\sin\pi\nu}{\pi\nu}F(\nu)\right]^2, \tag{4}$$

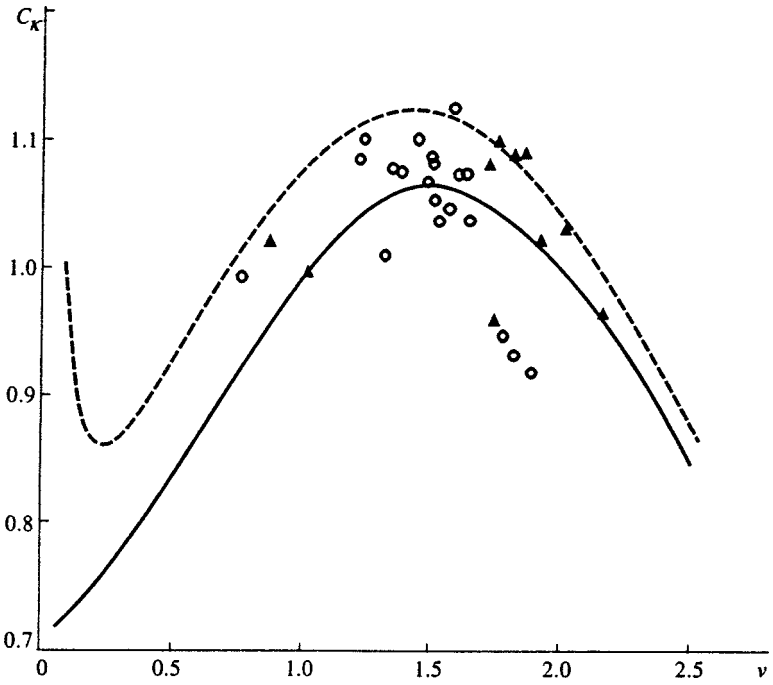

FIG. 1. The asymptotic coefficient at infinity for neutral atoms (○) and singly charged positive ions (▲): the solid curve corresponds to $C_\kappa^{(0)}$, the dashed curve corresponds to the quasiclassical approximation (30), and the points correspond to the results of Hartree–Fock calculations (Ref. 11).

$$F(\nu)=\left\{1-\left(\frac{\sin\pi\nu}{\pi\nu}\right)^2\left[\nu^2\psi'(\nu)-\nu-\frac{1}{2}\right]\right\}^{-1/2}, \qquad (4a)$$

with $\psi'(\nu)$ the trigamma function.[16] When the Coulomb interactio n is switched off ($\nu\to 0$), the range r_{cs} tends to r_s and

$$C_\kappa^{(0)}=\frac{1}{\sqrt{2}}(1+a_1\nu+\cdots), \quad c_1=1-2\nu+\cdots$$

($a_1=\ln 2-1+C=0.2704$, with $C=5772\ldots$), so that the expansion (3) assumes the form corresponding to a short-range potential:

$$\frac{1}{2C_\kappa^2}=1-\kappa r_s+c_3(\kappa r_s)^3+\cdots. \qquad (5)$$

As $\nu\to n=1,2,3,\ldots$, Eq. (4) yields

$$C_\kappa^{(0)}=\frac{2^{n-1}}{n!}[1+b_1\delta+O(\delta^2)], \quad c_1=\frac{1}{2}\delta^2-\delta^3+\ldots, \qquad (6)$$

where $\delta=(\nu-n)/n$, and $b_1=a_1\approx 0.270$ at $n=1$, $b_1\approx -0.459$ at $n=2$, etc. The values $C_\kappa^{(0)}=2^{n-2}/n!$ and $c_1=0$ correspond to a purely Coulomb ($\nu=n$) spectrum.[2)] Finally, for highly excited (Rydberg) states,

$$C_\kappa^{(0)}=\frac{2^{\nu-1}}{\Gamma(\nu+1)}\left[1+\frac{\sin^2\pi\nu}{12\pi^2\nu^3}+\cdots\right],$$

$$c_1=\frac{\sin^2\pi\nu}{2\pi^2\nu^2}[1+O(\nu^{-3})]. \qquad (6a)$$

In the limit $\nu\gg 1$ the coefficient c_1 is small numerically and the dependence of C_κ on the effective range r_{cs} becomes unimportant (the same is true for ν close to integral values $n=1,2,\ldots$).

What makes the expansions (3) and (5) so remarkable is that they contain no terms quadratic in the effective range.[3)] Furthermore, the coefficient s c_3, which we calculated for several model potentials $U(r)$, are small [see Eq. (19)]. This broadens the range of applicability of these expansions.

We compare the results of Hartree–Fock calculations[11] and the zeroth approximation $C_\kappa\approx C_\kappa^{(0)}$, i.e., $r_{cs}=0$ in (3). The values of the coefficients A in Ref. 11 differ from those of C_κ in normalization: $A=2\kappa^{\nu+1/2}C_\kappa$. The open circles in Fig. 1 refer (in order of increasing ν) to the ground states of neutral atoms:[4)] He($\nu=0.744$), Zn(1.203), Au(1.215), Cu(1.326), Ag(1.340), Ru(1.359), Cr(1.418), Y(1.460), Yb(1.475), U(1.484), Ca(1.493), Pa(1.515), Sr(1.546), La(1.563), Li(1.587), Ba(1.616), Na(1.626), K(1.770), Rb(1.805), and Cs(1.869).

In Fig. 1 the values of C_κ are also marked (▲) for several positive ions: $Li^+(\nu=0.848)$, $He^+(1.000)$, $Al^+(1.701)$, $Be^+(1.729)$, $Zn^+(1.741)$, $Cd^+(1.794)$, $Fe^+(1.833)$, $Mg^+(1.903)$, $Ti^+(2.002)$, and $Ca^+(2.141)$ (we considered only the cases where the valence electron is in the

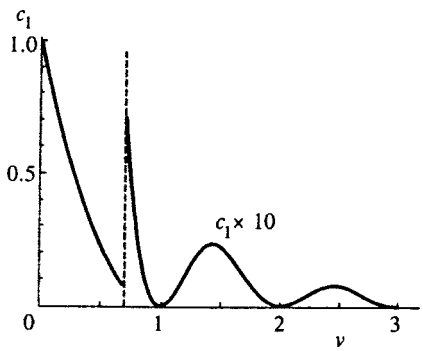

FIG. 2. The coefficient c_1 in the expansion (1) as a function of ν. For $\nu>0.7$, the scale along the vertical axis is increased by a factor of ten.

TABLE I. Low-energy parameters of negative ions.

Ion	κ	a_s	r_s	C_κ
H^-	0.235	6.17	2.65	1.15
Li^-	0.212	6.6	2.7	1.09
Na^-	0.201	7.1	3.0	1.12
K^-	0.192	7.1	2.7	1.03
Rb^-	0.189	6.7	2.2	0.92
Fe^-	0.171	8.4	3.5	1.12
Cu^-	0.301	4.7	1.9	1.09
Ag^-	0.309	4.7	2.1	1.17

Note: The quantities referring to the negative hydrogen ion have been taken from Ref. 10.

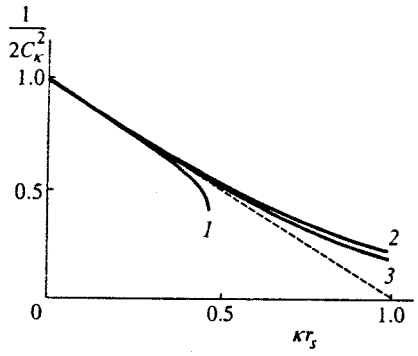

FIG. 3. The dependence of $1/2C_\kappa^2$ on κr_s (for the ground state). Curves *1* and *2* correspond to the Hulthén and Breit potentials, and curve *3* corresponds to a rectangular well.

s state). It is clear that even the zeroth approximation describes rather well the results of Hartree–Fock calculations and has roughly a 10% accuracy (even though the expansion parameter satisfies $\kappa r_{cs} \sim 1$). The explanation is that numerically the coefficients c_1 and c_3 in (3) and (5) are small. Thus, we have $c_1 = 0.0458$ for $\nu = 0.75$, and its maximum value in the region $1 < \nu < \infty$ is equal to $c_m = 0.0236$ ($\nu_m \approx 1.45$). The coefficients of c_3 of the cubic term depend on the shape of the potential $U(r)$, and $|c_3| \leqslant 0.15$ for all the models considered.

Note that the coefficients A have a much larger spread than the C_κ. For instance, $A^2 = 0.18$, 4.00, 8.24, and 42 for the atoms Cs, H, and He and the ion Li^+, respectively, while $C_\kappa^2 = 0.854$, 1.000, 0.986, and 1.04. The coefficients C_κ, in contrast to A, are scale invariant (i.e., do not change under the scaling transformation $r \to \alpha r$ in the Schrödinger equation), so that they depend on the shape of the potential $V(r)$ and not on the depth and range of the potential separately. In particular, in the Coulomb problem, $V(r) = -Z/r$, these coefficients are independent of Z.

Most cases have $C_\kappa > C_\kappa^{(0)}$, i.e., they correspond to positive values of r_{cs} (see Table I), but sometimes $C_\kappa < C_\kappa^{(0)}$ holds (the points in Fig. 1 lying below the solid curve). In the cases of the Be^+ ion and the Cs atom, this cannot be explained by the imprecision of the Hartree–Fock calculations, while for the other atoms, $C_\kappa \approx C_\kappa^{(0)}$ within two standard deviations.

Using (3) or (5) and the effective range expansion,[13,14] we can easily calculate the low-energy parameters, the scattering length and the effective range, from the values of κ and C_κ. For instance in the case where $Z = 0$ (negative ions),

$$\kappa r_s = 1 - \frac{1}{2C_\kappa^2}, \quad a_s^{-1} = \kappa - \frac{1}{2}\kappa^2 r_s, \tag{7}$$

which yields the values of a_s and r_s listed in Table I. This table shows that $a_s \gg r_0$, as it should be for loosely bound ($\kappa r_s \ll 1$) systems. Here the effective ranges r_s vary from two to three Bohr radii and the values of κr_s vary from 0.41 for Rb^- to 0.65 for Ag^-, which is within the limits of applicability of expansion (5) (see Fig. 3 and (19) below). The asymptotic coefficients C_κ listed in Table I have been recalculated from the data of Ref. 11 by the method mentioned earlier.

In the presence of Coulomb interaction, the formulas become somewhat more complicated:

$$\kappa r_{cs} = \frac{1}{c_1}\left[1 - \left(\frac{C_\kappa^{(0)}}{C_\kappa}\right)^2\right], \quad \frac{1}{a_{cs}} = \kappa\nu f(\nu) - \frac{1}{2}\kappa^2 r_{cs}, \tag{8}$$

where $f(\nu)$ is the function defined in (A4). Here, however, due to the smallness of the coefficient c_1 (see Fig. 2), the calculation of the effective range r_{cs} by this formula is not very reliable, and we employ a different calculation method. The experimental values of the quantum defect[5)] for *s* states, listed in Ref. 11, were used to calculate the value of a_{cs} by the Seaton formula[17,18]

$$a_{cs} = -(2\pi Z)^{-1} \tan \pi\delta_0,$$

after which the range r_{cs} was calculated by the second equation in (8).

Note that for alkali and alkali-earth elements, the values of a_{cs} are negative and small compared to the Bohr radius a_B, which corresponds to weak attraction of the highly excited *s* electrons to the atomic core. Here, due to the smallness of the scattering length, the effective ranges r_{cs} may become relatively large in this connection see formula (B7) and the formulas that follow in Appendix B). On the other hand, for positive ions (except Mg^+) the a_{cs} are positive, which can be interpreted as effective repulsion of the outer electron by the core. In this case the fractional part of the quantum defect $\{\delta_0\}$ is larger than 0.5 (note that $\{\delta_0\} \approx 0.5$ corresponds to the transformation region for the atomic spectrum[12]). The results of calculations are listed in Table II.

Expansions similar to (3) and (5) can be obtained for the moments of the electron density distribution,

$$\langle r^\alpha \rangle = \int_0^\infty \chi_\kappa^2(r) r^\alpha \, dr.$$

In particular,

$$\langle r \rangle = \frac{1}{2\kappa}\{(3\nu + c_1)(1 + c_1\kappa r_{cs}) + O((\kappa r_0)^2)\}. \tag{9}$$

For atoms of alkali elements even the zero-range ($\langle r \rangle_0$) approximation, i.e., $r_{cs} = 0$ in (9), has an accuracy of 10–20%. For instance, $\langle r \rangle / \langle r \rangle_0 = 1.021$, 1.058, 1.11, 1.15, and 1.20 for Li, Na, K, Rb, and Cs, respectively. Thus, the above

TABLE II. Parameters of neutral atoms and positive ions.

Atom or ion	ν	a_{cs}	r_{cs}	$C_\kappa^{(0)}$	C_κ	$c_1\times 100$
Li	1.589	0.490	0.59	1.063	1.076±0.03	1.87
Na	1.627	−0.308	2.71	1.060	1.040±0.03	1.64
K	1.770	−0.101	18.4	1.044	0.95±0.1	0.71
Rb	1.805	−0.070	40.5	1.038	0.94±0.09	0.52
Sr	1.546	−0.179	24.5	1.065	1.05±0.1	2.08
Ca	1.492	−0.290	13.0	1.066	1.06±0.1	2.29
Mg^+	1.903	−0.018	25.9	1.021	1.025±0.01	0.13
Ca^+	2.141	0.056	14.2	0.965	0.970±0.03	0.20
Sr^+	2.221	0.103	9.60	0.942	0.924±0.03	0.14
Ba^+	2.332	0.298	8.09	0.907	1.19±0.36	0.70

Note: As in Table I, the values of scattering lengths and effective ranges are measured in atomic units, and C_κ and c_1 are dimensionless factors.

expansions make it possible to easily estimate the values of C_κ and $\langle r^\alpha\rangle$ for multi-electron atoms and ions.

Simple analytical formulas similar to (9) have been obtained for the rms radius $\langle r^2\rangle^{1/2}$ and the higher moments of the *s*-electron density distribution. Allowance for corrections of order κr_{cs} to the effective-range values determined earlier (see Table II) is important and usually leads to good agreement with the results of numerical calculations. We limit ourselves to two examples: for the ground state of the Li atom we have $\langle r\rangle=3.82$ (3.87) and $\langle r^2\rangle^{1/2}=4.16$ (4.21), and for Na we have $\langle r\rangle=4.09$ (4.21) and $\langle r^2\rangle^{1/2}=4.40$ (4.55); the radii $\langle r\rangle$ and $\langle r^2\rangle^{1/2}$ are measured here in atomic units, and the figures in parentheses are the values calculated by the Hartree–Fock method.[11]

3. Comparison with exact solutions. Let us examine some potentials for which the Schrödinger equation with $l=0$ can be solved analytically. This will make it possible to establish the limits of applicability of the expansions (3) and (5).

(a) For a rectangular well (of range $R\equiv r_0$ and depth U_0, with $g=2U_0R^2$ the dimensionless coupling constant) we have[3,10]

$$C_\kappa=\left[\frac{1-(\kappa R)^2/g}{2(1+\kappa R)}\right]^{1/2}e^{\kappa R}, \tag{10}$$

$$a_s=R\xi,\quad r_s=\left(1-\frac{1}{g\xi}-\frac{1}{3\xi^2}\right)R,\quad \xi=1-\frac{\tan\sqrt{g}}{\sqrt{g}}.$$

Expanding these expressions for $\kappa\to 0$ and $g\to\tilde g_n=(n-1/2)^2\pi^2$, we obtain

$$\frac{1}{2C_\kappa^2}=1-\kappa R+\frac{(\kappa R)^2}{\tilde g_n}+\left(\frac{2}{3}-\frac{\tilde g_n+2}{\tilde g_n^2}\right)(\kappa R)^3+\cdots$$

$$=1-\kappa r_s+\left(\frac{1}{3}-\frac{1}{2\tilde g_n}\right)(\kappa r_s)^3+\cdots. \tag{11}$$

In accordance with (5), the expansion in powers of κr_s does not have a term $\propto r_s^2$ (in contrast to the expansion in powers of κR). For the ground level, $\tilde g_1=\pi^2/4$ and $c_3=1/3-2\pi^{-2}=0.1307$.

(b) The Hulthén potential

$$U(r)=-\frac{g}{2R^2[\exp(r/R)-1]} \tag{12}$$

resembles the Yukawa potential and is often used in atomic and nuclear physics. The bound state ns appears at $g=\tilde g_n=n^2$ (see Ref. 9):

$$\kappa R=\frac{g-\tilde g_n}{2n},$$

$$C_\kappa=\sqrt{\frac{n+\kappa R}{2(n+2\kappa R)}}\,\frac{\Gamma(n+1+2\kappa R)}{\Gamma(n+1)\Gamma(1+2\kappa R)}. \tag{13}$$

The *s*-scattering length and the effective range are[6)]

$$a_s=R\,a(g)=2R\sum_{n=1}^{\infty}n\left(\frac{1}{g-g_n}+\frac{1}{g_n}\right),$$

$$r_s=\frac{2}{3}R\left\{a-2a^{-2}\left[\frac{d^2(ag)}{dg^2}+4\zeta(3)\right]\right\}, \tag{14}$$

where $a(g)=\psi(1+\sqrt{g})+\psi(1-\sqrt{g})+2C$, with $\psi(z)=\Gamma'(z)/\Gamma(z)$, and $\zeta(3)=1.202$ (Ref. 16). It only remains to expand these expressions for $\kappa\to 0$, but for an arbitrary n the formulas become very complicated. We limit ourselves to the ground state, where

$$g=1+2\kappa R,\quad \tilde g_1=1,\quad C_\kappa=\sqrt{\frac{1}{2}(1+\kappa R)(1+2\kappa R)},$$

and (see Appendix B)

$$c_3=\frac{11-16\zeta(3)}{54}=-0.1525. \tag{15}$$

(c) The simplest model of $U(r)$ is the Breit potential, which is used to describe nucleon–nucleon interactions at low energies and in the theory of resonance nuclear reactions.[19,20] The potential is specified by the boundary conditions

$$\chi(r)\equiv 0,\quad 0<r<R;\quad \left.\frac{r\chi'(r)}{\chi(r)}\right|_{r=R}=-g. \tag{16}$$

In this case there is only one *s* level, for which $\kappa R=g>0$, and

$$C_\kappa=\frac{1}{\sqrt{2}}e^g, \quad a_s=(1+g^{-1})R,$$

$$\kappa r_s=\frac{2g(1+g+g^2/3)}{(1+g)^2}. \tag{17}$$

This implies that

$$g=\sum_{k=1}^{\infty} d_k(\kappa r_s)^k, \quad d_k=\frac{3^{k-1}\Gamma((2k+1)/3)}{2^k k!\,\Gamma((4-k)/3)} \tag{17a}$$

and, finally,

$$\frac{1}{2C_\kappa^2}=e^{-2g}=1-\kappa r_s+\sum_{k=3}^{\infty} c_k(\kappa r_s)^k, \tag{18}$$

where $c_3=1/6$, $c_4=1/12$, $c_5=-1/120$, $c_6=-83/1440$, ... The rapid decrease of these coefficients explains why the approximation $C_\kappa\approx[2(1-\kappa r_s)]^{-1/2}$ maintains good accuracy up to values $\kappa r_s\sim 0.5$ (Fig. 3).

(d) We also studied the delta-function interaction

$$U(r)=-\frac{g}{2R}\,\delta(r-R)$$

and the separable Yamaguchi potential.[21] Here are the results of calculations for the Hulthén potential, the Yamaguchi potential, the rectangular well, the delta-function potential, and the Breit potential, respectively:

$$c_3=-0.152,\ 2/27,\ 0.131,\ 3/20, \text{ and } 1/16 \tag{19}$$

(the values of c_3 for the Hulthén potential and the rectangular well refer to the ground state). Although c_3 is model-dependent, we may assume that its numerical smallness is retained for realistic potentials $U(r)$.

Figure 3 depicts C_κ^{-2} as a function of κr_s. The straight dashed line corresponds to a situation where all the expansions terms in (5) beginning with $(\kappa r_s)^3$ are dropped. This simple approximation proves to be remarkably accurate even outside the $\kappa r_s\ll 1$ range, so that in many cases we can limit ourselves to the first-order correction in κr_0.

(e) The formulas become much more complicated when there is Coulomb interaction. We limit ourselves to the Breit model (16), for which

$$\chi_\kappa(r)=2^{1-\nu}\kappa^{1/2}C_\kappa W_{\nu,1/2}(2\kappa r)\,\theta(r-R), \tag{20}$$

$$\frac{1}{2C_\kappa^2}=2^{-\nu}\int_{2\kappa R}^{\infty} W_{\nu,1/2}^2(x)\,dx=\frac{1}{2}[C_\kappa^{(0)}]^2 \times\left\{1-c_1\int_0^{2\kappa R} w^2(x)\,dx\right\}, \tag{20'}$$

where $w=\Gamma(1-\nu)W_{\nu,1/2}(x)$.

With the Breit model it is convenient to examine the special case $\nu=n$, so that the energy of the ns level is the same as in a purely Coulomb field. For $n=1$ we have

$$C_1=[e^{-2\rho}(1+2\rho+2\rho^2)]^{-1/2}=1-\frac{2}{3}\rho^3-\rho^4+\cdots, \tag{21}$$

$$g_1=-1+\rho,$$

where $\rho=R/a_B$, and $a_B=1/Z$ is the Bohr radius. For the general case $\nu=n$ we have, for $\rho\ll 1$,

$$C_n=C_\kappa^{(0)}\left(1+\frac{2\rho^3}{3n^3}-\frac{\rho^4}{n^3}+\cdots\right), \quad C_\kappa^{(0)}=\frac{2^{n-1}}{n!}, \tag{21'}$$

where the coupling constant in (16) corresponding to $\nu=n$ is

$$g_n=-1+\rho+\frac{n^2-1}{3n^2}\rho^2+\cdots. \tag{21''}$$

Thus, the corrections to the value $C_\kappa^{(0)}$ from the distortions of the Coulomb potential in the $r<R$ range begin with terms $\propto(\kappa R)^3$. But if the potential $V(r)$ is finite for $0<r<R$, the correction is even smaller, of order $(\kappa R)^4$.

4. Now we study the WKB approximation for the asymptotic coefficients at infinity:

$$C_\kappa^{\mathrm{WKB}}=\xi\left(\frac{2\kappa}{\mu}\right)\frac{1}{2\kappa\sqrt{T_r}}(\kappa r_t)^{-\nu} \times\exp\left\{\kappa r_0-\int_{r_t}^{\infty}\left(|p(r)|-\kappa+\frac{\nu}{r}\right)dr\right\}, \tag{22}$$

$$T_r=2\int_0^{r_t}\frac{dr}{p(r)}$$

$(l=0)$. Here the energy of the ns level is determined by the (modified) Bohr–Sommerfeld quantization rule:

$$\int_0^{r_t} p(r)\,dr=(n-\gamma)\pi,$$

$$p^2=-\kappa^2-2V(r), \quad n=n_r+1=1,2,3,\ldots, \tag{23}$$

in which the phase $\gamma\pi$ depends on the behavior of the potential at short distances. If $V(r)\propto r^\alpha$ for $r\to\infty$, then for attractive potentials we have[22]

$$\gamma=\begin{cases}\dfrac{1}{4}, & \alpha\geqslant 0,\\[2mm] (\alpha+1)/2(\alpha+2), & -2<\alpha\leqslant 0.\end{cases} \tag{24}$$

In particular, $\gamma=0$ for potentials with the Coulomb singularity at zero (Yukawa, Hulthén, and other potentials), and $\gamma=1/4$ for potentials that are finite at zero.

In Eqs. (22) and (23), $p(r)$ is the quasiclassical momentum, r_t is the turning point, T_r is the period of radial oscillations of a classical particle, and $\xi(x)\equiv\xi(x,0)$, where

$$\xi(x,a)=\sqrt{2\pi}\frac{x^{x+a-1/2}e^{-x}}{\Gamma(x+a)} =1-\frac{a_1}{2x}+\frac{a_2}{24x^2}+\cdots, \quad x\to\infty, \tag{25}$$

with $a_1=b^2-1/12$, $a_2=3a_1^2+4b^3-b$, and $b=a-1/2$.

The fact that (22) contains ξ as a factor reflects the need to modify the Kramers matching conditions near the turning point in the case of shallow levels. According to Refs. 23 and 24, this is because the quasiclassical region does not overlap

the region near point r_t where the linear expansion of the potential is valid (which is used to derive ordinary matching conditions[1]). As a result, the quasiclassical wave function in the subbarrier region $r>r_t$ should be multiplied by $\xi(2\lambda/\mu)$ if the potential $V(r)$ decreases as $\exp(-\mu r)$ at large distances (for potentials with a Coulomb ''tail'' at infinity, $\mu=0$ and $\xi(\infty)=1$).

We illustrate the expression (22) by the Hulthén potential, which allows an exact solution for states with $l=0$. In this case the quantization rule (23) yields[22] (as it does for the Coulomb potential[25]) the exact spectrum of s levels [see the first formula in (13)]. From (22) we obtain

$$C_{ns}^{\mathrm{WKB}}=\xi(n(\lambda-1))\sqrt{\frac{\lambda+1}{8\pi n(\lambda-1)}}\,a^n,$$

$$a=\frac{\lambda^\lambda}{(\lambda-1)^{\lambda-1}},\quad \lambda=\frac{g}{\tilde{g}_n}, \tag{26}$$

$$\rho_n\equiv\frac{C_{ns}^{\mathrm{WKB}}}{C_{ns}}=\frac{\xi(n\lambda)}{\xi(n)}=\frac{\xi(n+2\kappa R)}{\xi(n)}. \tag{27}$$

The approximation (22) has a high accuracy for $n\gg1$ and for an arbitrary level energy

$$\rho_n=1+\frac{\kappa R}{6n(n+2\kappa R)}+\frac{1}{72n^2}\left(\frac{\kappa R}{n+2\kappa R}\right)^2+\cdots;$$

for shallow levels this approximation has a high accuracy for all $n=1,2,\ldots$ [owing to the factor $\xi(2\kappa R)$ in (22)]:

$$\rho_n=1+d_1\kappa R+d_2(\kappa R)^2+\cdots,\quad \kappa R\ll1,$$

$$d_1=2[\ln n-\psi(n)]-\frac{1}{n},$$

$$d_2=\frac{1}{2}d_1^2+\frac{2n+1}{n^2}-2\psi'(n) \tag{28}$$

($d_1=0.154$ and $d_2=-0.278$ at $n=1$, $d_1=0.041$ and $d_2=-0.039$ at $n=2$, and $d_1\approx1/6n^2$ and $d_2\approx-1/3n^3$ for $n\gg1$). Even in the least favorable case of deep levels we have $\rho_n\to1/\xi(n)$, which amounts to 1.042, 1.028, and $1+0.083n^{-1}$ for $n=2$, $n=3$, and $n\to\infty$, respectively.

Similar results were obtained for other potentials. For instance, for $V(r)=-g/2R^2\cosh^2(r/R)$ we have $\kappa R=\sqrt{g+1/4}-\sqrt{g_n+1/4}$, $\tilde{g}_n=2n(2n-1)$, and

$$\rho_n=\sqrt{\frac{\xi(2(n+\kappa R),1/2)}{\xi(2n,1/2)}}, \tag{29}$$

which yields $\rho_n=1+\kappa R/48n(n+2\kappa R)\to1$ for $n\gg1$, and $\rho_n=1$ at the moment when the ns level appears, while in the case of deep levels ($\kappa\to\infty$) we have $\rho_n\to[\xi(2n,1/2)]^{-1/2}=0.9898$, 0.9948, and $1-0.0104n^{-1}$ for $n=1$, 2, and $n\gg1$, respectively. These examples show that the quasiclassical approximation modified in accordance with Refs. 22–24 is not only asymptotically exact in the limit $n\to\infty$ but remains so even for moderate quantum numbers $n\sim1$.

Finally, for $V(r)=-Z/r$ we have $\xi=1$, the turning point $r_t=2\nu^2/Z$, and

$$C_\kappa^{\mathrm{WKB}}=(8\pi\nu)^{-1/2}\left(\frac{2e}{\nu}\right)^\nu \tag{30}$$

(see the paper by Ammosov *et al.*,[6] in which $n^*\equiv\nu$ is the effective principal quantum number and the asymptotic coefficient C_{n^*0} differs from our coefficient by a factor of two). For $\nu\gtrsim1$ this expression is in good agreement with $C_\kappa^{(0)}$, as Fig. 1 shows. However, as $\nu\to0$, formula (30) does not become expansion (5) [in contrast to (3)], which expansion is valid for loosely bound states in a short-range potential. An advantage of (3) is also that (3) allows for the correction from the effective range of the system, which, in particular, makes it possible to find the parameters r_s and r_{cs} (see Tables I and II).

5. The case $l\neq0$. For arbitrary l the asymptotic behavior of the normalized wave function can be expressed by the formula

$$\psi_{\kappa lm}(\mathbf{r})\approx2^{l+3/2}C_{\kappa l}\sqrt{\frac{\kappa^3}{(2l+2)!}}\,e^{-\kappa r}(\kappa r)^{\nu-1}Y_{lm}\left(\frac{\mathbf{r}}{r}\right). \tag{31}$$

For the lowest state with angular momentum l in the Coulomb problem, $V(r)=-Z/r$, we have[1] $C_{\kappa l}=1$ for any $l=0,1,\ldots$, which explains the choice of the l-dependent factors in (31).

For $l\gtrsim1$, the solution of the Schrödinger equation that decreases at infinity is $W_{\nu,l+1/2}(2\kappa r)\propto r^{-l}$ when $r\to0$, so that the zero-range approximation is not valid. A fairly general result can be obtained for the case where there is a shallow l level in the ''inner'' potential $U(r)$. When the l level appears, its wave function decreases at infinity (due to the centrifugal barrier, $l\neq0$) and remains normalized: $\chi_l(r)\approx A_l r^{-l}r\gg r_0$. Applying the matching procedure in the region $r_0\ll r\ll\min(\kappa^{-1},a_B)$ and allowing for the coupling $1/A_l^2$ with the effective range of the system,[26] we get

$$C_{\kappa l}^2=(2l+2)!\left[\frac{\Gamma(l+1-\nu)}{2^{l+1-\nu}\Gamma(l+1)}\right]^2\left|\frac{\kappa^{2l-1}}{\tilde{r}_l}\right|,\quad l\gtrsim1, \tag{32}$$

where $\tilde{r}_l$ is the value of the effective range at the moment when the l level appears. Note that $\kappa^{2l-1}/\tilde{r}_l$ is a dimensionless parameter, with[7] $\tilde{r}_l<0$. For instance for a rectangular well of range R,

$$\tilde{r}_l=-\beta_l R^{1-2l} \tag{33}$$

($\beta_1=3$, $\beta_2=15$, $\beta_3=315$, ...), and similarly for other short-range potentials.

There is an important difference between (3) and (32): while in (3) the first term is universal and is determined entirely by the Coulomb interaction (and the energy of the level), (32) incorporates the effective range $\tilde{r}_l$, which depends on the short-range part of the potential $U(r)$. For states with $l\neq0$ there is no analog of expansion (3) that is insensitive to distortions in the Coulomb potential at distances $r\sim r_0\ll a_B$. Note that the presence of a shallow l level (not a Coulomb level, i.e., $\nu\neq n$) already points to the transformation of the atomic spectrum (the Zel'dovich effect) for states with $l\neq0$ (Refs. 28 and 29). Formula (32) belongs to this case.

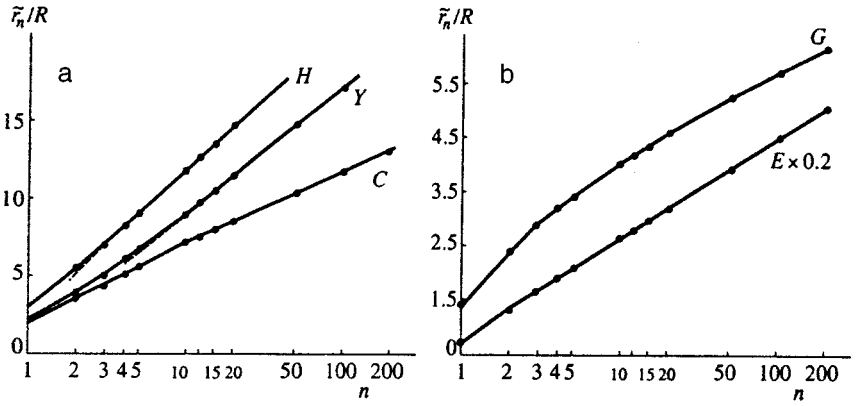

FIG. 4. Effective ranges $\tilde{r}_n$ at the time the ns level appears, $g=\tilde{g}_n$. Curves H, Y, C, E, and G refer, respectively, to $v(x)=(e^x-1)^{-1}$, e^{-x}/x, $1/\cosh^2 x$ [the potential is written in the form (C1)], $\exp(-x)$, and $\exp(-x^2)$. In the case E the numerical values of the ranges are reduced by a factor offive. The dots indicate the results of numerical calculations.

6. The expansions (3) and (5) relate the asymptotic coefficient at infinity to the effective range of the system. If the explicit expression for C_κ is known, its expansion for $\kappa \to 0$ yields the value of range, $\tilde{r}_n \equiv r_s(g=g_n)$ at the time when the ns level appears (and also the first correction $\propto \kappa$). Using (13), for the Hulthén potential we obtain

$$r_s(g)=\tilde{r}_n-\beta_n'\kappa\tilde{r}_n^2+\cdots, \quad g\to\tilde{g}_n, \tag{34}$$

where

$$\tilde{r}_n=\beta_n R, \quad \beta_n=4[\psi(n+1)-\psi(1)]-n^{-1}. \tag{35}$$

These expressions can also be derived by expanding the exact formula (14) near $g=\tilde{g}_n$. For the ground level we have $\beta_1=3$ and $\beta_1'=7/9$, for $n=2$ we have $\beta_2=5.5$ and $\beta_2'=79/121$, and for highly excited states we have

$$\tilde{r}_n=4R\left(\ln n+C+\frac{1}{4n}-\frac{1}{12n^2}+\cdots\right),$$

$$\beta_n'=\frac{1}{2}+\frac{\pi^2}{24(\ln n+C)^2}+\ldots. \tag{36}$$

Similarly, for $U(r)=-g/2\cosh^2(r/R)$ we have

$$\tilde{r}_n=2R[\psi(2n)+C]$$

$$=2R\left[\ln n+(C+\ln 2)-\frac{1}{4n}-\frac{1}{48n^2}+\cdots\right]. \tag{37}$$

We will show that the logarithmic increase in the effective ranges $\tilde{r}_n$ for highly excited ns states is a general property of short-range (but not finite) potentials (see also Fig. 4). Suppose that

$$U(r)\approx-\frac{1}{2}g\mu^2(\mu r)^{-2\beta}\exp\{-(\mu r)^\alpha\}, \quad r\gg r_0. \tag{38}$$

Then at the moment when the ns level appears, the wave function of this level, as $r\to\infty$, is

$$\chi_0(r)=1-g_n\alpha^{-2}(\mu r)^{-2(\alpha+\beta-1)}$$

$$\times\exp\{-(\mu r)^\alpha\}+\cdots, \quad r\to\infty \tag{38a}$$

[the parameter μ^{-1} refers to the asymptotic behavior of the potential $U(r)$ at large distances and does not necessarily coincide with the characteristic range r_0 introduced in (2)]. The difference $1-\chi_0^2(r)$ becomes exponentially small for $r>r_*$, where

$$(\mu r_*)^\alpha=\ln g_n-2(\alpha+\beta-1)\ln\ln g_n+\cdots.$$

For $n\gg 1$ the quasiclassical approximation can be used. Hence $g_n\propto n^2$ and, for $r_0\ll r<r_*$,

$$|\chi_0(r)|\lesssim[-U(r)]^{-1/4}\sim g_n^{-1/4}(\mu r)^{\beta/2}\exp\left\{\frac{1}{4}(\mu r)^\alpha\right\}$$

$$\propto\exp\left\{-\frac{\ln n}{2}\left[1-\left(\frac{r}{r_*}\right)\right]^\alpha\right\}\ll 1. \tag{39}$$

The integral in the Schwinger–Smorodinskiĭ formula (A9) is concentrated at distances $r\lesssim r_*$ (here the transient region $|r-r_*|\sim r_*/\ln n$ is narrow for $n\to\infty$), which means that[8)]

$$\tilde{r}_n=2^{(\alpha+1)/\alpha}\mu^{-1}\left\{(\ln n)^{1/\alpha}-\frac{\alpha+\beta-1}{\alpha}\right.$$

$$\left.\times(\ln n)^{(1-\alpha)/\alpha}\ln\ln n+\cdots\right\}. \tag{40}$$

In particular, for potentials with a purely exponential ($\alpha=1$ and $\beta=0$) ''tail'' at infinity, this asymptotic behavior agrees with the exact formulas (35) and (37), where the expansion is in reciprocal powers of N:

$$\tilde{r}_n=4\mu^{-1}\left(\ln N+k_0+\frac{k_1}{N}+\frac{k_2}{N^2}+\cdots\right), \tag{41}$$

where $N=n-\gamma$, with the constant γ, which dot depend on n, defined in (24). For instance, for a potentials with a Coulomb singularity at zero (e.g., the Hulthén potential), $\gamma=0$ and $N=n$, while for potentials that are finite at zero, $N=n-1/4$. Closed analytical expressions can be obtained for two coefficients, k_0 and k_1 (see Appendix C).

On the other hand, the Yukawa potential has $\alpha=1$ and $\beta=1/2$, so that the asymptotic formula (40) contains a double logarithm: see curve Y in Fig. 4a, which shows that for the Yukawa potential the asymptotic behavior specified by (40) sets in much later than for potentials with a purely exponential ''tail'' (curves H and C). Figure 4b shows that the increase in the effective ranges $\tilde{r}_n$ for the Gaussian potential ($\alpha=2$) is slower than at $\alpha=1$ and corresponds to a

$\tilde{r}_n \propto \sqrt{\ln n}$ dependence (the results depicted in Fig. 4 by points were obtained by solving the Schrödinger equation numerically).

Thus, the effective range of the ns state (at $g \approx g_n$) ''expands'' logarithmically with increasing n. We note a certain similarity between this result and the well-known Froissart constraint in high-energy physics.[30,31]

7. We concludw with some remarks. The expansion (5) is valid only in the resonance case: $\kappa r_0 \ll 1$ and $|a_s| \gg r_s$, i.e., in the presence of a shallow level (real or virtual) in a short-range potential $U(r)$. As for (3), the above condition is not necessary for this expansion to be true, since the particles are coupled not by $U(r)$ but by the Coulomb attraction ($Z > 0$). As (3) shows, the correction to $C_\kappa^{(0)}$ contains the factor c_1. For $\nu \gg 1$ we have $c_1 \propto \nu^2 \to 0$, which ensures that the zeroth approximation $C_\kappa \approx C_\kappa^{(0)}$ is accurate in the case of highly excited states.

The coefficient c_1 remains small in the region $\nu \sim 1$–2, to which the larger fraction of the s states of neutral atoms and positive ions belongs (see Fig. 1). However, $c_1(\nu) \to 0$ as $\nu \to n$, and the condition for the applicability of (3) changes, since the approximation (2) breaks down. Formally this manifests itself in the vanishing of $\varphi(r)/\Gamma(1-\nu)$ for $r \lesssim r_0$, but actually the reason is that the coefficients in the effective range expansion become anomalously large as $a_{cs} \to 0$: $r_{cs} \propto r_0^3/a_{cs}^2$, etc. [see (B7)]. This case must be examined separately. As implied by (21), here the correction to $C_\kappa^{(0)}$ is small but does not vanish at $\nu = n$. We have written the final expression for the expansion (3) in a form that is valid for all $\nu \gg r_0$, including the values of the parameter ν close to the integral values $n = 1,2,3, \ldots$.

Finally, we note that an expression of the form (3) for the asymptotic coefficient of the wave function also arises in the two-channel problem with Coulomb interaction in one channel and short-range interaction in the other. These problems were studied in Refs. 32 and 33 by employing the example of a proton–antiproton atom in which the $p\bar{p}$ and $n\bar{n}$ channels are coupled because of strong interaction at small ($r \sim r_0 \ll a_B$) distances. Here the problem effectively reduces to a one-channel problem, with formulas (3) and (4) remaining valid if the parameter r_{cs} in (3) is replaced with[9)]

$$r_{cs} \to (1+\rho^2) r_e - \rho^2 (2\Delta)^{-1/2},$$

$$\rho^2 = \left\{ \left(\frac{1}{a_0} - \frac{1}{a_1} \right) \Big/ \left(\frac{1}{a_0} + \frac{1}{a_1} - 2\Delta \right) \right\}^2, \tag{42}$$

where a_0 and a_1 are the scattering lengths with isospins 0 and 1, $\Delta = \sqrt{2m_n(m_n - m_p)} \approx 1/4\ \mathrm{fm}^{-1}$, $r_e = r_{11} = r_{22}$ are the effective-range matrix elements, and ρ^2 is the relative probability of finding the system in the $n\bar{n}$ state within the range of nuclear forces [what is important is that $c_1(\nu)$ retains its previous value (4)]. However, the question of whether the coefficient of the next term ($\propto \kappa^2$) in the expansion vanishes was not discussed in those papers.

The authors are grateful to I. L. Beĭgman, N. B. Delone, B. O. Kerbikov, V. P. Kraĭnov, L. B. Okun', and the reviewer for discussions and useful remarks, and to S. G. Pozdnyakov, A. V. Sergeev, and M. N. Markina for the help provided in doing the numerical calculations and in preparing the manuscript for publication. This work was made possible by a grant from the Russian Fund for Fundamental Research (Grant No. 98-02-17007).

APPENDIX A:

To derive effective range expansions we note that the wave function of the shallow s level in the Coulomb field $V(r) = -Zr^{-1} + U(r)$ distorted at small distances ($r \lesssim r_0 \ll \kappa^{-1}$) has the form (2), where

$$\Gamma(1-\nu) W_{\nu,1/2}(2\kappa r) \equiv w(r) = v(r,\kappa^2) - \frac{1}{2}\rho f(\nu) u(r,\kappa^2), \tag{A1}$$

$$u(r,\kappa^2) = \eta(\rho) + \frac{1}{6}\kappa^2 r^2 \tilde{\eta}(\rho) + O((\kappa r)^4), \tag{A2}$$

$$v(r,\kappa^2) = \xi(\rho) + \frac{1}{2}\kappa^2 r^2 \tilde{\xi}(\rho) + O((\kappa r)^4), \tag{A3}$$

$$f(\nu) = 2\psi(1-\nu) - 2\ln\nu + \nu^{-1}, \quad \psi(z) \equiv \Gamma'(z)/\Gamma(z), \tag{A4}$$

$\nu = Z/\kappa = (\kappa a_B)^{-1}$, $\rho = 2r/a_B$, $a_B = 1/Z$ is the Bohr radius, and $W_{\nu,1/2}$ is the Whittaker function, and $\varphi(r)$ is the wave function of the system corresponding to zero energy; here

$$\varphi(r) \approx \xi(\rho) - \frac{r}{a_{cs}}\eta(\rho), \quad r \gg r_0. \tag{A5}$$

The functions ξ, η, etc., in (A2) and (A3) can be expressed in terms of cylinder functions, with $\xi(0) = \tilde{\xi}(0) = \eta(0) = \tilde{\eta}(0) = 1$. For instance,[10)] $\eta(\rho) = \rho^{-1/2} J_1(2\sqrt{\rho})$ and $\xi(\rho) = -\pi\rho^{1/2} N_1(2\sqrt{\rho})$, so that, as $\rho \to 0$,

$$\eta(\rho) = 1 - \frac{\rho}{2} + \frac{\rho^2}{12} + \cdots,$$

$$\xi(\rho) = 1 - \rho\ln\rho - (2C-1)\rho + \frac{1}{2}\rho^2\ln\rho + \cdots,$$

$$\tilde{\eta}(\rho) = 1 - \frac{1}{12}\rho + \frac{1}{24}\rho^2 + \cdots,$$

$$\tilde{\xi}(\rho) = 1 - \frac{1}{3}\rho\ln\rho + \frac{2}{9}(2-3C)\rho + \frac{1}{9}\rho^2\ln\rho + \cdots \tag{A6}$$

($C = 0.577\ldots$) and, as $\rho \to \infty$,

$$\eta(\rho) \approx \pi^{-1/2}\rho^{-3/4}\sin\left(2\sqrt{\rho} - \frac{\pi}{4}\right),$$

$$\xi(\rho) = \pi^{1/2}\rho^{1/4}\cos\left(2\sqrt{\rho} - \frac{\pi}{4}\right),$$

etc. The function $f(\eta)$ in (A4) enters into the equation that determines the spectrum of shallow atomic levels in the Coulomb problem with short-range interaction:[12,35]

$$Zf\left(\frac{Z}{\kappa}\right) = \frac{1}{a_{cs}} + \frac{1}{2}\kappa^2 r_{cs} + O(\kappa^4), \tag{A7}$$

where a_{cs} and r_{cs} are the nuclear Coulomb scattering length and effective range;[12–15] in particular,

$$r_{cs}=2\int_0^\infty\left\{\left[\xi(2Zr)-\frac{r}{a_{cs}}\,\eta(2Zr)\right]^2-\varphi^2(r)\right\}dr \qquad \text{(A8)}$$

(in view of (A5) this integral converges at distances $r\sim r_0$). When the ns level appears, $a_{cs}=\infty$ (if $Z\neq 0$) or $a_s=\infty$ (if $Z=0$), and (A8) becomes

$$\tilde{r}_{cs}=2\int_0^\infty[\xi^2(2Zr)-\varphi_0^2(r)]\,dr,$$

$$\tilde{r}_s=2\int_0^\infty[1-\chi_0^2(r)]\,dr. \qquad \text{(A9)}$$

This formula was obtained by Schwinger[36] and Smorodinskiĭ;[37] see also p. 553 in Ref. 1. A tilde indicates quantities that refer to the time at which the bound s state appears.

By comparing (A1) and (A5) and allowing for Eq. (A7) we find that the functions $w(r)$ and $\varphi(r)$ are matched in the region $r_0\ll r\ll\min(\kappa^{-1},\ a_B)$. The normalization condition $\int_0^\infty\chi_\kappa^2(r)\,dr=1$ yields

$$\frac{2^{2\nu-2}\Gamma^2(1-\nu)}{\kappa C_\kappa^2}=\int_0^{r_1}\varphi^2(r)\,dr+\int_{r_1}^\infty w^2(r)\,dr$$
$$=\int_0^\infty w^2(r)\,dr-2r_{cs}$$
$$-\int_0^{r_1}\left[w^2(r)-\left(\xi-\frac{r}{a_{cs}}\,\eta\right)^2\right]dr, \qquad \text{(A10)}$$

where we have used (A8) and taken the matching point r_1 from the specified region. Equations (A1), (A6), and (A7) for $r<r_1$ yield

$$w(r)=\left(\xi-\frac{r}{a_{cs}}\right)\eta+\frac{1}{2}\,\kappa^2\left(r^2\tilde{\xi}-\frac{r^3}{3a_{cs}}\,\tilde{\eta}-r_{cs}r\,\eta\right)+O(\kappa^4),$$

so that the last integral in (A10) is of order $\kappa^2 r_0^3$ and contributes nothing to the terms $\propto(\kappa r_0)^2$ in the expansion of C_κ^{-2}. Since

$$\int_0^\infty[W_{\nu,1/2}(x)]^2\,dx=2\Gamma^2(\nu+1)\left\{1-\left(\frac{\sin\pi\nu}{\pi}\right)^2\right.$$
$$\left.\times\left[\psi'(\nu)-\frac{2\nu+1}{2\nu^2}\right]\right\}, \qquad \text{(A11)}$$

for $\kappa r_0\ll 1$ we arrive at (3). Note that the first two terms in this expansion have already been obtained in the theory of Coulomb systems with short-range interaction (hadronic atoms[12,32]). There a cruder method was used, however, and it was not understood that the terms $\propto r^2\ln r$, r^2, and $r^3\ln r$ (which actually exist[16] in the expansion of Coulomb radial wave functions as $r\to 0$) always cancel out without a remainder, in view of which the terms not taken into account in (3) are of order $(\kappa r_0)^3$.

APPENDIX B:

In the case of an s wave,

$$k\cot\delta_0(k)=-ik\frac{f(k)+f^*(k)}{f(k)-f^*(k)}=-\frac{1}{a_s}+\frac{1}{2}\,r_s k^2+\cdots, \qquad \text{(B1)}$$

where $\sigma_0(k)$ is the phase of s scattering, and $f(k)=|f|\exp(-i\delta_0)$ is the Jost function. The Schrödinger equation with $l=0$ and potential (12) has exact solutions, which can be expressed in terms of the Gauss hypergeometric function. By normalizing the discrete-spectrum wave function we arrive at (13), and for $f(k)$ we have[9]

$$f(k)=\prod_{n=1}^\infty\left(1-\frac{g}{n(n-2ikR)}\right)=\frac{\Gamma(1-2ikR)}{\Gamma(1-ikR+\sqrt{g-(kR)^2})}$$
$$\times\frac{1}{\Gamma(1-ikR-\sqrt{g-(kR)^2})}. \qquad \text{(B2)}$$

Expanding this expression with $k\to 0$, we arrive at formulas (14) for the scattering length and the effective range in the Hulthén potential. For the function $a(g)$ introduced in (14) we find that

$$a(g)=\pi\cot\pi z+2[\psi(1+z)+C]+(1+z)^{-1}$$
$$=z^{-1}+1-\sum_{m=1}^\infty a_m z^m, \qquad \text{(B3)}$$

where $z=\sqrt{g}-1$, $a_m=1$ at $m=1,3,5,\ldots$, $a_m=2\zeta(m+1)-1$ at $m=2,4,\ldots$, and $\zeta(s)$ is the Riemann zeta function.[16] We transform (14) into

$$r_s=\frac{1}{3}\left\{b''+\frac{2}{b}(1-b'^2)+\frac{3b'}{1+z}-16\zeta(3)b^2\right\}R, \qquad \text{(B4)}$$

where $b=1/a=z-z^2+2z^3-2cz^4+4cz^5+\cdots$, $b'=db/dz$, and $c=2-\zeta(3)$. As a result we have

$$\frac{r_s}{R}=3-7\kappa R+\frac{1}{2}[41-16\zeta(3)](\kappa R)^2+\cdots, \qquad \text{(B5)}$$

and for

$$\frac{1}{2C_\kappa^2}=1-\sum_{n=1}^\infty(-1)^n(2^{n+1}-1)(\kappa R)^n$$

we have the expansion (5) with the coefficient c_3 given by (15). Here, as expected, the term quadratic in the effective range is not present in (5).

For short-range potentials we have the following expansions: at the moment when the ns level appears ($g\to\tilde{g}_n$, $n=1,2,\ldots$) we have

$$\frac{a_s}{R}=\sum_{j=-1}^\infty\tilde{\alpha}_j(g-\tilde{g}_n)^j,\quad \frac{r_s}{R}=\sum_{j=0}^\infty\tilde{\beta}_j(g-\tilde{g}_n)^j, \qquad \text{(B6)}$$

and near the zeros of the scattering length, $g\to g_n$, we have

$$\frac{a_s}{R}=-\Sigma_{j=1}^\infty\alpha_j(g-g_n)^j,\quad \frac{r_s}{R}=\sum_{j=-2}^\infty\beta_j(g-g_n)^j, \qquad \text{(B7)}$$

with $0<g_1<\tilde{g}_1<g_2<\tilde{g}_2<\cdots$. For instance, for a rectangular well $\sqrt{\tilde{g}_n}=(n-1/2)\pi$,

$$\tilde{\alpha}_{-1}=2, \quad \tilde{\alpha}_0=1-\frac{1}{2\tilde{g}_n},$$

$$\tilde{\beta}_0=1, \quad \tilde{\beta}_1=-\frac{1}{2\tilde{g}_n}, \ldots, \tag{B8}$$

and the values of g_n can be found from the equation $\tan\sqrt{g}=\sqrt{g}$, from which it follows that

$$\sqrt{g_n}=\left(n+\frac{1}{2}\right)\pi-\frac{1}{n\pi}+\cdots \quad (n\gg 1),$$

$$\alpha_1=\frac{1}{2}, \quad \alpha_2=\frac{1}{4}\left(1-\frac{1}{2g_n}\right),$$

$$\beta_{-2}=-\frac{1}{3\alpha_1^2}=-\frac{4}{3}, \quad \beta_{-1}=\frac{4}{3}\left(1+\frac{1}{g_n}\right), \ldots. \tag{B9}$$

For the Hulthén potential,

$$\tilde{g}_n=n^2, \quad \tilde{\alpha}_1=2n, \quad \tilde{\alpha}_0=\frac{1}{2}\,\tilde{\beta}_0=2[\psi(n)+C]+\frac{3}{2n}. \tag{B10}$$

For highly excited states ($n\gg 1$) we obtain

$$\tilde{\alpha}_0=2(\ln n+C)+\frac{1}{n}+\cdots,$$

$$\tilde{\beta}_1=-\frac{(\ln n)^2}{n}\left(1+\frac{2C}{\ln n}+\cdots\right), \tag{B11}$$

$$\sqrt{g_n}=n+1-\frac{1}{2(\ln n+C)}+\cdots,$$

$$\alpha_1\approx 2\frac{(\ln n)^2}{n}, \quad \alpha_2\approx 2\frac{(\ln n)^3}{n^2}, \quad \beta_{-2}\approx\frac{2n^2}{\ln n}. \tag{B12}$$

When the coupling constant satisfies $g=g_n$, the scattering length vanishes, while the effective range $r_s(g)$ has second-order poles,[11] and it may become either $+\infty$ (say, for the Hulthén potential) or $-\infty$ (rectangular well). This explains the presence of the corresponding curves in Fig. 3.

Finally, for $g\to 0$ we have

$$\frac{a_s}{R}=-\sum_{j=1}^{\infty}\alpha_j' g^j, \quad \frac{r_s}{R}=\sum_{j=-1}^{\infty}\beta_j' g^j. \tag{B13}$$

For instance, for a rectangular well we have

$$\alpha_j'=\frac{2^{2j+3}-2}{\pi^{2j+2}}\zeta(2j+2), \quad \beta_{-1}'=\frac{6}{5}, \quad \ldots \tag{B14}$$

($\alpha_1'=1/3$, $\alpha_2'=2/15$, $\alpha_3'=17/315$, etc.), and for the Hulthén potential we have

$$\alpha_j'=2\zeta(2j+1), \quad \beta_{-1}'=4\zeta(5)[\zeta(3)]^{-2}=2.871 \tag{B15}$$

(here we have used $\psi^{(n)}=(-1)^{n+1}n!\,\zeta(n+1)$ for $n\geqslant 1$).

APPENDIX C:

We write the short-range attractive potential

$$U(r)=-\frac{g}{2R^2}v\left(\frac{r}{R}\right)\equiv -U_0 f(\mu r), \tag{C1}$$

where $R\equiv r_0$ is the range of the forces, and the parameter μ^{-1} refers to the asymptotic region $r\gg R$, in which $U(r)\to 0$. The function $v(r/R)$, or $f(\mu r)$, specifies the shape of the potential. We impose the condition

$$f(x)=e^{-x}+O(e^{-\lambda x}), \quad \lambda>3/2, \tag{C2}$$

as $x\to\infty$, which ensures the convergence of the integral J_1 in (C4). It is common practice to assume $R=\mu^{-1}$ and $v(x)\equiv f(x)$, but this is not always convenient (e.g., in the case of the Woods–Saxon potential).

Using the WKB method, we can derive formulas for the expansion coefficients in (41):

$$k_0=C-\ln J_0, \quad k_1=\frac{1}{2\pi}J_0(1-J_1), \tag{C3}$$

where $C=0.5772\ldots$ is Euler's constant, and

$$J_0=\frac{1}{\pi}\int_0^{\infty}\sqrt{f(x)}\,dx, \quad J_1=\frac{1}{2}\int_0^{\infty}([f(x)]^{-1/2}-e^{x/2})\,dx \tag{C4}$$

(see p. 147 in Ref. 38). We discuss several examples. For the Hulthén potential (12), $\mu^{-1}=R$ and

$$N=n, \quad k_0=C, \quad k_1=1/4, \tag{C5}$$

while for $v(x)=1/\cosh^2 x$ we get $\mu^{-1}=R/2$ and

$$N=n-1/4, \quad k_0=C+\ln 2=1.270, \quad k_1=-1/4. \tag{C6}$$

These values fully agree with the expansions (36) and (37), which follow from the exact formulas for $\tilde{r}_n$. For the exponential potential, $\mu^{-1}=R$ and

$$N=n-1/4, \quad k_0=C+\ln(\pi/2)=1.029, \quad k_1=1/\pi^2 \tag{C7}$$

(the difference in the slopes of the curves in Fig. 4a can be explained by the different values of the parameter μ in these cases). Finally, assuming that

$$f(x)=(e^x+a)^{-1}, \quad -1\leqslant a<\infty, \tag{C8}$$

we have

$$J_0=\frac{2}{\pi}\phi(a), \quad J_1=a[\phi(a)-(1+\sqrt{1+a})^{-1}],$$

$$\phi(a)=\begin{cases}\arcsin\sqrt{-a}/\sqrt{-a}, & -1\leqslant a<0,\\ a^{-1/2}\ln(\sqrt{a}+\sqrt{1+a}), & a>0,\end{cases} \tag{C9}$$

with

$$\phi(a)=\begin{cases}\pi/a-\sqrt{1+a}+\cdots, & a\to -1,\\ 1-(1/6)a+(3/40)a^2+\cdots, & a\to 0,\\ (1/2)a^{-1/2}(\ln 4a+1/2a+\cdots), & a\to\infty.\end{cases}$$

The family of potentials (C8) includes the Hulthén potential ($a=-1$), the exponential potential ($a=0$), and the

Woods–Saxon potential known from nuclear physics ($a=\exp(\mu R)$, with R the nuclear radius, and μ^{-1} is the diffusivity of the edge of the nucleus) . In the latter case,

$$\tilde{g}_n = n^2\pi^2\left\{1+\frac{2\ln 2}{\rho}+O(e^{-\rho})\right\}^{-2}, \quad \rho=\mu R\gg 1,$$

$$k_0=\frac{1}{2}\rho-\ln\rho+O(1), \quad k_1=-\frac{\rho^2}{4\pi^2}+\cdots. \tag{C10}$$

In the limit $\mu\to\infty$ the potential becomes finite (a rectangular well of range R) and the term in (41) leading to a logarithmic increase in the effective range vanishes.

In conclusion we note that (41) is usually a divergent series. For instance, in the case of the Hulthén potential, the expressions in (35) imply that

$$k_j=-j^{-1}B_j\propto(-1)^{j/2}(j-1)!\,(2\pi)^{-j}, \quad j\to\infty, \tag{C11}$$

i.e., the coefficients (with even j) increase as factorials. The same is true of the example specified by (37).

*E-mail: karnak@theor.mephi.msk.su

[1)]For example, the probability of an atom or ion being ionized in electric and magnetic fields is proportional to C_κ^2 (Refs. 2–8). The coefficients C_κ are also present in calculations of the interaction of atomic particles over large distances and in the theory of other peripheral processes.[8]

[2)]Here we ignore the sign of C_κ, which in the case of a Coulomb field is determined by the value of the factor $(-1)^{n-1}$ (the common definition[1]).

[3)]As for the next terms in the expansion, they are generally finite; see the example (18).

[4)]The figures in parentheses are the values of parameter ν.

[5)]The energy E_{nl} of highly excited (Rydberg) states is $-Z^2/2(n-\delta_l)^2$, where the quantum defect δ_l depends on the orbital angular momentum l but is almost entirely independent of n.

[6)]To our knowledge, the literature does not contain the expressions for C_κ, a_s, and r_s (these expressions are derived in Appendix B). Here and in what follows a tilde indicates quantities that refer to the moment when a level appears.

[7)]For states with $l\geqslant 1$ in a short-range potential.[26] On the other hand, for s states the effective range $\tilde{r}_s$ is positive only if the potential $U(r)$ does not contain a wide barrier.[27]

[8)]It is assumed that $U(r)$ is a smooth function and that the quasiclassical approximation is applicable. The obvious counterexample is finite potential, say, a rectangular potential well (for which $\tilde{r}_s=R$ irrespective of the number of the level; see, e.g., Ref. 10) or the delta-function potential.

[9)]The definitions of the scattering lengths a_0 and a_1 given in Refs. 32 and 33 differ in sign. Here we follow Ref. 32.

[10)]The explicit expressions for the functions $\tilde{\eta}$ and $\tilde{\xi}$ are cumbersome and can be found, say in Ref. 34.

[11)]In contrast, the effective range $r_s(g)$ has a first-order pole when $g\to 0$.

[1] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-relativistic Theory*, 3rd ed., Pergamon Press, Oxford (1977).
[2] B. M. Smirnov and M. I. Chibisov, Zh. Éksp. Teor. Fiz. **49**, 841 (1965) [Sov. Phys. JETP **22**, 585 (1966)].
[3] A. M. Perelomov, V. S. Popov, and M. V. Terent'ev, Zh. Éksp. Teor. Fiz. **50**, 1393 (1966) [Sov. Phys. JETP **23**, 924 (1966)]; **51**, 309 (1966) [**24**, 207 (1967)].
[4] V. S. Popov, V. P. Kuznetsov, and A. M. Perelomov, Zh. Éksp. Teor. Fiz. **53**, 331 (1967) [Sov. Phys. JETP **26**, 222 (1968)].
[5] T. Yamabe, A. Tachibana, and H. J. Silverstone, Phys. Rev. A **16**, 877 (1977).
[6] M. V. Ammosov, N. B. Delone, and V. P. Kraĭnov, Zh. Éksp. Teor. Fiz. **91**, 2008 (1986) [Sov. Phys. JETP **64**, 1191 (1986)].
[7] V. S. Popov, B. M. Karnakov, and V. D. Mur, Phys. Lett. A **229**, 306 (1997); Zh. Éksp. Teor. Fiz. **113**, 1579 (1998) [JETP **86**, 860 (1998)].
[8] V. M. Galitskiĭ, E. E. Nikitin, and B. M. Smirnov, *Theory of Collisions of Atomic Particles* [in Russian], Nauka, Moscow (1981).
[9] R. G. Newton, *Scattering Theory of Waves and Particles*, McGraw-Hill, New York (1966).
[10] Ta-You Wu and T. Ohmura, *Quantum Theory of Scattering*, Prentice Hall, New York (1962).
[11] A. A. Radtsig and B. M. Smirnov, *Reference Data on Atoms, Molecules, and Ions*, Springer, Berlin (1985).
[12] V. S. Popov, A. E. Kudryavtsev, and V. D. Mur, Zh. Éksp. Teor. Fiz. **77**, 1727 (1979) [Sov. Phys. JETP **50**, 865 (1979)]; **80**, 1271 (1981) [**53**, 650 (1981)].
[13] H. A. Bethe, Phys. Rev. **76**, 38 (1949).
[14] J. D. Jackson and J. M. Blatt, Rev. Mod. Phys. **22**, 77 (1950).
[15] G. E. Brown and A. D. Jackson, *The Nucleon–Nucleon Interaction*, North-Holland, Amsterdam (1976).
[16] *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun, National Bureau of Standards Applied Mathematics Series 55, Washington, D.C. (1964).
[17] M. J. Seaton, Mon. Not. R. Astron. Soc. **118**, 504 (1958).
[18] P. G. Burke, *Potential Scattering in Atomic Physics*, Plenum Press, New York (1977).
[19] G. Breit and W. G. Bouricius, Phys. Rev. **74**, 1546 (1948).
[20] G. Breit, *Theory of Resonance Reactions*, Springer, Berlin (1959).
[21] Y. Yamaguchi, Phys. Rev. **95**, 1628 (1954).
[22] M. S. Marinov and V. S. Popov, Zh. Éksp. Teor. Fiz. **67**, 1250 (1974) [Sov. Phys. JETP **40**, 621 (1975)]; J. Phys. A **8**, 1250 (1975).
[23] B. M. Karnakov, V. D. Mur, V. S. Popov, Zh. Éksp. Teor. Fiz. **107**, 1768 (1995) [JETP **80**, 983 (1995)].
[24] V. S. Popov, B. M. Karnakov, and V. D. Mur, Phys. Lett. A **210**, 402 (1996).
[25] A. B. Migdal, *Qualitative Methods in Quantum Theory*, Addison-Wesley, Reading, Mass. (1977).
[26] V. D. Mur and V. S. Popov, Teor. Mat. Fiz. **27**, 204 (1976).
[27] A. B. Migdal, A. M. Perelomov, and V. S. Popov, Yad. Fiz. **14**, 874 (1972) [Sov. J. Nucl. Phys. **14**, 488 (1972)].
[28] B. M. Karnakov, V. D. Mur, and V. S. Popov, Dokl. Akad. Nauk SSSR **279**, 345 (1984) [Sov. Phys. Dokl. **29**, 938 (1984)].
[29] B. M. Karnakov, A. E. Kudryavtsev, V. D. Mur, and V. S. Popov, Zh. Éksp. Teor. Fiz. **94**(7), 65 (1988) [Sov. Phys. JETP **67**, 1333 (1988)].
[30] M. Froissart, Phys. Rev. **123**, 1053 (1961).
[31] A. Martin, Preprint CERN/LHCC 97-49 (1997).
[32] V. D. Mur, A. E. Kudryavtsev, V. S. Popov, Preprint No. 46, Institute of Theoretical and Experimental Physics, Moscow (1980); JETP Lett. **31**, 401 (1980).
[33] J. Jaenicke, B. Kerbikov, and H.-J. Pirner, Z. Phys. A **339**, 297 (1991).
[34] V. D. Mur and V. S. Popov, Yad. Fiz. **42**, 1470 (1985) [Sov. J. Nucl. Phys. **42**, 930 (1985)].
[35] T. L. Trueman, Nucl. Phys. **26**, 57 (1961).
[36] J. Schwinger, Phys. Rev. **72**, 742 (1947); **78**, 135 (1950).
[37] Ya. A. Smorodinskiĭ, Dokl. Akad. Nauk SSSR **60**, 217 (1948).
[38] B. M. Karnakov and V. P. Kraĭnov, *Quasiclassical Approximation in Quantum Mechanics (WKB Method)* [in Russian], Moscow Engineering Physics Institute Press, Moscow (1992).

Translated by Eugene Yankovsky

Inductive interaction of rapidly rotating conductive bodies with a magnetized plasma

R. R. Rafikov, A. V. Gurevich,*) and K. P. Zybin

I. E. Tamm Theoretical Physics Division, P. N. Lebedev Physics Institute, Russian Academy of Sciences, 117924 Moscow, Russia

(Submitted 26 June 1998)
Zh. Éksp. Teor. Fiz. **115**, 542–563 (February 1999)

This paper discusses the magnetohydrodynamic flow of a supersonic plasma flux around a thin conductive body. The specific features of the inductive interaction of the body with the plasma and the process by which plasma waves are generated when the body rotates are investigated. The structure of the magnetic-field perturbations and the distribution of the plasma currents that result from the inductive interaction are studied. Expressions are obtained for the forces that act on a plate and the torques produced by these forces. A simple model is used to take into account the kinetic effects associated with the finiteness of the absorption and emission currents that transport charge from the plate to the plasma and back. A surface-potential distribution is found that can substantially accelerate particles in the neighborhood of the body. © *1999 American Institute of Physics.* [S1063-7761(99)01202-0]

1. INTRODUCTION

In ordinary gas dynamics, a moving body is decelerated by the direct interaction of the incident gas flow with the surface of the body. When the body is in supersonic motion, this interaction causes a shock wave to be generated.

The picture is very different in a highly disperse magnetized plasma. Direct interaction of the particles with the surface of the body is less substantial here. On the other hand, a new form of interaction appears, associated with the induction of an electric field and currents on the surface of the body. This interaction electromagnetically decelerates the body and causes an Alfvén wave and other magnetohydrodynamic (MHD) waves to be generated in the plasma. Such electromagnetic interaction of conductive bodies with a magnetized plasma has become known as inductive interaction.

Drell, Foley, and Ruderman[1] were the first to indicate that inductive interaction could accompany streamline flow around a body under ideal magnetohydrodynamic conditions. The detailed theory of this phenomena was developed by Gurevich, Krylov, and Fedorov.[2] They showed that it can play a substantial role in a plasma in outer space. Recently obtained experimental data confirm that it is important to consider inductive effects.[3–5] A number of studies[6–8] have been devoted to applications of the theory to various specific problems, for example, the Jupiter–Io interaction and the motion of a highly elongated conductive filament in the ionosphere (theta ring). We should emphasize that only translational motion of the body around which flow occurs was considered in all these papers.

At the same time, bodies around which a space plasma flows, for instance planets, rotate, and this rotation can substantially affect the inductive-interaction effects. This paper is devoted to an investigation of this question.

It should be emphasized that the actual process of inductive interaction of a body with a magnetized plasma consists of two independent problems: the generation of waves in ideal magnetohydrodynamics by a current flowing over the surface of the body,[2,9] and the structure of the transition layer that determines how a current flows from the plasma to the body and back. The latter problem is associated with the possibility that the surface can emit and absorb charged particles of the plasma.[10] It has a largely kinetic character and strongly depends on the physical structure of the body around which the flow occurs—this can be, for example, a solid surface, or even a fairly dense near-surface collisional plasma (like the lower layers of the ionosphere or the lower region of the solar corona). Significant longitudinal (i.e., along the magnetic field) electric fields can appear in the near-surface emission and absorption layer because of inductive interaction, causing the particles to be accelerated and generating high-frequency radiation as well as other auroral processes.

This paper will discuss how the rotation of a body around which a plasma flows affects both the generation of MHD waves and the structure of the near-surface layer.

2. FLOW AROUND BODIES IN IDEAL MAGNETOHYDRODYNAMICS

2.1. Fundamental equations

Let us consider the flow around a rotating body by a supersonic flux of plasma with a frozen-in magnetic field. To do this, we investigate the inductive interaction caused by the generation of a current on a body with finite conductivity.

Specifically, in a coordinate system connected to the moving body there is an induction electric field

$$\mathbf{E}=-\frac{\mathbf{v}\times\mathbf{B}}{c}, \tag{1}$$

where $\mathbf{v}$ is the velocity at which the plasma flows relative to the body. The electric current that appears in this case in a

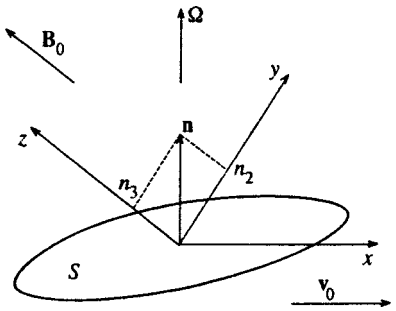

FIG. 1. Spatial disposition of the plate relative to the magnetic field.

body with finite conductivity perturbs the external field and the plasma velocity, resulting in the generation of MHD waves.

To describe steady-state streamline flow, we use the equations of ideal magnetohydrodynamics:

$$\operatorname{div}\rho\mathbf{v}=0, \quad \operatorname{div}\mathbf{B}=0, \quad \operatorname{curl}(\mathbf{v}\times\mathbf{B})=0, \tag{2}$$

$$(\mathbf{v}\cdot\nabla)\mathbf{v}=-\frac{1}{\rho}\operatorname{grad}p-\frac{1}{4\pi\rho}\mathbf{B}\times(\operatorname{curl}\mathbf{B}). \tag{3}$$

In order to investigate purely inductive effects, we assume that the body is an infinitely thin circular plate S of radius R, located along the incoming flux, with normal $\mathbf{n}=n_2\mathbf{e}_2+n_3\mathbf{e}_3$. We assume that $\mathbf{e}_3$ is directed along magnetic field $\mathbf{B}_0$, which in turn has no component along the translational velocity of the plate. Unit vector $\mathbf{e}_1$ is directed along this velocity, while $\mathbf{e}_2$ forms with them an orthogonal coordinate system. Here in general we consider a plate tilted relative to the field, i.e., $n_2\neq 0$. It follows from this that field $\mathbf{B}_0$ has no component along the x axis and that the y axis in general does not lie in the plane of the plate (see Fig. 1).

The surface conductivity $\hat{\Sigma}$ of the plate is written as

$$\hat{\Sigma}=\begin{pmatrix}\Sigma_p & \Sigma_h\\ -\Sigma_h & \Sigma_p\end{pmatrix}, \tag{4}$$

where Σ_p is the Pedersen conductivity, and Σ_h is the Hall conductivity.

The structure of the perturbations is studied in the most natural coordinate system, connected to the center of the plate and moving along with it. In this system, the unperturbed values of the field, the velocity, and the plasma density are

$$\mathbf{B}_0=B_0\mathbf{e}_3, \quad \mathbf{v}_0=v_0\mathbf{e}_1, \quad \rho=\rho_0.$$

The perturbations of these quantities are sought in the form

$$\mathbf{B}=\mathbf{B}_0+M_A\mathbf{b}, \quad \mathbf{v}=\mathbf{v}_0+M_A\mathbf{u}, \quad \rho=\rho_0+M\rho, \tag{5}$$

where

$$M=v_0/v_s, \quad v_s=\left(\frac{\partial p}{\partial\rho}\right)_s, \quad M_A=v_0/v_A.$$

The corrections are assumed to be small, so that the problem can be studied by linearizing the MHD equations, Eqs. (2) and (3), with respect to these corrections. As a result of linearization we find

$$\frac{\partial b_x}{\partial x}+\frac{\partial b_y}{\partial y}+\frac{\partial b_z}{\partial z}=0, \tag{6}$$

$$\frac{\partial\rho}{\partial x}+\frac{1}{M}\left(\frac{\partial u_x}{\partial x}+\frac{\partial u_y}{\partial y}+\frac{\partial u_z}{\partial z}\right)=0, \tag{7}$$

$$\frac{\partial u_y}{\partial x}+\frac{1}{M}\frac{\partial\rho}{\partial y}+\frac{1}{M_A}\left(\frac{\partial b_z}{\partial y}-\frac{\partial b_y}{\partial z}\right)=0, \tag{8}$$

$$\frac{\partial u_z}{\partial x}+\frac{1}{M}\frac{\partial\rho}{\partial z}=0, \tag{9}$$

$$\frac{\partial u_x}{\partial x}+\frac{1}{M}\frac{\partial\rho}{\partial x}+\frac{1}{M_A}\left(\frac{\partial b_z}{\partial x}-\frac{\partial b_x}{\partial z}\right)=0, \tag{10}$$

$$\frac{\partial b_x}{\partial x}-\frac{1}{M_A}\frac{\partial u_x}{\partial z}=0, \tag{11}$$

$$\frac{\partial b_y}{\partial x}-\frac{1}{M_A}\frac{\partial u_y}{\partial z}=0, \tag{12}$$

$$\frac{\partial b_z}{\partial x}-\frac{1}{M_A}\frac{\partial u_z}{\partial z}-\frac{M}{M_A}\frac{\partial\rho}{\partial x}=0. \tag{13}$$

The boundary conditions on the plate are determined from the following considerations. Naturally, the normal component of the plasma velocity at the plate must vanish:

$$(\mathbf{u}\cdot\mathbf{n})|_S=0. \tag{14}$$

Then, because a surface current is present, the magnetic field parallel to the surface of the plate experiences a jump:

$$\{\mathbf{n}\times\mathbf{b}\}_S=\frac{4\pi}{cM_AB_0}\mathbf{I}. \tag{15}$$

Here $\{f\}_S$ denotes the jump of the value of f as one goes through the surface of the plate, and $\mathbf{I}$ is the surface current. Finally, the tangential component of the electric field is also continuous:

$$E_t=-c^{-1}(\mathbf{v}\times\mathbf{B})_t=-c^{-1}B_nv_t$$

(this equation is valid because the normal velocity v_n vanishes). However, the normal component of field $\mathbf{B}$ is also continuous, and hence we find that the the tangential velocity of the plasma is continuous as one goes through the plate. In the general case of arbitrary tilt of the plate, this means that

$$\{u_x\}_S=\{u_y\}_S=\{u_z\}_S=0. \tag{16}$$

Under the conditions

$$1\ll M_A\ll M \tag{17}$$

the equations split, giving three waves: an Alfvén wave, and fast and slow magnetodynamic waves:

$$\frac{\partial u_x}{\partial x}-\frac{1}{M_A}\frac{\partial b_x}{\partial z}=0, \quad \frac{\partial b_x}{\partial x}-\frac{1}{M_A}\frac{\partial u_x}{\partial z}=0, \tag{18}$$

$$\frac{\partial u_y}{\partial x}+\frac{1}{M_A}\left(\frac{\partial b_z}{\partial y}-\frac{\partial b_y}{\partial z}\right)=0, \quad \frac{\partial b_y}{\partial y}+\frac{\partial b_z}{\partial z}=0, \tag{19}$$

$$\frac{\partial b_y}{\partial x}-\frac{1}{M_A}\frac{\partial u_y}{\partial z}=0, \quad \frac{\partial b_z}{\partial x}-\frac{M}{M_A}\frac{\partial\rho}{\partial x}=0, \tag{20}$$

$$\frac{\partial \rho}{\partial x}+\frac{1}{M}\frac{\partial u_z}{\partial z}=0, \quad \frac{\partial u_z}{\partial x}+\frac{1}{M}\frac{\partial \rho}{\partial z}=0. \tag{21}$$

The accuracy of this system is the same as that of the system considered in Ref. 2, which describes the induction effects caused by translational motion of a plate. Our paper investigates inductive-interaction phenomena caused by the rotation of a body. As we can see, the basic system of equations (18)–(21) maintains its form in this case. The introduction of rotation of the body only changes the boundary conditions of the problem.

In what follows, we shall assume that the conditions given by inequalities (17) are satisfied, which allows us to consider the excitation of the different types of waves separately.

2.2. The Alfvén wave

For the Alfvén wave we have Eq. (18) with the boundary conditions

$$\{b_x\}_S=4\pi I_y/cM_AB_0, \quad \{u_x\}_S=0. \tag{22}$$

We shall write out in detail the expression for the surface current. To do this, we introduce the $x'y'z'$ coordinate system, with the x' axis along the direction of motion, the z' axis normal to the plate, and the y' axis in the plane of the plate and perpendicular to x' and z'. In this system, field $\mathbf{B}_0$ has the form $(0,n_2B_0,n_3B_0)$. The velocity of a point on the plate with coordinates $(x',y',0)$ is

$$\mathbf{v}=v_0\mathbf{e}_{x'}+\mathbf{\Omega}\times\mathbf{r}'.$$

A surface induction field of

$$\mathbf{E}=-\frac{\mathbf{v}\times\mathbf{B}}{c}=-\frac{B_0n_3}{c}\begin{pmatrix}\Omega x\\ \Omega y-v_0\end{pmatrix} \tag{23}$$

is generated at this point. The result of this is the surface current

$$\mathbf{I}=\hat{\Sigma}\mathbf{E}=-\frac{B_0n_3}{c}\begin{pmatrix}\Sigma_p\Omega x+\Sigma_h\Omega y'-\Sigma_hv_0\\ \Sigma_p\Omega y'-\Sigma_h\Omega x-\Sigma_pv_0\end{pmatrix}. \tag{24}$$

Since $y'=y/n_3$, we finally obtain

$$I_x=-\frac{B_0n_3}{c}\left(\Sigma_p\Omega x+\Sigma_h\Omega\frac{y}{n_3}-\Sigma_hv_0\right), \tag{25}$$

$$I_{y'}=-\frac{B_0n_3}{c}\left(\Sigma_p\Omega\frac{y}{n_3}-\Sigma_h\Omega x-\Sigma_pv_0\right). \tag{26}$$

By simple transformations, we obtain from Eq. (18) together with Eq. (26) the following boundary-value problem:

$$\frac{\partial^2 b_x}{\partial x^2}-\frac{1}{M_A^2}\frac{\partial^2 b_x}{\partial z^2}=0, \tag{27}$$

$$\{b_x\}_s=f_p(x,y)=\frac{4\pi}{c^2}\Sigma_pv_An_3+\frac{\Omega R}{v_0}\frac{4\pi v_A}{c^2}\times\left(\Sigma_p\frac{y}{R}-n_3\Sigma_h\frac{x}{R}\right), \quad \left\{\frac{\partial b_x}{\partial z}\right\}_s=0. \tag{28}$$

Here the rotation of the body causes the presence of the second term in Eq. (28), proportional to the angular frequency Ω.

The solution of Eq. (27) with the boundary conditions of Eq. (28) can immediately be written out:

$$b_x=\frac{1}{2}\begin{cases}f_p(x^+,y), & n_2y+n_3z>0,\\ -f_p(x^-,y), & n_2y+n_3z<0,\end{cases} \tag{29}$$

where

$$x^{\pm}=x\mp M_A\frac{n_2y+n_3z}{n_3}.$$

It can be seen that a nonzero solution exists only inside the Alfvén cylinders F given by the condition

$$(x^{\pm})^2+\frac{y^2}{n_3^2}=R^2,$$

and that the perturbations characteristically are communicated along the magnetic field, as expected for an ordinary Alfvén wave. In marked contrast with the case of purely translational motion, the magnetic field in the resulting solution is no longer constant inside the cylinder, but depends strongly on position.

The perturbation of velocity u_x in this case is written in the form

$$u_x=-\frac{1}{2}\begin{cases}f_p(x^+,y), & n_2y+n_3z>0,\\ f_p(x^-,y), & n_2y+n_3z<0.\end{cases} \tag{30}$$

To study how rotation affects the flow pattern, we consider the structure of the currents in space. Their volume density equals

$$\mathbf{j}=\frac{c}{4\pi}\left(0,\ \frac{\partial b_x}{\partial z},\ -\frac{\partial b_x}{\partial y}\right). \tag{31}$$

Substituting Eq. (29) into this formula, we find that a homogeneous volume current

$$\mathbf{j}_{\pm}=\left[0,\ -\frac{\Omega\Sigma_hn_3}{2c},\ \frac{\Omega}{2cM_A}(n_2M_A\Sigma_h\pm\Sigma_p)\right] \tag{32}$$

exists inside the cylinder. The presence of a nonzero volume current inside the Alfvén cylinders fundamentally differs from the case of a nonrotating plate, where there was no internal current and the entire current flowed along the surfaces of discontinuity, passing through the edges of the plate. The presence of an internal current is a direct consequence of the rotation and the resulting dependence of the internal magnetic field on the coordinates (see Ref. 2).

It can also be seen from Eq. (32) that, owing to rotation, Alfvén waves are generated not only because of the Pedersen conductivity Σ_p, but also because of the Hall conductivity Σ_h. This is true only when rotation enters the picture; it is absent for purely translational motion. The volume current that flows inside the Alfvén cylinders and the surface current on the plate are closed through the surface of the Alfvén cylinders, where a delta-function current flows, associated with the magnetic-field jump at the boundary (see Fig. 2).

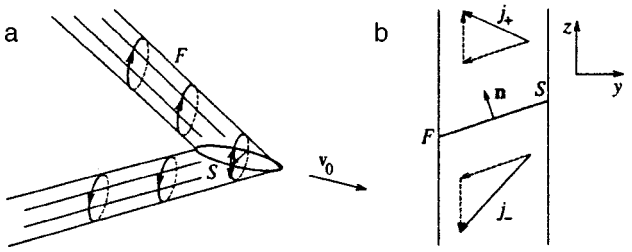

FIG. 2. Volume current inside the Alfvén cylinders: (a) side view, (b) front view; F is the surface of the Alfvén cylinders.

Such a singular current appears because in the ideal case we are considering currents can be transmitted through the plasma with as large a density as we like. This assumption of ideal magnetohydrodynamics is of course not valid in an actual plasma, where the velocity of the charge carriers and their spatial density are limited. This circumstance will be taken into account in what follows by considering the kinetic effects associated with the fact that current is transported from the plate to the plasma via electron emission and absorption, with the corresponding currents being finite.

2.3. The fast magnetosonic wave

To consider the fast magnetosonic wave, we use the system of Eqs. (19) and (20) with the boundary condition following from Eq. (25):

$$\frac{\partial u_y}{\partial x}+\frac{1}{M_A}\left(\frac{\partial b_z}{\partial y}-\frac{\partial b_y}{\partial z}\right)=0, \quad \frac{\partial b_y}{\partial y}+\frac{\partial b_z}{\partial z}=0, \tag{33}$$

$$\frac{\partial b_y}{\partial x}-\frac{1}{M_A}\frac{\partial u_y}{\partial z}=0, \quad \rho=\frac{M_A}{M}b_z, \tag{34}$$

$$\{n_2 b_z-n_3 b_y\}_S=f_B(x,y)=\frac{4\pi}{c^2}\Sigma_h v_A n_3+\frac{\Omega R}{v_0}\frac{4\pi v_A}{c^2}\times\left(\Sigma_p n_3\frac{x}{R}+\Sigma_h\frac{y}{R}\right), \quad \{u_y\}_S=0. \tag{35}$$

We introduce a function $\psi(x,y,z)$ such that

$$b_y=\frac{\partial\psi}{\partial z}, \quad b_z=-\frac{\partial\psi}{\partial y}, \quad u_y=M_A\frac{\partial\psi}{\partial x}. \tag{36}$$

Then Eqs. (33) and (34) yield the two-dimensional wave equation

$$M_A^2\frac{\partial^2\psi}{\partial x^2}-\frac{\partial^2\psi}{\partial y^2}-\frac{\partial^2\psi}{\partial z^2}=0, \tag{37}$$

$$\left\{n_2\frac{\partial\psi}{\partial y}+n_3\frac{\partial\psi}{\partial z}\right\}_S=f_B, \quad \left\{\frac{\partial\psi}{\partial x}\right\}_S=0. \tag{38}$$

Its solution can be written as

$$\psi=-\frac{1}{2\pi n_3}\int_{S'}f_B(\xi_1,\xi_2)\,d\xi_1 d\xi_2 \times\frac{\theta(x-\xi_1-M_A\sqrt{(y-\xi_2)^2+(z+n_2\xi_2/n_3)^2})}{\sqrt{(x-\xi_1)^2-M_A^2[(y-\xi_2)^2+(z+n_2\xi_2/n_3)^2]}}, \tag{39}$$

where S' is the projection of S onto the xy plane, and $\theta(z)$ is the Heaviside function. It can be seen that a nonzero solution exists only within the Mach cone $y^2+z^2<x^2/M_A^2$.

Because the boundary conditions are linear, we can represent the solution as a sum of two parts, ψ^0 and $\psi^{\rm rot}$, caused by translational and rotational motion, respectively. Let us investigate the asymptotic behavior of these parts. For a nonrotating disk, i.e., for $\Omega=0$, an explicit expression can be obtained for ψ^0 in the approximation $R\ll x,y,z$ with (x,y,z) lying fairly deep inside the Mach cone. In this case, by expanding the expression under the integral in powers of $\xi_1/x,\xi_2/y,\xi_2/z$ and integrating, we find that

$$\psi^0(x,y,z)=\frac{2v_A\Sigma_h\pi R^2 n_3}{c^2\sqrt{x^2-M_A^2r^2}}, \tag{40}$$

where $r=\sqrt{y^2+z^2}$.

The magnetic-field perturbations have a toroidal structure:

$$b_y^0=\frac{2v_A\Sigma_h\pi R^2 n_3 M_A^2}{c^2(x^2-M_A^2r^2)^{3/2}}z, \tag{41}$$

$$b_z^0=-\frac{2v_A\Sigma_h\pi R^2 n_3 M_A^2}{c^2(x^2-M_A^2r^2)^{3/2}}y. \tag{42}$$

Starting from these expressions, it is easy to analyze the structure of the currents. Specifically, we find $\mathbf{b}=\mathrm{curl}(\psi\mathbf{e}_1)$, while

$$\mathbf{j}=\frac{cM_AB_0}{4\pi}\mathrm{curl}\,\mathrm{curl}(\psi\mathbf{e}_1) =\frac{cM_AB_0}{4\pi}\left(-\frac{\partial^2\psi}{\partial^2 y}-\frac{\partial^2\psi}{\partial^2 z},\frac{\partial^2\psi}{\partial y\partial x},\frac{\partial^2\psi}{\partial z\partial x}\right). \tag{43}$$

As a result, we see that there is no toroidal component of the current, and

$$j_x^0=\frac{v_A\Sigma_h\pi R^4 n_3 M_A^3 B_0}{2\pi c(x^2-M_A^2r^2)^{5/2}}(2x^2+M_A^2r^2), \tag{44}$$

$$j_r^0=\frac{v_A\Sigma_h\pi R^4 n_3 M_A^3 B_0}{2\pi c(x^2-M_A^2r^2)^{5/2}}\,3rx. \tag{45}$$

The equation for the current lines $dx/dr=j_x/j_r$ looks like

$$\frac{dx}{dr}=\frac{2x^2+M_A^2r^2}{3xr} \tag{46}$$

and has as its solution $x^2=C_0r^{4/3}+M_A^2r^2$. It can be seen that, at large distances from the body, these are simply cones.

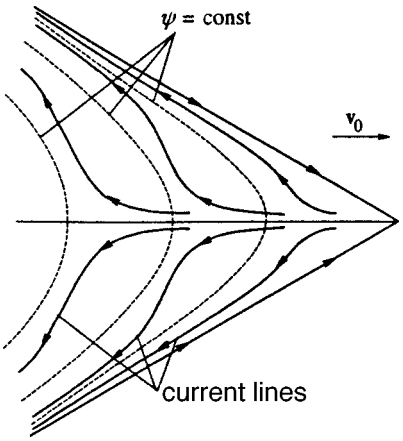

FIG. 3. Current surfaces and surfaces of constant potential ψ in the absence of rotation.

For comparison, we note that the surfaces of constant ψ are determined by the equation $x^2=M_A^2r^2+C_1^2$ (see Fig. 3).

Under the same assumptions $R\ll x,y,z$, we can find the asymptotic form of the second term in Eq. (39), which describes the rotation effects:

$$\psi^{\rm rot}=n_3\frac{\pi R^3[v_A\Sigma_p x-v_0\Sigma_h n_3M_A(y-n_2z)]}{2c^2(x^2-M_A^2r^2)^{3/2}}. \tag{47}$$

In summary, we obtain at large distances from the plate

$$\psi(x,y,z)=\frac{2v_A\pi R^2n_3}{c^2\sqrt{x^2-M_A^2r^2}} \times\left[\Sigma_h+\frac{\Omega R}{v_0}\frac{R(\Sigma_p x-\Sigma_h M_A^2(n_3y-n_2z))}{4(x^2-M_A^2r^2)}\right]. \tag{48}$$

It can be seen from this equation that perturbations caused by rotation decrease with distance more rapidly than do translational perturbations, so that the current structure at distances of

$$x\gg R\frac{\Sigma_P}{\Sigma_h}\frac{\Omega R}{v_0},\quad \text{for}\quad \Sigma_p>\Sigma_hM_A,$$

$$x\gg RM_A\frac{\Omega R}{v_0},\quad \text{for}\quad \Sigma_p<\Sigma_hM_A,$$

will mainly be determined by the translational motion.

If $\Sigma_h=0$ holds, however, the fast magnetosonic wave caused by translational motion does not appear at all, and the perturbation structure will therefore be determined only by the rotation of the plate. In this case

$$\psi(x,y,z)=\frac{v_A\Sigma_p\pi R^4\Omega n_3x}{2c^2v_0(x^2-M_A^2r^2)^{3/2}}. \tag{49}$$

As in the case of purely translational motion, the field for $\Sigma_h=0$ is toroidal:

$$b_y=\frac{v_A\Sigma_p\pi R^4\Omega n_3M_A^2x}{2cv_0(x^2-M_A^2r^2)^{5/2}}z, \tag{50}$$

$$b_z=-\frac{v_A\Sigma_p\pi R^4\Omega n_3M_A^2x}{2cv_0(x^2-M_A^2r^2)^{5/2}}y. \tag{51}$$

Consequently, the current has no toroidal component, and we have for the remaining components of the current

$$j_x=\frac{3\pi R^4n_3\Omega M_A}{2c(x^2-M_A^2r^2)^{7/2}}x\Sigma_p(2x^2+3M_A^2r^2), \tag{52}$$

$$j_r=-\frac{3\pi R^4n_3\Omega M_A}{2c(x^2-M_A^2r^2)^{7/2}}r\Sigma_p(4x^2+M_A^2r^2). \tag{53}$$

The equation of the current lines

$$\frac{dx}{dr}=\frac{x}{r}\frac{2x^2+3M_A^2r^2}{4x^2+M_A^2r^2} \tag{54}$$

gives the following form of the current surfaces:

$$(x^2-M_A^2r^2)^5=C_2x^2r^4.$$

It can be seen that in this case, in the limit of large distances, these surfaces approach a cone $x^2=M_A^2r^2$.

Note also that, for nonzero rotational velocity Ω, the structure of the fast magnetosonic wave explicitly depends not only on the Hall conductivity Σ_h (as was true in the case of purely translational motion), but also on the Pedersen conductivity Σ_p.

2.4. The slow magnetosonic wave

The behavior of the slow magnetosonic wave is determined by the velocity u_y, which enters into the boundary conditions for Eqs. (21):

$$u_z\Big|_S=-\frac{n_2}{n_3}u_y\Big|_S=w(x,y),\quad \{u_z\}_S=0. \tag{55}$$

It can be seen that, as in the case of a nonrotating plate, slow sound is not excited for $n_2=0$ when the plane of the plate is orthogonal to the magnetic field.

We obtain the following boundary-value problem from Eq. (21):

$$\frac{\partial^2u_z}{\partial x^2}-\frac{1}{M^2}\frac{\partial^2u_z}{\partial z^2}=0, \tag{56}$$

$$u_z\Big|_S=-\frac{n_2}{n_3}u_y\Big|_S=w(x,y),\quad \{u_z\}_S=0. \tag{57}$$

These equations are quite analogous to the Alfvén case, so that it is easy to obtain

$$u_z=\frac{1}{2}\begin{cases}w(x^+,y), & n_2y+n_3z>0,\\ w(x^-,y), & n_2y+n_3z<0,\end{cases} \tag{58}$$

$$\rho=\frac{1}{2}\begin{cases}w(x^+,y), & n_2y+n_3z>0,\\ -w(x^-,y), & n_2y+n_3z<0,\end{cases} \tag{59}$$

where

$$x^{\pm}=x\mp M\frac{n_2y+n_3z}{n_3}.$$

As in the case of an Alfvén wave, here all the perturbations are concentrated inside divergent cylinders, which are now tilted at an angle of $\arctan(1/M)$, and not $\arctan(1/M_A)$.

The solution of the problem inside the cylinders is completely determined by velocity u_y, and hence by ψ on the surface of the plate. However, on the surface of the body itself, at distances less than R, the integration of Eq. (39) to find ψ is appreciably more complicated because now it must be carried out not over the entire surface of the plate, as was done when obtaining the asymptotic behavior given by Eqs. (40) and (47), but only over that part of it where the argument of the Heaviside function is positive.

3. THE TRANSITION LAYER AT THE SURFACE OF THE BODY

3.1. Fundamental equations

As is well known, because of the processes of electron absorption and emission and ion recombination, a body in a plasma does not remain neutral. It acquires an electric charge, which causes an irrotational electric field to appear on the surface of the body.

$$\mathbf{E}=-\nabla\phi, \tag{60}$$

where ϕ is the surface potential. After rewriting the continuity equation for the current on the surface of the plate, we get

$$\frac{\partial I_x}{\partial x}+\frac{\partial I_y}{\partial y}=\{j_z\}|_S. \tag{61}$$

The current on the surface now equals

$$I=-\hat{\Sigma}\left(\nabla\phi+\frac{\mathbf{v}\times\mathbf{B}}{c}\right). \tag{62}$$

Substituting this expression into Eq. (61), we obtain the well-known equation (See Refs. 12, 9, and 10) that determines the surface potential distribution:

$$\mathrm{div}_S\left[\hat{\Sigma}\left(\nabla\phi+\frac{\mathbf{v}\times\mathbf{B}}{c}\right)\right]=\{j_z\}_S. \tag{63}$$

By knowing the potential distribution, it is possible, using Eq. (62), to construct a picture of how the surface current I is distributed, which can then be used to determine the structure of the MHD waves.

It is important to take into account the potential that appears on the body, because that is what determines the longitudinal potential difference in the transition layer between the ideal MHD flux and the surface of the body. The longitudinal electric field can accelerate the plasma particles to rather large energies. Thus, for example, the decimeter radiation of the auroral zones of Jupiter is usually affected by precisely such a mechanism, when it is assumed that particles accelerated in the neighborhood of Io collide with the upper layers of the polar regions of the Jovian atmosphere. The same effects, but in a more complex form because of the presence of the magnetosphere, affect the particle acceleration and the auroral processes in the near-polar regions of the earth.

We should emphasize that the absorption and emission currents j_z themselves usually strongly depend on the value and distribution of potential ϕ on the body. It is determined by kinetic processes close to the surface, by the structure of the surface itself or by the near-surface plasma, by the absorption or recombination conditions, and also by charged-particle emission processes. Establishing the surface layer between the ideal MHD plasma and the surface of the body thus constitutes a separate problem, which depends on the specific conditions and is determined by many physical parameters. We consider here the simplest case of free absorption and emission of electrons. Assuming that the translational velocity v_0 of the body satisfies the conditions

$$v_{T_i}\ll v_0\ll v_{T_e} \tag{64}$$

($v_{T_i}=\sqrt{T/M}, v_{T_e}=\sqrt{T/m}$ are the thermal velocities of the ions and electrons, respectively), the ion current can be neglected. The free absorption of electrons is described in this case by (see, for example, Ref. 11)

$$j_z=-j_0\begin{cases}1, & \phi>0,\\ \exp\left(\dfrac{e\phi}{T}\right), & \phi<0,\end{cases} \tag{65}$$

$$j_0=\sqrt{\frac{T}{2\pi m}}\,eN. \tag{66}$$

Here N is the plasma density. For free emission of electrons, the dependence of the current on the surface potential has a similar form:

$$j_z=j_0\begin{cases}1, & \phi<0,\\ \exp\left(-\dfrac{e\phi}{T}\right), & \phi>0.\end{cases} \tag{67}$$

Everywhere in what follows we shall confine ourselves to just these laws when specifically considering absorption and emission currents.

It was shown earlier[9,10] that, with purely translational motion, a substantial surface potential arises only in narrow regions at the edge of a body, where the surface conductivity varies sharply.

Actually, the gradient $\nabla\phi$ far from the edge of a plate is small if the conductivity of the plate is constant along its surface and satisfies the condition

$$\epsilon=\frac{\Sigma v_0B_0}{cj_0R}\ll 1 \tag{68}$$

(see Ref. 10). Neglecting it in Eq. (63), we see that $j_z=0$, i.e., the potential is $\phi=0$.

However, at the very edge of the plate, where the surface conductivity abruptly falls off to zero, a term arises with a singular value of $\partial\Sigma/\partial r$. Because of this, a current $j_z\neq 0$ appears close to the edge, and $\phi\neq 0$. A thin layer (by comparison with the size of the plate), in which $\phi\gg T/e$ holds, is formed in this region. If we consider a region of positive

potential, this causes the emission current to be suppressed and the absorption current to be amplified to its maximum value of j_0. In a region of negative potential, the situation is accordingly the reverse. It is immediately possible to estimate the order of magnitude of the thickness of the current layer. The induction potential creates a total current of $I \sim E_{\rm ind} l \Sigma$ on the plate, where l is the characteristic scale of the body, in our case the radius R of the plate. In order to pump such a current from the plasma to the plate by emission or absorption, Eqs. (65), (67), it is necessary to have an area of at least $S \sim E_{\rm ind} l \Sigma / j_0$. If it is assumed that this area S extends along the edge of the plate, $S \sim l x_0$, we obtain an estimate of the thickness of the current layer:

$$x_0 \sim E_{\rm ind} \Sigma / j_0 .$$

If the conditions of the problem are such that the layer is actually narrow, $x_0 \ll R$ [for this, as can easily be seen, the condition given in inequality (68) must be satisfied], we can simplify the problem by regarding it as quasi-one-dimensional. More precisely, we can assume in this case that, inside the layer, all the derivatives along the layer are negligible by comparison with those across the layer, because of which the problem becomes one-dimensional.

3.2. The geometry of the transition layer

We shall study the problem for the simplest case, $n_2 = 0$. We consider the region right at the edge of the plate at scales of the order of the layer thickness. Since this thickness is assumed to be small (the conditions necessary for this will be obtained subsequently), we can in general consider the edge of the plate to be flat (we let the radius of the plate tend to infinity). We then need only solve the problem of plasma flow onto a semi-infinite plate, examine the structure of the transition layer, and then generalize the result.

Consider the case of $\phi > 0$; the results for $\phi < 0$ can be obtained from simple symmetry considerations. We introduce at the edge of the plate a rectangular coordinate system with its x axis directed into the plate perpendicular to the edge, while the y axis is parallel to the edge. It can then be seen that the x axis is antiparallel to the radius of the polar coordinate system with origin at the center of the plate, while the y axis is antiparallel to φ at the same point.

We shall assume that the normal to the edge of the plate and the velocity of the incoming plasma make a certain angle α between them. In the case of a circular plate, this angle is $\alpha = -\varphi$, and therefore it is angle φ that will be used below. When $\Omega \neq 0$ holds, an additional component of the velocity, ΩR, appears along $\mathbf{e}_\varphi$ or, in the local system of the edge of the plate, in the negative direction of the y axis. The velocity of the inflowing flux in such a geometry is

$$\mathbf{v} = (v \cos\varphi, -v \sin\varphi - \Omega r, 0).$$

Assuming that the plate is semi-infinite, all the quantities are independent of y and, measuring x from the edge of the plate, we can rewrite Eq. (63) in a one-dimensional version as

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{2\Omega B}{c} = \frac{2 j_0}{\Sigma_p} \tag{69}$$

(taking into account that the absorption current in the layer takes the maximum value).

The solution of this equation in our one-dimensional version is

$$\phi(x) = \left(\frac{j_0}{\Sigma_p} - \frac{\Omega B}{c} \right) x^2 + C_1 x + C_2 . \tag{70}$$

We define the boundary conditions for the potential equation in the following natural way. First, according to Eq. (66), the current density in space now cannot exceed j_0. As a result, the surface current I_x perpendicular to the edge of the plate must go to zero; i.e., the following condition must be satisfied:

$$\left(\hat{\Sigma} \left(\nabla \phi + \frac{\mathbf{v} \times \mathbf{B}}{c} \right) \right) \Bigg|_{x=0} = 0. \tag{71}$$

Hence, using the explicit expression for $\mathbf{v}$, we have

$$\left. \frac{\partial \phi}{\partial x} \right|_{x=0} = - \frac{\Sigma_h B_0 v_0}{\Sigma_p c} \cos\varphi - \frac{B_0}{c} (v_0 \sin\varphi - \Omega R). \tag{72}$$

The second condition is given at the inner limit x_0 of the boundary layer, where $\phi(x_0) = 0$. Because of the same bound on the spatial current, it is natural to assume that the x component of the surface current is continuous at this point. For this to be true, it is obviously necessary that

$$\phi(x_0) = 0, \quad \phi'(x_0) = 0, \tag{73}$$

where the prime denotes differentiation with respect to x.

The three boundary conditions of Eqs. (72) and (73) make it possible to determine the constants C_1 and C_2 in the expression for the potential and the position of the edge of the boundary layer:

$$C_1 = \frac{B_0 \Omega R \Sigma_p}{c j_0} - \frac{B_0 v_0}{c j_0} (\Sigma_p \sin\varphi + \Sigma_h \cos\varphi), \tag{74}$$

$$C_2 = \left(\frac{C_1^2}{4} \right) \left(\frac{j_0}{\Sigma_p} - \frac{\Omega B_0}{c} \right), \tag{75}$$

$$x_0 = - \frac{B_0 \Omega R \Sigma_p}{c j_0} + \frac{B_0 v_0}{c j_0} (\Sigma_p \sin\varphi + \Sigma_h \cos\varphi). \tag{76}$$

The expression for the potential can then be rewritten as

$$\phi(x) = \left(\frac{j_0}{\Sigma_p} - \frac{\Omega B}{c} \right) (x - x_0)^2 . \tag{77}$$

Taking this into account, we obtain for the surface current in the coordinate system of the edge of the plate that

$$I_x = -2 \left(j_0 - \frac{\Omega B \Sigma_p}{c} \right) (R - r - x_0(\varphi)) + \frac{vB}{c} (\Sigma_p \sin\varphi + \Sigma_h \cos\varphi) - \frac{\Sigma_p B_0}{c} \Omega R, \tag{78}$$

$$I_y = 2 \left(j_0 \frac{\Sigma_h}{\Sigma_p} - \frac{\Omega B \Sigma_h}{c} \right) (R - r - x_0(\varphi)) + \frac{vB}{c} (\Sigma_p \cos\varphi - \Sigma_h \sin\varphi) + \frac{\Sigma_h B_0}{c} \Omega R. \tag{79}$$

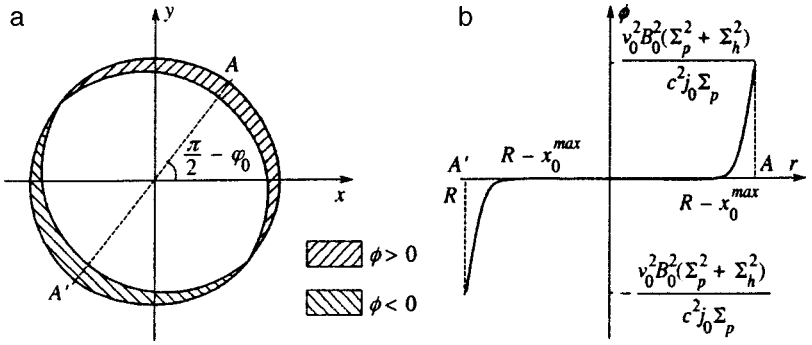

FIG. 4. Potential distribution on the plate ifor $\Omega = 0$: (a) position of the transition layers, (b) dependence of the potential on the distance to the edge of the plate.

In the region of negative potential, the thickness of the layer behaves like

$$x_0 = \frac{B_0 v_0}{c j_0}(\Sigma_p \sin\varphi + \Sigma_h \cos\varphi) + \frac{B_0 \Omega R \Sigma_p}{c j_0}, \tag{80}$$

and the potential itself varies inside the layer as

$$\phi(x) = -\left(\frac{j_0}{\Sigma_p} + \frac{\Omega B}{c}\right)(x - x_0)^2. \tag{81}$$

It is now easy to generalize these results to the case of a plate of arbitrary shape. To do this, it is simply necessary to take the same formulas for the potential ϕ and the layer thickness x_0^{Ω}, where these quantities are now considered dependent on φ as the angle that determines the position of the point under consideration at the edge of the plate. The boundary layer itself will now have a variable width which depends on angle φ according to Eq. (76).

We first examine what will occur in our case of a circular plate in the absence of rotation. When $\Omega = 0$ holds, the solution takes the form

$$\phi(x) = \frac{j_0}{\Sigma_p}(x - x_0)^2. \tag{82}$$

The surface current in the boundary layer in this case equals

$$I = -\begin{pmatrix} \Sigma_p & \Sigma_h \\ -\Sigma_h & \Sigma_p \end{pmatrix} \begin{pmatrix} \frac{2 j_0}{\Sigma_p}(x - x_0) - \frac{v B}{c}\sin\varphi \\ -\frac{v B}{c}\cos\varphi \end{pmatrix}$$

$$= -\begin{pmatrix} 2 j_0 (x - x_0) - \frac{v B}{c}(\Sigma_p \sin\varphi + \Sigma_h \cos\varphi) \\ -2 j_0 \frac{\Sigma_h}{\Sigma_p}(x - x_0) - \frac{v B}{c}(\Sigma_p \cos\varphi - \Sigma_h \sin\varphi) \end{pmatrix}. \tag{83}$$

The boundary layer has a thickness that depends on φ according to the law

$$x_0 = \frac{B_0 v_0}{c j_0}(\Sigma_p \sin\varphi + \Sigma_h \cos\varphi). \tag{84}$$

Its maximum thickness

$$x_0^{\max} = \frac{v B}{c j_0}\sqrt{\Sigma_p^2 + \Sigma_h^2}$$

is attained when $\varphi_{\max} = \pi/2 - \arctan(\Sigma_h/\Sigma_p)$. However, if the angle equals $-\varphi_0 = -\arctan(\Sigma_h/\Sigma_p)$ or $\pi - \varphi_0$, the thickness of the boundary layer goes to zero. The thickness $x_0^{\max}$ must be small by comparison with R, and consequently

$$v_0 B_0 \sqrt{\Sigma_p^2 + \Sigma_h^2}/c j_0 R \ll 1; \tag{85}$$

i.e., inequality (68) is in fact satisfied. As already mentioned, the surface potential outside the layer (far from the edge of the plate) is very small in this case, and the absorption and emission currents in this region exactly cancel. In the region $(-\varphi_0, \pi - \varphi_0)$, we obtain a boundary layer with a positive potential where the absorption current dominates. In the interval $(\pi - \varphi_0, 2\pi - \varphi_0)$, a layer is formed with a negative potential and, accordingly, an emission current. Because of symmetry, the total currents flowing through these layers are equal to each other in magnitude and consequently completely compensate each other, thus closing the current circuit (see Fig. 4).

Let us examine what will occur when $\Omega \neq 0$ holds. Here we no longer have such clear symmetry as we had for $\Omega = 0$. The geometry of the boundary layer is substantially altered: the extents of the layers of the positive and negative potentials will be different. Specifically, the layer of positive potential now does not extend to π but is included within the angles

$$\left(-\varphi_0 + \arcsin\left(\frac{\Omega R}{v_0}\frac{\Sigma_p}{\sqrt{\Sigma_p^2 + \Sigma_h^2}}\right),\right.$$

$$\left.\pi - \phi_0 - \arcsin\left(\frac{\Omega R}{v_0}\frac{\Sigma_p}{\sqrt{\Sigma_p^2 + \Sigma_h^2}}\right)\right).$$

A layer of negative potential accordingly takes up the rest of the range of angles. The widths of the layers also change: the positive layer becomes thinner; the negative layer, conversely, widens. The maximum width of the former is now

$$x_0^{\max} = \frac{v B}{c j_0}\sqrt{\Sigma_p^2 + \Sigma_h^2} - \frac{B_0 \Omega R \Sigma_p}{c j_0}, \tag{86}$$

while that of the latter is

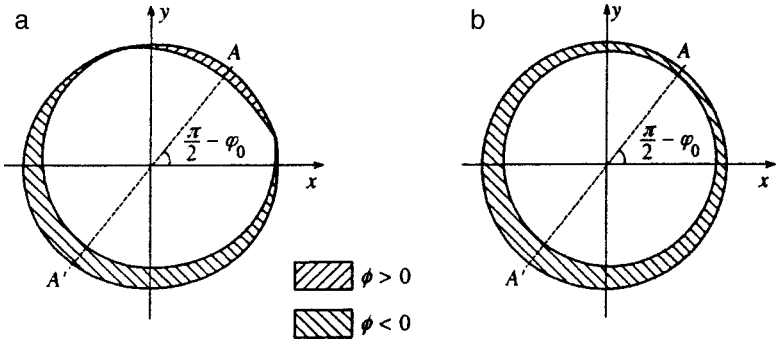

FIG. 5. Position of the transition layers on the surface of the plate for $\Omega\neq 0$: (a) the case $\Omega<\Omega^{cr}$, (b) the case $\Omega>\Omega^{cr}$.

$$x_0^{\max}=\frac{vB}{cj_0}\sqrt{\Sigma_p^2+\Sigma_h^2}+\frac{B_0\Omega R\Sigma_p}{cj_0}. \tag{87}$$

It can be seen that, as Ω approaches $\Omega^{cr}=v_0\sqrt{\Sigma_h^2+\Sigma_p^2}/R\Sigma_p$ in magnitude, the range of angles within which the layer of positive potential lies decreases, and the layer itself becomes narrower. At $\Omega=\Omega^{cr}$, this layer disappears altogether. The situation is reversed with the layer of negative potential—it broadens, increasing in the longitudinal direction. In the limit $\Omega\to\Omega^{cr}$, it occupies the entire edge of the plate and, as the angular velocity increases further, is a margin of variable width at the edge of the plate (see Fig. 5).

A question arises: How are the currents on the plate closed, since the excess emission current will now flow through the boundary layer without being balanced by an absorption current in the positive layer? To answer this question, it is necessary to turn to the region outside the boundary layer. When rotation is brought into consideration, we can no longer neglect the potential generated in this part of the plate. Specifically, we take Eq. (63) and, as was done earlier, we neglect the derivatives of the potential in the region far from the edge. Then we obtain

$$\frac{\Omega B}{c}=\frac{j_z}{\Sigma_p}. \tag{88}$$

Assuming that the potential difference in this region is small by comparison with T_e/e, we can expand the current j_z in powers of $e\phi/T$ in Eq. (88), which is the sum of the absorption current, Eq. (65), and the emission current, Eq. (67). We obtain as a result a constant positive potential

$$\phi=\frac{T_e}{e}\frac{\Sigma_p\Omega B_0}{cj_0} \tag{89}$$

(see Fig. 6). The solution given by Eq. (89) is valid only for sufficiently large currents j_0 in the plasma:

$$\Omega B_0\Sigma_p/cj_0\ll 1.$$

This potential creates an additional absorption current with density

$$j^+=j_0(1-\exp(e\phi/T_e))=B_0\Omega\Sigma_p/c. \tag{90}$$

This current, according to inequality (68), is much less than current j_0 in the boundary layer. On the other hand, however, it is collected from essentially the entire surface of the plate, whereas the boundary layer where current flows with a large density is small. This additional absorption current gives a total current equal to $\pi R^2B_0\Sigma_p\Omega/c$, whereas the excess of the emission current over the absorption current in the boundary layer is, as shown by an elementary calculation,

$$\Delta I=j_0(S^--S^+)=j_0\int_{S^-}x_0^-(\varphi)d\varphi-j_0\times\int_{S^+}x_0^+(\varphi)d\varphi=\frac{B_0\Omega\Sigma_p}{c}\pi R^2, \tag{91}$$

i.e., the same amount. This is how the current is closed on a body around which flow occurs as it rotates: the emission current flows only in a narrow band given by Eq. (76) close to the edge, while the absorption current that compensates it flows onto the entire surface.

3.3. The case of small emission and absorption currents

There is substantial interest in considering the limit of small emission and absorption currents or of a high rotational rate of a body when the relationship inverse to inequality (68) is satisfied:

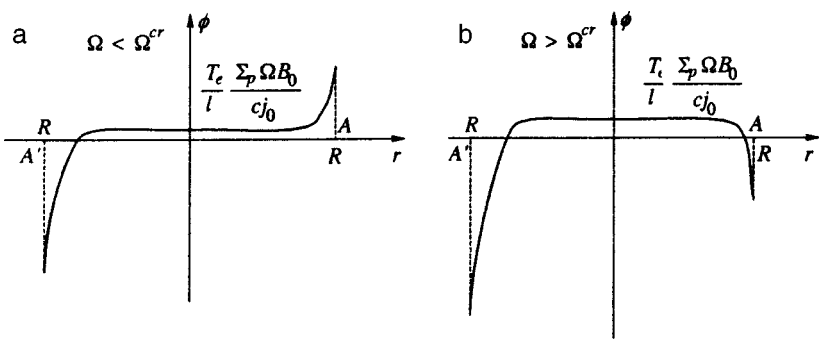

FIG. 6. Dependence of the potential on the distance to the edge of the plate for $\Omega\neq 0$: (a) the case $\Omega<\Omega^{cr}$, (b) the case $\Omega>\Omega^{cr}$.

$$\epsilon_1=\frac{\Sigma v_0 B_0}{c j_0 R}\gg 1, \quad \epsilon_2=\frac{\Sigma\Omega B_0}{c j_0}\gg 1. \tag{92}$$

Such a situation can occur in a disperse plasma or at low temperature when the thermal velocity of its particles does not give rise to large current densities.

If inequalities (92) are satisfied, we can no longer assume that the main current in our case is concentrated in the boundary layer, where the current density equals j_0. Specifically, as the angular velocity Ω increases, the picture of the potential distribution over the surface of the plate qualitatively changes approximately as follows: the edge layer of negative potential gradually broadens, while the positive potential in the internal regions of the plate linearly increases with increasing Ω. When this potential is comparable in magnitude with T_e/e (this occurs when parameter ϵ_2 becomes ~ 1), the current in the internal regions will behave in a nonlinear manner. In this case, as can easily be seen, the size of the boundary layer becomes about R, and our assumption concerning one-dimensionality breaks down (see Ref. 9).

When inequalities (92) are satisfied, however, it is comparatively simple to find the potential distribution over the surface. Specifically, we now must solve the complete Eq. (63) without assuming that it is one-dimensional. However, because of inequalities (92), the emission and absorption currents on its right-hand side can be neglected. Then the equation for the potential is rewritten in polar coordinates as

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\phi}{\partial r}\right)+\frac{1}{r^2}\frac{\partial^2\phi}{\partial\varphi^2}=-2\frac{\Omega B_0}{c}. \tag{93}$$

The boundary condition for Eq. (93), i.e., the vanishing of the normal component of the surface current at the edge of the plate, is now rewritten as

$$\Sigma_p\frac{\partial\phi}{\partial r}+\frac{\Sigma_h}{R}\frac{\partial\phi}{\partial\varphi}=\frac{B_0 v_0}{c}(\Sigma_p\sin\varphi+\Sigma_h\cos\varphi)-\frac{B_0\Sigma_p\Omega R}{c}. \tag{94}$$

This boundary-value problem is similar to a classical Neumann problem. Its solution is found in the form

$$\phi=\sum_{m=1}^{\infty}\left[\alpha_m(r)\cos(m\varphi)+\beta_m(r)\sin(m\varphi)\right]. \tag{95}$$

Substituting this into Eq. (93) and using the boundary condition of Eq. (94) and the condition that there is no singularity in the solution at $r=0$, we find nontrivial functions $\alpha_m\beta_m$ only for $m=0,1$. The complete solution is

$$\phi(r,\varphi)=-\frac{\Omega B_0}{2c}r^2+r\frac{B_0 v_0}{c}\sin\varphi+C, \tag{96}$$

where C is an arbitrary constant, to determine which one more boundary condition is necessary. It can be obtained by balancing the total current on the plate: since the absorption and emission currents virtually everywhere have a density j_0 (since $\phi\gg T/e$), it is necessary that the area occupied by the positive potential be equal to the area occupied by the negative potential. The constant C strongly depends on the ratio between ΩR and v_0. The expression for C involves a very unwieldy integration in the general case, but it is much simpler in the limiting cases: in the limit $\Omega R\ll v_0$ it obviously is simply equal to zero, but in the opposite case of $\Omega R\gg v_0$, we find that $C=\Omega B_0 R^2/4c$.

It is easy to physically interpret the resulting picture. Actually, if the inequalities (92) are valid, the current transmitted through the plate by the induction potential only is of the order of $E_{\rm ind}\Sigma R$, which is much larger than the current that can flow out of the plasma. In summary, the surface potential is aligned on the surface of the body so as to cancel the induction potential difference on the plate as much as possible and consequently to reduce the current through it to about $j_0 R^2$.

3.4. The field and currents in the transition layer

Now, after clarifying the details of the geometrical structure of the transition layer, we can proceed to find the currents and fields in this layer.

Using Eqs. (78) and (79) for the components of the surface current, we obtain the current component along the y axis in a Cartesian coordinate system with its origin at the center of the plate:

$$J_y=I_r\sin\varphi+I_\varphi\cos\varphi \tag{97}$$

$$=2(x-x_0^\Omega)\left(\frac{j_0}{\Sigma_p}-\frac{\Omega B}{c}\right)(\Sigma_p\sin\varphi-\Sigma_h\cos\varphi)$$
$$-\frac{Bv\Sigma_p}{c}-\frac{\Omega r B}{c}(\Sigma_p\sin\varphi-\Sigma_h\cos\varphi). \tag{98}$$

Note that the current J_y does not go to zero at the very edge of the plate, since, although $I_r=0$ holds at $r=R$, we have $I_\varphi\neq 0$. This causes a jump of b_x to appear in the field at the edge of the plate and, as a consequence, produces delta-function Alfvén currents in the plasma. An apparent paradox appears, since these currents must be closed by the normal component of the current flowing over the plate, but there is no such component. This contradiction is easy to resolve if we recall that, in the presence of a nonzero current I_φ, a fast magnetosonic wave will be generated in the space around the plate. It naturally also creates its own currents, which will just compensate the delta-function Alfvén currents at the boundary of the cylinders. As a result, the current balance is restored, and the apparent contradiction associated with it is removed—there are no delta-function currents in the total solution.

Substituting Eq. (98) into Eq. (15), we obtain the boundary condition for the field component b_x, after which it is easy to immediately write out the solution by analogy with the earlier solution:

$$b_x^+ = -\frac{2\pi v_0 \Sigma_p}{c^2 M_A} - \frac{\Omega R}{v_0}\frac{2\pi v_A}{c^2}\frac{\Sigma_p y - \Sigma_h x^+}{R}$$
$$+\delta_l\left(2\frac{j_0}{\Sigma_p}R\frac{\Sigma_p y-\Sigma_h x^+}{\sqrt{y^2+(x^+)^2}} - 2\frac{j_0}{\Sigma_p}(\Sigma_p y-\Sigma_h x^+)\right.$$
$$\left. -\frac{vB}{c\Sigma_p}\frac{\Sigma_p^2 y^2-\Sigma_h^2 (x^+)^2}{y^2+(x^+)^2} + \frac{\Omega R}{v_0}\frac{2\pi v_A}{c^2}\frac{\Sigma_p y-\Sigma_h x^+}{R}\right),$$

where δ_l equals unity in the layer and zero everywhere outside it.

Having the expression for the field, the spatial current density can be found in the layer neighborhoods of the Alfvén cylinders:

$$j_y^+ = -\Omega R\frac{\Sigma_h}{2cR} + \frac{\delta_l}{2B}\left[-2j_0\frac{x^+}{\Sigma_p R^2}(\Sigma_h x^+ - \Sigma_p y)\right.$$
$$-4\frac{v_0 B_0}{c\Sigma_p}\frac{y^2 x^+(\Sigma_p^2+\Sigma_h^2)}{R^4} - \frac{\Omega R}{v_0}\frac{v}{2c}$$
$$\left.\times\left(-2\frac{\Sigma_h}{R}+\frac{y(\Sigma_p x^+ + \Sigma_h y)}{R^3}\right)\right], \quad (99)$$

$$j_z^+ = \frac{\Omega R}{M_A}\frac{\Sigma_p}{2cR} + \frac{\delta_l}{2BM_A}\left[2j_0\left(1+\frac{x^+(\Sigma_h x^+ + \Sigma_p y)}{R^2\Sigma_p}\right)\right.$$
$$-4\frac{vB}{c\Sigma_p}\frac{y(x^+)^2(\Sigma_p^2+\Sigma_h^2)}{R^4} - \frac{\Omega R}{v_0}\frac{v_A}{2c}$$
$$\left.\times\left(2\frac{\Sigma_p}{R}-\frac{x^+(\Sigma_p x^+ + \Sigma_h y)}{R^3}\right)\right]. \quad (100)$$

All the formulas for the case $z<0$ are written in a similar way. It can be seen that the spatial current that flowed earlier along the surface of the cylinders now flows in a layer of finite thickness given by Eq. (76).

Similarly, by computing the x component of the current on the plate in terms of I_r and I_φ, the field and the currents in the fast magnetosonic wave can be found.

3.5. The forces and torques acting on the body

The currents that appear on a moving body interact with the external magnetic field, and this results in the appearance of forces acting on the body. The value of the forces is about JB/c, where J is the total current on the body, and substantially depends on the current density j_0 in the plasma. If j_0 is high, corresponding to ideal magnetohydrodynamics, J is determined by the induction mechanism of current generation on the body and depends only on its conductivity, the velocity of the body, and field B. If, conversely, inequalities (92) are satisfied, the current J depends only on j_0 and on the geometrical size of the body. We shall consider the former case here, in which current density j_0 is assumed to be large and inequality (68) is consequently satisfied.

To compute the forces and torques acting on a rotating body, it is sufficient to use the unperturbed values for the magnetic field and the speed of the plasma. The calculation is fairly simple when field $\mathbf{B}_0$ is arbitrarily oriented relative to the body. We therefore go to a coordinate system with its z axis normal to the plate, and assume that field $\mathbf{B}_0$ has all three components B_x, B_y, B_z.

In this case, as shown by a calculation carried out similarly to Ref. 2, the total force is made up of a deceleration force, a shearing force, and a buoyancy force and equals

$$\mathbf{F} = \frac{v_0 B_z \pi R^2}{c^2}\begin{pmatrix} -\Sigma_p B_z \\ -\Sigma_h B_z \\ B_y\Sigma_h - B_x\Sigma_p \end{pmatrix}. \quad (101)$$

Comparison with the results obtained in Ref. 2 for purely translational motion shows that rotation does not contribute to the forces acting on the body. It turns out, however, that it produces a torque that did not exist in the case of translational motion. The torque equals

$$\mathbf{M} = \int_S \frac{1}{c}[\mathbf{r}\times(\mathbf{I}\times\mathbf{B}_0)]dS = -\int_S \frac{1}{c^2}\{\mathbf{r}\times[\hat{\Sigma}(\mathbf{v}_0\times\mathbf{B}_0)\mathbf{B}_0]\}dS$$
$$= \frac{\Omega B_z \pi R^4}{4c^2}\begin{pmatrix} \Sigma_h B_y - \Sigma_p B_x \\ -\Sigma_h B_z - \Sigma_p B_y \\ -2B_z\Sigma_p \end{pmatrix}. \quad (102)$$

If we consider the various components of the torque separately, the following picture appears: The torque M_z decelerates the rotation of the body according to the law

$$\Omega = \Omega_0 \exp\left(-\frac{B_z^2\Sigma_p \pi R^4}{2I_z c^2}t\right), \quad (103)$$

where I_z is the moment of inertia of the body around the axis coinciding with the normal.

The terms in M_x and M_y proportional to Σ_p tend to rotate the plate so that its normal becomes parallel to the magnetic field, with the angle between the field direction and the normal decreasing proportionally to $\exp(-B_z^2\pi R^4\Sigma_p t/4c^2 I_z)$. It can be seen that the time constant of this process is essentially the same as for the deceleration of the rotation, which means that, generally speaking, these two processes in fact occur together.

Finally, the terms in M_s, M_y that are proportional to Σ_h cause the plate to precess relative to the magnetic-field direction. The nature of this precession is easy to understand. In fact, a nonzero magnetic moment is generated by the surface current on the plate. A calculation using Eqs. (25) and (26) gives

$$\mathbf{m} = \frac{\pi R^4 \Sigma_h B_z}{4c^2 I_z}\mathbf{L} \quad (104)$$

for this moment, where $\mathbf{L}$ is the angular momentum of the plate around the normal. The precession of this magnetic moment in field $\mathbf{B}$ also results in precession of the plate

$$\frac{d\mathbf{L}}{dt} = \mathbf{m}\times\mathbf{B} = \mathbf{\Omega}_{pr}\times\mathbf{L}, \quad \mathbf{\Omega}_{pr} = -\frac{\pi\Sigma_h B_z R^4}{4c^2 I_z}\mathbf{B}. \quad (105)$$

Thus, we have found that the resulting motion of the plate under the action of the torques associated with the inductive forces caused by rotation is a superposition of precession, deceleration of the main rotation, and alignment of the plate perpendicular to the field. The variation of the torque considered here is qualitatively similar to that appearing in the magnetosphere of pulsars.[13]

This work was carried out with the partial financial support of the Russian Fund for Fundamental Research (Grant No. 96-02-18217) and the Soros Foundation (Grant No. s98-1545).

*)E-mail: alex@td.lpi.ac.ru

[1] F. D. Drell, H. M. Foley, and A. M. Ruderman, J. Geophys. Res. **70**, 3131 (1965).
[2] A. V. Gurevich, A. L. Krylov, and E. N. Fedorov, Zh. Éksp. Teor. Fiz. **75**, 2132 (1978) [Sov. Phys. JETP **48**, 1074 (1978)].
[3] L. A. Frank *et al.*, Science **274**, 394 (1996).
[4] M. G. Kivelson *et al.*, Science **274**, 396 (1996).
[5] D. J. Williams *et al.*, Science **274**, 401 (1996).
[6] F. M. Neubauer, J. Geophys. Res. **85**, 1171 (1980).
[7] D. J. Southwood, M. G. Kivelson, R. J. Walker, and J. A. Slavin, J. Geophys. Res. **85**, 5959 (1980).
[8] K. Papadopoulos, C.-L. Chang, and A. Drobot, Geophys. Res. Lett. **25**, 745 (1998).
[9] A. V. Gurevich and N. T. Pashchenko, Geomagn. Aeron. **21**, 225 (1981).
[10] A. V. Gurevich and N. T. Pashchenko, Acta Astronautica **10**(2), 91 (1982).
[11] Ya. L. Al'pert, A. V. Gurevich, and L. P. Pitaevskiĭ, *Space Physics with Artificial Satellites*, Consultants Bureau, New York (1965) [Russian original, Nauka, Moscow (1964)].
[12] A. V. Gurevich, A. L. Krylov, and E. E. Tsedilina, Space Sci. Rev. **19**, 59 (1976).
[13] V. S. Beskin, A. V. Gurevich, and Ya. N. Istomin, *Physics of the Pulsar Magnetosphere*, Cambridge University Press (1993).

Translated by W. J. Manthey

Self-preservation of large-scale structures in a nonlinear viscous medium described by the Burgers equation

S. N. Gurbatov*) and G. V. Pasmanik

N. I. Lobachevskiĭ Nizhniĭ Novgorod State University, 603000 Nizhniĭ Novgorod, Russia
(Submitted 22 May 1998)
Zh. Éksp. Teor. Fiz. **115**, 564–583 (February 1999)

We use the asymptotic solution of the one-dimensional Burgers equation to study the self-preservation of large-scale random structures. We show that in the process of their evolution, large-scale structures remain stable against small-scale perturbations for the case of a continuous initial spectrum with a spectral index smaller than unity. We study both analytically and numerically the correlation coefficient of a large-scale structure and of the same structure with a high-frequency perturbation and show that with the passage of time the coefficient tends to unity. Using the asymptotic formulas of the theory of random excursion of stochastic processes, we study the statistical properties of the perturbing field and find that the effect of high-frequency perturbations is equivalent to the introduction of effective viscosity. © *1999 American Institute of Physics.* [S1063-7761(99)01302-5]

1. INTRODUCTION

The nonlinear diffusion equation, first proposed by Burgers[1,2] as a model of hydrodynamic turbulence, plays an important role in studies of the evolution of nonlinear waves in nondispersive media. Indeed, this equation describes two main effects inherent in turbulence: the nonlinear redistribution of energy in the spectrum, and the decrease in energy in the region of small-scale components of the spectrum. When the initial conditions or the external forces are random, this equation describes turbulence of a gas without pressure,[3] commonly known as Burgers turbulence.

What makes the Burgers equation so remarkable is that it has an exact analytical solution.[4,5] This leads to several rigorous statistical results (see, e.g., Refs. 2,6–19), which may serve as tests for verifying approximate method used in studies of hydrodynamic turbulence. Later it was found that the Burgers equation also describes a broad class of nonlinear effects in wave propagation theory, plasma physics, and acoustics.[9,20,21] The Burgers equation and its multidimensional vector generalization arise in the problem of surface growth.[22–25] In this case the vector field satisfying the given equation describes the gradient of a growing surface. The three-dimensional Burgers equation also forms the basis of the adhesion model, which is used to describe the formation of the large-scale structure of the universe.[9,26–29]

A characteristic feature of solutions of the Burgers equation for infinite Reynolds numbers is the existence of local and statistical self-similarity.[9] The term ''local self-similarity'' means that because of the combined action of nonlinearity and viscosity, each field realization exhibits universal behavior: over large time intervals the field becomes a sequence of triangular pulses with the same slope $v_x' = 1/t$. When the initial conditions are periodic, the relative discontinuity velocities are zero, with the result that the periodicity of the field is preserved. Here, because of the combined action of nonlinearity and viscosity at the discontinuities, the information about the amplitude and shape of the initial perturbation is completely lost. When the initial field is noise, the coordinates and velocities of the discontinuities are random quantities, which leads to merging of the shock fronts.

Due to multiple merging of the discontinuities, over long enough time intervals the statistical characteristics of turbulence also become self-similar and are determined by a single scale $l(t)$, the integral turbulence scale.[6,9] In the region of large wave numbers of the signal energy spectrum a universal power-law asymptotic behavior $E(k,t) \propto k^{-2}$ related to discontinuity formation sets in. Here the law of growth of $l(t)$ owing to discontinuity merging, and the shape of the probability distributions and turbulence spectra strongly depend on the behavior of the initial spectrum $E_0(k) \propto k^n$ in the region of small wave numbers k. For $n > 2$ the parametric generation effect gives rise to a universal low-frequency asymptotic behavior of the spectrum, and the law of growth of $l(t)$ is very sensitive to the type of statistics of the initial perturbation. In particular, when the statistics of the initial perturbation is Gaussian, the external scale $l(t)$ increases in proportion to $t^{1/2}$ with a certain logarithmic correction.[6,9] Here $l(t)$ is determined only by the integral characteristics of the initial spectrum. For $1 < n < 2$ the low-frequency part of the spectrum is ''frozen,'' $E(k,t) = E_0(k) \propto k^n$, but the larger fraction of the energy is concentrated in the frequency range $k > k_s(t)$, and the external scale $l(t)$ is proportional to $t^{1/2}$, as it is in the case when $n > 2$ (see Ref. 15). When $n < 1$, the low-frequency part of the spectrum is preserved, and it is this part that determines the law of growth of the external scale $l(t) \propto t^{2/(n+3)}$. Here the law of growth of $l(t)$ and the shape of the spectrum do not depend on the behavior of the high-frequency part of the spectrum.

In the present paper we study the stability of the evolution of nonlinear structures $v(x,t)$ generated by an initial field $v_0(x)$ with respect to external perturbations $v_h(x)$. When $v_0(x)$ is a periodic signal, random perturbations lead

to fluctuations in the velocity of the discontinuities, discontinuities merge and, as a result, total destruction of the periodic structure occurs.[30] When the initial field $v_0(x)$ is noise, the highly nonlinear structures continuously interact and, due to the merging of discontinuities, their characteristic scale $l(t)$ constantly increases. The presence of a noise perturbation $v_h(x)$ results in additional fluctuations $\Delta x_k(t)$ in the discontinuity coordinates, and these fluctuations increase in strength with the passage of time.

Thus, the final result of the evolution of the field is determined by the competition of two factors, the increase in the external scale $l(t)$ of the structures and the increase in the strength $\Delta x_k(t)$ of discontinuity fluctuations, the latter being related to the perturbation $v_h(x)$. If the condition $|\Delta x(t)| \ll l(t)$ is met for all values of t, such structures remain stable against random perturbations.

We then examine the stability of large-scale structures against small-scale perturbations that have the same spectrum $E_0(k) \propto k^n$ but are located in the high-frequency part of the spectrum. Preliminary estimates made by Aurell *et al.*[31] have shown that for $n<1$ large-scale structures remain stable against small-scale perturbations. In this paper we calculate, both analytically and numerically, the coefficient $r(t)$ of the correlation of the unperturbed and perturbed fields. This coefficient is the quantitative measure of closeness of the given structures. We show that, for $n<1$, even in the case where the initial correlation coefficient r_0 is much smaller than unity and the perturbation $v_h(x)$ substantially exceeds in amplitude the field $v_0(x)$, i.e., $|v_h| \gg |v_0|$, the evolution of nonlinear structures leads to an increase in the correlation coefficient $r(t)$, and asymptotically $r(t)$ tends to unity. As is known, the interaction of the low- and high-frequency components can be interpreted as the manifestation of turbulent viscosity.[32] In our research we establish that the asymptotic stability of large-scale structures is equivalent to the effect of the increase in Reynolds numbers in a medium with finite viscosity.[9]

2. THE BASIC PROPERTIES OF BURGERS TURBULENCE

The nonlinear diffusion equation, i.e., the Burgers equation

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} = \mu \frac{\partial^2 v}{\partial x^2}, \quad v(x,t=0) = v_0(x), \tag{1}$$

reduces to the linear diffusion equation (and hence has an exact solution) if we use the Hopf–Cole change-of-variables procedure.[4,5]

However, statistical analysis is difficult if we use the exact equation, especially in the case of large Reynolds numbers, which corresponds to letting μ approach zero. On the other hand, in the case of vanishingly small viscosity, the solution of the Burgers equation has the asymptotic representation[2,4,9]

$$v(x,t) = \frac{x - y(x,t)}{t}, \quad v(x,t) = -\frac{\partial S(x,t)}{\partial x}, \tag{2}$$

$$G(x,y,t) = S_0(y) - \frac{(x-y)^2}{2t}, \quad S_0(y) = -\int^y v_0(x)\,dx, \tag{3}$$

where $y(x,t)$ is the coordinate of the absolute maximum of the function $G(x,y,t)$ in y, which is known as the Lagrangian coordinate, and $S(x,t) = G(x,y(x,t),t)$ is the Eulerian potential of the velocity field.[9] It is on the basis of this solution that the majority of the statistical results concerning Burgers turbulence were obtained.

Let us introduce the characteristic time for the manifestation of nonlinear effects, $t_n = l_0/\sigma_0$, where σ_0 and l_0 are, respectively, the characteristic amplitude and the spatial scale of the initial velocity field. For times $t \gg t_n$ a global maximum in y of the function $G(x,y,t)$ lies within a small neighborhood of a local maximum of the initial potential $S_0(y)$, i.e., the zero of the initial field $v_0(x)$. Thus, in the limit $\mu \to 0$ the Lagrangian coordinate $y(x,t)$ is a step function of x, with the result that the velocity field function $v(x,t)$ behaves in the universal manner

$$v(x,t) = \frac{x - y_k}{t}, \quad x_{k-1} < x < x_k \tag{4}$$

in each cell between the discontinuities. The positions of the discontinuities are determined by the condition that two absolute maxima are equal, $G(x_k, y_{k-1}, t) = G(x_k, y_k, t)$, at point x_k:

$$x_k = \frac{y_{k+1} + y_k}{2} + V_k t, \quad V_k = \frac{S_0(y_k) - S_0(y_{k+1})}{y_{k+1} - y_k}. \tag{5}$$

If the initial conditions are specified in the form of a random field, the velocities of the shock fronts formed in the process of evolution of the field are also random quantities. This implies that the discontinuities merge, which means that the characteristic distance between them increases and, respectively, the number of such discontinuities decreases. We call this characteristic distance $l(t)$ the external turbulence scale, which can be estimated by Eq. (3). The spatial interval $|x-y|$ where a global maximum of $G(x,y,t)$ may be located is approximately $l(t)$ long and is determined by the condition that the increments of the initial potential and of the parabola are of the same order:

$$|\Delta S_0| \simeq |S_0(y+l) - S_0(y)| \simeq \sqrt{d_s(l)} = \frac{l^2}{t}, \tag{6}$$

where $d_s(z)$ is the structure function of the initial potential, $d_s(\rho) = \langle (S_0(x+\rho) - S_0(x))^2 \rangle$. Equation (6) implies that the asymptotic law of the increase of the external turbulence scale is determined by the behavior of the structure function at large distances, $d_s(\rho \gg l_0)$. We postulate that the energy spectrum of the initial perturbation of the velocity has a power-function asymptotic behavior in the region of small wave numbers:

$$E_v(k) = \alpha_n^2 k^n b_0(k), \quad b_0(0) = 1, \quad b_0(k > k_0) = 0, \tag{7}$$

where $b_0(k)$ is the high-frequency cutoff factor.

Depending on the value of the index n, the structure function exhibits the following asymptotic behavior:[9]

$$d_s(\rho)\simeq\begin{cases}\beta_n^2\rho^{1-n}, & \rho\gg l_0, \quad n<1,\\ 2\sigma_s^2, & \rho\gg l_0, \quad n>1,\end{cases} \tag{8}$$

$$\beta_n^2=\frac{2\pi\alpha_n^2}{\Gamma(n-2)\sin(\pi(1-n)/2)}, \quad \sigma_s^2=\langle S_0^2\rangle.$$

Then Eqs. (6) and (8) yield the following estimate of the external turbulence scale:

$$k_c^{-1}=l(t)\simeq\begin{cases}(\alpha_n t)^{2/(n+3)}, & n<1,\\ (\sigma_s t)^{1/2}, & n>1,\end{cases} \tag{9}$$

where $k_c(t)$ is a characteristic spatial frequency near which there is a crossover of the energy spectrum $E_v(k,t)\propto k^n$ to the next section $E_v(k,t)\propto k^{-2}$. Here the turbulence energy $\sigma^2(t)=\langle v^2(x,t)\rangle$ decreases according to the law

$$\sigma^2(t)=\frac{l^2}{t^2}\propto\alpha_n^2 k_c(t)\simeq\begin{cases}\alpha_n^{4/(n+3)}t^{-(2n+2)/(n+3)}, & n>1,\\ \sigma_s^2/t^{-1}, & n<1.\end{cases} \tag{10}$$

Thus, the properties of the solution of the Burgers equation are different for different values of n. For $n>1$ the increase in the external turbulence scale is determined by certain integral characteristics of the spectrum of the initial potential, $E_S(k)$. As shown in Refs. 7 and 9, in this case the statistical properties of the Burgers turbulence are self-similar for $n>2$ and $t\gg t_n$, and the energy spectrum exhibits universal behavior in the region of small ($E_v(k,t)\propto l/k^2t^2$) and large ($E_v(k,t)\propto l^4k^2/t^2$) wave numbers. For $n<1$ it can also be shown that the statistical properties of turbulence become self-similar irrespective of the value of the initial Reynolds number.[9,33] The increase in the external scale is due to the local behavior of the initial spectrum in the region of large values of k, and the energy spectrum is preserved in the interval $k<k_c(t)$.

We assume that not only is the low-frequency part of the spectrum preserved, but the large-scale components of field realizations also remain stable against a high-frequency perturbation. A simple model of Burgers turbulence for the case where $n<1$ was proposed in Refs. 33 and 34. The initial perturbation was represented by an infinite sum of the harmonics $k_m=k_0\varepsilon^{-m}$ (the Weierstrass spectrum) with very sparse frequencies ($\varepsilon\gg1$). The main assumption was that the energy of each component decreases independently from the other component energies and that the total turbulence energy is approximately equal to the sum of the energies of all the components. It was found that the external turbulence scale and energy in the given discrete model are described approximately by the same power function as for the continuous-noise spectrum and are independent of the number of high-frequency modes in the initial spectrum.

The objective of the present work is to show that the assumption that high-frequency components of the spectrum have a small effect on the low-frequency components is also true for the continuous spectrum.

3. PRESERVATION OF THE LARGE-SCALE STRUCTURE OF A RANDOM FIELD: A NUMERICAL EXPERIMENT

Let us consider the evolution of two random initial perturbations $v_0(x)$ and $\tilde{v}_0(x)$:

$$\tilde{v}_0(x)=v_0(x)+v_h(x). \tag{11}$$

We assume that the energy spectra of both processes are described by Eq. (7) but that for the process $v_0(x)$ the cutoff frequency is k_*, while for the process $\tilde{v}_0(x)$ the cutoff frequency is $k_0>k_*$. This means that the initial velocities $v_0(x)$ and $\tilde{v}_0(x)$ have a common part of their Fourier transforms in the region $k\in[0,k_*]$ and that the process $\tilde{v}_0(x)$ contains an additional high-frequency perturbation $v_h(x)$ whose spectrum is in the frequency interval $k\in[k_*,k_0]$. In this paper we employ an algorithm based on the fast Legendre transform to solve the one-dimensional Burgers equation numerically. Using the fact that $y(x,t)$ is a nondecreasing function and the definition of the fast Legendre transform, $\varphi(x)=\max_a[\Phi(a,0)+xa]$, we can construct the solution of the Burgers equation in $O(N\log_2 N)$ steps, where N is the number of points of the discrete grid on which the initial conditions are specified.[35] Figure 1 shows the realizations of an unperturbed field (a) and two cases (b) and (c) of a perturbed field for different values of the cutoff wave number $k_0\gg k_*$. Clearly, in both cases the variance of the unperturbed field is much smaller than the variances of the perturbed fields. However, for $t\gg t_n$, when the fields have well-developed discontinuities, the realizations of the initial large-scale field and those of the perturbed fields differ little from each other, which supports the hypothesis that large-scale structures remain stable against small-scale perturbations (Figs. 2 and 3).

To prove this assumption, we investigated both analytically and numerically the perturbed–unperturbed field correlation coefficient

$$r(t)=\frac{B(t)}{\sigma(t)\tilde{\sigma}(t)}, \quad B(t)=\langle v(x,t)\tilde{v}(x,t)\rangle,$$
$$\sigma^2(t)=\langle v^2(x,t)\rangle, \quad \tilde{\sigma}^2(t)=\langle\tilde{v}^2(x,t)\rangle, \tag{12}$$

the quantitative characteristic of the similarity of these processes, which have a common low-frequency part of the spectrum at input. For initial perturbations with an energy spectrum (7), the initial correlation coefficient is

$$r(t=0)=r_0=\left(k_*/k_0\right)^{(n+1)/2}, \tag{13}$$

which means that for $k_0\gg k_*$ the correlation is weak.

The characteristic time for the nonlinear effects of a signal with an initial spectrum of the form (7) and a cutoff frequency k_* to manifest themselves is

$$t_n=\frac{1}{k_*\sigma_0}\simeq[\alpha_n k_*^{(n+3)/2}]^{-1}.$$

For $t<t_n$ the nonlinear distortions of the wave are small and $v(x,t)\simeq v_0(x)$, while for $t\gg t_n$ the wave becomes a sequence of triangular pulses with equal slopes. The power spectrum $E_v(k,t)$ at this stage becomes self-similar,[9]

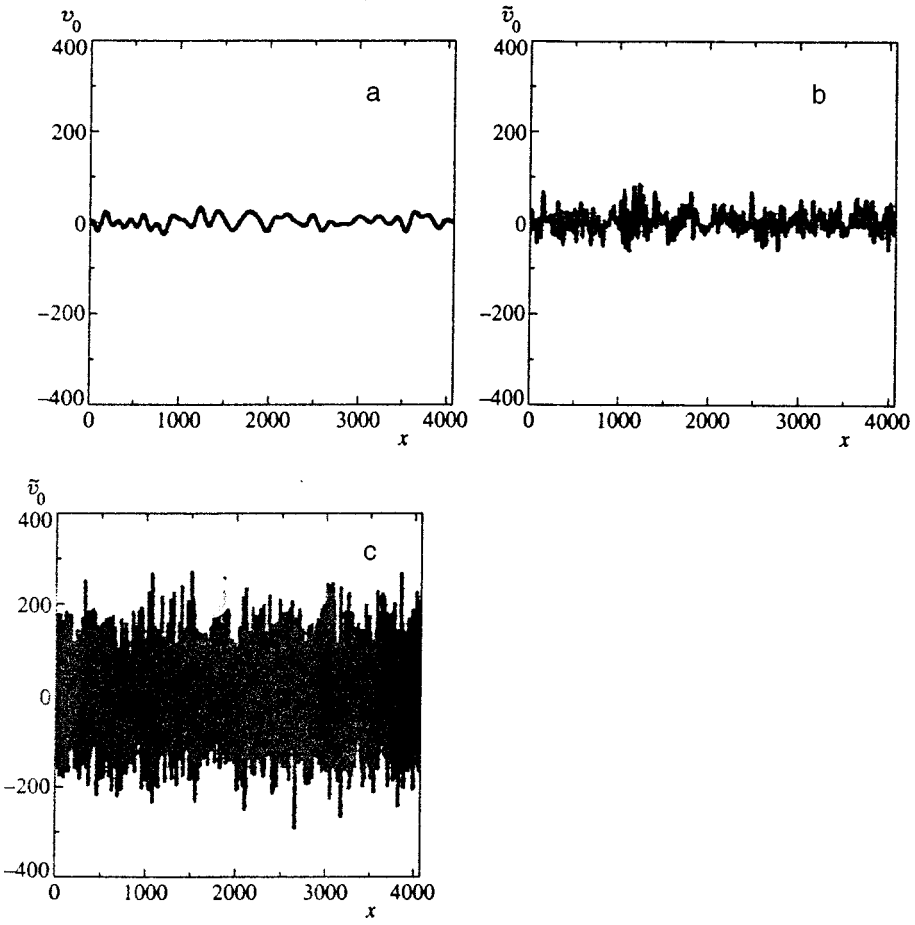

FIG. 1. The process $v_0(x)$ with a cutoff wave number $k_*=32$ (a); the process $\tilde{v}_0(x)$ with a cutoff wave numbers $k_0=128$ and $k_0=2048$ [(b) and (c), respectively] at time $t=0$.

$$E_v(k,t)=\frac{l^3}{t^2}\tilde{E}(k\,l)$$

and reproduces the initial spectrum $E_v(k)$ in the spatial frequency range $k<k_c(t)\simeq(\alpha_n t)^{-2/(n+3)}$, with the energy decreasing according to the law (10), $\sigma^2(t)\simeq\alpha^2 k_c^{n+1}(t)$. Figure 4 shows the energy spectra of the processes $v(x,t)$ and $\tilde{v}(x,t)$ at $t=0$ and $t_1\gg t_n$ for white noise, with the cutoff frequencies being k_* and $k_0=8k_*$, respectively. Clearly, for $t\gg t_n$ the spectra of both signals are almost the same and have a high-frequency asymptotic behavior $E_v(k,t)\propto k^{-2}$. The fact that the spectra of the perturbed and unperturbed signals are almost the same at the stage of well-developed discontinuities is one confirmation of the assumption that high-frequency components have a small effect on the low-frequency components.

Three stages can be distinguished in the behavior of the correlation coefficient (12). For $t<\tilde{t}_n=[\alpha_n k_0^{(n+3)/2}]^{-1}$, the nonlinear distortions of both signals are small and $r(t)\simeq r_0$. In the interval $\tilde{t}_n<t<t_n$ only the distortions of field $\tilde{v}(x,t)$ are significant. If we assume that the low-frequency part of the signal $\tilde{v}(x,t)$ remains almost the same, we have

$$\langle v(x,t)\tilde{v}(x,t)\rangle\simeq\langle v_0^2(x)\rangle=\sigma_0^2, \tag{14}$$

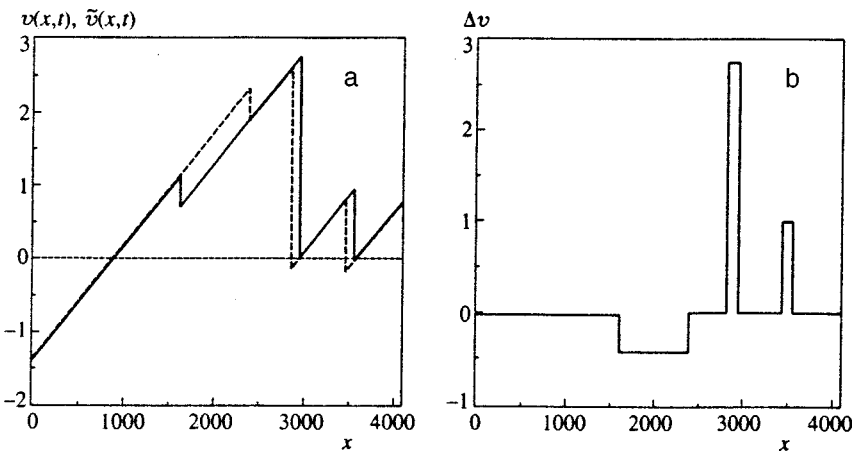

FIG. 2. Evolution of the process $v(x,t)$ with a cutoff wave number $k_*=32$ (dashed curve) and that of the process $\tilde{v}(x,t)$ with a cutoff wave number $k_0=128$ (solid curve) (a) and their difference Δv (b) at time $t=0.125$ ($k_c=4$).

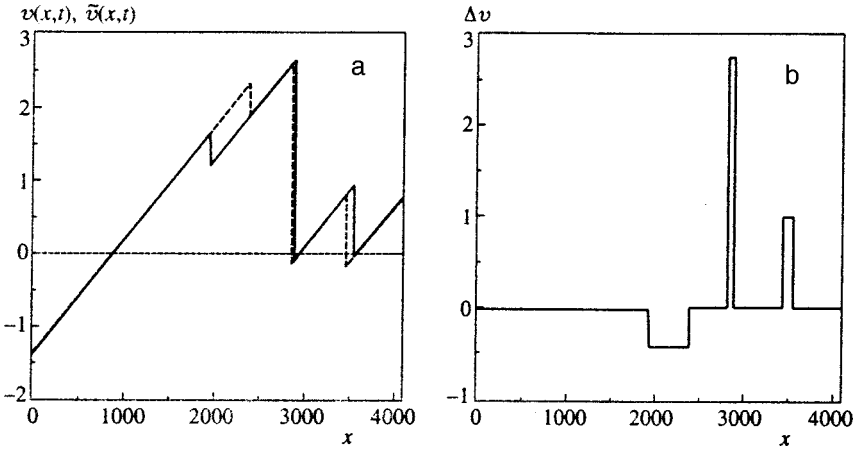

FIG. 3. The same as in Fig. 2 with the parameter values $k_*=32$ and $k_0=2048$ (a) and (b).

and the correlation coefficient (12) can be estimated in the following way:

$$r(t)\simeq\frac{\sigma_0^2}{\sigma_0\tilde{\sigma}(t)}\simeq\left(\frac{k_*}{k_c(t)}\right)^{(n+1)/2}\simeq(k_*l(t))^{(n+1)/2}\propto t^{(n+1)/(n+3)}. \tag{15}$$

The results of the numerical experiments confirming this assumption are illustrated by Fig. 5. We show the dependence of the correlation coefficient r on the ratio k_0/k_* of spatial frequencies for two values of time, t_1 and $t_2<t_1[k_c(t_2)=4k_c(t_1)]$. It is assumed that for t_1 and t_2 the condition $k_*\ll k_c(t)\ll k_0$ is met, i.e., we can ignore the nonlinear distortions of the initial perturbation of $v_0(x)$. In the initial section of the diagram ($k_0\ll k_c(t)$), where we can ignore the nonlinear distortions of both initial fields, the correlation coefficient decreases with increasing k_0 and is described by the expression (13), $r(t)\simeq r_0\propto k_0^{1/2}$ (Fig. 5b clearly shows this). For $k_*\ll k_c(t)\ll k_0$ only the perturbed field is distorted, and the nonlinear decay of high-frequency components reduces the variance of the perturbed signal. Figure 5 shows that under the given conditions the correlation coefficient reaches a plateau and becomes independent of k_0 [see Eq. (15)], which confirms the hypothesis that the low-frequency part of the field does not change [see Eq. (14)]. The results of the numerical experiment suggest that in the asymptotic region $r(t_1)/r(t_2)=1.97$ holds, in good agreement with the theoretical prediction that $r(t_1)/r(t_2)=[k_c(t_2)/k_c(t_1)]^{1/2}=2$.

In the last stage, where $t\gg\tilde{t}_n$ holds, both signals become sequences of the triangular pulse (4) and nonlinear absorption of the low-frequency part of the initial spectrum occurs. The results of numerical experiments designed to calculate the correlation coefficient in this stage are shown in Fig. 6. In this case, as k_0 increases, the correlation coefficient also reaches a plateau, which is close to unity.

Figure 7 depicts the behavior of the correlation coefficient in time for fixed values of the parameters k_* and k_0. Clearly, as time passes the correlation coefficient increases and tends to a value close to unity. The reason is that although the positions of the shock fronts of the perturbed and unperturbed fields are close, they do not coincide because of the presence of a high-frequency perturbation of the initial potential of the process $\tilde{v}_0(x)$. Below we present the analytical results and confirm the computer-modeling data both qualitatively and quantitatively.

4. STABILITY OF LARGE-SCALE STRUCTURES AGAINST SMALL PERTURBATIONS

We consider the behavior of realizations in the limit $t\gg t_n$ by employing the asymptotic solution of the Burgers equation [see Eqs. (1)–(3)]. In accordance with this solution, $y(x,t)$ in (2) is the coordinate of the absolute maximum of the function $G(x,y,t)$ (Eq. (3)). The curvature of the parabola in (3) is $1/t$, and the curvature of the initial potential S_0 can be estimated at

$$\sigma_{v_0'}=\langle(S_0'')^2\rangle^{1/2}=\frac{\sqrt{2}\,\alpha_n k_*^{(n+3)/2}}{(n+3)^{1/2}}. \tag{16}$$

Thus, for $t\gg t_n\simeq 1/\sigma_{v_0'}$ the parabola in (3) is a smooth function on the scale of the initial potential. This means that a global maximum of the function $G(x,y,t)$ lies within a small neighborhood of a local maximum of the initial potential

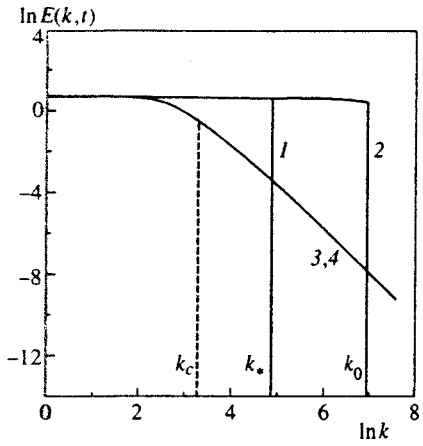

FIG. 4. Energy spectra of the process $v_0(x)$ with a cutoff wave number $k_*=128$ (curve *1*) and the process $\tilde{v}_0(x)$ with a cutoff wave number $k_0=1024$ (curve *2*) at time $t=0$, and their evolution at time $t=0.006\simeq 10t_n$ with $k_c=32$ (curves *3* and *4*) on a log–log scale at $n=0$ (averaging is over 1000 realizations).

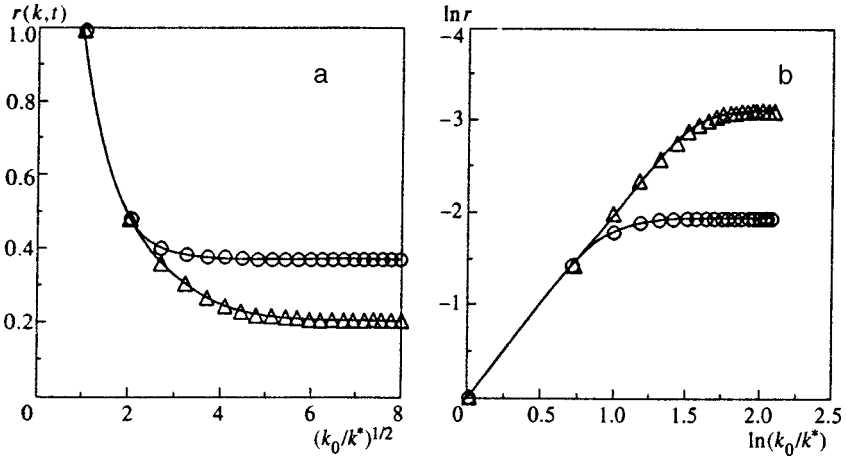

FIG. 5. Correlation coefficients of the perturbed and unperturbed velocity fields as functions of the cutoff wave number ratio (k_0/k_*) at different times (for two different values of k_c) on linear (a) and log–log (b) scales at $k_*=32$: ○, $k_c(t_1)=8k_*$; and △, $k_c(t_2)=32k_*$ (statistical averaging is over 100 realizations, and every tenth point is identified).

$S_0(y)$ and that the solution (2) exhibits universal behavior between two discontinuities [Eq. (4)], with y_k the coordinates of the local maxima of $S_0(y)$.

The occurrence of a high-frequency perturbation $v_h(x)$ leads to fluctuations in the ''zeros'' of a sawtooth wave. With the new component $v_h(x)$ in the initial velocity [Eq. (11)], a perturbation is added to the initial potential:

$$\tilde{S}_0(y)=S_0(y)+S_h(y). \tag{17}$$

From (5) we see that the perturbation $S_h(y)$ gives rise to fluctuations in the discontinuity velocity, owing to which the strength of the fluctuations of the discontinuity positions, $\Delta x_k(t)$, increases linearly with time. For a periodic unperturbed signal the discontinuities are at rest: $S_0(y_{k+1})=S_0(y_k)$ and $V_k=0$ [Eq. (5)]. Hence an arbitrarily small perturbation sets the discontinuities in motion, merges them, and, as a result, completely destroys the regular periodic sawtooth structure.[30] When the initial perturbation is noise, the discontinuities of the large-scale structure have random velocities by themselves, which causes them to merge and results in an increase in the external turbulence scale $l(t)$. If the rate of merging of the discontinuities of the large-scale structure is high enough and at all times such that $|\Delta x| \ll l(t)$ holds, we can speak of self-preservation of the evolution of the large-scale structure.

The high-frequency perturbation v_h can be characterized by the parameters

$$\sigma_{v_h}^2=\langle v_h^2(x)\rangle=\frac{2\alpha_n^2 k_0^{n+1}}{n+1}\left[1-\left(\frac{k_*}{k_0}\right)^{n+1}\right], \tag{18}$$

$$\sigma_{S_h}^2=\langle S_h^2(x)\rangle=\frac{2\alpha_n^2}{(1-n)k_*^{1-n}}\left[1-\left(\frac{k_*}{k_0}\right)^{1-n}\right], \tag{19}$$

$$l_h^2=\frac{\sigma_{S_h}^2}{\sigma_{v_h}^2}=\frac{1+n}{1-n}\frac{1}{k_*^{1-n}k_0^{n+1}}\left[1-\left(\frac{k_*}{k_0}\right)^{1-n}\right]\times\left[1-\left(\frac{k_*}{k_0}\right)^{n+1}\right]^{-1}, \tag{20}$$

where l_h is the spatial scale of the high-frequency perturbation $S_h(y)$.

Thus, as the cutoff frequency k_0 increases, the velocity variance $\sigma_{v_h}^2$ increases monotonically, the potential variance $\sigma_{S_h}^2$ increases up to a value determined by the spatial frequency k_* and then becomes constant, and the spatial scale of the signal decreases with increasing k_0 and for $k_0 \gg k_*$:

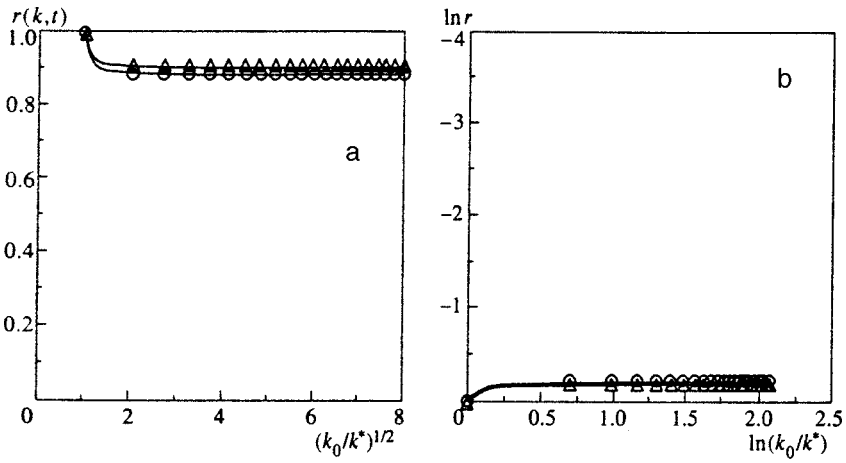

FIG. 6. The same as in Fig. 5 at $k_*=32$ with the following parameters: ○, $k_c=k_*/4$; and △, $k_c=k_*/8$.

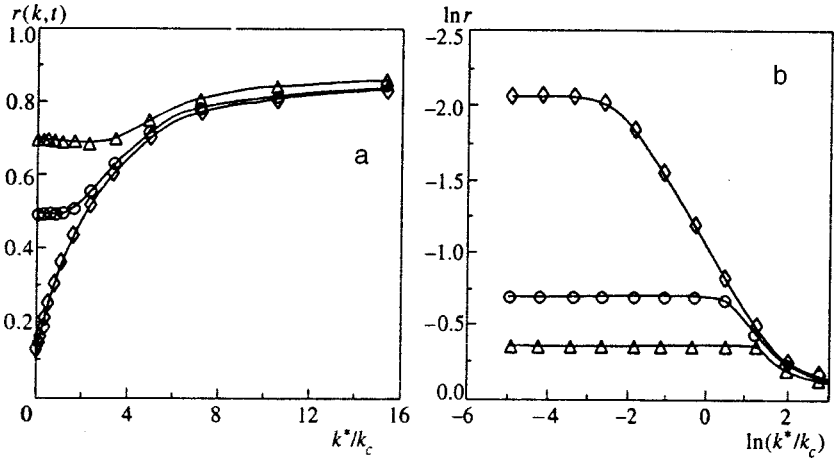

FIG. 7. Correlation coefficients of the perturbed and unperturbed velocity fields as functions of the ratio $k_*/k_c \sim t^{2/3}$ on linear (a) and log–log (b) scales with $k_*=16$: $\diamond$, $k_0=64k_*$; $\circ$, $k_0=4k_*$; and $\triangle$, $k_0=2k_*$ (statistical averaging is over 100 realizations, and every tenth point is identified).

$$l_h \simeq \sqrt{\frac{1+n}{1-n}}\,(k_* k_0)^{-1/2}\left(\frac{k_*}{k_0}\right)^{n/2}.$$

We begin by examining the case of small perturbations, $\sigma_{v_h}^2 \ll \sigma_0^2$. Such perturbations cause fluctuations in the positions of the zeros [Eq. (4)] and to fluctuations in the positions x_k of the discontinuities [Eq. (5)]:

$$\tilde{y}_k = y_k + \Delta y_k, \quad \tilde{x}_k = x_k + \Delta x_k. \tag{21}$$

In a small neighborhood of a local maximum of $S_0(y)$ the fluctuations in the positions of the zeros are determined by the equation

$$\tilde{S}_0'(y_k+\Delta y_k) = \tilde{v}_0(y_k+\Delta y_k) = 0, \tag{22}$$

$$v_0(y_k+\Delta y_k) + v_h(y_k+\Delta y_k) \simeq v_0'(y_k)\Delta y_k + v_h(y_k) = 0,$$

$$\Delta y_k = \frac{v_h(y_k)}{v_0'(y_k)}.$$

Using this equation, we can estimate the fluctuation:

$$\overline{\Delta y_k^2} = \langle \Delta y_k^2 \rangle \simeq \frac{\sigma_{v_h}^2}{\sigma_{v_0'}^2} = \frac{n+3}{n+1}\frac{1}{k_*^2}\left[\left(\frac{k_0}{k_*}\right)^{n+1} - 1\right]. \tag{23}$$

Clearly, their strength increases with the cutoff frequency k_0 and does not depend on time t.

In accordance with Eq. (5), there are two reasons for fluctuations in the positions of the shock fronts: the fluctuations in the zeros y_k and the fluctuations in the velocity V_k of the discontinuities:

$$\Delta x_k = \Delta y_k + \Delta V_k t. \tag{24}$$

The main contribution to the fluctuations in the velocity of a discontinuity is provided by the perturbation of the initial potential,

$$\Delta V_k = \frac{S_h(y_k) - S_h(y_{k+1})}{\eta_k}, \quad \eta_k = y_{k+1} - y_k, \tag{25}$$

while Δx_k can be estimated as follows:

$$\overline{\Delta x_k} = \langle \Delta x_k^2 \rangle^{1/2} = \Delta y_k + \Delta V_k t, \quad \Delta V_k \simeq \frac{\sigma_{S_h}}{l}. \tag{26}$$

Here $\sigma_{S_h}^2$ is the is the variance of the potential of the high-frequency perturbation [Eq. (19)], and $l(t) \simeq \langle \eta_k \rangle$ is the external turbulence scale. Thus, the size of the fluctuations in the positions of the discontinuities increases with time but reaches a plateau as the cutoff spatial frequency k_0 increases.

To calculate the correlation coefficient $r(t)$ [Eq. (12)], we examine the difference between the two processes $v(x,t)$ and $\tilde{v}(x,t)$:

$$\Delta v = \tilde{v}(x,t) - v(x,t), \tag{27}$$

$$\langle \Delta v^2 \rangle = \tilde{\sigma}^2(t) + \sigma^2(t) - 2r(t)\tilde{\sigma}(t)\sigma(t). \tag{28}$$

For $k_c(t) \ll k_*, k_0$ the variance $\sigma(t) \simeq \tilde{\sigma}(t)$ and the correlation coefficient can be written as follows:

$$r(t) = 1 - \frac{\varepsilon(t)}{2}, \quad \varepsilon(t) = \frac{\langle \Delta v^2 \rangle}{\sigma^2(t)}, \tag{29}$$

where $\varepsilon(t)$ is the relative deviation of the two processes. The difference Δv is the sum of rectangular pulses of two types (Figs. 2b and 3b). If the processes $v(x,t)$ and $\tilde{v}(x,t)$ at point x belong to the same cell containing the zero y_k, then $\Delta v = \Delta y_k/t$, and the pulse width $x_{k+1} - x_k$ is approximately equal to $l(t)$, where $l(t)$ is the external turbulence scale. If at the point x these processes are inside different cells, then $\Delta v = (y_{k+1} - y_k)/t$, and the pulse amplitude is proportional to $l(t)$/t, while its length is equal to the shift in the position of the discontinuity, Δx_k.

Using the expression (10) for the turbulence energy, we can estimate the relative deviation ε as follows:

$$\varepsilon = \frac{\overline{\Delta y_k^2}}{l^2} + \frac{\overline{\Delta x_k}}{l} = \frac{\overline{\Delta y_k^2}}{l^2} + \left(\frac{\overline{\Delta y_k}}{l} + \frac{\Delta V_k t}{l}\right). \tag{30}$$

Such estimates can also be made in the case of a periodic initial signal with a high-frequency perturbation. For a periodic signal, $l = l_0 = \text{const}$ and ε increases with time. This means that the presence of high-frequency noise destroys periodic structure.[30]

But if we have noise as the initial signal $v_0(x)$, the final result depends on the rate of growth of the external scale $l(t)$. Using (9), (18), (22), and (25), we arrive at an estimate for the relative deviation ε:

$$\varepsilon(t)=\varepsilon_1(t)+\varepsilon_2(t)+\varepsilon_3(t), \tag{31}$$

$$\varepsilon_1(t)=\frac{n+3}{n+1}\left[\left(\frac{k_0}{k_*}\right)^{n+1}-1\right]\left[\frac{k_c(t)}{k_*}\right]^2, \quad \varepsilon_2=\sqrt{\varepsilon_1},$$

$$\varepsilon_3=\sqrt{2}(1-n)^{-1/2}\left[1-\left(\frac{k_*}{k_0}\right)^{1-n}\right]^{1/2}\left[\frac{k_c(t)}{k_*}\right]^{(1-n)/2}. \tag{32}$$

Thus, when k_* is fixed the relative deviation tends to zero with the passage of time, decreasing for $n<1$ in proportion to $t^{(n-1)/(n+3)}$. This means that the solution $v(x,t)$ remains stable against small perturbations. But if $t=\text{const}$ and the variance of the high-frequency perturbation $v_h(x)$ decreases with increasing k_0, then from (32) it formally follows that the relative deviation $\varepsilon_1(t)$ increases without limit. However, rigorous analysis shows that even in the case of strong fluctuations in $v_h(x)$ the large-scale structure remains stable against the high-frequency perturbation $v_h(x)$ (see Figs. 2 and 3).

5. SELF-PRESERVATION OF LARGE-SCALE STRUCTURES AGAINST STRONG PERTURBATIONS

Now we turn to the limit $k_0 \gg k$, where the initial energy of the perturbation is much higher than the initial energy of the unperturbed signal, $\sigma_{v_h}^2 \gg \sigma_0^2$, and the initial correlation coefficient r_0 is much smaller than unity. But since the asymptotic behavior of a realization of the velocity field is determined by the properties of the initial potential in (3), we can assume that for $n<1$ the relative deviation $\varepsilon(t)$ tends to zero with the passage of time due to the finiteness of the variance of the perturbation potential [Eq. (19)]:

$$\sigma_{S_h}^2=\frac{\alpha_n^2}{(1-n)k_*^{1-n}}, \quad k_0 \gg k_*. \tag{33}$$

For $t \gg t_n$ a global maximum of the function $G(x,y,t)$ lies within a small neighborhood of a local maximum of $S_0(y)$, so that we must account for the effect of the random process $S_h(y)$ on the position $\tilde{y}_k$ and height $\tilde{H}_k$ of the maximum of the process $\tilde{S}(y)$ within a small neighborhood of a local maximum of $S_0(y)$:

$$\tilde{S}_0(y)=S_0(y)+S_h(y)$$

$$=\frac{1}{2}S_0''(y_k)(y-y_k)^2+S_0(y_k)+S_h(y). \tag{34}$$

The perturbation action $S_h(y)$ in Eqs. (17) and (34) is a statistically homogeneous process with a finite variance $\sigma_{S_h}^2$ [Eq. (19)] and a spatial scale l_h [Eq. (20)] that decreases with increasing k_0.

We would like to make some preliminary estimates before we discuss the results obtained by the asymptotic theory.

We write the second derivative of the unperturbed initial potential as

$$S_0''(y_k)=-\gamma_k\sigma_{S''}, \quad \sigma_{S''}^2=\frac{2\alpha_n^2 k_*^{n+3}}{n+3}, \tag{35}$$

where γ_k is a dimensionless parameter characterizing the size of the second derivative in the neighborhood of the local maximum of $S_0(y)$, and $\sigma_{S''}$ is the variance of the second derivative of the potential. For a rough estimate we can take $\tilde{H}_k$ and the dispersion of $S_h(y)$ to be of the same order. Then the deviation in the position of the absolute maximum, $\Delta y_k = y-y_k$, in Eq. (34) can be obtained from the condition $\Delta y_k^2 \simeq S_h/|S_0''|$. This yields estimates for $\tilde{H}_k$ and Δy_k:

$$\tilde{H}_k \simeq \sigma_{S_h}=\sqrt{\frac{2\alpha_n^2}{(1-n)k_*^{1-n}}},$$

$$\langle \Delta y_k^2 \rangle \simeq \frac{\sigma_{S_h}}{\gamma_k\sigma_{S''}}=\sqrt{\frac{n+3}{1-n}}\,\frac{1}{k_*^2\gamma_k}. \tag{36}$$

Thus, for $n<1$, when the variance of the perturbed potential is finite, even in the case of a strong perturbation of the velocity field, v_h, the relative fluctuations of the coordinates of the zeros of the sawtooth wave,

$$\varepsilon_1=\frac{\langle \Delta y_k^2 \rangle}{l^2} \simeq \sigma_{S_h} t^{-4/(n+3)}, \tag{37}$$

and the relative fluctuations of the positions of the shock fronts,

$$\varepsilon_3=\frac{\Delta V_k t}{l} \simeq \frac{\tilde{H}_k t}{l^2} \simeq \sigma_{S_h} t^{-(1-n)/(n+3)}, \tag{38}$$

remain finite as the variance σ_{v_h} of the perturbation increases, i.e., as k_0 increases, and decrease in strength with the passage of time.

Let us now study in greater detail the statistical properties of the coordinate $\tilde{y}_k$ and height $\tilde{H}_k$ of an absolute maximum of the process $\tilde{S}_0(y)$ in a small neighborhood of a local maximum of $S_0(y)$ in Eq. (34). Clearly, solving this problem means finding the statistical characteristics of the absolute maximum of $G(x,y,t)$ given by Eq. (3) in the asymptotic solution (2) of the Burgers equation with the time variable t replaced by $(-S_0''(y_k))^{-1}$, the curvature radius of the unperturbed potential. The latter problem was thoroughly studied in Refs. 7 and 9, where it was found that over long time intervals the absolute maximum has a Gaussian distribution and its value, a double exponential distribution. Construction of an asymptotic theory is possible if the time intervals are such that the parabola in (3) is a fairly smooth function on the scale of the initial action $S_0(y)$ and a large number of local maxima of the initial action contend for the right to be the absolute maximum of $G(x,y,t)$.

Returning to our problem (34), we note that we must estimate the characteristic number N of the local maxima of the perturbation $S_h(y)$ within the region Δy_k [Eq. (36)] where the local maxima contend for the right to be absolute maxima. Equations (20) and (36) imply that for $k_0 \gg k_*$ this number is

$$N \simeq \frac{\Delta y_k}{l_h} \sim \left(\frac{k_0}{k_*}\right)^{(1+n)/2} \gg 1$$

and increases with k_0. Thus, for $k_0 \gg k_*$ we can use the results of the asymptotic theory developed in Refs. 7and 9 to find the statistical characteristics $\tilde{y}_k$ and $\tilde{H}_k$ in (36). For this we must formally replace the time variable t in the appropriate formulas of Refs. 7 and 9 with the curvature of the unperturbed action $S_0(y)$, i.e., $t \mapsto 1/\gamma_k \sigma_{S''}$.

The asymptotic theory of Refs. 7 and 9 contains a large dimensionless parameter τ, which is the ratio of the time t to the characteristic time of manifestation of nonlinear effects. In our case it is

$$\tau = \frac{\sigma_{v_h}}{2\pi l_h S_0''} = \frac{\sigma_{v_h}}{2\pi l_h \gamma_k \sigma_{S''}} = \frac{M}{\gamma_k}, \tag{39}$$

where

$$M = \frac{1}{2\pi} \frac{[(n+3)(1-n)]^{1/2}}{(1+n)} \left(\frac{k_0}{k_*}\right)^{n+1} = A\left(\frac{k_0}{k_*}\right)^{n+1}. \tag{40}$$

Note that $M \sim N^2$, where N is the number of local maxima contending for the right to be the absolute maximum of $\tilde{S}_0(y)$. The results of Refs. 7 and 9 imply that for $M \gg 1$ the height $\tilde{H}_k$ of the absolute maximum [Eq. (36)] is

$$\tilde{H}_k = \sigma_{S_h} \xi_k \left(1 + \frac{\eta}{\xi_k^2}\right) = \sigma_{S_h} \xi_k + \sigma_{S_h} \frac{\eta}{\xi_k}, \tag{41}$$

where η is a dimensionless random quantity obeying the double exponential distribution

$$F_\infty(\eta) = e^{-e^{-\eta}}, \tag{42}$$

with $\langle \eta \rangle \sim \langle \eta^2 \rangle \sim 1$. Here ξ_k is the solution of the transcendental equation

$$\sqrt{\frac{M}{\xi_k \gamma_k}} \exp\left(-\frac{\xi_k^2}{2}\right) = 1, \quad \xi_k \simeq \sqrt{\ln M - \ln \gamma_k}. \tag{43}$$

We see that for $k_0/k_* \gg 1$ the height of the absolute maximum is weakly dependent on the local curvature of the potential, $S_0''(y_k) = \gamma \sigma_{S''}$ and for all practical purposes is a deterministic function with a mean

$$\langle \tilde{H} \rangle = \sigma_{S_h} \xi_k \simeq \sigma_{S_h} \sqrt{\ln M} \tag{44}$$

and a small variance

$$\langle \Delta \tilde{H}^2 \rangle = \frac{\sigma_{S_h}^2}{\xi_k^2} \simeq \frac{\sigma_{S_h}^2}{\ln M}. \tag{45}$$

The coordinate $\tilde{y}_k$ of a local maximum has a Gaussian distribution with a mean $\langle \tilde{y}_k \rangle = y_k$ and a variance Δy_k:

$$\langle \Delta y_k^2 \rangle = \frac{\sigma_{S_h}}{\xi_k \sigma_{S''} \gamma_k} \simeq \sqrt{\frac{n+3}{1-n}} \frac{1}{k_*^2 \gamma_k} \frac{1}{\sqrt{\ln M}}, \tag{46}$$

which decreases with the increasing ratio k_0/k_* of the spatial frequencies.

Thus, for $k_0 \gg k_*$ and $n < 1$, all global maxima of $\tilde{G}(x,y,t)$ are approximately within a small neighborhood of the corresponding absolute maxima of $G(x,y,t)$, and the potential $\tilde{S}(x,t)$ satisfies the following expression:

$$\tilde{S}(x,t) \simeq S(x,t) + \langle \tilde{H} \rangle. \tag{47}$$

The asymptotic solution (2) shows that the behavior of $\tilde{v}(x,t)$ asymptotically reproduces the behavior of the unperturbed field $v(x,t)$. In the case of a strong perturbation, the relative deviation of the processes $\tilde{v}$ and v is still described by Eq. (30), where

$$\varepsilon_1 = \left(\frac{n+3}{1-n}\right)^{1/2} \left(\frac{k_c(t)}{k_*}\right)^2 \frac{1}{\sqrt{\ln M}}, \quad \varepsilon_2 = \sqrt{\varepsilon_1},$$
$$\varepsilon_3 = (1-n)^{-1/2} \left(\frac{k_c(t)}{k_*}\right)^{(1-n)/2} \frac{1}{\sqrt{\ln M}}. \tag{48}$$

Thus, for $k_c(t) \gg k_*$, the deviation is also due to fluctuations in the positions of the discontinuities in the velocity field and tends to zero with the passage of time.

The above analysis provides a good qualitative and quantitative explanation of the results of computer simulation of the dependence of the correlation coefficient on the spatial frequency k_0 of the high-frequency perturbation and on time t. In particular, the formulas give a good description of the process in which the correlation coefficient r reaches a plateau, where its value depends weakly on the bandwidth of the perturbation spectrum k_0. For instance, in Fig. 6 the theoretical value of r at $k_c = k_*/4$ obtained by the formula $r = 1 - \varepsilon/2 \simeq 1 - \varepsilon_3/2$ [Eq. (48)] is equal to 0.8925, while experiment yields $r = 0.89$. The results of the theoretical treatment imply that as t increases, the correlation coefficient of the perturbed and unperturbed fields must tend to unity. In the numerical experiments the evolution of the fields was studied in finite-size regions. Asymptotically, only a single triangular pulse remains in such a region, and the fluctuations of its zero are weak but finite. This is the reason why r tends to a value close to unity.

6. SMALL-SCALE PERTURBATIONS AND TURBULENT VISCOSITY

It is well known that the effect of small-scale perturbations on large-scale flows can be described by introducing an effective turbulent diffusion coefficient. In connection with the Burgers equation, this problem was studied in Refs. 32 and 36. Here we briefly discuss the features of the effect of viscosity and small-scale perturbations on the evolution of noise signals. For infinite Reynolds numbers, the initial field becomes a sequence of triangular pulses, all having the same slope $v_x' = 1/t$ and completely determined by the positions of the zeros, y_k, and of the discontinuities, x_k. For a small but finite dissipation coefficient μ, viscosity is important only near the discontinuities and broadens the shock front. In the neighborhood of a discontinuity,

$$v(x,t) \simeq \Delta v_k \tanh \frac{x - x_k}{\delta}, \tag{49}$$

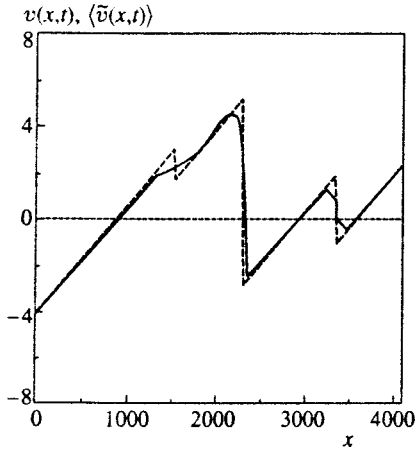

FIG. 8. Evolution of the unperturbed field $v(x,t)$ with a cutoff wave number $k_*=32$ (dashed curve) and that of the mean field $\langle\tilde{v}(x,t)\rangle$ with an addition high-frequency part in the energy spectrum and a cutoff wave number $k_0=1024$ (solid curve) at time $t=0.125$ ($k_c=4$); averaging is over 1000 realizations of the high-frequency perturbation.

where Δv_k is the discontinuity amplitude, and $\delta=2\mu/\Delta v_k$ is the discontinuity width. In the discontinuity stage,

$$\Delta v_k=\frac{y_{k+1}-y_k}{t}\sim\frac{l}{t},$$

where l is the external turbulence scale, and $\delta\sim\mu t/l$. The degree of nonlinearity of the waves can be characterized by a dimensionless parameter known as the Reynolds number

$$\mathrm{Re}(t)\simeq\frac{\Delta v(t)l(t)}{\mu}=\frac{l(t)}{\delta(t)}\simeq\frac{l^2(t)}{\mu t}, \tag{50}$$

equal to the ratio of the external turbulence scale $l(t)$ to the internal scale $\delta(t)$.

For a periodic signal, $l=\mathrm{const}$. Due to the increase in the width of the shock front, the Reynolds number diminishes, and at $\mathrm{Re}(t)\sim 1$ the wave reaches the linear regime. With a noise perturbation there are two competing factors: the increase in the external scale $l(t)$ owing to the merging of discontinuities, and the increase in the internal scale owing to dissipation. For random signals with the spectrum (7) for $n<1$ we find that Eqs. (9) and (50) yield

$$\mathrm{Re}(t)\sim t^{(1-n)/(n+3)},$$

i.e., the Reynolds number increases with time and the shape of the wave becomes more and more nonlinear.

We arrive at the same conclusions if we examine the evolution of the mean field $\langle\tilde{v}(x,t)\rangle$, assuming that the averaging is done over the ensemble of realizations of the high-frequency perturbation $v_h(x)$. Indeed, the perturbed coordinates of the zeros have means $\langle\tilde{y}_k\rangle=y_k$, so that Eq. (4) implies that far from the discontinuities the mean of the perturbed field is equal to the unperturbed field, $\langle\tilde{v}(x,t)\rangle=v(x,t)$, a fact that Fig. 8 clearly demonstrates. Near a discontinuity we must allow for the effect of perturbations on the position of the discontinuity coordinates [Eq. (5)].

Bearing in mind that for $k_0\gg k_*$ the perturbed action has a double exponential distribution, we can show that the mean field is still described by (49), where the effective viscosity is

$$\mu_{\mathrm{eff}}=\left(\frac{2\pi}{\alpha_n}\right)^{2/3}\frac{2\sigma_{S_h}}{\xi},\quad \xi\simeq\sqrt{\ln M}\,.$$

Here the shock-front width is determined by the fluctuations in the positions of the shock fronts, and the relative deviations (38), which are inversely proportional to the Reynolds number, decrease with the passage of time.

7. CONCLUSION

Let us discuss the effect of high-frequency (small-scale) perturbations on the evolution of the large-scale components of the field for different indices n of he spectrum of the initial perturbation (Eq. (7)). We assume that the energy spectrum of the initial field $v_0(x)$ and the perturbation $v_h(x)$ obeys a power law: $E_0(k)\propto\alpha_n^2k^n$. For $v_0(x)$ the cutoff frequency is k_*, and the perturbation spectrum occupies the interval $k\in[k_*,k_0]$, with $k_0\gg k_*$

In Refs. 6–10, 14 and 15 it was shown that if the initial field obeys Gaussian statistics, the turbulence energy diminishes for $n>1$ as

$$\sigma^2(t)\simeq\sigma_S t^{-1}ln^{-1/2}(t\sigma_v^2/\sigma_S), \tag{51}$$

where $\sigma_S^2=\langle S_0^2\rangle$ is the variance of the initial potential (see Eq. (3)), and $\sigma_v^2=\langle v_0^2\rangle$ is the initial-velocity variance. For the unperturbed field $\tilde{v}(x,t)$ from Eq. (7) we have $\sigma_S^2\simeq k_*^{n-1}$, while for the perturbed field $\tilde{v}(x,t)$ we have $\sigma_{\tilde{S}}^2\simeq k_0^{n-1}$. Thus, for $k_0\gg k_*$ the variance of the perturbed field is much larger than the variance of the unperturbed field for $n>1$, which means that large-scale structures are unstable with respect to small-scale structures in the case being discussed here. The physics of this is that for $n>1$ large-scale structures form because of parametric generation of low-frequency components, which in turn is due to the nonlinear interaction between the spectral components of the entire spectrum. Here, even in the nontrivial case $1<n<2$, when the low-frequency part of the spectrum is preserved ($E(k,t)=E_0(k)\sim\alpha_n^2k^n$, $k\to 0$), the main fraction of the energy of the field is determined by the processes of parametric generation of low-frequency components, and the decrease in energy is described by Eq. (51) (see Ref. 15). Note that for $n>1$, the evolution of the low-frequency components is sensitive not only to small-scale perturbations but also to the shape of the probability distribution of the initial potential.[12,17,19]

The case $-1<n<1$ and $k_0,k_*\to\infty$ corresponds to a situation in which the initial potential is a fractal Brownian process. In this case the field $v(x,t)$ is strictly statistically self-similar and it is possible to find the velocity distribution, the discontinuity amplitudes, etc., for it.[2,16,18] Here the low-frequency part of the spectrum is preserved, and it is this part that determines the asymptotic behavior of the field. In Ref. 33 it was shown, with a rigor that would be accepted by most physicists, that at a finite cutoff frequency k_* the laws governing the evolution of the energy spectrum are the same as

those that govern the evolution of purely fractal Brownian motion ($k_* \to \infty$). Avellaneda *et al.*[13] found that the shape of the probability distribution of velocity and the derivatives of the velocity at large deviations are also independent of the cutoff frequency k_* of the initial spectrum.

Thus, as a result of our research we found that for $-1 < n < 1$ not only statistical characteristics but also the random-field realizations proper remain stable against initial small-scale perturbations of both small and large amplitudes.

For $-3 < n < -1$, when the velocity itself is fractal Brownian motion,[11,29] it is obvious that the evolution of large-scale structures also remains stable against small-scale perturbations. It would be interesting to know how sensitive the probability distribution of the discontinuity amplitudes and the distances between the discontinuities are to the value of the cutoff frequency k_*.

In conclusion we note that the asymptotic stability of large-scale structures for $n < 1$ is equivalent to the effect of the increase in Reynolds numbers in a medium with finite viscosity.[9,10,36]

The authors are grateful to E. Aurell and A. I. Saichev for the fruitful discussions during research. This work was made possible by grants from the Russian Fund for Fundamental Research (RFBR Project 96-02-19303) and a RFBR–INTASproject (Grant No. 95-IN-RU-723).

*)E-mail: gurb@rf.unn.runnet.ru

[1] J. M. Burgers, Proc. R. Neth. Acad. Sci. **17**, 1 (1939).
[2] J. M. Burgers, *The Nonlinear Diffusion Equation*, Reidel, Dordrecht (1974).
[3] A. M. Polyakov, Phys. Rev. E **52**, 6183 (1995).
[4] E. Hopf, Commun. Pure Appl. Math. **3**, 201 (1950).
[5] J. D. Cole, Q. Appl. Math. **9**, 225 (1951).
[6] S. Kida, **93**, 337 (1979).
[7] S. N. Gurbatov and A. I. Saichev, Zh. Éksp. Teor. Fiz. **80**, 689 (1981) [Sov. Phys. JETP **53**, 347 (1981)].
[8] J. D. Fournier and U. Frisch, J. Mec. Theor. Appl. **2**, 699 (1983).
[9] S. N. Gurbatov, A. N. Malakhov, and A. I. Saichev, *Nonlinear Random Waves in Nondispersive Media* [in Russian], Nauka, Moscow (1990), p. 215.
[10] S. N. Gurbatov, A. N. Malakhov, and A. I. Saichev, *Nonlinear Random Waves and Turbulence in Nondispersive Media: Waves, Rays, Particles*, Manchester Univ. Press (1991), p. 1.
[11] Ya. Sinai, Commun. Math. Phys. **148**, 601 (1992).
[12] S. Alberio, S. A. Molchanov, and D. Surgalis, Probab. Theory Relat. Fields **100**, 457 (1994).
[13] A. Avellaneda, R. Ryan, and E. Weinan, Phys. Fluids **7**, 3067 (1995).
[14] S. A. Molchanov, D. Surgalis, and W. A. Woyczhynski, Commun. Math. Phys. **168**, 209 (1995).
[15] S. N. Gurbatov, S. I. Simdyankin, E. Aurell *et al.*, J. Fluid Mech. **344**, 339 (1997).
[16] G. N. Molchan, J. Stat. Phys. **88**, 1139 (1997).
[17] T. J. Newman, Phys. Rev. E **55**, 6989 (1997).
[18] R. Ryan, Commun. Pure Appl. Math. **LI**, 47 (1998).
[19] S. Gurbatov and U. Frisch, in *Advances in Turbulence VII*, U. Frisch, Kluwer Academic Publishers, Dordrecht (1998), p. 437.
[20] O. V. Rudenko and S. I. Soluyan, *Theoretical Foundations of Nonlinear Acoustics*, Consultants Bureau, New York (1977).
[21] G. Whitham, *Linear and Nonlinear Waves*, Wiley, New York (1974).
[22] M. Kardar, G. Parisi, and Y. Zhang, Phys. Rev. Lett. **56**, 889 (1986).
[23] A.-L. Barabasi and H. E. Stanley, *Fractal Concepts in Surface Growth*, Cambridge Univ. Press, Cambridge (1995).
[24] J.-P. Bouchaud, M. Mezard, and G. Parisi, Phys. Rev. E **52**, 3656 (1995).
[25] V. Gurarie and A. Migdal, Phys. Rev. E **54**, 4908 (1996).
[26] S. N. Gurbatov, A. I. Saichev, and S. F. Shandarin, Mon. Not. R. Astron. Soc. **236**, 385 (1989).
[27] S. F. Shandarin and Ya. B. Zeldovich, Rev. Mod. Phys. **61**, 185 (1989).
[28] D. Weinberg and J. Gunn, Mon. Not. R. Astron. Soc. **247**, 260 (1990).
[29] M. Vergassola, B. Dubrulle, U. Frisch *et al.*, Astron. Astrophys. **289**, 325 (1994).
[30] S. N. Gurbatov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **26**, 3, (1983); **26**, 283 (1983).
[31] E. Aurell, S. Gurbatov, and I. Wertgeim, Phys. Lett. A **182**, 1 (1993); **182**, 109 (1993).
[32] I. G. Yakushkin, Zh. Éksp. Teor. Fiz. **81**, 967 (1981) [Sov. Phys. JETP **54**, 513 (1981)].
[33] S. N. Gurbatov, I. Yu. Demin, and A. I. Saichev, Zh. Éksp. Teor. Fiz. **87**, 497 (1984) [Sov. Phys. JETP **60**, 284 (1984)].
[34] S. N. Gurbatov and D. G. Crighton, Chaos **5**, 3 (1995); **5**, 524 (1995).
[35] A. Noullez and M. Vergassola, J. Sci. Comput. **9**, 259 (1994).
[36] S. N. Gurbatov, I. Yu. Demin, and N. V. Pronchatov-Rubtsov, Zh. Éksp. Teor. Fiz. **91**, 1352 (1986) [Sov. Phys. JETP **64**, 797 (1986)].

Translated by Eugene Yankovsky

Negative ions in liquid xenon

A. G. Khrapak*) and K. F. Volykhin

Institute for High Temperatures, Russian Academy of Sciences, 127412 Moscow, Russia
(Submitted 15 June 1998)
Zh. Éksp. Teor. Fiz. **115**, 584–592 (February 1999)

The structure and mobility of negative ions of oxygen in liquid Xe is investigated. It is shown that the strong exchange interaction of the outer, weakly bound electron of the negative ion with the surrounding liquid leads to a partial compensation of the electrostriction effect, and it prevents the formation of a solid cluster around the negative ion. A simple perturbative model describing the structure of the negative ion in the liquid matrix is developed. The mobility of O_2^- ions in liquid Xe on the saturation line is estimated. The reasons for the difference in mobility of negative and positive ions are discussed. © *1999 American Institute of Physics.* [S1063-7761(99)01402-X]

1. INTRODUCTION

In liquefied rare gases, the structure and properties of admixed ions are considered, as a rule, in the context of Atkins' model.[1] This model takes into consideration the effect of electrostriction, and it predicts the formation of a solid cluster around the ion. The cluster then determines the transport properties of such an ion in the condensed phase. The consequences of this theory are in good agreement with experiments on the mobility of positive ions. However, Atkins' theory does not depend on the sign of the ionic charge, and it predicts the same effects for positive and negative ions.

Recently, experimental data on the mobility of certain negative and positive ions in liquid xenon became available which indicate a difference in the transport properties of positive and negative ions. It was found that negative ions exhibit higher mobility than positive ions of comparable size.[2,3]

The difference in transport properties of positive and negative ions is, above all, due to the fact that the outer electron of the negative ion is localized in a spatial region with a characteristic size appreciably greater than that of the inner electron shells of the parent atom or molecule of the ion. When a negative ion is located in a dense gas or liquid it perturbs the surrounding medium. The character of the perturbation depends on the character of interaction of the outer electron with the medium. One important factor that determines the qualitative difference in the perturbation near positive and negative ions is the repulsive exchange interaction of the weakly bound outer electron of the negative ion with the electron shells of the surrounding atoms or molecules.

The structure of negative ions in dense gases and liquids with atoms of low polarizability (He and Ne) were investigated recently.[4,5] It was shown that the competition between the short-range exchange repulsion and the long-range polarization attraction of the outer electron with the atoms of the medium leads to the formation of a microcavity around the negative ion, which is surrounded by a denser layer of atoms. This effect largely determines the negative ion mobility. Similar complexes, in which a light quantum particle is localized inside the liquid density fluctuation and kept stationary, were first suggested and investigated by Ferrell[6] to explain the anomalously long life-time of positronium in liquid helium due to creation of a cavity (''bubble'') around the positronium. Bubbles in nonpolar liquids can also be created around free electrons, excited atoms, and positive and negative ions. Detailed discussions of this problem can be found in the special issue «Ions and Atom in Superfluid Helium», Zeitschrift für Physik B, Cond. Matter **98**, No 3 (1995).

A difference in the mobilities of positive and negative ions was also observed in liquefied rare gases with high polarizability. For example, in liquid xenon, negative ions of O_2 and SF_6 exhibit mobility several times that of positive ions.[2] As a result of the high polarizability of atoms of the medium, no microcavity is created in this case, but the strong exchange interaction leads to partial cancellation of the electrostriction and it prevents the formation of a solid cluster around the negative ion. The determination of the structure of the negative ion becomes complex because of the more complicated nature of the interaction of the outer electron of the ion with the liquid.

In this paper, we propose a simplified model of the ionic structure that considers the influence of the weakly bound electron of the negative ion on the liquid matrix. From these results, we estimate the mobility of O_2^- in liquid Xe by means of a modified Stokes equation.

2. SELF-CONSISTENT FIELD MODEL OF A NEGATIVE ION IN A NONPOLAR LIQUID

We consider a lone electron bound by the polarization potential to the molecule that formed the negative ion. The asymptotic form of the electron wave function, $\psi(r)$, in the polarization potential of the molecules is very well known. At large distances r, the wave function varies as $\psi(r) \propto r^{-1}\exp(-r/\lambda)$. The characteristic size λ of the spatial region where the electron is localized is determined by its ground state energy ε in the negative ion,[7] $\lambda = \hbar/\sqrt{2m\varepsilon}$.

In the following, we consider weakly bound electrons for which λ is appreciably greater than the size of the parent molecule. The weakly bound electron spends a significant time far from the molecule. It moves with an average kinetic energy $\varepsilon_k \propto \hbar^2/2m\lambda^2$. The motion of such an electron has many features in common with the motion of a quasifree electron of the same energy. In rarefied gases this interaction results in scattering of the electron by isolated molecules. The scattering potential consists of a short-range repulsive part due to exchange interaction and an attractive long-range polarization part.

In a liquid, the situation is significantly more complicated. Here, the long-range part of the polarization potential is determined by the atomic polarization potentials of the atoms around the ion. In this case the potential of a scatterer can be represented by a superposition of the electron-atom interaction potential and the screened polarization potentials of all other atoms around the ion. The total scattering potential of an electron is described as a ''muffin-tin'' potential.[8]

It is worthwhile to note that the potential decays faster than $1/r^4$ as a result of the screened polarization. This leads to a reduction in the characteristic size of the interaction region. In rare gases, in the supercritical region at liquid densities and on the saturation line in liquids, the elastic electron-atom cross-section depends weakly on electron energy and scattering angle. There is essentially no Ramsauer–Townsend effect, and the scattering cross-section of a slow electron is determined by an effective scattering length $L_{\text{eff}}(N)$.

The electron effective scattering length is a function of the liquid density N. It can be determined from experimental data on the quasifree electron mobility.[9] In the low-density limit, the effective scattering potential is transformed into the potential of an isolated atom, and $L_{\text{eff}}(N)$ becomes the electron scattering length of an isolated atom

$$L_a = \sqrt{\frac{\alpha}{a_0}} \cot \sqrt{\frac{\alpha}{a_0 R_a^2}},$$

where α is the atomic polarizability, a_0 is the Bohr radius, and R_a denotes the effective radius of the short-range hard core of the electron–atom interaction potential.[10] With increasing liquid density, the relative role of attraction in the scattering process decreases as a result of polarization screening. This leads to a decrease in scattering length, and even to a sign reversal at sufficiently high densities.

In the optical approximation,[7,8] the energy of the bottom of the electron conduction band is $V_0 = 2\pi\hbar^2 N L_{\text{eff}}(N)/m + U_p(N)$, where $U_p = -\alpha e^2 F_L N/2\sigma$ is the shift in electron energy as a result of the screened polarization interaction. The Lorentz screening factor[8] $F_L(N) = [1 + 8\pi\alpha N/3]^{-1}$ takes into account the weakening of the polarization field of a particular atom by its interaction with the induced dipoles of all other atoms of the medium; σ is the radius of the first peak of the pairwise correlation function of the liquid.

We model the motion of the weakly bound electron of a negative ion in a nonpolar liquid as the motion of the electron in the field of point scatterers with an effective potential $V_{\text{eff}} = 2\pi\hbar^2 L_{\text{eff}}(N)\delta(\mathbf{r} - \mathbf{R}_i)/m$, where $\delta(\mathbf{r})$ is the Riemann delta function, $\mathbf{r}$ and $\mathbf{R}_i$ are the coordinates of the electron and scattering center, respectively. We also assume that the conditions for the optical model are satisfied, i.e.,

$$L_{\text{eff}}(N) \ll \lambda, \quad N L_{\text{eff}}^3(N) \ll 1. \tag{1}$$

The Schrödinger equation for the wave function of a weakly bound electron in the liquid acquires the form

$$-\frac{\hbar^2}{2m}\Delta\psi(\mathbf{r}) + U(\mathbf{r})\psi(\mathbf{r}) + \sum_i \frac{2\pi\hbar^2}{m} L_{\text{eff}}(N)\,\delta(\mathbf{r} - \mathbf{R}_i)\,\psi(\mathbf{r}) = [E - U_p(N)]\psi(\mathbf{r}), \tag{2}$$

where E is the electron ground state energy of the negative ion in the liquid and $U(\mathbf{r})$ is the potential of the molecule on which the electron is localized. The shift of the electron energy $U_p(N)$ was taken into account in Eq. (2). The electron energy is then

$$\varepsilon = E - U_p(N).$$

We assume that the effect of the medium on ε is small,

$$\varepsilon \gg \Delta\varepsilon = \langle \psi(\mathbf{r}) | \sum_j \frac{2\pi\hbar^2}{m} L_{\text{eff}}(N)\,\delta(\mathbf{r} - \mathbf{R}_j) | \psi(\mathbf{r}) \rangle. \tag{3}$$

First-order perturbation theory supports this notion.[7] Integrating Eq. (3) over the electron coordinate and averaging over the atomic coordinates, we obtain the following relationship between the de Broglie wavelength of the electron λ, the electron effective scattering length in the medium $L_{\text{eff}}(N)$, and the number density N of the liquid:

$$\frac{\lambda}{L_{\text{eff}}(N)} \gg 4\pi\lambda^3 N \exp\left(-\frac{2\sigma}{\lambda}\right). \tag{4}$$

This places an additional constraint on the electron binding energy of the ion. In fact, when $\lambda \gg \sigma$, the right-hand side of Eq. (4) is proportional to λ^3, but its left-hand side is linear in λ. At very large λ, the number of atoms of the medium $N\lambda^3$ in the vicinity of the electron orbit can be large. In this case, the interaction of the electron with the medium, as determined by its scattering length, cannot be considered weak. The contribution of the medium to the electron binding energy is significant, and inequality Eq. (4) is violated. In conjunction with Eq. (1), Eq. (4) determines the range of applicability of the current approximation.

We now turn to a discussion of the free energy functional $\Delta F\{N(\mathbf{r}), \psi(\mathbf{r})\}$ of the negative ion-liquid system. The presence of the negative ion in the liquid leads to a change in the local structure of the liquid near the ion. In other words, the perturbation of the liquid affects the electron spectrum of the negative ion. The structure of the complex is determined by the minimum free energy, which can be represented as a sum of two terms,

$$\Delta F\{N(\mathbf{r}), \psi(\mathbf{r})\} = \Delta E\{N(\mathbf{r}), \psi(\mathbf{r})\} + \Delta F_l\{N(\mathbf{r})\}.$$

The first represents the change in electron ground-state energy due to the presence of the liquid. An accurate calculation of ΔE is complicated, and requires a self-consistent calculation of the local liquid density $N(\mathbf{r})$ and the electron

wave function $\psi(\mathbf{r})$. If the conditions of the perturbation theory are satisfied, however, the problem can be simplified significantly, and ΔE can be written

$$\Delta E\{\psi(\mathbf{r}),N(\mathbf{r})\}\simeq\langle\psi(\mathbf{r})|\sum_j \frac{2\pi\hbar^2}{m} \times L_{\rm eff}(N)\,\delta(\mathbf{r}-\mathbf{R}_j)|\psi(\mathbf{r})\rangle+U_p\{N(\mathbf{r})\}. \tag{5}$$

The second term in the free energy functional $\Delta F_l\{N(\mathbf{r})\}$ represents the change in free energy of the liquid as a result of a change in its local density. The relation between the optimum density profile $N(\mathbf{r})$ and the electron wave function $\psi(\mathbf{r})$ is determined by minimizing the free energy, $\delta\Delta F\{N(\mathbf{r}),\psi(\mathbf{r})\}/\delta N(\mathbf{r})=0$. A knowledge of $N\{\psi(\mathbf{r})\}$ enables one to determine the potential energy of electron interaction with atoms of medium. This leads to the nonlinear Schrödinger equation. Its solution, which can be obtained numerically, is usually very complicated, and we consider only the simplest case of small changes in the local density $|N(\mathbf{r})-N|/N\ll 1$.

In this case, the relation between $N(\mathbf{r})$ and $\psi(\mathbf{r})$ becomes very simple:

$$N(\mathbf{r})/N=\exp[-\gamma\tilde{V}(\mathbf{r})/Ms^2], \tag{6}$$

where M is the atomic mass, s is the velocity of sound in the liquid, and $\gamma=C_p(T)/C_V(T)$ is the ratio of specific heats of the liquid. The function $\tilde{V}(\mathbf{r})=\tilde{V}_1(\mathbf{r})+\tilde{V}_2(\mathbf{r})$ represents the effective interaction potential between the electron and liquid. In first-order perturbation theory, the first term has the form

$$\tilde{V}_1(\mathbf{r})=\langle\psi(\mathbf{r}')|\frac{2\pi\hbar^2}{m}L_{\rm eff}(N)\,\delta(\mathbf{r}-\mathbf{r}')|\psi(\mathbf{r}')\rangle = \frac{2\pi\hbar^2}{m}L_{\rm eff}(N)|\psi(\mathbf{r})|^2. \tag{7}$$

This takes electron scattering by the effective potential into consideration. The second term in $\tilde{V}(\mathbf{r})$ takes into account the fact that when the electron is near a scattering center, it interacts with the screening potentials of the other atoms of the liquid. These contribute to $U_p(\mathbf{r})$, but they do not participate in the scattering process:

$$\tilde{V}_2(\mathbf{r})=\langle\psi(\mathbf{r}')|V_p(\mathbf{r}-\mathbf{r}')g(\mathbf{r}-\mathbf{r}')F_l(N)|\psi(\mathbf{r}')\rangle. \tag{8}$$

Here, $g(\mathbf{r}-\mathbf{r}')$ is the pairwise correlation function of the atoms comprising the liquid. In liquids made up of atoms with small polarizability, where the atomic polarization potentials are essentially unscreened, $\overline{V}_2(\mathbf{r})$ is negligible in comparison with $\overline{V}_1(\mathbf{r})$. Note that in the limit $r\gg\sigma$, $\overline{V}_2(\mathbf{r})$ takes the natural form of the screening polarization interaction: $\overline{V}_2(\mathbf{r})=-\alpha e^2F_l(N)/2r^4$.

3. RESULTS AND DISCUSSION

The present model was used to investigate the structure of the negative ion of the oxygen molecule, O_2^-, in liquid Xe on the saturation line, with the aim of comparing calculated and experimental results.[2] Interest in the O_2^- ion stems from the presence of this ion as an admixture in essentially all experiments on electron and ion mobilities in liquefied rare gases. Furthermore, experimental data on the mobility of O_2^- in dense gases and liquids are available. The binding energy of the electron in O_2^- in vacuum (electron affinity) is $\varepsilon\simeq 0.46$ eV, which corresponds to $\lambda\simeq 5.5a_0$. In the atomic density range considered here, the electron–atom, scattering cross-section depends weakly on the electron energy, and there is no Ramsauer–Townsend effect.[11] Therefore, we assume for simplicity that the cross-section is independent of energy, and that it is determined by the effective scattering length $L_{\rm eff}(N)$. The applicability condition for perturbation theory is then satisfied for O_2^- in liquid Xe. The effect of the medium on the weakly bound electron of the negative ion O_2^- at the triple point results in a correction to the electron ground-state energy of $\Delta\varepsilon\simeq 0.05$ eV, which is significantly less than the electron affinity of O_2 in vacuum.

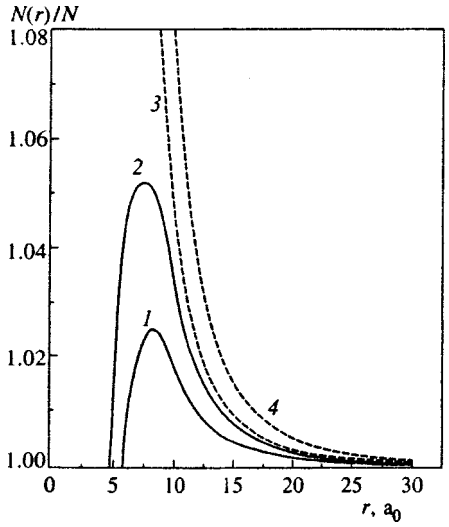

FIG. 1. Profiles of local density as a function of the distance to the center of the O_2^- ion in Xe on the saturation curve. The solid curves were calculated in accordance with (6) for different temperatures: *1*) 161.4, *2*) 200 K. The dashed curves were calculated in accordance with Atkins' model for the same temperatures: *3*) 161.4, *4*) 200 K.

Liquid density profiles were calculated at the triple point of Xe, at $T=161.4$ K, $N=1.36\times10^{22}$ cm^{-3}, and on the saturation curve at $T=200$ K, $N=1.23\times10^{22}$ cm^{-3}. For these conditions we adopted effective electron scattering lengths $L_{\rm eff}(N=1.36\times10^{22}\,{\rm cm}^{-3})=0.6a_0$ and $L_{\rm eff}(N=1.23\times10^{22}\,{\rm cm}^{-3})=0.3a_0$. Results of the calculations are shown in Fig. 1. In the same figure, density profiles calculated by means of Atkins' model are shown. Far from the ion, the long-range polarization part of the interaction potential plays the main role. Closer to the center of the ion, the exponentially increasing term [see Eq. (7)], which takes the exchange interaction into consideration, becomes dominant. In liquid Xe, in spite of its high polarizability $\alpha=27.11$ a.u., the exchange interaction is so strong that as the liquid density increases, $L_{\rm eff}(N)$ reverses sign from negative to positive at $N^*\simeq 1.1\times10^{22}$ cm^{-3}. The increasing importance of $\overline{V}_1(\mathbf{r})$ results in a slower rise in the local density near the center of the ion. The density profiles $N(r)$, shown in Fig. 1 for small distances of the order of the radius of the first coordination

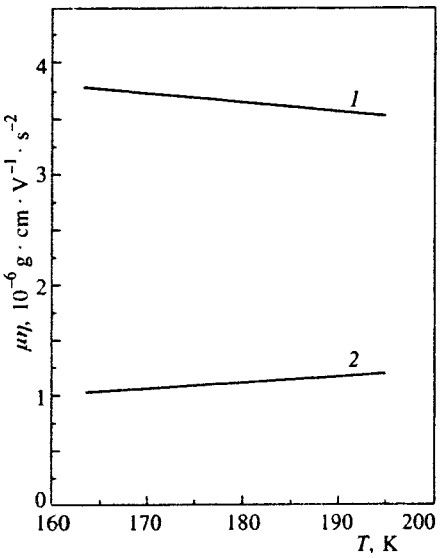

FIG. 2. Experimental results[2] for $\mu(T)\eta(T)$: *1*) O_2^- ion; *2*) positive ion of *n*-pentane in liquid Xe on the saturation curve.

sphere, cannot be correct, since the applicability of the present model would be violated. But we speculate that an accurate calculation would not change the qualitative picture, because the probability of finding an electron in this region is low.

As can be seen from Fig. 1, the assumption of small changes in liquid density near the ion is entirely justified (note the expanded scale of the ordinate). In contrast, the Atkins' model predicts a large increase in liquid density near the ion. As the result of the strong exchange interaction, the changes are so small that the creation of a solid-like cluster around a negative ion is not observed. We believe that this is the main reason for the experimentally measured[2] mobility difference between the O_2^- ion and positive ions. Measurements of the mobility of an O_2^+ ion are not available, but the mobility of the positive ions of *n*-pentane and tetramethylsilane in liquid Xe were found to be almost three times less than the O_2^- mobility.

For a simple estimate of the negative ion mobility in a liquid, we assume that during the motion of the ion in a weak electric field, a viscous flow of the liquid occurs that is determined by its local density near the ion. It is then possible to model the motion of both negative and positive ions as the motion of spheres with effective radii $R_{\rm eff}^{\pm}$ that differ for the two kind of ions. From the Stokes formula, the mobility is

$$\mu=\frac{e}{cR_{\rm eff}^{+}\eta}, \tag{9}$$

where η is the viscosity of the undisturbed liquid and the numerical constant $c=4\pi$ for a bubble or 6π for a cluster.[12] It follows from Eq. (9) that the product $\mu\eta$ is independent of the kinetic characteristics of the liquid. It is determined solely by the structure of the resulting ionic complex.

In Fig. 2 the experimental dependence of $\mu\eta$ on the temperature of liquid Xe along the saturation curve is shown. It is obvious that besides the quantitative differences in $\mu\eta$ of the positive and negative ions, there is also a qualitative difference between them. This is exhibited by the differing slopes of $\mu(T)\eta(T)=$const for the two types of ions, and it demonstrates the differing dependencies of the structures of the positive and negative ions on liquid density and temperature.

Ionic motion in a liquid results in energy dissipation. The liquid layer with maximum viscosity is attached directly to the surface of the ionic sphere. This layer contributes most actively to dissipation. With this assumption, it is possible to estimate the effective radii of the charged complexes,

$$R_{\rm eff}^{\pm}(N)\simeq R_c^{\pm}(N)(\eta_m/\eta), \tag{10}$$

where $R_c^{\pm}(N)$ are the radii of positive and negative ion complexes, respectively; η_m are the maximum viscosities near the ion, and η is, as before, the viscosity of the bulk liquid. Expanding η_m in Taylor series up to the second order in the relative variation in liquid density, we can write

$$\eta_m\simeq\eta(N,T)[1+\xi_T(N_m-N)], \tag{11}$$

where N_m denotes the maximum density near the ion. The effective ionic radii are then

$$R_{\rm eff}^{\pm}(N)=R_c^{\pm}(N)(1+\xi_T\Delta N),\quad \Delta N=N_m-N. \tag{12}$$

We now analyze the situations for negative and positive ions separately. In case of the negative ion, it is evident from the calculations (Fig. 1) that beyond the maximum, ΔN decreases with N. In other words, the characteristic size of the localization region of the weakly bound electron λ determines the size of the negative ion. For a simple estimate, R_c^- can be set to λ and can be considered independent of the liquid density. Then, according to Eq. (12), $R_{\rm eff}^-(N)$ is an increasing function of ΔN, and consequently a decreasing function of the liquid density N, which is consistent with the experimental data shown in Fig. 2.

We next consider the mobility of the negative ion, using the present model for the region of enhanced density and pressure near the ion. This region results in a local increase in shear viscosity. As the viscosity of the liquid $\eta(N,T)$ depends on density and temperature, $\eta(r)$ can be evaluated as a function of distance from the center of the ion.

As can be seen from Fig. 1, the local density increase is small. Moreover, the viscosity of the liquid Xe under these conditions is approximately a linear function of the liquid pressure. The local viscosity as a function of distance from the center of the ion can be written as $\eta(r)=\eta_\infty[1+(R\eta/r)^4]$, where η_∞ is the viscosity of the unperturbed liquid and $R_\eta=[(\partial\eta/\partial P)_T(\alpha F_L(N)e^2N/2\eta_\infty)]^{1/4}$ is the characteristic distance. R_η depends on properties of the liquid, and does not depend on any properties of the ion.

Previous results[13] were used to calculate negative ion mobility. The effective ion radius $R_{\rm eff}=R_c^-\varphi$, where φ is a dimensionless number that includes corrections to the Stokes formula. For a bubble (perfect slip boundary conditions) of radius R_c^- and $y_0=(R_\eta/R_c^-)^4>1$, we use the approximate expression[13]

$$\varphi(y_0)\cong\frac{8\sqrt{2}}{2\pi}y_0^{1/4}\left[1-\frac{8\sqrt{2}}{3\pi}y_0^{-3/4}\right]. \tag{13}$$

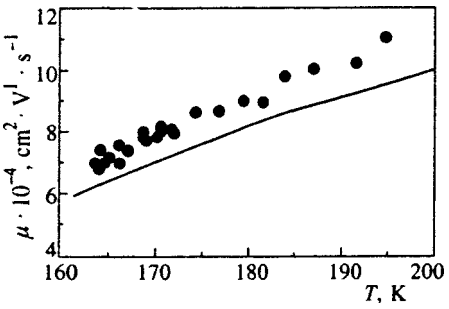

FIG. 3. Mobility of the O_2^- ion in liquid Xe on the saturation curve. The curve is calculated in accordance with the Stockes formula, and the points are the experimental results.[2]

The estimates of R_η were made at $T=170\,\mathrm{K}$ and $T=200\,\mathrm{K}$. The results were very similar for the two points. Based on the present approach for these two cases, $R_\eta \approx 8.6a_0$ and $\varphi \approx 1.56$. Our calculation of negative ion mobility by means of the modified Stokes formula (9) and (13) is presented in Fig. 3. We obtain satisfactory agreement with experiment,[2] although we consider these results to be preliminary.

For a positive ion, $R_c^+(N)$ denotes the radius of the cluster created from the surrounding atoms as a result of electrostriction. $R_c^+(N)$ is a sharp function of the liquid density, and it can be determined using the modified Atkins' model.[2] Near the triple point, the cluster size is determined by the surface tension of the liquid–solid interface, and can be written

$$[R_c^+(N)]^3 \simeq \frac{\alpha e^2}{16\pi\varepsilon_s\varepsilon_l\sigma_{sl}^t(N_s-N)^3}. \tag{14}$$

Here α is the polarizability of the Xe atom, ε_s and ε_l are permittivity of solid and liquid Xe, respectively, N_s is the density of solid Xe, and σ_{sl}^t is the surface tension of the liquid–solid interface at the triple point. Substituting (14) into (12), we have $R_{\mathrm{eff}}^+=(\zeta_1/\Delta N)+\zeta_2$, where ζ_1 and ζ_2 are weak functions of N.

Thus, the effective radius of a positive ion is a decreasing function of ΔN, and consequently it is an increasing function of the liquid density N. We obtain the same condition on the saturation curve far from the triple point, where it is possible to neglect the effect of surface tension on cluster properties. In this case, the cluster radius is given by

$$[R_c^+(N)]^4=\frac{\alpha e^2 N}{8\pi\varepsilon_s\varepsilon_l(P_s-P)} \propto \frac{\alpha e^2 N\gamma}{8\pi\varepsilon_s\varepsilon_l s^2\Delta N}. \tag{15}$$

Substitution of Eq. (14) into Eq. (12) results in an increasing function of N at small ΔN. We obtain $R_{\mathrm{eff}}^+=\zeta_3(N/\Delta N)^{1/4}+\zeta_4 N^{1/4}\Delta N^{3/4}$, where ζ_3 and ζ_4 are also weak functions of N. From these qualitative considerations it can be seen that the increase in R_{eff}^+ with N is related to the increase in the cluster radius R_c^+ despite a simultaneous decrease in the distortion of the local liquid density ΔN near the cluster. This result is in good agreement with the experimental data[2] shown in Fig. 2.

The radius of the negative ion depends weakly on liquid density, and is mainly determined by the inner state of the ion. As result, the distortion of the liquid density near the negative ion is the principal factor determining $R_{\mathrm{eff}}^-(N)$. With increasing liquid density, the role of the exchange interaction of the weakly bound electron with atoms is augmented, which results in an increase in $L(N)$ and a larger decrease in ΔN near the ion. Therefore, the present model enables us to interpret qualitative differences in the structure and mobility of negative and positive ions.

The authors are grateful to W. F. Schmidt for stimulating discussions. This work was supported by Russian Fund for Fundamental Research (Grant No 96-02-00230) and by the Deutsche Forschungsgemeinschaft [Grant No RUS 113/433/O (R,S)]. One of the authors (A.G.K.) thanks the NATO for a traveling grant.

*)E-mail: khrapak@inmech.msu.su

[1] K. R. Atkins, Phys. Rev. **116**, 1339 (1959).
[2] O. Hilt, W. F. Schmidt, and A. G. Khrapak, IEEE Trans. Dielectr. Electr. Insul. **1**, 648 (1994).
[3] M. F. Cole and R. A. Bachman, Phys. Rev. B **15**, 1388 (1977).
[4] A. G. Khrapak, W. F. Schmidt, and K. F. Volykhin, Phys. Rev. E **51**, 484 (1995).
[5] K. F. Volykhin, A. G. Khrapak, and W. F. Schmidt, Zh Éksp. Teor. Fiz. **81**, 1642 (1995) [JETP **81**, 1642 (1995)].
[6] R. A. Ferrell, Phys. Rev. **108**, 167 (1958).
[7] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics. Nonrelativistic Theory*, Pergamon Press, Oxford (1985).
[8] J. Lekner, Phys. Rev. **158**, 130 (1967).
[9] I. T. Iakubov, in *Linking the Gaseous and Condensed Phases of Matter: The Behavior of Slow Electrons*, ed. by L. G. Christophorou, E. Illenberger, and W. F. Schmidt, Plenum Press, New York (1994), p. 96.
[10] A. G. Khrapak, in *Linking the Gaseous and Condensed Phases of Matter: The Behavior of Slow Electrons*, ed. by L. G. Christophorou, E. Illenberger, and W. F. Schmidt, Plenum Press, New York (1994), p. 121.
[11] V. M. Atrazhev, I. T. Iakubov, and V. V. Pogosov, Phys. Lett. A **204**, 393 (1995).
[12] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, Pergamon Press, Oxford (1984).
[13] J. Mathews, Phys. Fluids **21**, 876 (1978).

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

Negative Ca^- and Ba^- ions of large radius on the surface and in the volume of liquid helium

P. D. Grigor'ev*) and A. M. Dyugaev

L. D. Landau Institute of Theoretical Physics, Russian Academy of Sciences, 142432 Chernogolovka, Moscow Region, Russia
Grenoble High Magnetic Field Laboratory MPI-FRF and CNRS BP 166, F-38042 Grenoble Cedex 09, France
(Submitted 17 September 1998)
Zh. Éksp. Teor. Fiz. **115**, 593–604 (February 1999)

Negative Ca^- and Ba^- ions of large radii on the surface of and in bulk liquid helium have been studied. Our results indicate that these ions are adsorbed on the helium surface. Ions on free liquid helium surfaces have not been studied previously because it was thought impossible to confine them on the surface. Ca^- and Ba^- ions have very low binding energies, therefore, like electrons, they form a bubble of large radius in bulk helium, whose energy is higher than on the surface. The behavior of ions on the surface exhibits a number of previously unknown features owing to their large masses and strong localization in the horizontal plane. Even in the absence of confining electric field, a hole is formed under an ion due to the polarization attraction between the liquid helium and the charged ion. This hole formation reduces the ion mobility by several orders of magnitude and increases its effective mass severalfold. The critical density of electrons and ions is approximately the same on the surfaces of thin and thick helium films.
[S1063-7761(99)01502-4]

1. INTRODUCTION

Numerous studies have been dedicated to properties of electrons on a helium surface (see, e.g., the monograph cited as Ref. 1). This intense interest stems from the great number of beautiful and novel effects that occur in systems of surface electrons, and also because some effects observed in semiconducting heterostructures and interesting from the standpoint of technical applications can be more conveniently studied using the electron gas on a helium surface. Ions, however, until now have been studied only inside bulk helium, because almost all the ions placed on a helium surface ''drown.'' The reason is that the polarization energy associated with the attraction of an ion to helium overwhelms other contributions unless the ion size is very large. The latter case is discussed in this paper. The binding energy of an extra electron in such an ion is negligible, so the localization radius of this electron, which determines the ion size, is very large.

Recently research into effects on helium surfaces has undergone a revival sparked by the creation of an electron gas of high density $n_e \sim 10^{11}$ cm^2 on thin films of liquid helium. Mitsura *et al.*[2] detected quantum melting of an electron Wigner crystal, which takes place when the electron density increases. The interpretation of these experimental data,[2] however, is far from certain, since other scenarios are possible. At low electron densities n_e, owing to imperfection of the substrate below the thin helium film, they may form isolated ''pools'' superposed on a long-wavelength potential relief generated by the substrate. At higher densities, however, these pools may overlap, and a connected two-dimensional electronic system is produced. In order to discriminate between these two scenarios, it might be useful to have an opportunity to switch quantum effects on and off in the electron system on the helium surface, i.e., to vary the electron mass. This trick can be used in real experiments, since there are negative Ca^- and Ba^- ions with negligible binding energies and enormous dimensions (see Table I). These ions are adsorbed on the helium surface and, owing to their large masses, can be treated as classical particles at all reasonable temperatures.

A mixture of electrons and Ca^- and Ba^- ions on a helium surface in the regime of the quantum Hall effect is a unique object for experiment.[3] The point is that the roles of shallow and charged impurities in formation of plateaus on curve of Hall resistance versus magnetic field have remained unclear thus far. It is impossible to separate the contributions of shallow and charged impurities to parameters of solid-state heterostructures tested in the regime of quantum Hall effect. Given that the helium surface is pure, i.e., it has no impurities, one can separate the effects of charged impurities by varying the temperature or density of negative ions in the mixture of electrons and Ca^- or Ba^- ions. Specifically, at low temperatures Ca^- and Ba^- ions form a Wigner crystal, which generates a periodic potential for surface electrons, thereby lifting the infinite-fold degeneracy of the electron Landau levels in a magnetic field without broadening the levels. At higher temperatures the Wigner crystal melts and the electron Landau levels are broadened.

TABLE I. Numerical calculations of bubble energy and optimal radius in bulk liquid helium and on its surface for several typical ions of large radius, with characteristics of these ions.

Ion	Electron affinity in vacuum E_0, eV	Ion size in vacuum $1/\kappa_0$, Å	Atom or molecule polarizability β, a_B^3	Turning point $r_0=(\alpha/a_B\kappa_0^2)^{1/4}$, Å	Optimal bubble radius in helium R, Å
O_2^-	−0.46	2.87	10.6	2.23	8.0
Ba^-	−0.145	4.87	270	6.5	12.0
Ca^-	−0.0245	11.84	170	9	14.1

Ion	Correction to ion energy in helium volume $E_{in}-E_0$, eV	Optimal separation between ion and surface h, Å (estimate)	Correction to ion energy on surface E_{sur} taking into account hole shape, eV (estimate)	Potential barrier for ion on surface of thick helium film Δ, eV
O_2^-	−0.024	–	–	≈0
Ba^-	0.024	14	−0.011	0.04
Ca^-	0.060	20	−0.003	0.065

2. ENERGY OF A NEGATIVE ION IN BULK HELIUM

Let us calculate the energy of negative ions in bulk helium and prove that Ca^- and Ba^- ions are adsorbed on helium surface, whereas ions of smaller radius (such as O_2^- or H^-) submerge or ''drown'' in the helium.

In bulk helium an ion forms a bubble whose energy is composed of three components:[1]

$$E_{in}(R)=4\pi\alpha R^2-\frac{\varepsilon-1}{2\varepsilon}\nu\frac{e^2}{R}+E_e(R). \tag{1}$$

The first term on the right is the energy of surface tension in the bubble (α is the surface tension coefficient). The second term is due to the polarization interaction between an electron and liquid helium with dielectric constant ϵ. The dimensionless factor ν is close to unity and depends on the charge distribution within the bubble. We have $\nu=1$ if the charge is concentrated at its center. For an electron delocalized inside the bubble, $\nu=1.35$. Taking into account the approximate shape of the outer electron wave function, let us assume $\nu=1.16$, which is accurate to within 5%. The most difficult part is the calculation of the last term, $E_e(R)$, which is the ground-state energy of the outer electron in the potential of the polarization interaction with the atom and confined within the bubble by the spherically symmetrical wall of liquid helium. This potential has not been evaluated accurately because it results from the complicated interaction of the outer electron with the rest of the electrons. We know only the binding energy of this electron to its atom and the potential shape at large separations from the atom center:

$$U(r)=-\frac{\beta e^2}{2r^4}, \tag{2}$$

where β is the atom polarizability. In their calculation of the energy of an O_2^- ion in liquid helium Volykhin *et al.*[4] used the model potential

$$U(r)=\begin{cases}+\infty, & r<b,\\ -\dfrac{\beta e^2}{2r^4}, & b<r<R,\\ +U_0, & r>R,\end{cases} \tag{3}$$

where the core radius b is derived from the binding energy E_0 of the outer electron in vacuum, and U_0 can be set to $+\infty$. Next, following Ref. 4, the segment $[b,R]$ is divided in two: one segment extending to the turning point r_0 and one beyond it. The turning point is given by

$$r_0=\left(\frac{\beta}{a_B\kappa_0^2}\right)^{1/4}, \quad \kappa_0^2\equiv-\frac{2mE_0}{\hbar^2}. \tag{4}$$

Here a_B is the Bohr radius and m is the electron mass. On the first segment $[b,r_0]$ we neglect the binding energy in comparison with the potential, while on the second segment $[r_0,R]$ we neglect the potential in comparison with the binding energy. Afterwards the solutions of the Schrödinger equations are matched at the point r_0.

In this approximation, one can derive the function $E_e(R)$ without solving the Schrödinger equation on the interval $(0,r_0)$, and an important point is that no assumption is made about the potential shape at $r<r_0$, thus the solution is made notably simpler and more general (as will be shown below, the potential described by Eq. (3) cannot be used in the case of Ca^- and Ba^- ions). Immersion of an ion in helium does not affect the potential for $r<r_0$ but shifts the outer electron energy. If on the interval $(0,r_0)$ we neglect in comparison with the potential not the energy but only the energy shift due to the helium environment, leaving alone the total binding energy, the helium surface does not change the wave function at $r<r_0$. Thus, the logarithmic derivative of the wave function $\psi(r)$ to the left of the joining point does not change as a result of immersion in the helium:

$$\frac{\chi'(r_0-0)}{\chi(r_0-0)}=-\kappa_0=-\frac{\sqrt{-2mE_0}}{\hbar},$$

where $\chi(r)=\psi(r)r$. To the right of the matching point, with due account of the condition $|E_e|\gg|U(r)|$, the wave function has the form

$$\psi=\frac{1}{r}\sinh[\kappa(R-r)],$$

and turns to zero on the bubble surface. By equating the logarithmic derivatives at point r_0 we obtain an equation for the energy $E_e(R)$:

$$\frac{\chi'(r_0)}{\chi(r_0)}=-\kappa_0=-\kappa\coth[\kappa(R-r_0)],$$

$$E_e(R)\equiv-\frac{\hbar^2\kappa^2}{2m}. \tag{5}$$

The neglect of the binding energy is legitimate at all $r<r_0$ only if $r_0\kappa\ll1$ holds. In fact, we have $r_0\kappa\sim1$, so Eq. (5) is only approximately correct. More accurate calculations can be performed numerically, given the exact shape of the potential. Before proceeding to these calculations, let us try to refine Eq. (5) on the basis of the following considerations. The change in the logarithmic derivative of the wave function on the left of the turning point due to the energy shift has the form

$$\Delta\frac{\chi'}{\chi}\bigg|_{r=r_0-0}=\frac{\Delta\chi'}{\chi}-\frac{\chi'\Delta\chi}{\chi^2}\bigg|_{r=r_0-0}. \tag{6}$$

The first term

$$\frac{\Delta[\chi'(0)+\int_0^{r_0}\chi''dr]}{\chi(r_0)}=\frac{\Delta\left[\int_0^{r_0}\frac{2m}{\hbar^2}(U(r)-E)\chi dr\right]}{\chi(r_0)}$$

$$\approx-\frac{2m}{\hbar^2}\frac{\int_0^{r_0}\chi dr}{\chi(r_0)}\Delta E.$$

The last integral on the right requires knowledge of the wave function on the segment $[0,r_0]$. Gribakin *et al.*[5] performed an accurate numerical calculation of interaction between an extra electron and the rest of the electrons in Ca^-, Ba^-, and Sr^- ions in vacuum. As a result, they determined the wave functions of the extra electrons in these ions in vacuum. The calculated binding energies[5] are in reasonable agreement with experimental data.[6,7] Assume that the wave function on the interval $(0,r_0)$ changes little when an ion is immersed in helium. Then, using the outer electron wave function,[5] we obtain

$$\frac{\int_0^{r_0}\chi dr}{\chi(r_0)}\approx r_0.$$

The second term on the right of Eq. (6) is small, since the point r_0 is close to the maximum of $\chi(r)$, where the wave function is determined largely by the normalization condition and is little affected by the energy shift. With the correction given by Eq. (6), we replace Eq. (5) by

$$\frac{\chi'(r_0)}{\chi(r_0)}=-\kappa_0-(\kappa_0^2-\kappa^2)r_0=-\kappa\coth[\kappa(R-r_0)]. \tag{7}$$

In the limit $\kappa(R-r_0)\gg1$ we can find an analytical solution of this equation:

$$E_e(R)-E_0=-E_0\frac{4\exp[-2\kappa_0(R-r_0)]}{1+2\kappa_0r_0}. \tag{8}$$

For arbitrary $\kappa(R-r_0)$ Eq. (7) cannot be solved analytically with respect to κ. We can find only the inverse function

$$R(\kappa)=r_0+\frac{1}{\kappa}\tanh^{-1}\left(\frac{\kappa_0+r_0(\kappa_0^2-\kappa^2)}{\kappa}\right).$$

The only difference between Eq. (8) and the corresponding expression derived from Eq. (5) is the factor $(1+2\kappa_0r_0)^{-1}$. This factor is 0.3 for Ba^- and 0.4 for Ca^-. Thus, Eq. (7) yields more accurate results, although it is also approximate.

At low temperatures the helium density on the bubble boundary around O_2^- increases locally,[4] which leads to a higher polarization contribution in Eq. (1). This effect, however, need not be taken into account because we have already neglected an effect of the same order of magnitude but with the opposite sign due to the difference between the radii R in the polarization and exchange interactions with the surrounding liquid helium. Moreover, the polarization pressure $P=\beta e^2/2vR^4$ is a rapidly decaying function with range R, and for the large Ca^- and Ba^- ions of interest to us these effects are very weak.

In order to improve the reliability of our results, we have performed a numerical calculation of the energy $E_e(R)$. In this connection, let us discuss the selection of the potential acting on the outer electron. Equation (3) is a poor approximation for Ca^- and Ba^- ions. Specifically, since these atoms have large polarizabilities $\beta_{Ba}=270a_B^3$ and $\beta_{Ca}=170a_B^3\gg\beta_{O_2}=10.6a_B^3$, with a potential like that defined by Eq. (3) we should set the boundary of the hard core at a distance $b\approx4$ Å from the atom center so that the outer electron energy be equal to the binding energy in vacuum. It is clear that the range of the exchange repulsion should be approximately equal to the atom size, i.e., ~1 Å. Elaborate numerical calculations[5] show that the outermost node of the wave function is in fact at 1.2 Å in Ba^- and 0.8 Å in Ca^-. So, the inner wall in the model potential should be set at these distances from the atom center. In order to satisfy this condition, we selected the potential in the form

$$U(r)=\begin{cases}+\infty, & r<b,\\ -\dfrac{\beta e^2}{2(r^2+a^2)^2}, & b<r<R,\\ +\infty, & r>R,\end{cases} \tag{9}$$

where $b_{Ba}=1.2$ Å, $b_{Ca}=0.8$ Å, and $b_{O_2}=0.482$ Å. The parameter a is derived by equating the electron energy at $R=\infty$ in this potential to the ion binding energy in vacuum: $a_{Ba}=2.335$ Å, $a_{Ca}=2.593$ Å, and $a_{O_2}=0$. This potential is closer to the real one than that defined by Eq. (3) and also admits an analytical solution in the limit $E\to0$. Then, by

solving numerically the Schrödinger equation with this potential, we obtain $E_e(R)$. It is well approximated by the empirical formula

$$E_e(R)-E_0=-E_0\frac{4\exp[-2\kappa_0(R-r_0)]}{1+2\kappa_0 r_0}\coth\frac{\kappa_0 R^2}{3r_0}. \tag{10}$$

In order to determine the optimal bubble radius R_0 and the ion energy in helium $E_e(R_0)$, we should solve the equation

$$\frac{\partial E_{\rm in}}{\partial R}=0,$$

where $E_{\rm in}$ is given by Eq. (1) and $E_e(R)$ by Eq. (10). The results are listed in Table I for several typical negative ions of large radii alongside parameters of these ions. The parameters $1/\kappa_0$ and r_0 were derived from available measurements[6,7] of binding energy E_0 of negative ions and atomic polarizabilities (the latter are the same as in Refs. 4 and 5). The discrepancy between energy corrections calculated from Eq. (7) and numerically is 10 to 15%, while the corresponding dsicrepancy between calculations of the optimal radius is less than 5%. These are the error estimates in our calculations.

In order to avoid confusion, note that the ion radius in the third column of Table I is the separation from the atom center at which the absolute value of the wave function drops by a factor $\sim e$, whereas the exponential tail of this function extends much further. The wave function goes to zero on the bubble boundary. Therefore the optimal bubble radius is the distance at which the wave function tail is cut off. This leads to the considerable difference between $1/\kappa_0$ and R, which increases with the ion binding energy in vacuum, i.e., with the energy gained by cutting off the wave function tail.

Table I clearly shows that the energies of Ca^- and Ba^- ions in helium are higher than in vacuum. This result is easily understandable. The electron affinities in these atoms are negligible (see Table I), and the outer electron is almost free (in Ca^- the localization radius is ~ 12 Å). Like free electrons, these electrons form a large bubble in bulk helium, and the main contribution to the bubble energy is due not to the polarization attraction to helium (as for small ions), but to the energy of electron compression by the bubble boundary and surface tension energy. The additional energy of a free electron in helium is $\Delta\sim 0.15$ eV, and for Ca^- $\Delta\approx 0.06$ eV. The radius of the bubble around a free electron is $R_-=17$ Å, while around Ca^- it is $R\approx 14$ Å. Thus, the bubble parameters around a free electron and around a Ca^- ion are similar.

We have thus proven that Ca^- and Ba^- are adsorbed on a helium surface, where their energy is lower than in vacuum owing to long-range polarization forces of attraction to the gas–liquid helium interface. Owing to their large masses and initial localization in the horizontal plane, surface ions have a number of properties which are unusual compared with those of surface electrons. Let us proceed to a discussion of these properties.

3. PROPERTIES OF NEGATIVE IONS ON A HELIUM SURFACE

Even in the absence of a confining electric field, a negative ion on a helium surface generates a static surface deformation around it, which we term a hole, by analogy with the deformation produced by electrons. Recall that an electron produces a hole only in a strong electric field.[1] Hole formation in a zero confining field is caused by the polarization interaction between the ion and liquid helium.

Let us assume first that the helium surface deformation is small and calculate the height at which the ion levitates. This effective ion size h can be calculated similarly to the optimal radius of the bubble in helium, i.e., by minimizing the full energy on the surface $E_{\rm sur}$, which contains, in the case of a plane surface, only two terms, namely, the energy of polarization interaction between the ion and helium $(e^2/4z)(\varepsilon-1)/(\varepsilon+1)$ and the energy E_b due to the cut-off of the outer electron wave function on the helium surface. In estimating the latter, let us use the approximation of a uniform distribution over solid angle:

$$E_b(h)\approx\frac{1}{4\pi}\int[E_e(R)-E_0]d\Omega,\quad R(\Omega)=\frac{h}{\cos\theta}, \tag{11}$$

whence

$$E_b(h)=\frac{1}{2}\int_0^1\left[E_e\left(\frac{h}{\cos\theta}\right)-E_0\right]d(\cos\theta)$$

$$\approx\frac{1}{4\kappa_0 h}[E_e(h)-E_0],$$

where $E_e(R)$ is given by Eq. (10). By solving the equation

$$\frac{\partial E_{\rm sur}}{\partial h}=0,$$

we calculate the optimal separation between the ion and helium surface: $h_{\rm Ba}=14$ Å and $h_{\rm Ca}=21$ Å. Let us analyze the surface shape $\xi(r)$. The hole shape under the ion should satisfy the conditions

$$\xi'(0)=0,\quad \xi''(0)\leqslant\frac{1}{h}.$$

It should also be localized mostly within the ion radius, since only in this case can the energy of polarization attraction be significant. The surface tension energy E_α for such a hole is proportional to the square of its depth a^2 at the center, $E_\alpha\propto a^2$, and is little affected by the hole shape. For example, if the hole shape is fitted to the ion radius h:

$$\xi(r)=\begin{cases}h-a-\sqrt{h^2-r^2}, & r<\sqrt{h^2-(h-a)^2},\\ \xi=0, & r\geqslant\sqrt{h^2-(h-a)^2},\end{cases}$$

its surface tension energy is $E_\alpha=\pi a^2\alpha$. Such a deformation is not smooth around the edges. For a smooth hole, which is energetically preferable, of the form

$$\xi(r)=-a\exp\left(-\frac{r^2}{2h^2}\right),$$

the surface energy loss due to attraction is smaller by only a factor of 2.

When a hole forms the outer electron wave function is additionally compressed. In order to minimize the energy change due to this compression, it suffices to have an almost flat surface at the center of the hole (this region makes the major contribution to the correction E_b to the outer electron energy). This condition is satisfied by holes with $\xi''=0$. For example, a surface deformation described by the function

$$\xi(r)=-a\exp\left(-\frac{r^4}{2h^4}\right)$$

generates surface energy $E_\alpha=\pi a^2\alpha$. The increase in the energy of the polarization interaction of the ion charge in such a hole is

$$E_{\rm el}\approx -a\int_h^\infty 2\pi r\,dr\,\frac{e^2(\varepsilon-1)}{4\pi r^4}=-e^2a\,\frac{\varepsilon-1}{4h^2}.$$

Since $E_\alpha\propto a^2$ and $E_{\rm el}\propto -a$, at small a one always has $E_\alpha+E_{\rm el}<0$. Let us calculate the optimal hole depth a by minimizing

$$E_{\rm tot}=E_\alpha+E_{\rm el}\approx\pi a^2\alpha-\frac{e^2a}{4h^2}(\varepsilon-1),$$

$$E'_{\rm tot}(a)=2\pi a\alpha-e^2\frac{\varepsilon-1}{4h^2}=0,$$

whence it follows that

$$a=\frac{e^2(\varepsilon-1)}{8\pi\alpha h^2}\approx\begin{cases}7.4\ \text{Å} & \text{for Ba}^-,\\ 3.3\ \text{Å} & \text{for Ca}^-.\end{cases}\qquad(12)$$

The total change in the ion energy is

$$E_{\rm tot}=-\frac{e^2a(\varepsilon-1)}{8h^2}\approx\begin{cases}4\times10^{-3}\ \text{eV} & \text{for Ba}^-,\\ 0.8\times10^{-3}\ \text{eV} & \text{for Ca}^-.\end{cases}$$

Let us estimate the mobility and effective mass of Ca^- and Ba^- ions on helium surface. Here we use the approximation suggested by Shikin and Monarkha[8] for studying dynamic properties of surface electrons. The local deformation of the helium surface adiabatically follows the moving ion and generates a certain field of hydrodynamic velocities as the ion moves across the surface. Although the lateral size l of the hole under the ion is smaller than under a free electron, at temperatures above the superfluid transition ($T>2.18$ K) this size is, nonetheless, larger than the mean free path of thermal excitations. Therefore the hydrodynamic approximation applies.

In a reference frame traveling with the ion at velocity V_0, the full velocity of the liquid is $V=V_0+v$, where v is the field of velocities generated by the hole. The hole is rather flat [$\xi'(r)\ll1$], so we have $v\ll V_0$ and the boundary condition

$$\mathbf{V}\cdot\mathbf{n}=0$$

transforms to

$$v_z|_{z=0}=V_0\xi'(r)\cos\theta.$$

In the limit of large Reynolds numbers, when viscosity can be neglected in the hydrodynamic equations, the velocity field is determined by the equations

$$v=\nabla\varphi,\quad \Delta\varphi=0,\quad \varphi_{r,z\to\infty}\to0,$$

$$\left.\frac{\partial\varphi}{\partial z}\right|_{z=0}=V_0\xi'(r)\cos\theta.\qquad(13)$$

In our calculations, we have selected the hole shape

$$\xi(r)=\frac{a}{1+r^2/h^2},$$

which satisfies all necessary conditions for the hole under the ion. The calculation uses the expansion in terms of Bessel functions:

$$\xi(r)=\int_0^\infty G(\omega)J_0(\omega r)\omega\,d\omega,$$

where

$$G(\omega)=\int_0^\infty\xi(r)J_0(\omega r)r\,dr=ah^2K_0(\omega h).$$

The solution of Eq. (13) has the form

$$\chi(r,z)=V_0\cos\theta\int_0^\infty G(\omega)e^{-\omega z}J_1(\omega r)\omega\,d\omega.$$

Here J_0, J_1, and K_0 are Bessel functions.

The aggregate mass of a surface ion is given by the expression

$$M=\frac{\rho}{V_0^2}\int v^2 d^3V$$

$$=\frac{\rho}{V_0^2}\int \nabla(\varphi\nabla\varphi)d^3V=-\frac{\rho}{V_0^2}\int \varphi\frac{\partial\varphi}{\partial\mathbf{n}}dS$$

$$\approx-\frac{\rho}{V_0^2}\int \varphi\frac{\partial\varphi}{\partial z}dS=\rho\int r\,dr\int_0^{2\pi}d\theta\cos^2\theta$$

$$\times\int_0^\infty G(\omega)J_1(\omega r)\omega\,d\omega\int_0^\infty G(\omega')J_0(\omega' r)\omega'^2 d\omega'$$

$$=\pi\rho\int_0^\infty G^2(\omega)\omega^2 d\omega=\pi\rho a^2h^4$$

$$\times\int_0^\infty K_0^2(\omega h)\omega^2 d\omega=\rho a^2 h\frac{\pi^3}{32}$$

$$\approx\begin{cases}10^{-22}g & \text{for } Ba^-\\ 0.3\times10^{-22}g & \text{for } Ca^-,\end{cases}$$

where ρ is the liquid helium density and S is a surface element. Thus, the polarization correction to the ion mass is approximately equal to the ion mass.

In the mobility calculation we use the energy balance equation (Ref. 9, p. 79) $-eE_{||}V_0=\dot{W}$. If $\Delta\varphi=0$ holds, the energy dissipation is described by the formula[9]

$$\dot{W}=\eta\int\frac{\partial v^2}{\partial\mathbf{n}}dS\approx 2\eta V_0^2$$

$$\times\int_0^\infty(-\omega)G(\omega)\omega^2 d\omega\int_0^\infty G(\omega')\omega'^2 d\omega'$$

$$\times\int_0^{2\pi}d\theta\int_0^\infty r\,dr\Bigg[J_1'(\omega r)J_1'(\omega' r)\cos^2\theta$$

$$+J_1(\omega r)J_1(\omega' r)\frac{\sin^2\theta}{\omega\omega' r^2}+J_1(\omega r)J_1(\omega' r)\cos^2\theta\Bigg]$$

$$\approx-4\pi\eta V_0^2\int_0^\infty G^2(\omega)\omega^4 d\omega,$$

where η is the viscosity. In our specific case

$$\int_0^\infty G^2(\omega)\omega^4 d\omega=a^2h^4\int_0^\infty K_0^2(\omega h)\omega^4 d\omega=\frac{27\pi^2}{2^9}.$$

Thus, the mobility is

$$\mu=\frac{V_0}{eE_\parallel}=\left(4\pi\eta\int_0^\infty G^2(\omega)\omega^4 d\omega\right)^{-1}=\frac{128}{27\pi^3}\frac{h}{\eta a^2}$$

$$\approx\begin{cases}1.3\times10^{11}\ \text{cm/g} & \text{for } Ba^-\\ 10^{12}\ \text{cm/g} & \text{for } Ca^-\end{cases}$$

$$=\begin{cases}0.21\ \text{cm}^2/\text{s}\cdot\text{V} & \text{for } Ba^-\\ 1.6\ \text{cm}^2/\text{s}\cdot\text{V} & \text{for } Ca^-,\end{cases}$$

where the helium viscosity at $T=4$ K is assumed to be $\eta=3\cdot10^{-5}$ g/cm·s.

Note that, if we consider a hole of a different shape but with the equal depth and lateral size, the result will be very close. For example, for $\xi(r)=a\exp(-r/h)$ we have

$$G(\omega)=\frac{a/h}{[(1/h)^2+\omega^2]^{3/2}}.$$

Further calculations similar to those described above yield

$$M=\frac{\pi^2}{16}\rho h a^2,\quad \mu=\frac{4h}{3\pi^2\eta a^2}.$$

The difference between the parameters of holes of different shapes is a factor close to unity. However, if we had used the same hole shape as that under the electron confined by external electric field but with different parameters so that the vertical and lateral sizes of the hole under the ion were identical, we would obtain markedly different mobilities and effective masses. This fact clearly shows the difference between the shapes of the helium surface deformations induced by electrons and ions.

The external confining field generates additional surface deformation under an ion. As in the case of an electron, this deformation has a lateral size of order the gravitational capillary length:

$$\frac{1}{\kappa}=\sqrt{\frac{\alpha}{\rho g}}\gg h,$$

i.e., many times larger than the hole size. Therefore, the surface tension energy of these two deformations are simply added, and the deformations are superimposed independently. The electrostatic hole depth is given by Eq. (13) in Ref. 8 by substituting $L\to h$, i.e., the ion size, which is the only difference from the case of electrons, affects the hole depth weakly, since this parameter enters in the argument of the logarithm.

The expression for the critical density of charged particles on the surface of a thick helium film,

$$n^{\max}=\sqrt{\frac{\sqrt{\rho g\alpha}}{2\pi e^2}},\tag{14}$$

was obtained in the metallic approximation.[10] The only condition that was used in deriving Eq. (14) is that the helium surface should be equipotential. In the case of ions, this condition holds, because the surface vibration frequency near the critical concentration $n^{\max}$ (i.e., the energy of surface excitations divided by Planck's constant) is so small that ions, notwithstanding their low mobility, are mobile enough to follow changes in the potential. Therefore, Eq. (14) also applies to ions on a helium surface.

On a thin film surface, electrons are strongly localized in a hole owing to the strong field of image forces and, like ions, have a low mobility. Therefore, the difference between electrons and ions becomes insignificant. The minimal thickness of a charged helium film,

$$d_{\min}=\left(\frac{27\Lambda_1}{8\pi\alpha}\ln\frac{1}{\tilde{\kappa}l_\xi}\right)^{1/3}, \tag{15}$$

is approximately the same for ions and electrons.[1)] The small difference due to the difference between electron and ion sizes l_ξ is offset by the ions' inability to tunnel through the film. Along with $d_{\min}$, the critical densities of charged particles on a thin helium film are also very similar.

We are indebted to V. B. Shikin and Yu. N. Ovchinnikov for helpful discussions. The work was supported by Russian Fund for Fundamental Research (Grant 96-02-18168) and INTAS-RFBR (Grant 95-553).

*)E-mail: pashag@itp.ac.ru

1)Equation (15) was derived in Ref. 1, Sec. 15.3 using the same notation.

[1] V. B. Shikin and Yu. P. Monarkha, *Two-Dimensional Charged Systems in Helium* [in Russian], Nauka, Moscow (1989).
[2] G. Mistura, T. Gunzler, B. Bitnar *et al.*, Surf. Sci. **361–362**, 831 (1996).
[3] *The Quantum Hall Effect*, ed. by R. E. Prange and S. M. Girvin, Springer (1987).
[4] K. F. Volykhin, A. G. Khrapak, and V. F. Shmidt, Zh. Éksp. Teor. Fiz. **108**, 1642 (1995) [JETP **81**, 901 (1995)].
[5] G. F. Gribakin, B. V. Gul'tsev *et al.*, J. Phys. B **23**, 4505 (1990).
[6] V. V. Petrunin, H. H. Andersen, P. Balling, and T. Andersen, Phys. Rev. Lett. **76**, 744 (1996).
[7] V. V. Petrunin, J. D. Voldstad *et al.*, Phys. Rev. Lett. **75**, 1911 (1995).
[8] V. B. Shikin and Yu. P. Monarkha, Zh. Éksp. Teor. Fiz. **65**, 751 (1973) [Sov. Phys. JETP **38**, 373 (1973)].
[9] L. D. Landau and E. M. Lifshitz, *FLuid Mechanics*, Pergamon (1980).
[10] L. P. Gor'kov and D. M. Chernikova, JETP Lett. **18**, 68 (1973).

Translation provided by the Russian Editorial office.

Effect of a weak magnetic field on the state of structural defects and the plasticity of ionic crystals

Yu. I. Golovin*) and R. B. Morgunov

G. R. Derzhavin State University, 392622 Tambov, Russia
(Submitted 23 January 1998)
Zh. Éksp. Teor. Fiz. **115**, 605–623 (February 1999)

The aim of this work is to examine the influence of a weak (on the energy scale) magnetic field on the state of dislocations and point defects in ionic crystals. It is found that complex point defects existing in a metastable state are sensitive to a magnetic field $B\sim 1$ T. The contributions are identified, and the kinetics of various types of reactions within the structural defects and between them leading to plastification of the crystals in a magnetic field are determined. The effect of light on the sensitivity of the point defects to a magnetic field is described, and the spectral characteristics of this effect are determined. A resonant effect of the combined action of a weak constant magnetic field and a high-frequency magnetic field on the dislocation mobility is found to occur when these fields satisfy the conditions of electron paramagnetic resonance. © *1999 American Institute of Physics.* [S1063-7761(99)01602-9]

1. INTRODUCTION

In recent years it has been reliably determined that a magnetic field $B\sim 1$ T can have a strong effect on the dislocational plasticity of ionic crystals. The action of a magnetic field has been observed to shift dislocations in unloaded crystals,[1] increase their segment lengths in the presence of loading,[2] alter their microhardness,[3,4] increase the level of dislocational internal friction,[5] and lower their flow limit.[6] All these phenomena are commonly referred to as magnetoplastic effects. It has been found that these effects are sensitive to the type of dominant impurity in the crystal[7,8] and to x-ray irradiation.[9,10] This fact has elicited the hypothesis of an electronic nature of these effects in ionic crystals[11,12] and the suggestion of a possible mechanism of the action of a magnetic field on their plasticity.[13–15]

The difficulties in explaining the magnetoplastic effect are connected with the fact that the energy communicated by a field with $B\sim 1$ T to one paramagnetic center (which in ionic crystals can be identified with some impurity defect or an electron localized on a dislocation) is $\Delta U\approx \mu_B B\sim 10^{-4}$ eV (where μ_B is the Bohr magneton). At temperatures $T\sim 300$ K, which are typical for these experiments, $\Delta U\ll kT$. Therefore, under conditions near thermodynamic equilibrium it cannot be expected that a magnetic field with $B\sim 1$ T can affect the plastic properties of nonferromagnetic crystals.

However, at present a wide class of electronic processes are known which are sensitive to a magnetic field, in particular, magnetically sensitive chemical reactions involving radicals to which the foregoing thermodynamic arguments cannot be extended, by virtue of the short lifetimes of their intermediate spin-dependent steps, during which thermal fluctuations ''do not have enough time'' to destroy the correlation of the spins in the particle pairs.[16–20] The presence of paramagnetic properties in a number of structural defects in ionic crystals gives reason to believe that short-lived radical reactions between defects can take place during plastic deformation that are sensitive to a weak magnetic field.

The aim of the present work is to identify the thermodynamic reasons for the effect of a weak magnetic field on the state of the structural defects and the plasticity of ionic crystals, and determine the type of magnetically sensitive processes and the defects participating in them.

2. TECHNIQUE

To exhibit the various aspects of the effect of a magnetic field on the state of the defects and the plasticity of nominally pure crystals of NaCl, LiF, and KCl, we employed a set of complementary techniques. Changes in the characteristics of the mobility of individual edge dislocations as a result of the action of a magnetic field (mean velocity, size and number of the jumps executed by them, duration of the pauses between jumps, etc.) were observed using methods of continuous or double chemical etching by measuring 100–400 dislocation segment lengths and the sizes of the etch pits for each point on the graphs. Here we initiated motion of freshly introduced (by a scratch) dislocations by two fundamentally different techniques: by a mechanical compression pulse which was the same in all of the runs (duration 0.2 s and amplitude 0.5 MPa for NaCl) and by exposing the samples to a magnetic field in the absence of an external mechanical load. As will become clear from what follows, fundamentally different characteristics of the mobility of the dislocations are produced by these two different techniques.

Exposure of the sample to a magnetic field before and after introduction of the dislocations prior to application of the mechanical load or in combination with it, exposing the sample to light, and other variations of the conditions of the experiment allowed us to identify the contributions of the various factors to the magnetoplastic effect (in particular, the

role of changes in the states of the point defects and of the dislocations themselves in a magnetic field) and to bring out the possible multistep character of the processes of field-stimulated modification of defects.

We examined the effect of a magnetic field on the macroplastic characteristics of crystals in a specially designed deforming machine with quartz rods and supports providing conditions of ''ideally soft'' loading with linearly increasing (in time) mechanical compression loads $\sigma=\alpha t$, where the constant $\alpha=(1-10)\times 10^4$ Pa/s was held fixed during the course of a run. This technique, in contrast to the traditional technique of hard loading (in which the linearly time-dependent relative deformation ε is prescribed), allows one to eliminate negative feedback between the machine and the sample, in the presence of which the softening initiated by the external action reduces the load on the sample and upon cessation of plastic flow can ''mask'' the softening effect. In addition, the soft technique allows one to directly ascertain that in a magnetic field retraction of the rods, heating of the sample, and other side effects capable of distorting the measurement results are absent.

A constant magnetic field B up to 2.2 T was created with an electromagnet. The magnetic-field pulses had the shape of a half-period of a sine wave with amplitude 7 T and duration 10^{-2} s and were generated in a solenoid of few turns by a thyristor generator.

3. RESULTS

1. In the first series of experiments we investigated motion of dislocations in NaCl crystals not loaded by external forces, initiated by a magnetic field pulse with $B=7$ T. To elucidate the role of internal stresses and the possible effect on the mobility of the edge dislocations of the solenoidal electric field arising when the magnetic field is switched on and off, we measured the segment lengths L^* and the fraction of dislocations d shifted in the four possible crystallographic directions $[110]$, $[\bar{1}10]$, $[1\bar{1}0]$, and $[\bar{1}\bar{1}0]$. We found that regardless of the orientation of the magnetic field **B** relative to the crystallographic axes (we investigated three directions: $[100]$, $[110]$, and $[001]$) d and L^* were identical for the four groups of dislocations within the limits of accuracy of the experiments and were equal to $d=0.25\pm 0.03$ and $L^*=26\pm 1$ μm. Qualitatively, an analogous result was also obtained for motion of dislocations occurring in these crystals when not acted on by a magnetic field, as a result of two successive chemical etches. In this case $d=0.25\pm 0.03$ and $L_0=9\pm 1$ μm [Fig. 1b]. Of course, in the absence of a magnetic field the dislocations could move only under the action of the random internal stress fields, which should be cancel out in an unloaded crystal. The results obtained in our experiments with a magnetic field indicate that the action of the solenoidal electric field for the magnetic field varying in time at the rate $\sim 10^3$ T/s is negligibly small, and the magnetic field indeed exerts no motive force on the dislocations. It only facilitates depinning of dislocations from pinning centers, and their motion takes place as in the runs without a magnetic field under the action of internal stresses.

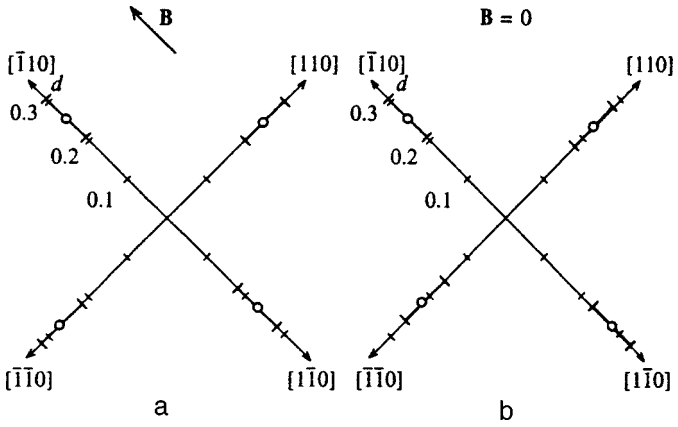

FIG. 1. Fraction of dislocations d shifted in the four possible directions in a crystal of NaCl: a) in crystals treated with a magnetic field pulse, b) in the control runs.

To reveal the kinetic features of magnetic-pulse stimulated motion of the dislocations, we used the continuous etching method. Toward this end, a sample with freshly introduced dislocations was placed in a capsule with a ''slow'' etchant in which a dislocation pit 3 μm in size was formed after a 10-second etch. The crystal, still immersed in the etchant, was exposed to a magnetic field pulse and 50 s later was pulled out of the etchant. In the control runs without a magnetic field, the crystal was kept in the etchant for the same amount of time as in the runs with the magnetic field. We found that in both cases during the entire time the sample was kept in the etchant the dislocations execute intermittent motion with several stops and starts (pauses). However, in the crystals subjected to magnetic-field treatment the velocity of the dislocations between pauses, the number of jumps, and the total time of the motion, measured by the technique described in Ref. 21, were substantially greater than in the control runs (Fig. 2). Each dislocation pit, standing as evidence of a serial pause of the dislocation, was formed during a time $\sim 5-10$ s, which is substantially exceeds the duration of a magnetic-field pulse (10^{-2} s). Consequently, the facilitated motion of the dislocations took place over an extended time after the magnetic field was already switched off, and almost the entire segment length was formed in its absence. An

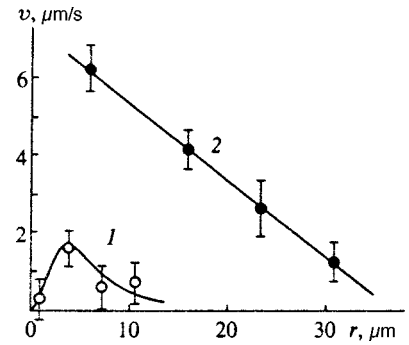

FIG. 2. Mean velocity of the dislocations v in NaCl crystals as a function of the traversed distance r: *1* — motion of dislocations in the control runs without a magnetic field, *2* — motion of dislocations stimulated by magnetic-field pulses with $B=7$ T.

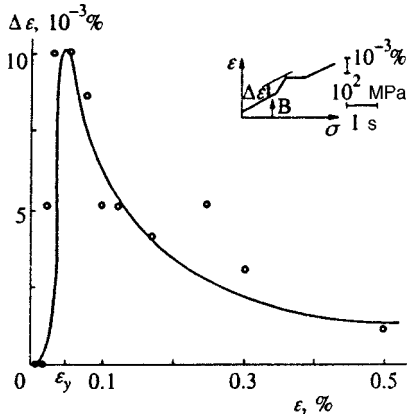

FIG. 3. The deformation jump $\Delta\varepsilon$ caused by a magnetic field pulse with $B=7$ T in NaCl crystals with flow limit ε_y versus the total deformation of the crystals ε. Inset shows a typical fragment of the deformation diagram at the instant the magnetic field is switched on.

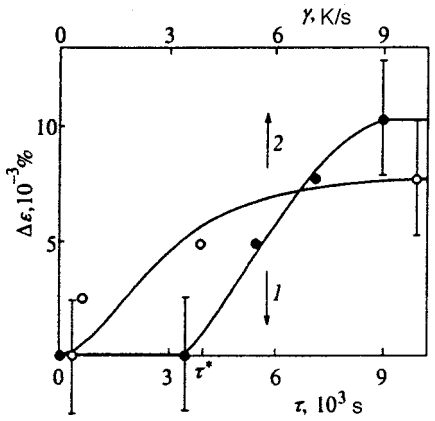

FIG. 4. Dependence of the relative deformation jump $\Delta\varepsilon$ caused by a magnetic field pulse with $B=7$ T in pre-quenched NaCl crystals having already reached a deformation of 0.15% on their preliminary storage time τ at $T=700$ K (average rate of subsequent cooling 9 K/s) (*1*) and on the average quench rate γ for $\tau=10^4$ s (*2*).

increased velocity of the dislocations between pauses, v, was also observed at distances $r\approx 10-30\ \mu$m from their original location (Fig. 2). This is indicative of magnetic-pulse-induced residual changes in the crystal, which even after the field is switched off continue to have an effect on the mobility of the dislocations.

2. To elucidate the role of a magnetic field in the depinning of dislocations from pinning centers under conditions of macro-deformation, we carried out a second series of experiments. The effect of a magnetic field on the macroplastic deformation of NaCl crystals created by a linearly increasing (in time) mechanical compression stress could be assessed from measurements of the deformation curve $\varepsilon(\sigma)$ made after switching on a magnetic-field pulse with strength $B=7$ T, directed along the compression axis. Switching on a magnetic field in the elastic region did not produce any changes in the deformation diagram. Switching on the field not long before reaching the flow limit ε_y decreased the latter. Switching on the field at the stage of easy slipping brought about a discontinuous increase in ε, i.e., a short-lived softening of the crystal took place. The maximum size of the jump $\Delta\varepsilon$ was observed right after the flow limit ε_y was reached (Fig. 3).

The deformation jump took place not while the field was acting, but only 100–150 ms after termination of the magnetic-field pulse. For 5–7 s after the magnetic-field-induced jump, the plastic flow rate was decreased in comparison with its value before the field was switched on. Switching the field on during this time had a significantly weaker effect on the plasticity (or simply no effect was recorded). The delay in the softening discontinuity relative to the magnetic-field exposure and the reduced sensitivity of the crystal to the field after the jump indicate that under conditions of macroplastic deformation an after-effect of the magnetic field is in fact observed, caused by residual changes in the state of the structural defects.

The narrow range of deformations near the flow limit in which the magnetoplastic effect is manifested implies that the magnetic field efficiently facilitates the surmounting of point defects by dislocations and has a weak effect (or no effect) on the interaction of slip dislocations with a forest dislocations. This is also indicated by the results of experiments in which we used two types of NaCl crystals: 1) crystals stored for several years at $T=293$ K, and 2) crystals quenched from 700 K with a cooling rate $\gamma>4$ K/s. During deformation of crystals of the first type, repeated switching of the magnetic field had no effect on the plastic flow rate of the crystals within the limits of accuracy of the experiments. If the crystals were subjected to quenching before being deformed, the process of their plastic flow became sensitive to switching the field on. We found that to elicit the effect of quenching, it was necessary to hold the crystals at $T=700$ K for a time $\tau>\tau^*=4\times10^3$ s. For $\tau<\tau^*$ the plastic properties of the crystals remained insensitive to a magnetic field (Fig. 4). The efficiency of temperature sensitization of the crystals to a magnetic field also depended on their average cooling rate γ: when it was lowered to 0.1 K/s heat treatment did not increase the sensitivity of the crystals to a magnetic field (Fig. 4). Note that in unquenched crystals stored under uncontrolled conditions a less systematic magnetoplastic effect was recorded and also without preliminary heat treatment; however, in this case it was considerably weaker.

3. Identification of the mechanisms of the softening action of a magnetic field requires that we determine the objects in the crystal that are exposed to its effect. Since it is unlikely that the physical properties of the defect-free regions of ionic crystals can be altered in a weak magnetic field, the softening of the crystals can be explained by several possible factors: a) the magnetic field alters the state of the point defects, b) a magnetic field alters the state of the dislocations, and c) the very process by which a dislocation becomes depinned from a pinning center is sensitive to a magnetic field.

In order to determine which of these possibilities are indeed realized, we performed a third series of experiments in which the plastic properties of the crystals were assessed on the basis of the segment lengths of individual dislocations acted on by a calibrated mechanical compression pulse, andt he crystals were exposed to a magnetic field with $B=1$ T I)

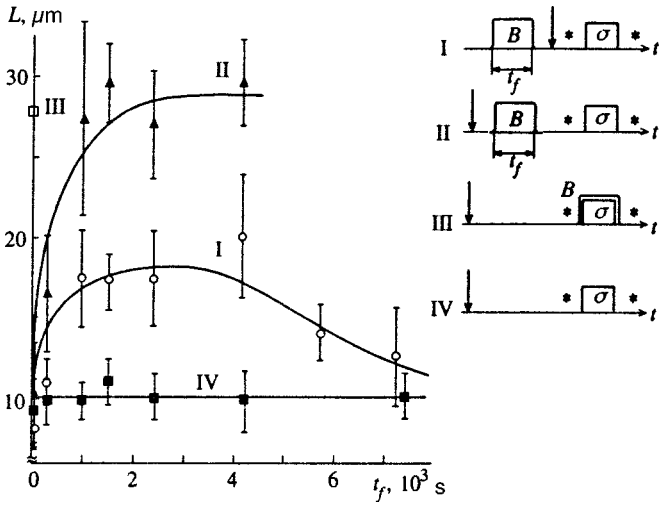

FIG. 5. Dependence of the dislocation mean free segment length L due to the mechanical loading of the NaCl crystals (the same in all of the runs) on the exposure time t_f to a constant magnetic field with $B=1$ T; I — dislocations introduced after exposure to a magnetic field with $B=1$ T; II — dislocations introduced before exposure to a magnetic field with $B=1$ T; III — crystals were simultaneously loaded and exposed to a magnetic field for 6 s; IV — magnetic field was absent during the pause between introduction of dislocations and loading ($B=0$ T). The diagram on the right shows the sequence of procedures in runs of each type (B — exposure to magnetic field, σ — mechanical loading, arrow — introduction of dislocations, asterisk — etching.

before introduction of fresh dislocations and commencement of loading, II) between the introduction of fresh dislocations and commencement of loading, and III) during loading of crystals with freshly introduced dislocations (see the diagram on the right side of Fig. 5).

In the runs of group I the crystal was kept in the magnetic field during a time interval $0<t_f<10^4$ s, dislocations were then introduced, and after the chemical etching procedure which exhibited their original arrangement the sample was subjected to compression with the same mechanical loading pulse in all runs. Measurement of the mean free segment length L of the dislocations in runs of this type enabled us to establish that it is greater in the crystals exposed to the magnetic field than in the control runs without a field for $t_f<8\times10^3$ s (Fig. 5). For $t_f>8\times10^3$ s the magnitude of L returned approximately to its original value. ''Memory'' in the crystals of their exposure to a magnetic field can also be revealed by measurements of the microhardness performed immediately after exposure to the the field. After the quenched crystals were treated in a magnetic field (duration 20 s and $B=2$ T) their microhardness fell by $\sim6\%$. In crystals subjected to an extended isothermal anneal at 700 K and subsequent slow cooling to $T=290$ K over the course of 6 h, the magnetic-field pulses left no residual changes.

In runs of group II (Fig. 5) the freshly introduced dislocations as well as point defects could be exposed to the action of the magnetic field with $B=1$ T since the dislocations were introduced into the crystal before it was exposed to the magnetic field. The segment lengths of the dislocations exposed to the magnetic field were measured, as in the first group of runs, after the field was switched off. It was determined that under the same experimental conditions (field strength, magnetic exposure time, and the loading-pulse parameters) the dislocation mean free segment length L in the group-II runs was greater than in the first group (Fig. 5).

In the group-I and group-II runs, creation of a noticeable change in the state of the crystal leading respectively to a 50% and a 100% increase in L required exposure of the crystals to a field with $B=1$ T for a duration $t_f\sim10^3$ s at $T=293$ K (Fig. 5). When, on the other hand, crystals with freshly introduced dislocations were subjected to the simultaneous action of a magnetic field and a mechanical load (as in the group-III runs), a twofold increase in L was reached even 5–10 s after exposure to the magnetic field (Fig. 5).

4. In the experiments of the fourth series, after exposure in a constant magnetic field dislocations were introduced into the crystal and then, to initiate their motion the crystal was subjected to a magnetic-field pulse ($B=7$ T, duration 10^{-2} s) instead of a mechanical load. In contrast to the runs of the third series, in which the mechanical load could stimulate depinning of the dislocations from all types of pinning centers (regardless of their sensitivity to a magnetic field), in the experiments of the fourth series the external testing action (magnetic-field pulse) could initiate depinning of dislocations only from magnetically sensitive pinning centers, and motion of the dislocations took place in the internal mechanical stress field. For exposure times $t_f<4\times10^3$ s in a constant magnetic field, in contrast to the analogous runs in the third series of experiments, L was observed to decrease with increasing t_f in comparison with the behavior of L^* in crystals not subjected to a preliminary exposure to a magnetic field (Fig. 6). For $t_f>4\times10^3$ s the value of L recovered, approaching a value somewhat less than L^* with further increase of t_f. The kinetics of recovery of the dislocation segment length is insensitive to the presence of a constant magnetic field on this rising segment of $L(t_f)$ (Fig. 6).

If a crystal without freshly introduced dislocations is subjected to the action of a brief field pulse ($B=7$ T, duration 10^{-2} s) instead of an extended exposure to a constant magnetic field, and dislocations are then introduced into it, then when a second pulse just like the first is switched on the mean free segment length L is found to be less than L^* (Fig. 7) as in the runs with a constant magnetic field. In special experiments we found that one magnetic-field pulse is enough so that repeated field pulses have no effect on the state of the crystal and do not alter the mobility of the subsequently introduced dislocations. Under these conditions, the action of the first magnetic-field pulse modifies the state of the crystal, and subsequent motion of fresh dislocations under the action of a second field pulse can serve as an indicator of the number of magnetically sensitive centers remaining in the crystal, from which dislocation depinning is facilitated by a magnetic field.

We found that the the mean free segment length of the dislocations initiated by a second field pulse depended on the time t_p between the two magnetic-field treatments. As t_p was increased, the mean free segment lengths of the dislocations initiated by the second field pulse increased (Fig. 7), i.e., the mobility of the dislocations recovered. Switching on an additional intermediate magnetic-field pulse just before introduction of dislocations in this series of experiments had no

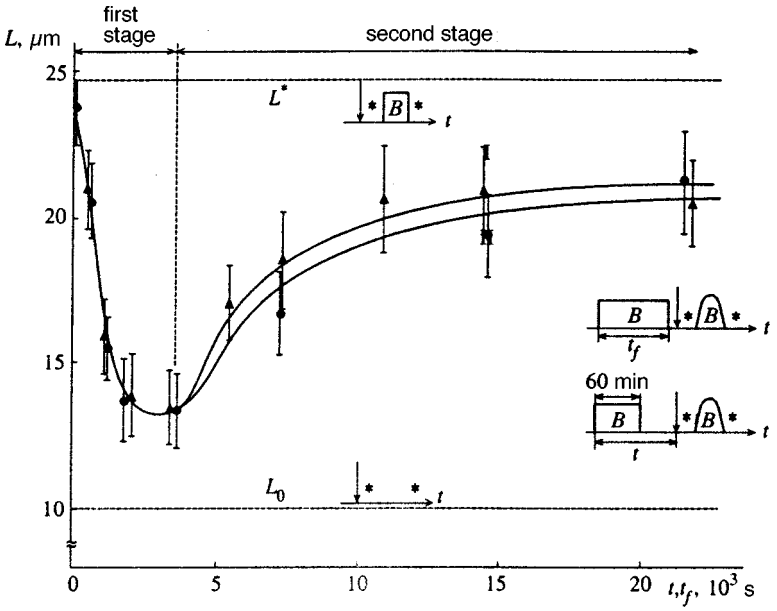

FIG. 6. Dependence of the dislocation mean free segment length L due to the magnetic field pulse (the same in all of the runs, $B=7$ T) in NaCl crystals on the duration of their preliminary exposure t_f to a constant magnetic field with $B=1$ T (●). Dependence of the dislocation mean free segment length L due to a magnetic field pulse (the same in all of the runs, $B=7$ T) in NaCl crystals subjected to a preliminary exposure to a magnetic field $B=1$ T for 60 min on the time t elapsed between the start of exposure of the crystals to a constant magnetic field and introduction of dislocations (▲). Notation the same as in Fig. 5, the magnetic field pulse is denoted by the half-period sine wave. A prime indicates values of the dislocation mean free segment length L^* initiated by a magnetic field pulse in crystals not subjected to preliminary exposure to a constant magnetic field, and of the mean free segment length L_0 due to double etching without a magnetic field.

effect on the kinetics of recovery of the mean free segment length or on the level to which L rose with further increase of t_p (Fig. 7).

5. Such experimental procedures as quenching, mechanical loading, and introduction of dislocations affect the state of all types of structural defects simultaneously. This hinders identification of the magnetically sensitive centers and determination of their physical characteristics. At the same time, it is well known that monochromatic light in the optical range is capable of selectively modifying certain types of point defects if the light frequency coincides with resonance frequencies of intra-center electron transitions or corresponds to transitions to the conduction band. A photostimulated change in the states of some types of point defects explains the photoplastic effects observed in Refs. 22 and 23. This suggests that the point defects, from which depinning is facilitated by the action of a magnetic field in our experiments, can also be selectively modified by light.

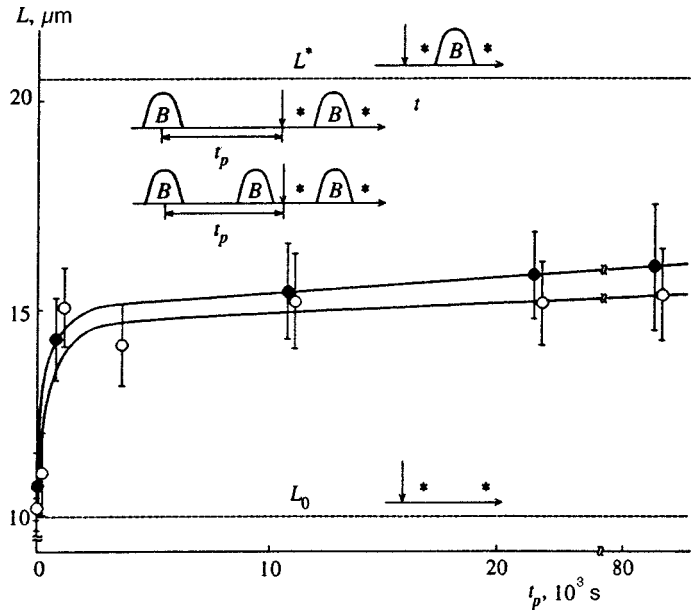

FIG. 7. The dislocation mean free segment length L initiated by a test magnetic field pulse in NaCl crystals subjected to a preliminary magnetic field pulse before introduction of dislocations as a function of the duration of the pause t_p between magnetic field pulses (●); the same thing under conditions where the sample is subjected to an additional intermediate magnetic field pulse before introduction of dislocations (○). A prime indicates values of the dislocation mean free segment length L^* initiated by a magnetic field pulse in crystals not subjected to preliminary exposure to a constant magnetic field, and of the mean free segment length L_0 due to double etching without a magnetic field. Notation is the same as in Figs. 5 and 6.

To test this conjecture, we carried out a fifth series of experiments in which crystals of NaCl, KCl, and LiF were illuminated for 15 min by a DKSSh-200 lamp through a diffraction-grating monochromator in the wavelength range 200–800 nm. We then introduced dislocations and subjected them to the action of magnetic-field pulse with $B=7$ T. Upon termination of the pulse, the crystals were again etched to determine the new locations of the dislocations. We found that the mean free segment length in the crystals exposed to light L' depends nonmonotonically on the photon energy E and differs from the mean free segment length L^* in the ''dark'' crystals (those not exposed to light) (Fig. 8). The greatest value of the difference $|L'-L^*|$ was reached at photon energies of the exciting light equal to $E_1=2.8$ eV for KCl, $E_2=3.6$ eV for NaCl, and as was determined by extrapolation of the Gaussian dependence of $L(E)$ in the short-wavelength region of the spectrum, $E_3=6.6$ eV for LiF. For all types of crystals L' was substantially less than L^*. Consequently, their illumination before magnetic-field exposure suppressed the magnetoplastic effect.

The photo-stimulated change in the sensitivity of the crystals to a magnetic field can be explained either by a general lowering of the mobility of the dislocations regardless of the reason for their motion in the crystal (e.g., as a result of a modification of the pinning centers that are insensitive to the magnetic field) or by a modification of the magnetically sensitive point defects. Of course, in the first case

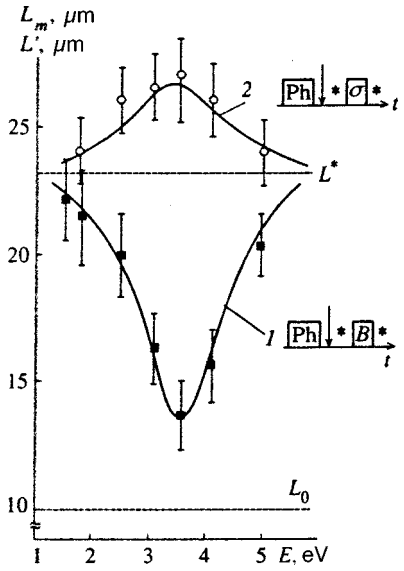

FIG. 8. Dependence of the dislocation mean free segment lengths due to a magnetic field pulse in NaCl crystals, L' (*1*), and due to mechanical loading, L_m (*2*), on the photon energy E of the preliminary photo-exposure (exposure to light). A prime indicates values of the mean free segment length L^* initiated by loading or by a magnetic field pulse in ''dark'' crystals, and of the mean free segment length L_0 due to etching. Notation is the same as in Fig. 5, *Ph* — photo-exposure.

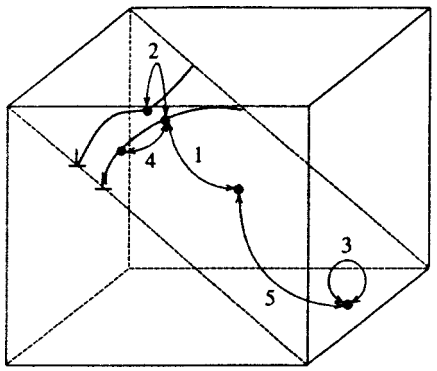

FIG. 9. Schematic depiction of types of reactions between crystal defects, during which a magnetic field can have an effect: 1) $R_D+R_P\leftrightarrow M_1$, 2) $R_{D1}+R_{D2}\leftrightarrow M_2$, 3) $R_{P'}+R_{P''}\leftrightarrow M_3$, 4) $R_{D'}+R_{D''}\leftrightarrow M_4$, $R_{P1}+R_{P2}\leftrightarrow M_5$ (R_D are paramagnetic centers localized in the dislocation nucleus, R_P are paramagnetic point defects; in the formulas the primed subscripts correspond to reagents belonging to the same structural defects, and the subscripts with numbers correspond to reagents belonging to different structural defects.

the dislocation segment lengths due to the action of external mechanical stresses will decrease after exposure to light (photo-exposure) the same as those due to the action of a magnetic field. In the second case, for mechanical loading of crystals exposed to light one can also expect an increase in L in comparison with L^*.

The choice between these two possibilities motivated a series of experiments in which the order of the procedures was similar, but the magnetic-field pulse was replaced by a mechanical compression pulse. The magnitude of the load was chosen so that the dislocation mean free segment length in the ''dark'' crystals L_m was equal to L^*. We found that in contrast to the experiments in which motion of the dislocations was initiated by a magnetic-field pulse, pre-exposure of the crystals to light increased L_m (Fig. 8). Consequently, the ''mechanical'' mobility of the dislocations in this case grew, and the susceptibility to an external magnetic field was lost. Taking the above facts into account, this means that the photo-stimulated change in the mobility of the dislocations is a consequence of the effect of light on specifically those point defects from which dislocation depinning is facilitated in a magnetic field.

4. DISCUSSION

In ionic crystals many point defects and dislocation nuclei containing dangling bonds or electrons and holes captured by the deformation potential, as is well known, possess paramagnetic properties.[24,25] Consequently, during structural rearrangements covalent bonds between them can be formed and broken. A set of examples of the effect of weak magnetic fields on the kinetics and the yields of these reactions is well known in the chemistry of radical reactions.[16–18] These effects have found a consistent description within the framework of the theory of spin-dependent intercombinational transitions observed in a reaction cell in the absence of thermodynamic equilibrium.[19,20] In Refs. 11 and 12 it was proposed to use elements of this theory to describe the interaction of dislocations with a local pinning center in the crystal. Obviously, such an approach can be applied only for those experimental conditions for which the magnetic field exists while the dislocation approaches the obstacle and when it overcomes it. As follows from the results presented in the preceding Section, after the magnetic field is switched off residual changes are preserved in the crystal which influence the mobility of the dislocations for an extended time. Since these changes can appear even if the crystal is exposed to the magnetic field in the absence of freshly introduced dislocations, we must broaden the spectrum of possible spin-dependent reactions in the crystal that can be affected by a magnetic field. Possible types of reactions capable of playing a role in the formation of plastic properties of crystals in a magnetic field are depicted schematically in Fig. 9. Motion of dislocations in the crystal is necessary for only two of these reaction types ($R_D+R_P\leftrightarrow M_1$ and $R_{D1}+R_{D2}\leftrightarrow M_2$; for an explanation of the notation see the caption to Fig. 9). The remaining reactions can occur in an unloaded crystal. Of course, to establish the mechanisms of the action of a field, it is necessary to determine the contribution of each of the reactions to the total effect of softening of the crystals in a magnetic field. The results of our study of the effect of a magnetic field on the mobility of individual dislocations allow us to do this.

Let us first discuss the possibility of the effect of a magnetic field on the state of the crystals in the absence of freshly introduced dislocations. In the group-I runs of the third series of experiments (Fig. 5) and in the runs of the fourth series of experiments (Fig. 6) in which the crystals were subjected to magnetic-field treatment before fresh dislocations were introduced, this field, in principle, could stimulate depinning of the growth dislocations present in the

crystal from the impurity atmospheres, thereby influencing the redistribution of long-range internal stresses. However, if this factor could have an effect on the mobility of subsequently introduced dislocations, it could not be expected that with increase of the duration of the exposure of the crystal to the magnetic field ($t_f > 4\times10^3$ s) the mobility of the subsequently introduced dislocations would spontaneously recover almost to their original value. In actual fact, the decrease in the mobility of the dislocations for $t_f < 4\times10^3$ s and its recovery for $t_f > 4\times10^3$ s cannot be explained by the same factor—relaxation of internal stresses upon depinning of growth dislocations.

This suggests that in the experiments in which the crystals were exposed to a magnetic field before fresh dislocations were introduced, the field could affect only the state of the point defects since during the time the crystals were exposed to a magnetic field fresh dislocations were absent, and the observed nonmonotonic changes in the plastic properties cannot be explained by relaxation of internal stresses upon depinning of growth dislocations in a magnetic field. Consequently, the following reactions in the subsystem of point defects could be susceptible to the action of a magnetic field: $R_{P'}+R_{P''}\leftrightarrow M_3$ and $R_{P1}+R_{P2}\leftrightarrow M_5$ (Fig. 9).

Although in the third series of experiments (Fig. 5) magnetic-field treatment of the crystals temporarily increased the mobility, and in the runs of the fourth series performed in the absence of external mechanical load it was observed to decrease (Fig. 6) and then recover with the passage of time in both cases. At first glance, this may seem to imply that the reactions $R_{P'}+R_{P''}\leftrightarrow M_3$ and $R_{P1}+R_{P2}\leftrightarrow M_5$ are reversible, i.e., that a magnetic field excites the point defects, which then relax to their initial state. However, the mobility of the dislocations recovers to a level lower than its original level, and the rising segment of $L(t_f)$ is insensitive to the presence of a constant magnetic field (Fig. 6). This may imply that the nonmonotonic variation of the mean free segment length after the action of a magnetic field can be explained by a multistep relaxation process in the subsystem of point defects in which the magnetic field plays the role of a ''trigger,'' opening up the possibility of the occurrence of reactions prohibited in its absence.

If this is not the case, then switching on an additional intermediate magnetic-field pulse just before introducing dislocations in the fourth series of experiments (Fig. 7) should make their mean free segment length the same as it was a short time after the first magnetic-field pulse. However, the experiments show that an intermediate pulse has no effect on the kinetics of recovery of the mean free segment length or on the level to which the dependence $L(t)$ rises (Fig. 7). Consequently, the state of the point defects is irreversibly altered, and the recovery of the sensitivity of the dislocation mean free segment length to a magnetic field may be a consequence of excitation of point defects during the introduction of dislocations or a consequence of motion of charged dislocations in the crystal.

The most likely reason for the sensitization of point defects to a magnetic field upon introduction of dislocations is action of the strong electric fields arising during the motion of charged dislocations at the moment the dislocations are introduced. Preliminary experiments show that an external variable electric field with field strength ~ 10 kV/cm leads to the appearance of sensitivity of point defects to a magnetic field in those crystals in which their sensitivity to a magnetic field was suppressed by isothermal annealing.

Thus, for $B=7$ T a magnetic-field pulse of duration 10^{-2} s proves sufficient for the occurrence of the first step of the relaxation process in the subsystem of metastable point defects, which proceeds in a magnetic field with $B=1$ T after 4×10^3 s. The second step, during which the thermodynamic potential of the crystal with point defects also continues to decrease, occurs after termination of the magnetic-field pulse. As a result of the second step, the point defects wind up in a state from which they can be transferred to an excited, magnetically sensitive state upon introduction of dislocations.

At present, the efficient action of a weak (on the energy scale) magnetic field on the relaxation rate of various systems from a metastable state is widely known. These include the rate of chemical reactions,[16–18] the photoconductivity[26] and viscosity of amorphous materials,[27] etc. The irreversibility of the change in the states of point defects in a magnetic field observed in our experiments, and also the necessity of pre-excitation of the crystals by introducing dislocations or quenching to elicit a magnetoplastic effect in them (Fig. 3) puts this effect into the same class with the phenomena described above and clarifies thermodynamic aspects of the effect of a magnetic field on the plasticity. Quenching or the electric field of the moving dislocations creates in the crystal a nonequilibrium distribution of structural defects over the available states, and a magnetic field facilitates their relaxation by accelerating only the rate of relaxation processes taking place in the subsystem of point defects and in the absence of a field. The possibility of a magnetic-field effect on the state of the point defects in crystals that have not been exposed to any preliminary excitation is probably connected with the presence in them of a biographic nonequilibrium distribution over degrees of aggregatedness, orientation, or some other property.

Note that other authors[6,10] have observed depinning of dislocations from pinning centers in a magnetic field in well-annealed crystals on the macroplastic as well as the microplastic level. This indicates that the plastification channel we have discovered of crystals in a magnetic field is not unique. In the absence of magnetically sensitive point defects, a magnetic field can also alter the kinetics of the other reactions enumerated in the caption to Fig. 9, which would also cause changes in the plasticity.

Let us now consider the runs in which both dislocations and point defects could be simultaneously exposed to the action of a magnetic field (runs of group II, see Fig. 4). The still greater softening of the crystals, in comparison with the runs of group I, arising in such experiments (Fig. 4) indicates that the state of the dislocations can also be exposed to the action of a magnetic field, i.e., reactions of the type $R_{D'}+R_{D''}\leftrightarrow M_4$ can occur and be sensitive to the presence of a magnetic field. An alternative explanation of this experimental fact may be that a magnetic field affects the depinning of dislocations from pinning centers from the start and, as a

consequence, facilitates their motion upon subsequent loading. To what extent each of these factors can contribute to the additional softening of the crystals in a magnetic field can be determined from an analysis of the experimental data on the motion of dislocations produced by continuous etching.

If depinning in a magnetic field, from the start, of a larger number of dislocations than in the control runs would produce a noticeable redistribution of the internal mechanical stresses and to a collective effect of growth of the mobility, then it may be expected that motion of the dislocations should be to a significant extent correlated—pinning and depinning of each dislocation should affect the others, and the waiting time for each dislocation would be interrelated with the parameters of its further motion from the start.

Measurement of the dimensions of the dislocation pits allowed us to determine the waiting times for 90 dislocations at pinning centers and how long they move between pauses by the technique described in Ref. 21. We found that the correlation coefficient Γ_1 between the time elapsed from the start of etching to the first depinning and the total dislocation mean free segment length is close to zero ($\Gamma_1 = -0.11$). The correlation coefficient between the waiting time until the first depinning and the time of motion of a dislocation until its first pause is also very low ($\Gamma_2 = -0.03$). This suggests that magnetic-field stimulated depinning of dislocations in fact has no noticeable effect on their subsequent motion from the start. In addition, for 90% of the moving dislocations the first depinning occurs after termination of the magnetic-field pulse. This all suggests that the increased velocity of the dislocations at large distances from their original location is a consequence of a ''memory'' in the structural defects of their exposure to a magnetic field. This viewpoint is also supported by the delay of the macroplastic deformation jump initiated by a short magnetic-field pulse (Fig. 3).

Thus, the effect of a magnetic field on the depinning of dislocations from paramagnetic centers (i.e., on reactions of the type $R_D + R_P \leftrightarrow M_1$) discovered by Al'shits *et al.*[1] cannot be the only reason for motion of dislocations in unloaded crystals. Another reason may be changes in the states of isolated structural defects, i.e., point defects and dislocation nuclei. Consequently, a complete treatment of the magnetoplastic effect in ionic crystals would have to take into account at least three types of magnetically sensitive reactions: 1) reactions between point defects in the bulk of the crystal, 2) reactions between point centers in the dislocation nucleus, and 3) reactions between a point defect and a paramagnetic center in a dislocation nucleus.

We can render this general picture of the effect of a magnetic field on the plasticity concrete by availing ourselves of data on the types of defects that can be affected by a magnetic field, and on the kinetics of processes taking place in the subsystem of point defects after exposure of the crystal to a magnetic field. Experimental data on the kinetics of magnetically stimulated processes together with qualitative estimates allow us to establish the direction of the reaction $R_{P'} + R_{P''} \leftrightarrow M_3$. Processes that could take place with point defects after exposure to a magnetic field can be nominally divided into two groups: 1) monomolecular—mutually independent local reactions, which include reorientations of defects, pair dissociation, transitions to another electron level inside a defect, etc., and 2) bimolecular—mutually dependent processes, examples of which include recombination of wandering defects into pairs, capture of electrons from the conduction band into traps, and other reactions requiring an encounter of reagent molecules of the same or different species. During a monomolecular process the number of pinning centers that have changed their state as of the time t after termination of excitation is given by $N(t) = N_0(1 - \exp(-t/\tau))$. Bimolecular processes obey a hyperbolic law of variation of the concentration of excited defects in time: $N(t) \sim 1/(1 + ct^n)$, $n > 0$ (Ref. 28). The well-known relation between L and N ($L \propto N^{-1/2}$) (Ref. 29) enables us to determine to which of these processes the reactions belong that are initiated in the crystals after termination of the magnetic pulse.

Fitting of the dependence of the quantity $(L - L^*)/L^*$ on t_f in the first step of the relaxation process (Fig. 6) by a straight line on a semi–log plot gives a root-mean-square error that is 30% less than the fit obtained by a straight line on a log–log plot. In the second step, in contrast, the root-mean-square error of the log–log fit is four times less than that of the semi-log fit. In light of the above, this may imply that the reaction occurring in the first step of the relaxation process in the subsystem of point defects in a magnetic field is a monomolecular reaction of the type $M_3 \rightarrow R_{P'} + R_{P''}$, and in the second step—a bimolecular reaction—of the type $R_{P1} + R_{P2} \rightarrow M_5$, for example. An additional argument in favor of this supposition is offered by the results obtained in Ref. 30, where it was established that a pulsed magnetic field is capable of leading to such a change in the picture of the decoration of the surface of ionic crystals, which corresponds to dissociation of an aggregated impurity. Bimolecularity of the reactions in the second step of the transformation of point defects can explain the diffusion times of their relaxation after termination of the magnetic-field pulse that are probably necessary to realize encounters of randomly wandering reagent molecules. This viewpoint is also supported by the increase in the rate of the transition of the defects to a new state when the temperature is raised,[31] which is probably due to a growth of the mobility of the reagent molecules R_{P1} and R_{P2}.

Thus, there are grounds to assume that in a magnetic field a monomolecular process of formation of reagent molecules occurs via dissociation of composite point defects (the magnetic field stimulates the reaction $M_3 \rightarrow R_{P'} + R_{P''}$). The products of this decay recombine with each other or combine with other defects after termination of the magnetic-field pulse.

We can make the step from a phenomenological treatment of the magnetoplastic effect to a determination of the mechanisms of the action of a magnetic field by availing ourselves of the results of experiments on optical quenching of the sensitivity of crystals to a magnetic field. Since the atomic rearrangement of the configuration of point defects occurring instantly under the action of light requires photon energies near or greater than the x-range, and since radiation from the far-infrared is commonly used to match the fre-

quency of the external radiation with the resonance frequency of the local natural oscillations of the defects, it can be asserted with a high degree of confidence that the use of light in the optical range with photon energy less than the width of the band gap will result in a rearrangement of the electronic structure of the magnetically sensitive centers. Consequently, an explanation of the magnetoplastic effect can be had by taking into account the effect of a magnetic field on the electronic properties of the defects. This gives reason to suppose that the process of electron interaction between defects is sensitive to a magnetic field, i.e., the reactions described above probably are of a chemical nature and consist of the formation of a spin–dependent exchange coupling between paramagnetic defects. The important role of solid-state chemical reactions between structural defects in the formation of the optical and electronic properties of crystals has been known for a long time.[32,33] However, treatments of the plastic properties of crystals, as a rule, take into account only the elastic or electrostatic interaction between the defects.[34] Such an approach, obviously, does not allow us to explain the entire panoply of experimental data on the effect of a magnetic field on the mobility of dislocations in ionic crystals.

Let us review the experimental facts which provide direct evidence of the need for an account of short-range exchange forces in any treatment of the interaction between point defects and dislocations. If the hypothesis of the spin nature of the magnetoplastic effect is valid, then the combined action of a weak constant magnetic field and a microwave magnetic field under certain conditions can give rise to a magnetic resonance in defect complexes, influence the efficiency of formation of secondary reaction products, and thus be reflected in the mobility of the dislocations. A test of this hypothesis would consist of a study of the mobility of individual edge dislocations under conditions of the simultaneous action of a weak constant magnetic field and a high-frequency magnetic field crossed with it. Such a formulation of the experiment is analogous to RYDMR, the well-known technique for studying chemical reactions in crossed magnetic fields from the final product yields.[35–37]

In the experiments we used NaCl single crystals annealed at 700 K and cooled back down to 293 K during 1 h with an impurity of bivalent metals (mainly Ca) at the level 0.01 at. % dislocations were placed inside a TE_{10} rectangular waveguide, connected to a klystron which worked at a frequency of 9.5 GHz and generated an electromagnetic field flux with power ~ 10 mW through the sample. The waveguide was located between the poles of an electromagnet which could create a constant magnetic field B_0 between 0.2 and 0.8 T in the vicinity of the crystal. The segment lengths of individual edge dislocations, which fulfilled the role of an indicator of magnetically sensitive processes in the crystal, were measured in the conventional way by double chemical etching. Each point on the graphs was constructed by averaging 100–400 segment lengths of individual dislocations measured under identical conditions. Double etching of the samples in the absence of external forces yielded a mean displacement of the dislocations $L_0 = 12 \pm 1$ μm. The displacement of the dislocations in crystals exposed for 15 min

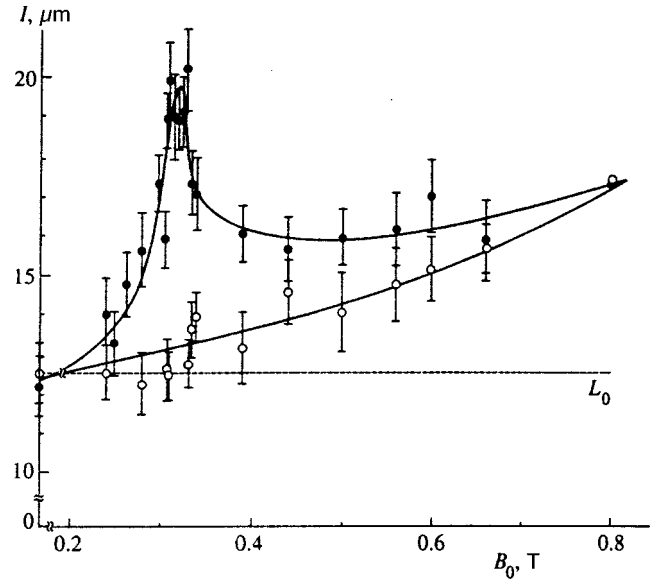

FIG. 10. Dependence of the dislocation mean free segment length L due to a constant magnetic field B_0 applied for 15 min with simultaneous application of a perpendicular high-frequency magnetic field ($\nu = 9.5$ GHz) (●) and without (○). A prime indicates a value of the mean free segment length L_0 due to etching in crystals not exposed to the action of external fields.

to the action of the microwave magnetic field in the absence of the constant magnetic field was the same in magnitude, i.e., treatment of the crystals with only the variable magnetic field did not affect the mobility of the dislocations (Fig. 10).

In the absence of the microwave field, exposure of the crystals to the constant magnetic field for 15 min caused an increase in the mean free segment length of the dislocations in comparison with L_0. As can be seen from Fig. 10, in the absence of the microwave field a monotonically increasing dependence of $L(B_0)$ was observed.

Exposure of the crystals to a constant magnetic field with simultaneous action of the microwave field for 15 min (in the configuration $\mathbf{B}_1 \perp \mathbf{B}_0$, where $\mathbf{B}_1$ is the amplitude of the microwave magnetic field) gave rise to a peak in the $L(B_0)$ dependence for $B_{res} = 0.33 \pm 0.01$ T (Fig. 10).

A similar result obtains for the frequency of the variable field equal to 152.2 MHz, where for $B'_{res} \approx 0.005$ T a resonant increase in the segment lengths of the dislocations is also observed, albeit weaker, but fully distinguishable. When the constant and microwave fields were applied in the configuration $\mathbf{B}_1 \| \mathbf{B}_0$, the peak in the $L(B_0)$ dependence disappeared and the dependence itself became similar to that observed in the absence of the microwave field. Thus, over a wide range of microwave frequencies by varying the strength the constant magnetic field it was possible to observe peaks in the $L(B_0)$ dependence.

Both in the combined action of a microwave and a constant magnetic field and when the crystals are exposed to a constant magnetic field in the absence of the microwave field, on average an equiprobable motion of the dislocations is observed in the four crystallographic directions [110], $[\bar{1}10]$, $[1\bar{1}0]$, and $[\bar{1}\,\bar{1}0]$ on all faces of the sample. Consequently, the role of external fields in our experiments reduces to a depinning of the dislocations from the pinning

centers, and the motion of the dislocations takes place under the action of random internal mechanical stresses.

Without touching upon the specific mechanisms of depinning of dislocations from pinning centers, let us discuss the general conclusions that follow from our results. The experimentally found value $B_{res}=0.33\pm0.01$ T corresponds to the field $B_{th}=h\nu/\mu_B g$ at which for the microwave frequency $\nu=9.5$ GHz resonant transitions take place between the spin sublevels of the electron with Landé factor $g=2.06\pm0.05$, which are split in the constant magnetic field. The possibility of recording a resonant increase in the mobility of the dislocations at the microwave frequency $\nu=152.2$ MHz for the corresponding value of B_0 is additional evidence of the resonant nature of the phenomenon.

Usually the role of the resonant microwave field is to increase the intensity of the transitions between the singlet and triplet sublevels of a nonequilibrium pair consisting of paramagnetic particles.[37] Intensification of such transitions in the short-lived pair D + T, as was shown theoretically in Ref. 38, can be reflected in the mobility of the dislocations. Thus, the combined influence of a constant magnetic field and a high-frequency magnetic field on the plasticity can distinguish the spin-dependent part of the forces braking the motion of the dislocations from the sum of elastic, electrostatic, and exchange interactions with the point defects.

5. CONCLUSIONS

We have shown experimentally that one of the prerequisites for a magnetic field to have an effect on structural defects is their pre-existence in a metastable state. This clarifies the thermodynamic aspects of the effect of weak electromagnetic fields on the physical properties of materials. We found that a magnetic field can alter the state of the dislocation nuclei as well as the state of the complexes. We have identified the types and order of the reactions between various defects and within one type of defect which are influenced by a magnetic field. We have uncovered the possibility of selective modification of magnetically sensitive point defects by light in the optical range, and also by the combined action of a constant magnetic field and a high-frequency magnetic field, making it possible to create experimental conditions for their identification. We have shown that to explain the sensitivity of intracrystalline reactions to a magnetic field it is necessary to take into account the electronic degrees of freedom of the defects which participate in the formation of chemical bonds between them. This points to a significant role for exchange forces in the formation of the plastic properties of ionic crystals. The suggested mechanisms by which a magnetic field can affect the state of defects in ionic crystals do not possess any specifics associated with the specific type of material, and under certain conditions can be used to explain magnetoplastic effects in metals, semiconductors, and polymers. The ensuing experimental situations can play the role of model conditions for the study of more complex objects sensitive to a magnetic field: polymer and protein macromolecules, and also biological systems.

The authors express their gratitude to Prof. V. I. Al'shits and Prof. M I. Molotskiĭ for fruitful discussions of this work.

This work was carried out with the financial support of the Russian Fund for Fundamental Research (Project No. 97-02-16074).

*)E-mail: golovin@main.tsu.tambov.ru

[1] V. I. Al'shits, E. V. Darinskaya, T. M. Perekalina, A. A. Urusovskaya, Fiz. Tverd. Tela (Leningrad) **29**, 467 (1987) [Sov. Phys. Solid State **29**, 265 (1987)].
[2] Yu. I. Golovin, O. L. Kazakova, R. B. Morgunov, Fiz. Tverd. Tela (St. Petersburg) **35**, 1384 (1993) [Phys. Solid State **35**, 700 (1993)].
[3] A. E. Smirnov and A. A. Urusovskaya, Fiz. Tverd. Tela (Leningrad) **29**, 852 (1987) [Sov. Phys. Solid State **29**, 485 (1987)].
[4] Yu. I. Golovin, R. B. Morgunov, D. V. Lopatin, and A. A. Baskakov, Phys. Status Solidi A **160**, R3 (1997).
[5] É. P. Belozerova, A. A. Svetashev, and V. L. Krasnikov, Izv. Ross. Akad. Nauk, Ser. Fiz. **61**, 291 (1997).
[6] A. A. Urusovskaya, V. I. Al'shits, A. E. Smirnov, and N. N. Bekkauer, JETP Lett. **65**, 497 (1997).
[7] V. I. Al'shits, E. V. Darinskaya, and E. A. Petrzhik, Izv. VUZov, Chern. Metall. **10**, 85 (1990).
[8] Yu. I. Golovin and R. B. Morgunov, JETP Lett. **61**, 596 (1995).
[9] V. I. Al'shits, E. V. Darinskaya, and O. L. Kazakova, JETP Lett. **62**, 375 (1995).
[10] V. I. Al'shits, E. V. Darinskaya, and O. L. Kazakova, Zh. Éksp. Teor. Fiz. **111**, 615 (1997) [JETP **84**, 338 (1997)].
[11] M. I. Molotskiĭ, Fiz. Tverd. Tela (Leningrad) **33**, 3112 (1991) [Sov. Phys. Solid State **33**, 1760 (1991)].
[12] V. I. Al'shits, E. V. Darinskaya, E. A. Petrzhik, Fiz. Tverd. Tela (Leningrad) **33**, 3001 (1991) [Sov. Phys. Solid State **33**, 1694 (1991)].
[13] M. I. Molotskii and V. N. Fleurov, Phys. Rev. Lett. **78**, 1 (1997).
[14] M. I. Molotskii and V. N. Fleurov, Phys. Rev. B **52**, 15829 (1995).
[15] M. I. Molotskii, R. E. Kris, and V. N. Fleurov, Phys. Rev. B **51**, 12531 (1995).
[16] A. L. Buchachenko, R. Z. Sagdeev, and K. Z. Salikhov, *Spin Polarization and Magnetic Effects in Radical Reactions*, Elsevier, Amsterdam, 1984 (Russian original Nauka, Novosibirsk, 1978).
[17] A. L. Buchachenko, Usp. Fiz. Nauk **62**, 1139 (1993).
[18] A. L. Buchachenko and I. V. Khudyakov, Usp. Khim. **60**, 1105 (1991).
[19] Ya. B. Zel'dovich, A. L. Buchachenko, and E. L. Frankevich, Usp. Fiz. Nauk **155**, 3 (1988) [Sov. Phys. Usp. **31**, 385 (1988)].
[20] A. L. Buchachenko and I. V. Berdinsky, Chem. Phys. Lett. **1**, 43 (1995).
[21] Yu. P. Pshenichnov, *Visualization of the Fine Structure of Crystals* [in Russian] (Metallurgiya, Moscow, 1974).
[22] Yu. A. Osip'yan and I. B. Savchenko, JETP Lett. **7**, 100 (1968).
[23] Yu. A. Osip'yan, V. F. Petrenko, and M. Sh. Shikhsaidov, JETP Lett. **20**, 163 (1974).
[24] P. W. M. Jacobs, J. Phys. Chem. Solids **51**, 35 (1990).
[25] S. Z. Shmurak and F. D. Senchukov, Fiz. Tverd. Tela (Leningrad) **15**, 2976 (1973) [Sov. Phys. Solid State **15**, 1985 (1974)].
[26] E. L. Frankevich, Zh. Éksp. Teor. Fiz. **50**, 1226 (1966) [Sov. Phys. JETP **23**, 814 (1966)].
[27] S. D. Savranskiĭ, Fiz. Khim. Stekla **13**, 659 (1987).
[28] Yu. R. Zakis, L. N. Kantorovich, and E. A. Kotomin, *Models of Processes in Wide-Gap Solids with Defects* [in Russian] (Zinatne, Riga, 1991).
[29] Zh. Fridel', *Dislocations* [in Russian] (Nauka, Moscow, 1966).
[30] G. I. Distler, V. M. Kanevskiĭ, V. V. Moskvin, *et al.*, Dokl. Akad. Nauk SSSR **268**, 591 (1983) [Sov. Phys. Dokl. **28**, 43 (1983)].
[31] Yu. I. Golovin and R. B. Morgunov, Fiz. Tverd. Tela (St. Petersburg) **37**, 1239 (1995) [Phys. Solid State **37**, 674 (1995)].
[32] S. Z. Shmurak, Izv. Akad. Nauk SSSR, Ser. Fiz. **40**, 1886 (1976).
[33] I. A. Parfianovich and É. É. Penzina, *Electron Color Centers in Ionic Crystals* [in Russian] (Vost.-Sib. Kn. Izd., Irkutsk, 1977).
[34] J. P. Hirth and J. Lothe, *Theory of Dislocations* (McGraw-Hill, New York, 1967).
[35] E. L. Frankevich, A. I. Pristupa, and V. I. Lesin, Chem. Phys. Lett. **47**, 304 (1977).
[36] E. L. Frankevich, V. I. Lesin, and A. I. Pristupa, Zh. Éksp. Teor. Fiz. **75**, 415 (1978) [Sov. Phys. JETP **48**, 208 (1978)].
[37] E. A. Ermakova and S. I. Kubarev, Khim. Fiz. **11**, 73 (1992).
[38] M. I. Molotskii and V. Fleurov, Philos. Mag. Lett. **73**, 11 (1996).

Translated by Paul F. Schippnick

Propagation of nonequilibrium phonons in ferroelectric ceramics and single crystals

S. N. Ivanov,*), A. V. Taranov, and E. N. Khazanov

Insitute of Radio-Engineering and Electronics, Russian Academy of Sciences, 103907 Moscow, Russia

E. P. Smirnova

A. F. Ioffe Physico-Technical Institute, Russian Academy of Sciences, 194021 St. Petersburg, Russia
(Submitted 9 July 1998
Zh. Éksp. Teor. Fiz. **115**, 624–631 (February 1999)

Phonon transport in ferroelectric ceramics and single crystals has been experimentally investigated. Our measurements indicate that, in the temperature range 1.7–3.8 K studied, the effective phonon diffusion coefficient behaves as $D_{\text{eff}} \propto T^{-5}$ in ferroelectrics with a broadened phase transition. This experimental dependence is in accord with the presence of a plateau in the thermal conductivity of such materials. The scattering by domain walls in $BaTiO_3$ single crystals has been identified, and our results are in quantitative agreement with calculations. © *1999 American Institute of Physics.* [S1063-7761(99)01702-3]

1. INTRODUCTION

In our experiments we have studied propagation of weakly nonequilibrium phonons at $\Delta T = T_h - T_0 \ll T_0$, where T_h is the heater temperature and T_0 is the ambient temperature, in a number of ferroelectric materials in single-crystal and ceramic modifications. When the thermal pulse technique[1] is used, i.e., the size of the samples in the phonon flux direction is larger than their diffusion length, the propagation of nonequilibrium phonons is described by the conventional heat equation. The peak amplitude of the bolometer signal as a function of temperature yields information about the mechanism of phonon scattering in the material, and in ceramics about the structure of grain boundaries.[2,3]

The interest in studies of ferroelectric materials is stimulated by at least two objectives.

1. Identification of the contribution of phonon scattering on grain boundaries. At present numerous theoretical calculations of coherent ultrasound reflection and refraction on grain boundaries are available.[4,5] On the other hand, no experimental data that are at all trustworthy have been obtained to date, primarily because the reflection of coherent ultrasound waves from domain boundaries is very low.[6,7] Suslov and Kagan[8,9] attempted to determine the contribution of domain boundaries to scattering of phonons in a thermal pulse in virtual ferroelectrics $SrTiO_3$ and $KTaO_3$ in an applied electric field, but the interpretation of the experimental data proved to be very complicated.

2. Determination of the main phonon scattering mechanism in a thermal pulse in ferroelectric ceramics and single crystals, which are characterized by anomalies in their low-temperature specific heat and thermal conductivity that can be interpreted in terms of the glass model.[10] This model presupposes that, in addition to low-energy excitations of crystalline materials, there are additional low-energy excitations in glasses, responsible for the low-temperature thermodynamics and phonon kinetics. With a view to interpreting measurements by the thermal pulse method, the transit time of a nonequilibrium phonon pulse as a function of temperature in such materials was calculated by Kozub *et al.*[3,11]

Thus, the aim of this work was the experimental investigation of phonon scattering by domain walls in single crystals and in ceramic grains of the conventional $BaTiO_3$ ferroelectric with a narrow phase transition, together with ceramics and single crystals with a diffusive phase transition, such as $PbMg_{1/3}Nb_{2/3}O_3$ (PMN), $PbSc_{1/2}Nb_{1/2}O_3$ (PSN), solid solutions based on these materials, and $Pb_{(1-1.5)x}Zr_{0.65}Ti_{0.35}La_x$ ($x = 0.06–0.1$) (PZTL).

2. EXPERIMENTAL TECHNIQUES

All ferroelectric ceramic and single-crystals samples used in our experiments were fabricated by conventional technologies. The ceramics were densely packed with a density of 97–98% of the theoretical single-crystal density at an average grain dimension of about 10^{-3} cm.

The structure of a sample cleaved surface was tested by a Joal JSM-840 scanning electron microscope. The micrographs characterized on the qualitative level a single-crystal domain structure or a pattern of grains and grain boundaries in a ceramic sample. The phonon kinetics at helium temperature was studied by the thermal pulse technique described in detail elsewhere.[1] Recall that a thin gold film is deposited on one side of a plate of tested material, and this film is heated by a very short ($\approx 10^{-7}$ s) current pulse, so that it acts as an injector of nonequilibrium phonons into the sample. On the opposite sample surface, a tin bolometer of a meandered shape with area 0.3×0.25 mm is fabricated. If the bolometer characteristic is biased by a weak magnetic field of $\approx 2 \times 10^2$ Oe, one can measure nonequilibrium phonon scattering versus temperature over a range of 1.7–3.8 K. The power dissipated in the heater is set at a sufficiently low level that the injected phonons could be described in terms of the ambient temperature in analyzing experimental data.

The basic parameter measured in the experiments is the transit time $t_{\max}$ of the nonequilibrium phonon peak detected

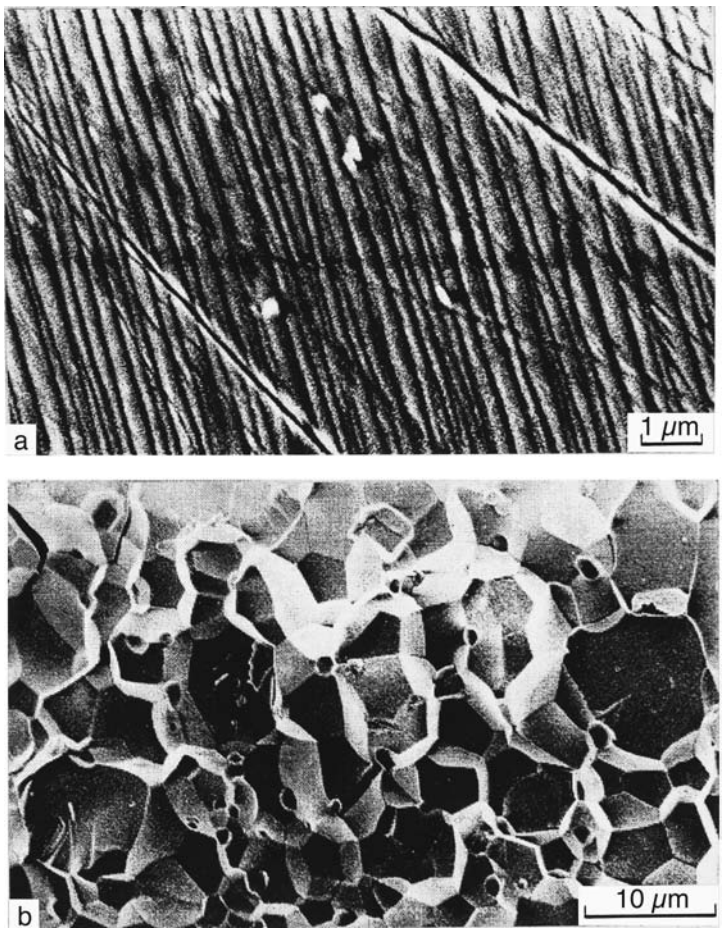

FIG. 1. Electron micrographs of (a) strip domains in the *xy* plane of a $BaTiO_3$ single crystal and (b) cleaved surface of a PMN ceramic sample fabricated by the hot-pressing technique.

by the bolometer. Its dependence on the temperature and features of the tested sample structure has been analyzed.

3. RESULTS AND DISCUSSION

Typical micrographs of sample surfaces are shown in Fig. 1. The $BaTiO_3$ single crystal is characterized by a fairly regular pattern of *a*-type domains (Fig. 1a). The boundary between the domains is a region where the order parameter is inhomogeneous,[12] and its width r_c can be treated as a domain wall boundary thickness. There is good reason to assume that r_c equals several lattice constants, i.e., in $BaTiO_3$ $r_c \approx 20-30$ Å, so that the phonon wavelength in our experiments is $\lambda_{ph} \gg r_c$. In real crystals a domain wall is pinned by defects, crystal inhomogeneities on crystal boundaries and in the wall region, i.e., elastic stress is present in the layer of thickness $\sim r_c$. The domain structure determines most of the ferroelectric properties of practical significance, and one aim of the reported work was to estimate its effect on the phonon kinetics at low temperature.

All ceramic samples under study were characterized by dense packing of most of their grains, of which the majority looked like crystallites (Fig. 1b). The statistical analysis of large sets of grains yielded the mean grain size R for a specific sample.

Before proceeding to the analysis of nonequilibrium phonon propagation, let us discuss features common to all the samples. Thus, the curves of bolometer signal versus time are bell-shaped with clearly defined peaks, which is typical of the diffusion propagation mode. Some data on propagation of thermal pulses in samples of different lengths from one material and at different temperatures are plotted in Fig. 2. The transit time t_{max} of the nonequilibrium phonon peak to the bolometer is proportional to L^2 with good accuracy, where L is the plate size in the phonon flux direction (see the inset to Fig. 2).

Given t_{max}, one can estimate the effective diffusion coefficient of the dominant phonon group,[2] $D_{eff} \approx L^2/t_{max}$. For most materials studied in our experiments, the thermal conductivity κ and specific heat C_v at helium temperature are known. This allowed us to calculate the phonon diffusion coefficient by the standard formula $D = \kappa/C_v$ and compare with our measurements. It turned out that the discrepancy in the single crystals and $BaTiO_3$ ceramic was within 10%.[10,13] In PMN and PZTL ceramics this discrepancy is notably higher, by a factor of up to two.[10,14–16] We do not think that this discrepancy is of major importance, since there are uncertainties in models of phonon propagation in ceramic materials,[2] and processes used to synthesize ceramic materials may vary.

In our opinion, observation of the proportionality $t_{max} \propto L^2$ in all the samples (Fig. 2) and absolute measurements of the diffusion coefficient for the dominant group of phonons in fairly good agreement with independent measurements provide solid evidence in favor of the applicability of the

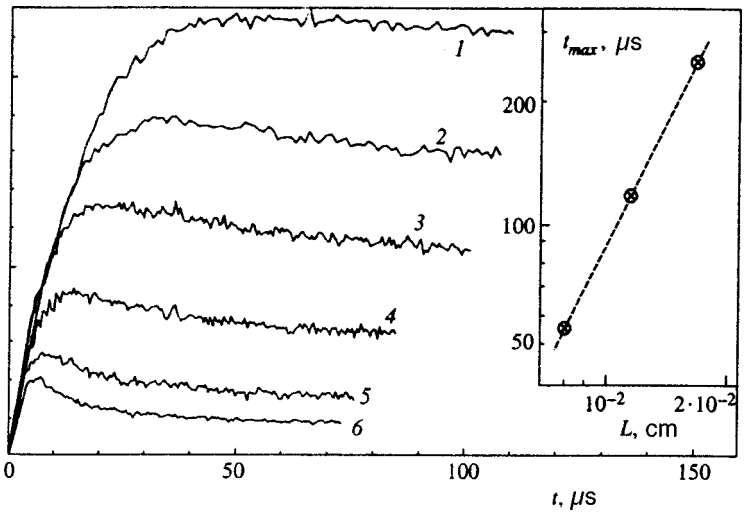

FIG. 2. Signals due to nonequilibrium phonons in a PNM ceramic sample of thickness $L=0.0085$ cm at (*1*) $T=3.82$ K; (*2*) 3.46 K; (*3*) 3.15 K; (*4*) 2.81 K; (*5*) 2.46 K; (*6*) 2.22 K. The inset shows $t_{\max}$ versus PMN sample length at $T=3.8$ K.

diffusion model[2,3] to our results, so we can proceed to the analysis of the $t_{\max}$ temperature dependence with a view to determining the phonon scattering mechanism in the ferroelectrics studied.

The curves of $t_{\max}$ versus temperature for the tested samples can be classified in two groups:[10] (1) ferroelectrics with a diffusive (broadened) phase transition, hence with glass-like properties; (2) ferroelectrics with a narrow phase transition, hence with crystal-like kinetic properties.

Figure 3 shows records of thermal pulses at several different temperatures in a typical ferroelectric of the first group, namely, $Pb_{(1-1.5)x}Zr_{0.65}Ti_{0.35}La_x$ $(x=0.08)$, where the ion disorder is considerable and translational symmetry is broken due to La doping,[10] so the kinetic properties are glass-like. For curves in Fig. 3 the law $t_{\max}\propto T^5$ is in accord with theoretical estimates[3,11] for the region of the thermal conductivity plateau in glassy ferroelectrics.

The inset to Fig. 3 shows the signal due to nonequilibrium phonons in pure PZT ceramic not doped with La. The propagation time in the pure material is considerably shorter, only a few microseconds. This case can be regarded as quasiballistic propagation. Recall that the transition in this ceramic is narrow, and its kinetic properties are those characteristic of crystals.[11]

Phonon scattering by low-energy vibrational excitations typical of dielectric glasses has been detected in samples of $PbMg_{1/3}Nb_{2/3}O_3-PbSc_{1/2}Nb_{1/2}O_3$ ceramic solid solutions. For samples of all compositions $t_{\max}$ follows a T^5 law.

All in all, the analysis of data on ferroelectrics with glass-like kinetic properties leads us to conclude that the thermal pulse technique unambiguously identifies such properties, since the peak position of the signal due to nonequilibrium phonons in the temperature range of 2 to 3.8 K follows the law $t_{\max}\propto T^5$. Figure 4 shows curves of thermal conductivity of PZTL and PMN ceramics and of the SiO_2 glass[10] versus temperature with the characteristic flat sections near 10 K. The graph also plots our measurements of effective diffusion coefficients for the tested materials. The temperature intervals corresponding to the flat sections of the thermal conductivity and to the diffusion coefficients described by the formula $D_{\text{eff}}\propto T^{-5}$ are fairly close to each other.

The assumption that scattering on grain boundaries is small and does not affect properties of glass-like ceramics in the low-temperature range is supported by direct measurements of the diffusion coefficient of nonequilibrium phonons in a PMN single crystal fabricated by the hot-pressing method (Fig. 1b). At a temperature $T=3.8$ K we have

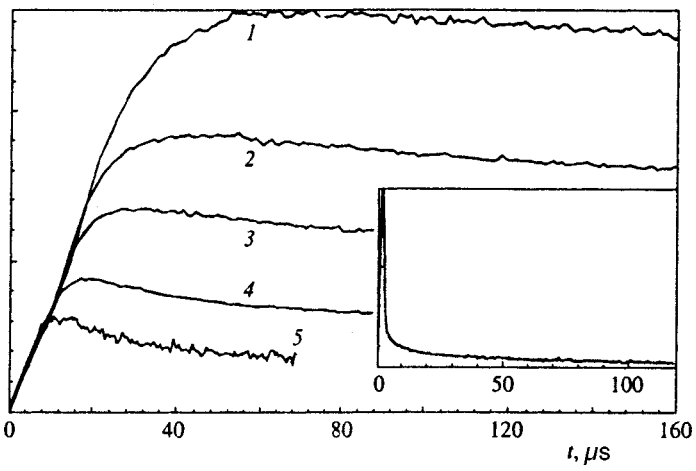

FIG. 3. Signals due to nonequilibrium phonons in a PZTL sample with $L=0.008$ cm at (*1*) $T=3.81$ K; (*2*) 3.43 K; (*3*) 3.17 K; (*4*) 2.8 K; (*5*) 2.49 K. The inset shows the nonequilibrium signal in a PZT ceramic sample with $L=0.1$ cm at $T=3.82$ K.

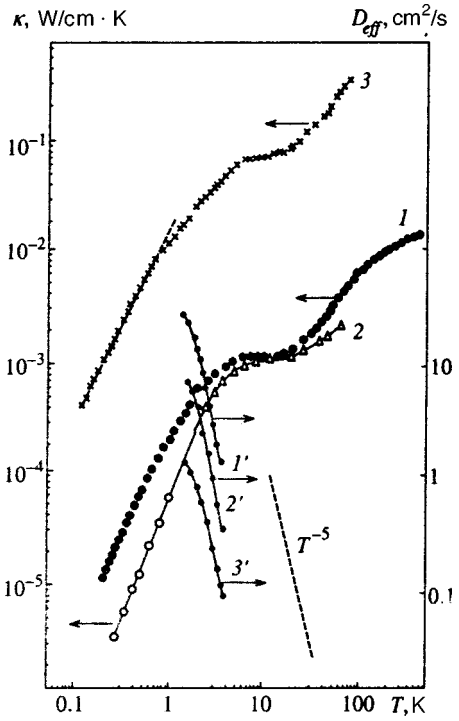

FIG. 4. Temperature dependence of thermal conductivity of ceramics and of glasses [(*1*) SiO_2; (*2*) PMN; (*3*) PZTL] and those of diffusion coefficients $D_{\text{eff}}=L^2/t_{\max}$ [(*1′*) SiO_2; (*2′*) PMN; (*3′*) PZTL] derived from measurements of this work.

$D_{\text{eff}}=2.5$ cm^2/s for both single-crystal and ceramic samples to within the accuracy of the measurements.

Now let us consider curves of the signal due to nonequilibrium phonons for ferroelectrics with a narrow phase transition. As was noted above, propagation of a thermal pulse in $PbZn_{0.65}Ti_{0.35}O_3$ not doped with La is quasi-ballistic (Fig. 2). Measurements of the typical ferroelectric $BaTiO_3$ are shown in Fig. 5. The main set of the curves was obtained in a ceramic sample with thickness $L=1$ mm. The main result, $\partial t_{\max}/\partial T<0$, is typical of ballistic propagation inside grains and scattering on thin interfaces between them, where the acoustic impedance is low owing to the low material density and concentration of micropores. In this situation phonons pass through grain boundaries more readily the higher the phonon frequency or, in this specific case, the sample temperature.[17] In our experiment $t_{\max}\propto T^k$, where $k\approx -1$.

The inset to Fig. 5 shows curves of the nonequilibrium phonon signal for a $BaTiO_3$ single crystal, which implies $t_{\max}\propto T^m$, where $m\approx 2$. In single crystals the effective diffusion coefficient for phonons is a factor of 10 to 20 higher, but, what is more important, it has a different temperature dependence.

This result unambiguously indicates that the contribution of domain boundaries to phonon scattering is minimal and does not affect our measurements. In the $BaTiO_3$ single crystal the phonon scattering is controlled by domain wall boundaries. Let us make some quantitative estimates. In the model with diffusion of nonequilibrium phonons[2,3] $t_{\max}$ is related to the effective phonon free path by the simple formula

$$t_{\max}=\frac{3}{2}\frac{L^2}{l_{\text{eff}}\bar{v}_s}, \tag{1}$$

where $\bar{v}_s$ is the mean phonon velocity in the sample. Using the data plotted in Fig. 5, we obtain $l_{\text{eff}}(3.8\ \text{K})\approx 10^{-3}$ cm, which increases with decreasing temperature.

Two neighboring domains in a ferroelectric are regions of the same crystal with different orientations related to one another by one or more symmetry operations. A phonon flux can undergo reflection on their boundary. This problem was solved[9] for the case of heat transfer in a system of N plane-parallel layers of thickness d. The effective free path is expressed by the formula[9]

$$l_{\text{eff}}=\frac{L(1-s)}{1+sL/d}, \tag{2}$$

where $L=Nd$ is the sample thickness, s is the phonon reflectivity on the boundary between two layers. At room temperature our samples have $d\approx 3\times 10^{-5}$ cm (see Fig. 1a), and we assume that this parameter is little changed as a result of the phase transition taking place as the temperature drops from room to helium. Using the experimental result $l_{\text{eff}}\approx 10^{-3}$ cm, we estimate the reflectivity $s(3.8\ \text{K})\approx 3\times 10^{-2}$. This measurement seems fairly reasonable in view

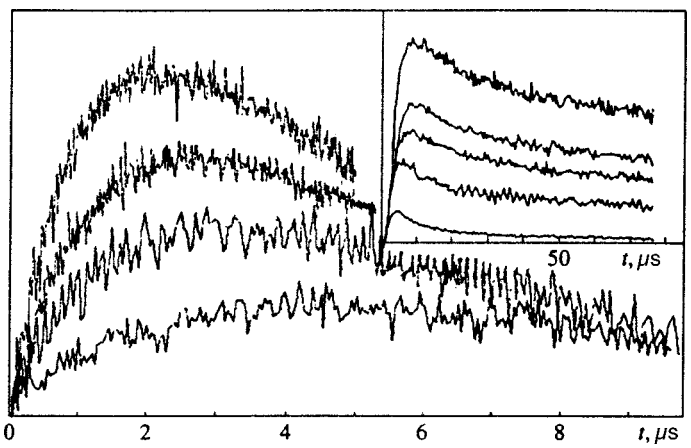

FIG. 5. Signals due to nonequilibrium phonons in a $BaTiO_3$ ceramic sample with $L=0.1$ cm. The inset shows nonequilibrium phonon signals in a $BaTiO_3$ single-crystal sample with $L=0.043$ cm.

of the results of Refs. 6 and 7 and an estimate of the phonon reflectivity on boundaries between grains in the ceramic corundum, $s=2.6\times10^{-2}$.[18]

These estimates have been made under the assumption that phonon scattering due to changes in the density and stiffness in domain walls can be neglected. Inclusion of reflections due to fluctuations in density $\Delta\rho/\rho$ and elasticity $\Delta c/c$ in the domain wall region[5,17] accounts for the temperature dependence of $t_{\max}$ observed in our samples, whereas quantitative estimates are difficult owing to the lack of data on $\Delta\rho/\rho$ and $\Delta c/c$ in the domain wall region.

In summary, the technique based on propagation of nonequilibrium phonons in ferroelectrics allows one to select materials with glass-like properties at helium temperatures and obtain quantitative estimates of phonon scattering by domain boundaries in single crystals.

The work was supported by the Russian Fund for Fundamental Research (Grant 97-02-16810).

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

[1] S. N. Ivanov, A. V. Taranov, and E. N. Khazanov, Zh. Éksp. Teor. Fiz. **99**, 1311 (1991) [Sov. Phys. JETP **72**, 731 (1991)].
[2] S. N. Ivanov, A. G. Kozorezov, E. N. Khazanov, and A. V. Taranov, Solid State Commun. **83**, 365 (1992).
[3] V. I. Kozub, A. M. Rudin, and H. Schober, Phys. Rev. B **50**, 6032 (1994).
[4] G. G. Kessenikh, D. G. Sannikov, and L. A. Shuvalov, Kristallografiya **16**, 350 (1971); **17**, 345 (1972) [Sov. Phys. Crystallogr. **16**, 287 (1971); **17**, 291 (1972)].
[5] G. G. Kessenikh and L. A. Shuvalov, Izv. Akad. Nauk SSSR, Ser. Fiz. **48**, 1168 (1984).
[6] S. Kh. Esayan, V. V. Lemanov, and G. A. Smolenskiĭ, Dokl. Akad. Nauk SSSR **217**, 83 (1974) [Sov. Phys. Dokl. **19**, 393 (1975)].
[7] V. V. Belov, O. Yu. Serdobol'skaya, and M. A. Suchkova, Fiz. Tverd. Tela **26**, 556 (1984) [Sov. Phys. Solid State **26**, 334 (1984)].
[8] A. V. Suslov, Fiz. Tverd. Tela **34**, 319 (1992) [Sov. Phys. Solid State **34**, 170 (1992)].
[9] V. D. Kagan and A. V. Suslov, Fiz. Tverd. Tela **36**, 2672 (1994) [Phys. Solid State **36**, 1457 (1994)].
[10] J. J. De Yoreo and R. O. Pohl, Phys. Rev. B **32**, 5780 (1985).
[11] V. I. Kozub and A. M. Rudin, Fiz. Tverd. Tela **38**, 337 (1996) [Phys. Solid State **38**, 189 (1996)].
[12] B. A. Strukov and A. P. Levanyuk, *Ferroelectric Phenomena in Crystals: Physical Foundations*, Springer, New York (1998) [Russian original, Nauka, Moscow (1983)].
[13] W. N. Lawless, Phys. Rev. B **14**, 134 (1976).
[14] E. Fisher, W. Hassler, E. Hegenbarth, and V. I. Fritsberg, Phys. Status Solidi A **66**, K169 (1981).
[15] I. Henning, P. Frach, E. Hegenbarth, and V. I. Fritsberg, Phys. Status Solidi A **70**, K7 (1982).
[16] D. A. Ackerman, D. May, R. C. Potter, and A. C. Anderson, Phys. Rev. B **23**, 3886 (1981).
[17] A. G. Kozorezov, J. K. Wigmore, C. Erd *et al.*, Phys. Rev. B **57**, 7411 (1998).
[18] A. A. Kaplyanskiĭ, M. B. Mel'nikov, and E. S. Feofilov, Fiz. Tverd. Tela **38**, 1434 (1996) [Phys. Solid State **38**, 792 (1996)].

Translation provided by the Russian Editorial office.

The Ginzburg–Landau expansion in the simple model of a superconductor with a pseudogap

A. I. Posazhennikova*) and M. V. Sadovskiĭ†)

Institute of Electrophysics, Ural Branch of the Russian Academy of Sciences, 620049 Ekaterinburg, Russia
(Submitted 16 June 1998)
Zh. Éksp. Teor. Fiz. **115**, 632–648 (February 1999)

We propose a simple model of the electron spectrum of a two-dimensional system with hot sections on the Fermi surface that significantly transforms the spectral density (pseudogap) in these sections. Using this model, we set up a Ginzburg–Landau expansion for *s* and *d* type Cooper pairing and analyze the effect of the pseudogap in the electron spectrum on the main properties of a superconductor. © *1999 American Institute of Physics.* [S1063-7761(99)01802-8]

1. INTRODUCTION

Among the various anomalies in the properties of high-T_c superconductors, the existence of a pseudogap in the electron spectrum of such materials at carrier concentrations below the optimum value has drawn much attention.[1,2] The most striking proof of the existence of this remarkable state has been obtained in measurements of photoemission spectra with angular resolution in the BSCCO system,[3,4] which demonstrated that the normal phase ($T>T_c$) exhibits essentially anisotropic variations in the spectral density of the current carriers. In particular, in these experiments the maximum pseudogap value was observed near the point $(\pi,0)$ in the Brillouin zone, while no pseudogap was observed along the diagonal. Correspondingly, the Fermi surface disintegrates near the point $(\pi,0)$, while along the diagonal the surface remains intact. In this sense it is common to speak of a *d* type pseudogap symmetry, which coincides with the symmetry of a superconducting gap in such systems. These anomalies exist up to temperatures $T\simeq T^*$ much higher than T_c.

There are many theoretical approaches that attempt to give an explanation of such anomalies. Two main groups of these approaches can be singled out: the pattern of formation of Cooper pairs above T_c (see Ref. 1, 5 and 6), and an alternation scheme based on the assumption that fluctuations of antiferromagnetic short-range order play the key role.[7–11]

Most papers on the subject deal mainly with the study of the pseudogap state of a high-T_c system in the normal phase ($T>T_c$). Our goal was to investigate the qualitative effects of the influence of a pseudogap in the electron spectrum on the main superconducting properties. We use the ideas developed in Refs. 7–11 but propose a very simple model of the pseudogap state in the normal phase, a model that allows a complete analytical investigation. On the basis of this model we do a microscopic derivation of the Ginzburg–Landau expansion for systems with *s* and *d* pairings and study the qualitative effects of the influence of a pseudogap (the disintegration of sections of the Fermi surface) on the main properties of the superconducting state.

2. ELEMENTARY MODEL OF A PSEUDOGAP STATE OF A TWO-DIMENSIONAL ELECTRON SYSTEM

As noted earlier, we adopt the simplest possible model of a pseudogap state, a model based on the picture of well-developed fluctuations of short-range antiferromagnetic order and close the model of ''hot points'' on the Fermi surface.[10,11] Let us assume that the Fermi surface of the two-dimensional electron system has the shape depicted in Fig. 1. A similar Fermi surface was proposed by Zheleznyak *et al.*,[12] who remarked that this Fermi surface resembles very closely the one observed by Dessau *et al.*[13,14] for some high-T_c systems. We assume that the short-range order fluctuations are static and Gaussian and define their correlation function as follows (cf. Ref. 7):

$$S(\mathbf{q})=\frac{1}{\pi^2}\,\frac{\xi^{-1}}{(q_x-Q_x)^2+\xi^{-2}}\,\frac{\xi^{-1}}{(q_y-Q_y)^2+\xi^{-2}} \tag{1}$$

for $-p_x^0\leqslant q_x\leqslant p_x^0$ and $-p_y^0\leqslant q_y\leqslant p_y^0$, where ξ is the correlation length of the fluctuations, and $Q_x=Q_y=2p_F$. For values of q_x and q_y that lie outside the specified ranges we assume that $S(\mathbf{q})=0$. The effective interaction between electrons and these fluctuations will be described by the quantity $(2\pi)^2W^2S(\mathbf{q})$, where the parameter W with the dimensions of energy defines the energy scale (width) of the pseudogap. Thus, we assume that only electrons belonging to the ''hot'' sections of the Fermi surface are scattered by the short-range fluctuations, with the scattering being actually one-dimensional.

The choice of the scattering vector $\mathbf{Q}=(2p_F,2p_F)$ presupposes a pattern of incommensurate fluctuations.Below we will consider the case of commensurate scattering with $\mathbf{Q}=(\pi/a,\pi/a)$, where a is the lattice constant. In the limit $\xi\to\infty$, such a model allows an exact solution by the methods proposed by Sadovskiĭ,[15,16] while for finite ξ one can employ the method developed by Sadovskiĭ and Timofeev[17,18] (with certain reservations; see Refs. 10, 11, and 19). Below we examine the simple case with $\xi\to\infty$, where the effective interaction with fluctuations (1) takes the simplest form

$$(2\pi)^2W^2\delta(q_x-2p_F)\,\delta(q_y-2p_F) \tag{2}$$

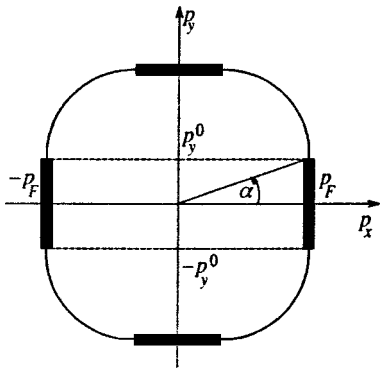

FIG. 1. The Fermi surface of a two-dimensional system. The hot sections are depicted by thick straight lines, whose width is of order $\sim\xi^{-1}$.

for $-p_x^0 \leqslant q_x \leqslant p_x^0$ and $-p_y^0 \leqslant q_y \leqslant p_y^0$. Here we can easily sum the entire perturbation series for an electron scattered by such fluctuations[15,16] and obtain the one-electron Green's function in the form

$$G(\epsilon_n, p) = \int_0^\infty d\zeta \exp(-\zeta) \frac{i\epsilon_n + \xi_p}{(i\epsilon_n)^2 - \xi_p^2 - \zeta W^2(\phi)}, \quad (3)$$

where $\xi_p = v_F(|\mathbf{p}| - p_F)$, with v_F the velocity at the Fermi surface, $\epsilon_n = (2n+1)\pi T$, and $W(\phi)$ is defined for $0 \leqslant \phi \leqslant \pi/2$ as follows:

$$W(\phi) = \begin{cases} W, & 0 \leqslant \phi \leqslant \alpha, \qquad \dfrac{\pi}{2} - \alpha \leqslant \phi \leqslant \dfrac{\pi}{2}, \\ 0, & \alpha \leqslant \phi \leqslant \dfrac{\pi}{2} - \alpha. \end{cases} \quad (4)$$

Here $\alpha = \tan^{-1}(p_y^0/p_F)$, and ϕ is the polar angle, which specifies the director of the vector $\mathbf{p}$ in the (p_x, p_y) plane. For other values of ϕ, the parameter $W(\phi)$ is determined quite similarly to (4) by symmetry considerations. Clearly, by varying α within the range $0 \leqslant \alpha \leqslant \pi/4$, we actually change the size of the hot sections on the Fermi surface, in which sections the nesting condition $\xi_{p-Q} = -\xi_p$ is satisfied. In particular, $\alpha = \pi/4$ corresponds to a square Fermi surface on which the nesting condition is satisfied everywhere. Outside the hot sections [the second inequality in (4)] the Green's function (3) simply coincides with the free-electron Green's function.

The spectral density corresponding to the Green's function (3), is

$$\rho(\epsilon\xi_p) = -\frac{1}{\pi} \operatorname{sgn}\epsilon \operatorname{Im} G(\epsilon\xi_p) \quad (5)$$

$$= \begin{cases} \dfrac{1}{W^2}(|\epsilon| + \xi_p \operatorname{sgn}\epsilon)\,\theta(\epsilon^2 - \xi_p^2) \exp\dfrac{\epsilon^2 - \xi_p^2}{W^2}, \\ \quad \text{if} \quad 0 \leqslant \phi \leqslant \alpha, \quad \dfrac{\pi}{2} - \alpha \leqslant \phi \leqslant \dfrac{\pi}{2}, \\ \delta(\epsilon - \xi_p), \\ \quad \text{if} \quad \alpha \leqslant \phi \leqslant \dfrac{\pi}{2} - \alpha, \end{cases} \quad (6)$$

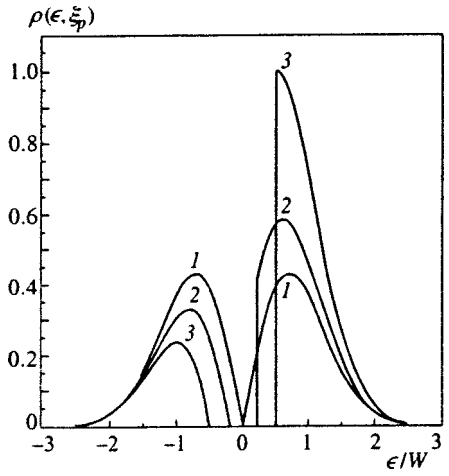

FIG. 2. Spectral density of the Green's function in a hot section of the Fermi surface: curve *1*, $\xi_p = 0$; curve *2*, $\xi_p = 0.1W$; and curve *3*, $\xi_p = 0.5W$.

and has a similar form in the other quadrants of the Brillouin zone. Equation (6) demonstrates the non-Fermi-liquid (pseudogap) behavior with a *d*-type symmetry in the vicinity of the hot sections of the Fermi surface and the free behavior in the cold sections. The behavior of the spectral density in a hot section of the Fermi surface is depicted schematically in Fig. 2. Allowing for the fact that the integral with respect to the polar angle ϕ of an arbitrary function $f[W(\phi)]$, with $W(\phi)$ defined in (4), is obviously

$$\int_0^{2\pi} d\phi\, f[W(\phi)] = 8\alpha f[W(\phi)] + (2\pi - 8\alpha) f(0), \quad (7)$$

we can use (6) to easily find the density of states:

$$\frac{N(E)}{N_0(0)} = -\frac{1}{\pi}\int_0^{2\pi} \frac{d\phi}{2\pi} \int_{-\infty}^{\infty} d\xi_p \operatorname{Im} G^R(\epsilon\xi_p)$$

$$= \frac{4}{\pi}\alpha N_W(\epsilon) + \left(\pi - \frac{4}{\pi}\alpha\right) N_0(0), \quad (8)$$

where $N_0(0)$ is the density of free-electron states at the Fermi level, and $N_W(\epsilon)$ is the density of states in the one-dimensional problem (a square Fermi surface) found in Refs. 15 and 16:

$$\frac{N_W(\epsilon)}{N_0(\epsilon)} = \left|\frac{\epsilon}{W}\right| \int_0^{\epsilon^2/W^2} d\zeta \frac{\exp(-\zeta)}{\sqrt{\epsilon^2/W^2 - \zeta}}$$

$$= 2\left|\frac{\epsilon}{W}\right| \exp\left(-\frac{\epsilon^2}{W^2}\right) \operatorname{Erfi}\frac{\epsilon}{W}, \quad (9)$$

where $\operatorname{Erfi} x$ is the probability integral (error function) of imaginary argument.

Figure 3 depicts the density-of-state curves in our model for different values of the parameter α, i.e., for hot sections of different size. We see that the pseudogap in the density of states becomes obscured rather quickly as the area of the hot sections decreases and generally is not very distinct. In a certain sense the effect of a decreasing α is similar to the effect of a decreasing correlation length ξ of the fluctuations,[17,18] so that in this sense the above approxima-

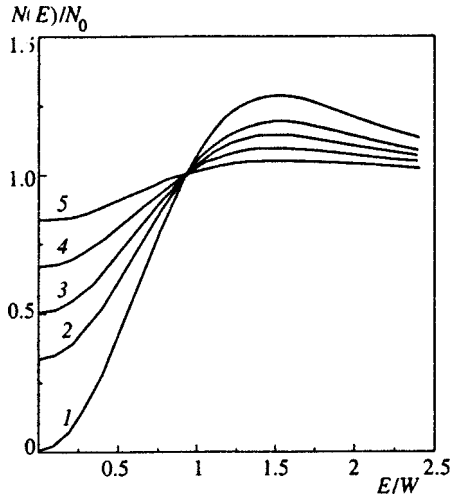

FIG. 3. Density of electron states for hot sections of different size: curve *1*, $\alpha=\pi/4$; curve *2*, $\alpha=\pi/6$; curve *3*, $\alpha=\pi/8$; curve *4*, $\alpha=\pi/12$; and curve *5*, $\alpha=\pi/24$.

tion $\xi\to\infty$ may not be a stringent restriction on the applicability of the model. One advantage of this approximation is the possibility of obtaining all the results in analytical form.

Concluding Sec. 2, we examine briefly the case of commensurate fluctuations, $\mathbf{Q}=(\pi/a,\pi/a)$. Figure 4 depicts the model of the Fermi surface used in this problem. The hot sections touch the boundaries of a new Brillouin zone that appears after long-range order (e.g., antiferromagnetic) has set in, and the strong scattering by fluctuations occurs at $\mathbf{Q}=(\pi/a,\pi/a)$. In this geometry the pseudogap opens in the direction of the diagonals of the Brillouin zone, which does not correspond to experiments involving high-T_c superconductors but is of certain theoretical interest. The problem is solved in the same way as in the previous case and generalizes the solution of the one-dimensional model first found by Wonneberger and Lautenschlager.[19] The one-electron Green's function is similar to (3), and $W(\phi)$ is again a function with a period $\pi/2$, but ''turned'' with respect to the

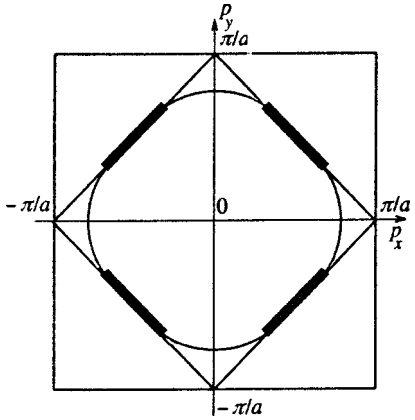

FIG. 4. The Fermi surface in the Brillouin zone of a two-dimensional system in the hot-section model for the case of short-range order fluctuations corresponding to period doubling. Also shown are the boundaries of the new Brillouin zone, which arises after long-range order sets in (e.g., due to an antiferromagnetic transition).

previous model through an angle of $\pi/4$ for $-\pi/4+\alpha\leqslant\phi\leqslant\pi/4+\alpha$:

$$W(\phi)=\begin{cases} W, & \pi/4-\alpha\leqslant\phi\leqslant\pi/4+\alpha, \\ 0, & -\pi/4+\alpha\leqslant\phi\leqslant\pi/4-\alpha, \end{cases} \tag{10}$$

where $0\leqslant\alpha\leqslant\pi/4$. Moreover, in the present case we must allow for a different combination of the Feynman diagrams, which must correspond to electron scattering by commensurate fluctuations.[19] As a result, in (3) we must replace

$$\int_0^\infty d\zeta\exp(-\zeta) \tag{11}$$

with

$$\int_0^\infty \frac{d\zeta}{2\sqrt{\pi\zeta}}\exp\left(-\frac{\zeta}{4}\right). \tag{12}$$

3. THE EQUATION FOR T_c

Let us now investigate the problem of superconductivity in the adopted model. We assume that the potential for Cooper pairing has the usual separable form[20]

$$V(\mathbf{p},\mathbf{p}')=V(\phi,\phi')=-Ve(\phi)e(\phi'), \tag{13}$$

where as before ϕ is the angle specifying the direction of electron momentum $\mathbf{p}$ in the plane, and $e(\phi)$ obeys the following model:

$$e(\phi)=\begin{cases} 1 & (s\text{ pairing}), \\ \sqrt{2}\cos 2\phi & (d\text{ pairing}). \end{cases} \tag{14}$$

As usual, the attractive constant V is assumed finite in a certain strip of width $2\omega_c$ in the vicinity of the Fermi level (ω_c is the characteristic frequency of the photons ensuring the attraction of electrons). In this case the superconducting gap (the order parameter) has the form

$$\Delta(\mathbf{p})\equiv\Delta(\phi)=\Delta e(\phi). \tag{15}$$

The equation for the transition temperature T_c can be obtained from the ordinary equation for Cooper instability,

$$1-\chi(0,0)=0, \tag{16}$$

where the generalized Cooper susceptibility $\chi(0,0)$ can be calculated by exact summation of the entire series of diagrams that allow for scattering by the short-range order fluctuations (2), in the same way the polarization operator was calculated by Sadovskiĭ.[15,16] As a result the equation for T_c becomes

$$\frac{1}{V}=-\int_0^\infty d\zeta\exp(-\zeta)\,T_c\sum_n\int_0^\infty\frac{d^2p}{(2\pi)^2}e^2(\phi)\{G_{\zeta W^2}\times(\epsilon_n;\mathbf{p},\mathbf{p})G_{\zeta W^2}(-\epsilon_n;-\mathbf{p},-\mathbf{p})+F_{\zeta W^2}\times(\epsilon_n;\mathbf{p},\mathbf{p}-\mathbf{Q})F_{\zeta W^2}(-\epsilon_n;-\mathbf{p},-\mathbf{p}+\mathbf{Q})\}, \tag{17}$$

where

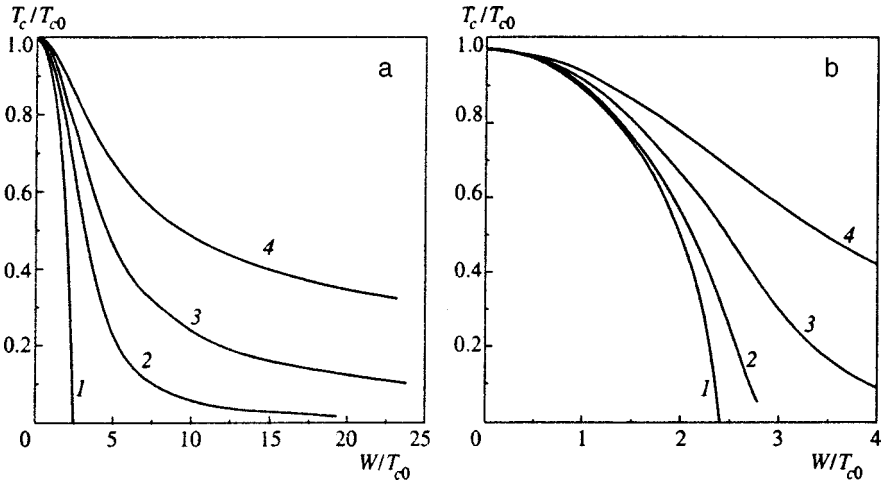

FIG. 5. T_c/T_{c0} as a function of the effective pseudogap width W/T_{c0} for hot sections of different size in the incommensurate fluctuation model for (a) *s* pairing (curve *1*, $\alpha=\pi/4$; curve *2*, $\alpha=\pi/6$; curve *3*, $\alpha=\pi/8$; and curve *4*, $\alpha=\pi/12$), and (b) *d* pairing (curve *1*, $\alpha=\pi/4$; curve *2*, $\alpha=\pi/6$; curve *3*, $\alpha=\pi/8$; and curve *4*, $\alpha=\pi/12$).

$$G_{\zeta W^2}(\epsilon_n;\mathbf{p},\mathbf{p})=\frac{i\epsilon_n+\xi_p}{(i\epsilon_n)^2-\xi_p^2-\zeta W^2(\phi)},$$

$$F_{\zeta W^2}(\epsilon_n;\mathbf{p},\mathbf{p}-\mathbf{Q})=\frac{\sqrt{\zeta}\,W(\phi)}{(i\epsilon_n)^2-\xi_p^2-\zeta W^2(\phi)} \tag{18}$$

are, respectively, the normal and anomalous Green's functions of a system with a dielectric gap.[15,16]

Applying standard transformations to (17), we get

$$\frac{1}{V}=\int_0^\infty d\zeta \times\exp(-\zeta)\,T_c\sum_n\int_0^\infty\frac{d^2p}{(2\pi)^2}\,\frac{e^2(\phi)}{\epsilon_n^2+\xi_p^2+\zeta W^2(\phi)}. \tag{19}$$

Summing over the frequencies yields

$$\frac{1}{V}=\frac{N(0)}{2\pi}\int_0^\infty d\zeta\exp(-\zeta)\int_{-\infty}^\infty d\xi \times\int_0^{2\pi}\frac{d\phi\,e^2(\phi)}{2\sqrt{\xi^2+\zeta W^2(\phi)}}\tanh\frac{\sqrt{\xi^2+\zeta W^2(\phi)}}{2T_c}. \tag{20}$$

If we now integrate with respect to ϕ as we did in (7), we arrive at the following formulas:

$$\frac{1}{g}=\frac{4\alpha}{\pi}\int_0^\infty d\zeta\exp(-\zeta)\int_0^{\omega_c}\frac{d\xi}{\sqrt{\xi^2+\zeta W^2}}\tanh\frac{\sqrt{\xi^2+\zeta W^2}}{2T_c} +\left(1-\frac{4\alpha}{\pi}\right)\int_0^{\omega_c}\frac{d\xi}{\xi}\tanh\frac{\xi}{2T_c} \tag{21}$$

for *s* pairing, and

$$\frac{1}{g}=\frac{4\alpha+\sin 4\alpha}{2\pi}\int_0^\infty d\zeta\exp(-\zeta)\int_0^{\omega_c}\frac{d\xi}{\sqrt{\xi^2+\zeta W^2}} \times\tanh\frac{\sqrt{\xi^2+\zeta W^2}}{2T_c}+\frac{\pi-4\alpha-\sin 4\alpha}{2\pi}\int_0^{\omega_c}\frac{d\xi}{\xi}\tanh\frac{\xi}{2T_c} \tag{22}$$

for *d* pairing. Here $g=N(0)V$ is the dimensionless Cooper-pairing constant. Figure 5 depicts curves representing the dependence of T_c/T_{c0} on the parameter W/T_{c0}, which specifies the effective pseudogap width, for different values of α (here T_{c0} is the transition temperature of an ideal system without a pseudogap). We see that for both types of pairing the occurrence of a pseudogap in the hot sections of the Fermi surface causes significant suppression of T_c, and the larger these hot sections are the stronger the suppression. Naturally, the suppression of T_c is stronger in the case of *d* pairing than in the case of *s* pairing, since the dielectrization of the spectrum (pseudogap) is in antiphase with the pairing interaction.

For commensurate fluctuations (Fig. 4) and *d*-type pairing, the equation for T_c becomes

$$\frac{1}{g}=\frac{4\alpha-\sin 4\alpha}{2\pi}\int_0^\infty\frac{d\zeta\exp(-\zeta/4)}{2\sqrt{\pi\zeta}}\int_0^{\omega_c}\frac{d\xi}{\sqrt{\xi^2+\zeta W^2}} \times\tanh\frac{\sqrt{\xi^2+\zeta W^2}}{2T_c}+\frac{\pi-4\alpha+\sin 4\alpha}{2\pi}\int_0^{\omega_c}\frac{d\xi}{\xi}\tanh\frac{\xi}{2T_c}. \tag{23}$$

Curves presenting the dependence of T_c/T_{c0} on the parameter W/T_{c0} for different values of α in this case are depicted in Fig. 6. Here the suppression of T_c by the pseudogap is less noticeable, since the superconducting gap reaches its maximum on the cold sections of the Fermi surface, where there is no pseudogap.

4. THE GINZBURG–LANDAU EXPANSION

The standard Ginzburg–Landau expansion for the difference in the free-energy densities of the superconducting and normal states is

$$F_s-F_n=A|\Delta_q|^2+q^2C|\Delta_q|^2+\frac{B}{2}|\Delta_q|^4, \tag{24}$$

where Δ_q is the Fourier transform of the order parameter:

$$\Delta(\phi,q)=\Delta_q e(\phi). \tag{25}$$

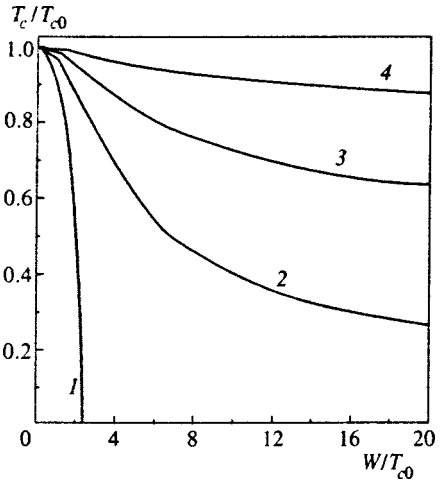

FIG. 6. T_c/T_{c0} as a function of the effective pseudogap width W/T_{c0} for hot sections of different size in the commensurate fluctuation model for the case of *d* pairing: curve *1*, $\alpha=\pi/4$; curve *2*, $\alpha=\pi/6$; curve *3*, $\alpha=\pi/8$; and curve *4*, $\alpha=\pi/12$.

Expansion (24) can be represented by the diagrams of the loop expansion for the free energy in the field of the order parameter fluctuations with a small wave vector **q**. These diagrams are depicted in Fig. 7, where all processes of scattering by short-range order fluctuations (2) are summed exactly in all loops (this can easily be done if we use the method developed in Refs. 15 and 16). In all other respects the method of calculation is similar to that used Ref. 20.[1)] As in Ref. 20, subtraction of the second diagram in Fig. 7 ensures the vanishing of the coefficient A at the transition point $T=T_c$. As a result, the Ginzburg–Landau coefficients can be written

$$A=A_0K_A, \quad C=C_0K_C, \quad B=B_0K_B, \tag{26}$$

where by A_0, C_0, and B_0 we denote the expressions for the case of a two-dimensional isotropic *s* superconductor in the absence of a pseudogap ($\alpha=0$),

$$A_0=N(0)\frac{T-T_c}{T_c}, \quad C_0=N(0)\frac{7\zeta(3)}{32\pi^2}\frac{v_F^2}{T_c^2},$$

$$B_0=N(0)\frac{7\zeta(3)}{8\pi^2T_c^2}, \tag{27}$$

and all the features of the models are reflected in the dimensionless coefficients K_A, K_C, and K_B. In the absence of a pseudogap, all these coefficients are equal to unity, while in the case of *d* pairing only K_B differs from unity, or $K_B=3/2$.

In particular, straightforward calculations yield

$$A=N(0)\frac{T-T_c}{2T_c^2}\frac{1}{2\pi}\int_0^\infty d\zeta\exp(-\zeta)\int_0^{\omega_c}d\xi \times\int_0^{2\pi}\frac{d\phi\, e^2(\phi)}{\cosh^2(\sqrt{\xi^2+\zeta W^2(\phi)}/2T_c)}, \tag{28}$$

so that after integrating with respect to ϕ we get

$$K_A=\frac{1}{2T_c}\beta_a\int_0^\infty d\zeta\exp(-\zeta) \times\int_0^{\omega_c}\frac{d\xi}{\cosh^2(\sqrt{\xi^2+\zeta W^2(\phi)}/2T_c)}+1-\beta_a, \tag{29}$$

where

$$\beta_a=\begin{cases}\dfrac{4\alpha}{\pi} & (s \text{ pairing}),\\[2ex] \dfrac{4\alpha+\sin 4\alpha}{\pi} & (d \text{ pairing}).\end{cases} \tag{30}$$

Figure 8 depicts curves representing the dependence of K_A on the effective pseudogap width W/T_{c0} for different values of α. Here we show only the curves for the case of *s* pairing. Qualitatively the corresponding curves for *d* pairing are similar, but all variations are on essentially smaller scales of W/T_{c0}, as in Fig. 5.

To calculate C, we must perform an expansion in a Taylor series in powers of **q** in the expression

$$-\int_0^\infty d\zeta\exp(-\zeta)\,T_c\sum_n\int_0^\infty\frac{d^2p}{(2\pi)^2}e^2(\phi)\{G_{\zeta W^2(\phi)} \times(\epsilon_n;\mathbf{p}_+,\mathbf{p}_+)G_{\zeta W^2(\phi)}(-\epsilon_n;-\mathbf{p}_-,-\mathbf{p}_-)+F_{\zeta W^2(\phi)} \times(\epsilon_n;\mathbf{p}_+,\mathbf{p}_+-\mathbf{Q})F_{\zeta W^2(\phi)}(-\epsilon_n;-\mathbf{p}_-,-\mathbf{p}_-+\mathbf{Q})\}, \tag{31}$$

where $\mathbf{p}_\pm=\mathbf{p}\pm\mathbf{q}/2$, and select the terms with $\mathbf{q}^2$. To simplify presentation, from now on we will use the notation

$$G_{\zeta W^2(\phi)}(\epsilon_n;\mathbf{p},\mathbf{p})\equiv G_{pp},$$

$$F_{\zeta W^2(\phi)}(\epsilon_n;\mathbf{p},\mathbf{p}-\mathbf{Q})\equiv F_{pp-Q}.$$

After lengthy calculations we arrive at an expression for the coefficient C:

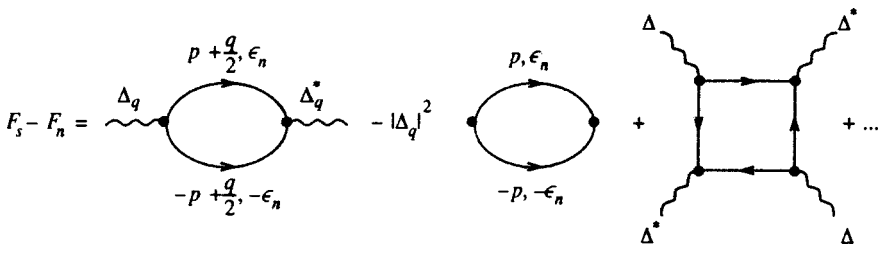

FIG. 7. The diagrammatic representation of the Ginzburg–Landau expansion in the field of short-range order fluctuations. The electron lines represent Nambu matrices composed of normal and anomalous Green's functions (18), and the loops are averaged over the parameter ζ with a distribution (11) or (12). The second loop is calculated for $q=0$ and $T=T_c$.

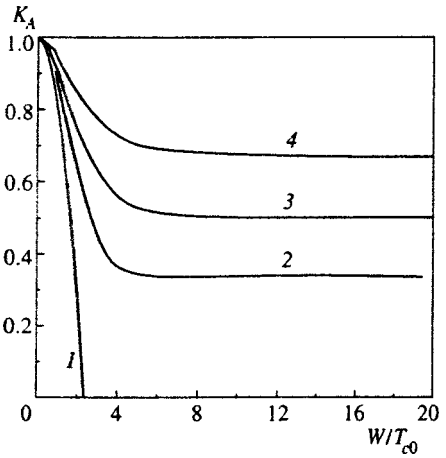

FIG. 8. The coefficient K_A as a function of the effective pseudogap width W/T_{c0} for hot sections of different size in the incommensurate fluctuation model for the case of *s* pairing: curve *1*, $\alpha=\pi/4$; curve *2*, $\alpha=\pi/6$; curve *3*, $\alpha=\pi/8$; and curve *4*, $\alpha=\pi/12$.

$$C=-T_c\frac{N(0)}{2\pi}v_F^2\sum_n \int_0^\infty d\zeta\exp(-\zeta)\int d\xi \times\int_0^{2\pi}\frac{d\phi\, e^2(\phi)(\xi^2-3\epsilon_n^2-3\zeta W^2(\phi))\cos^2\phi}{2(\epsilon_n^2+\xi^2+\zeta W^2(\phi))^3}. \quad (32)$$

Accordingly, after integrating over ξ and the angle ϕ, we arrive at an expression for the dimensionless coefficient K_C:

$$K_C=\beta_c\frac{4\pi^3T_c^3}{7\zeta(3)}\int_0^\infty d\zeta\exp(-\zeta)\sum_n \frac{1}{(\sqrt{\epsilon_n^2+\zeta W^2})^3}+1-\beta_c, \quad (33)$$

where $\beta_c=\beta_a$ [see Eq. (30)]. The respective relations between K_C and the parameter W/T_{c0} for the case of *s* pairing are depicted in Fig. 9. The pattern is similar for *d* pairing, but all variations are on essentially smaller scales of W/T_{c0}.

Examining the fourth-order term in the Ginzburg–Landau expansion is even more difficult technically. To obtain an expression for the coefficient B, we must find the trace of the product of four Green's functions $\hat{\mathbf{G}}_\mathbf{p}$, each of which is a Nambu matrix composed of normal and anomalous Green's functions (18):

$$\hat{\mathbf{G}}_\mathbf{p}=\begin{pmatrix} G_{p,p} & F_{p,p-Q} \\ F_{p-Q,p} & G_{p-Q,p-Q}\end{pmatrix}.$$

After we find the trace of the matrix $\hat{\mathbf{G}}_\mathbf{p}\hat{\mathbf{G}}_{-\mathbf{p}}\hat{\mathbf{G}}_\mathbf{p}\hat{\mathbf{G}}_{-\mathbf{p}}$, we can write an expression for B:

$$B=N(0)T_c\sum_{\epsilon_n}\int_0^\infty d\zeta\exp(-\zeta)\int_0^\infty\frac{d^2p\, e^4(\phi)}{(2\pi)^2}\times\{(G_{p,p}G_{-p,-p}+F_{p,p-Q}F_{-p,-p+Q})^2+G_{p,p}G_{-p,-p}F_{-p+Q,p}F_{p-Q,p}+G_{-p+Q,-p+Q}G_{-p,-p}F_{p,p-Q}F_{p-Q,p}+G_{p,p}G_{p-Q,p-Q}F_{-p+Q,-p}F_{-p,-p+Q}+G_{p-Q,p-Q}G_{-p+Q,-p+Q}F_{p,p-Q}F_{-p,-p+Q}\}. \quad (34)$$

Here we can directly verify that the sum of the last two terms in (34) yields a zero contribution, so that

$$B=N(0)T_c\sum_{\epsilon_n}\int_0^\infty d\zeta\exp(-\zeta)\int_0^\infty\frac{d^2p\, e^4(\phi)}{(2\pi)^2}\times(G_{p,p}G_{-p,-p}+F_{p,p-Q}F_{-p,-p+Q})^2. \quad (35)$$

This implies that

$$B=\frac{N(0)T_c}{2\pi}\sum_n\int_0^\infty d\zeta\exp(-\zeta)\times\int_{-\infty}^\infty d\xi\int_0^{2\pi}\frac{d\phi\, e^4(\phi)}{(\epsilon_n^2+\xi^2+\zeta W^2(\phi))^2}, \quad (36)$$

and after integrating with respect to ξ and ϕ we arrive at an expression for K_B similar to (33):

$$K_B=\beta_b\frac{4\pi^3T_c^3}{7\zeta(3)}\int_0^\infty d\zeta\exp(-\zeta)\sum_n\frac{1}{(\sqrt{\epsilon_n^2+\zeta W^2})^3}+1-\beta_b, \quad (37)$$

where

$$\beta_b=\begin{cases}\dfrac{4\alpha}{\pi}, & (s\text{ pairing}),\\ \dfrac{4\alpha}{\pi}+\dfrac{4\sin 4\alpha}{3\pi}+\dfrac{\sin 8\alpha}{6\pi}, & (d\text{ pairing}).\end{cases} \quad (38)$$

Thus, for *s* pairing the coefficients K_B and K_C simply coincide.

To conclude Sec. 4 we give the explicit expressions for the dimensionless Ginzburg–Landau coefficients for the case of *d* pairing in the model of commensurate short-range order fluctuations:

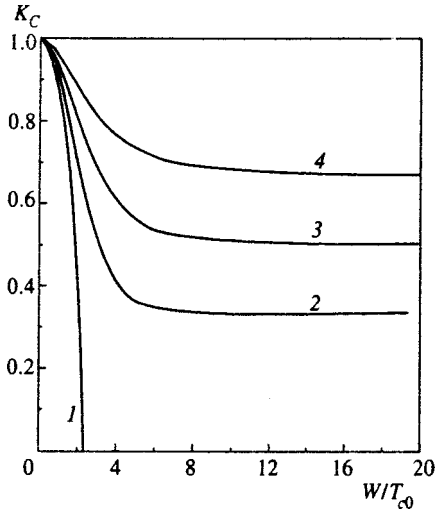

FIG. 9. The coefficient K_C as a function of the effective pseudogap width W/T_{c0} for hot sections of different size in the incommensurate fluctuation model for the case of *s* pairing: curve *1*, $\alpha=\pi/4$; curve *2*, $\alpha=\pi/6$; curve *3*, $\alpha=\pi/8$; and curve *4*, $\alpha=\pi/12$.

$$K_A=\beta_a\,\frac{1}{2T_c}\int_0^{\infty}\frac{d\zeta\,\exp(-\zeta/4)}{2\sqrt{\pi\zeta}}$$

$$\times\int_0^{\omega_c}\frac{d\xi}{\cosh^2(\sqrt{\xi^2+\zeta W^2}/2T_c)}+1-\beta_a\,, \tag{39}$$

$$K_{C,B}=\beta_{c,b}\frac{4\pi^3T_c^3}{7\zeta(3)}\int_0^{\infty}\frac{d\zeta\,\exp(-\zeta/4)}{2\sqrt{\pi\zeta}}$$

$$\times\sum_n\frac{1}{(\sqrt{\epsilon_n^2+\zeta W^2})^3}+1-\beta_{c,b}\,, \tag{40}$$

where

$$\beta_a=\beta_c=\frac{4\alpha-\sin 4\alpha}{\pi},$$

$$\beta_b=\frac{4\alpha}{\pi}-\frac{\sin 4\alpha}{6\pi}(5+\cos 4\alpha). \tag{41}$$

It is also easy to write the formulas reflecting the dependence of these coefficients on W/T_{c0} and different values of α. Qualitatively these expressions are similar to those in the incommensurate case, and the main difference are due to a different scale along the W/T_{c0} axis (cf. Fig. 6).

5. PHYSICAL CHARACTERISTICS OF SUPERCONDUCTORS WITH A PSEUDOGAP

As is known, the Ginzburg–Landau equations determine two characteristic lengths, the coherence length and the penetration depth for the magnetic field.

The coherence length at a given temperature, $\xi(T)$, is the characteristic scale of inhomogeneity in the order parameter Δ, which means it is actually the size of the Cooper pair:

$$\xi^2(T)=-\frac{C}{A}. \tag{42}$$

In ordinary superconductors (in the absence of a pseudogap),

$$\xi^2_{\rm BCS}(T)=-\frac{C_0}{A_0}, \tag{43}$$

$$\xi_{\rm BCS}(T)\approx 0.74\frac{\xi_0}{\sqrt{1-T/T_c}}, \tag{44}$$

where $\xi_0=0.18v_F/T_c$. For our case we have

$$\frac{\xi^2(T)}{\xi^2_{\rm BCS}(T)}=\frac{K_C}{K_A}. \tag{45}$$

The corresponding dependence of $\xi^2(T)/\xi^2_{\rm BCS}(T)$ on the parameter W/T_{c0} for the case of d pairing and incommensurate short-range order fluctuations is depicted in Fig. 10.

The penetration depth for the magnetic field in an ordinary superconductors is given by the formula

$$\lambda_{\rm BCS}(T)=\frac{1}{\sqrt{2}}\,\frac{\lambda_0}{\sqrt{1-T/T_c}}, \tag{46}$$

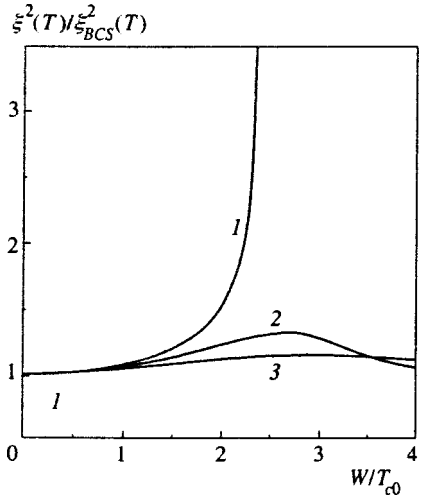

FIG. 10. The coherence length $\xi^2(T)/\xi^2_{\rm BCS}(T)$ as a function of the effective pseudogap width W/T_{c0} in the model of d pairing: curve *1*, $\alpha=\pi/4$; curve *2*, $\alpha=\pi/8$; and curve *3*, $\alpha=\pi/12$.

where $\lambda_0^2=mc^2/4\pi ne^2$ determines the penetration depth at $T=0$. For the general case we have an expression for the penetration depth in terms of the Ginzburg–Landau coefficients:

$$\lambda^2(T)=-\frac{c^2}{32\pi e^2}\,\frac{B}{AC}. \tag{47}$$

Then in the adopted model we have

$$\frac{\lambda(T)}{\lambda_{\rm BCS}(T)}=\left(\frac{K_B}{K_AK_C}\right)^{1/2}. \tag{48}$$

Curves representing the dependence of this parameter on the effective pseudogap width for the case of d pairing are depicted in Fig. 11.

Now let us calculate the Ginzburg–Landau parameter

$$\kappa=\frac{\lambda(T)}{\xi(T)}=\frac{c}{4eC}\sqrt{\frac{B}{2\pi}}. \tag{49}$$

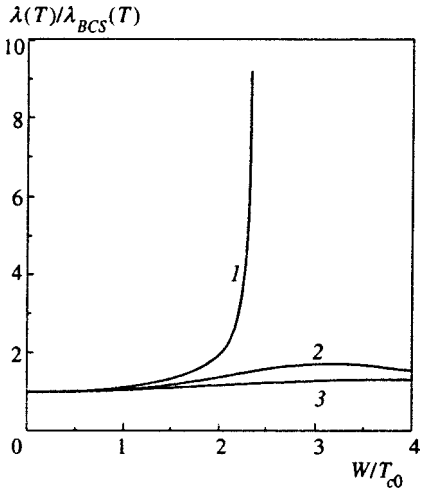

FIG. 11. The penetration depth $\lambda(T)/\lambda_{\rm BCS}(T)$ as a function of the effective pseudogap width W/T_{c0} in the model of d pairing: curve *1*, $\alpha=\pi/4$; curve *2*, $\alpha=\pi/8$; and curve *3*, $\alpha=\pi/12$.

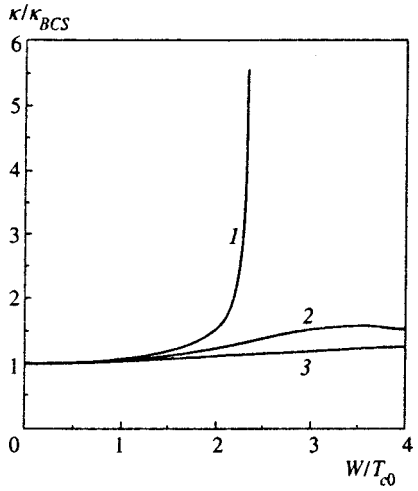

FIG. 12. The Ginzburg–Landau parameter $\kappa/\kappa_{\rm BCS}$ as a function of the effective pseudogap width W/T_{c0} in the model of d pairing: curve *1*, $\alpha=\pi/4$; curve *2*, $\alpha=\pi/8$; and curve *3*, $\alpha=\pi/12$.

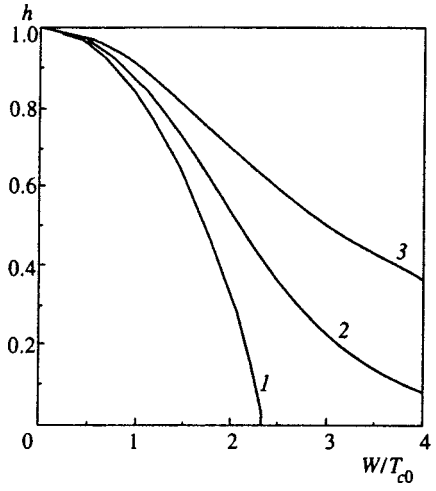

FIG. 13. The normalized slope of the curves for the upper critical field as a function of the effective pseudogap width W/T_{c0} in the model of d pairing: curve *1*, $\alpha=\pi/4$; curve *2*, $\alpha=\pi/8$; and curve *3*, $\alpha=\pi/12$.

In this model of a superconductor,

$$\frac{\kappa}{\kappa_{\rm BCS}}=\frac{\sqrt{K_B}}{K_C}, \tag{50}$$

where

$$\kappa_{\rm BCS}=\frac{3c}{\sqrt{7\zeta(3)}\,e}\,\frac{T_c}{v_F^2\sqrt{N(0)}} \tag{51}$$

is the Ginzburg–Landau parameter for the ordinary case. Curves representing the dependence of $\kappa/\kappa_{\rm BCS}$ on W/T_{c0} for the case of d pairing are depicted in Fig. 12.

Near T_c the upper critical field H_{c2} is expressed in terms of Ginzburg–Landau coefficients:

$$H_{c2}=-\frac{\phi_0}{2\pi}\,\frac{A}{C}, \tag{52}$$

where $\phi_0=c\pi/e$ is the quantum of magnetic flux. Then the slope of the curve for the upper critical field near T_c is

$$\left|\frac{dH_{c2}}{dT}\right|_{T_c}=\frac{24\pi\phi_0}{7\zeta(3)v_F^2}\,T_c\frac{K_A}{K_C}. \tag{53}$$

Curves representing the dependence of the slope of the curves for the field, $|dH_{c2}/dT|_{T_c}$, normalized to the slope of the curves for the field at T_{c0}, on the effective pseudogap width W/T_{c0} for the case of d pairing are depicted in Fig. 13. We see that the slope rapidly decreases with increasing pseudogap width.

We can also calculate the size of the heat-capacity discontinuity at the transition point, which is generally calculated by the formula

$$\frac{C_s-C_n}{\Omega}=\frac{T_c}{B}\left(\frac{A}{T-T_c}\right)^2, \tag{54}$$

where C_s and C_n are the heat capacities of the superconducting and normal states, respectively, and Ω is the volume. This readily yields a formula for the size of the heat-capacity discontinuity at T_{c0} $(W=0)$:

$$\left(\frac{C_s-C_n}{\Omega}\right)_{T_{c0}}=N(0)\,\frac{8\pi^2T_{c0}}{7\zeta(3)}. \tag{55}$$

Then the size of the heat-capacity discontinuity in our model can be expressed in terms of the dimensionless coefficients K_A and K_B as follows:

$$\frac{(C_s-C_n)_{T_c}}{(C_s-C_n)_{T_{c0}}}=\frac{T_c}{T_{c0}}\,\frac{K_A^2}{K_B}. \tag{56}$$

Curves representing the dependence of the size of the heat-capacity discontinuity on the effective pseudogap width for the case of d pairing are depicted in Fig. 14. We see that the discontinuity diminishes as the pseudogap widens.

Curves representing the dependence of the above quantities for the case of s pairing and for the model of commensurate fluctuations are more or less (qualitatively) similar to

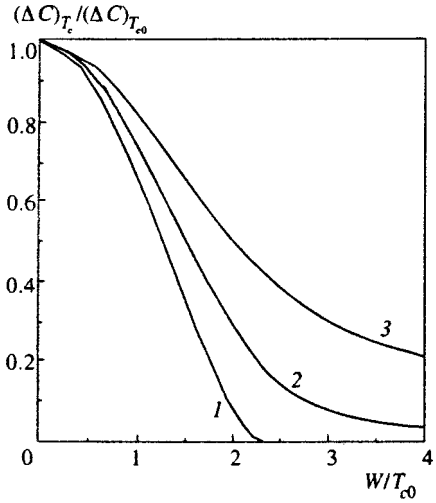

FIG. 14. The normalized size of the heat-capacity discontinuity as a function of the effective pseudogap width W/T_{c0} in the model of d pairing: curve *1*, $\alpha=\pi/4$; curve *2*, $\alpha=\pi/8$; and curve *3*, $\alpha=\pi/12$.

those depicted in Figs. 10–14, differing in the scale along the W/T_{c0} axis, in accordance with Figs. 5 and 6.

6. CONCLUSION

We have studied a very simple model of a pseudogap in a two-dimensional electron model, which nevertheless qualitatively explains a number of observed features of the electron structure of underdoped high-T_c superconducting systems. In particular, with this model one can easily obtain the *d* symmetry of the pseudogap state, a symmetry that is due to the pattern of the hot sections on the Fermi surface caused by strong scattering by fluctuations of short-range (antiferromagnetic) order. Naturally, the model can be directly generalized to the case of a large number of hot sections, and it can reformulated in a way that is closer to the model of hot points (Refs. 10 and 11); other generalizations can also be made fairly easily.

The main simplifying assumption (and the main drawback) of the model is that we use the $\xi\to\infty$ limit for the fluctuation correlation length, due to which the main results can be written as formulas. In reality ξ is not very large and depends on the temperature and the degree of doping, so that it is an important parameter that controls the physical picture of all phenomena. Our model allows, at least in principle, a generalization to finite ξ in the sense of Refs. 17 and 18, but all calculations becomes extremely involved. At the same time it is clear that the effect of a finite ξ reduces mainly to a situation in which the pseudogap becomes closed,[17,18] so that in this sense (as noted earlier) it simulates a decrease in the size of the hot sections in our model. This is true for effects basically controlled by the density of states (an example of a corresponding quantity is the transition temperature T_c). At the same time, this is not true of ''kinetic'' quantities (determined by the two-particle Green's function), such as the coefficient C of the gradient term in the Ginzburg–Landau expansion.

Another radical simplification of our model is the assumption that short-range order fluctuations are static and Gaussian. The validity of this assumption can be justified in the high-temperature limit $T\gg\omega_{sf}$, where ω_{sf} is the characteristic frequency of spin fluctuations.[9–11] Accordingly, the validity of the assumption that the fluctuations are static is questionable at temperatures near T_c. Nevertheless, our investigation shows that the Ginzburg–Landau expansion provides a good description of the influence of the main effect of ''disintegration'' of certain sections of the Fermi surface on the main characteristics of a superconductor with a pseudogap, and demonstrates the important role of pseudogap anomalies in the formation of a superconducting state in the region of the phase diagram of high-T_c systems where these effects manifest themselves already in the normal phase. More realistic models will be analyzed later.

The authors are grateful to É. Z. Kuchinskiĭ for useful discussions. The present work was partially supported by the Russian Fund for Fundamental Research (Project 96-02-16065), by the Statistical Physics State Program (Project IX.1), and by the High-T_c Superconductivity State Program of the Russian Ministry of Science (Project 96-0.51).

*)E-mail: posazhen@ief.uran.ru
†)E-mail: sadovski@ief.uran.ru
1)A similar treatment of the Ginzburg–Landau expansion in a model of the Peierls transition was carried out by McKenzie.[21]

[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. Emery, S. A. Kivelson, and O. Zachar, Phys. Rev. B **56**, 6120 (1997).
[7] A. P. Kampf and J. R. Schrieffer, Phys. Rev. B **41**, 6399 (1990); **42**, 7967 (1990).
[8] V. Barzykin and D. Pines, Phys. Rev. B **52**, 13 585 (1995).
[9] D. Pines, Tr. J. of Physics **20**, 535 (1996).
[10] J. Schmalian, D. Pines, and Stojkovic, Phys. Rev. Lett. **80**, 3839 (1998).
[11] J. Schmalian, D. Pines, and Stojkovic, E-prints archive cond-mat/9804129.
[12] A. T. Zheleznyak, V. M. Yakovenko, and I. E. Dzyaloshinskii, Phys. Rev. B **55**, 3200 (1997).
[13] D. S. Dessau, Z.-X. Shen, D. M. King, D. S. Marshall, L. W. Lombardo, P. H. Dickinson, A. G. Loeser, J. DiCarlo, C.-H. Park, A. Kapitulnik, and W. E. Spicer, Phys. Rev. Lett. **71**, 2781 (1993).
[14] Z.-X. Shen and D. S. Dessau, Phys. Rep. **253**, 1 (1995).
[15] M. V. Sadovskiĭ, Zh. Éksp. Teor. Fiz. **66**, 1720 (1974) [Sov. Phys. JETP **39**, 845 (1974)].
[16] M. V. Sadovskiĭ, Fiz. Tverd. Tela (Leningrad) **16**, 2504 (1974) [Sov. Phys. Solid State **16**, 1632 (1975)].
[17] M. V. Sadovskiĭ, Zh. Éksp. Teor. Fiz. **77**, 2070 (1979) [Sov. Phys. JETP **50**, 989 (1979)].
[18] M. V. Sadovskii and A. A. Timofeev, Sverkhprovodimost': Fiz., Khim., Tekhnol. **4**, 11 (1991) [Supercond., Phys. Chem. Technol. **4**, 9 (1991)]; J. Mosc. Phys. Soc. **1**, 391 (1991).
[19] W. Wonneberger and R. Lautenschlager, J. Phys. C: Solid State Phys. **9**, 2865 (1976).
[20] A. I. Posazhennikova and M. V. Sadovskiĭ, Zh. Éksp. Teor. Fiz. **112**, 2124 (1997) [JETP **85**, 1162 (1997)].
[21] R. H. McKenzie, Phys. Rev. B **52**, 16 428 (1995).

Translated by Eugene Yankovsky

Pseudogap and symmetry of superconducting order parameter in cuprates

A. A. Ovchinnikov, M. Ya. Ovchinnikova,*) E. A. Plekhanov

Institute of Chemical Physics, Russian Academy of Sciences, 117977 Moscow, Russia
(Submitted 16 June 1998)
Zh. Éksp. Teor. Fiz. **115**, 649–674 (February 1999)

The phase diagram, nature of the normal state pseudogap, type of the Fermi surface, and behavior of the superconducting gap in various cuprates are discussed in terms of a correlated state with valence bonds. The variational correlated state, which is a band analogue of the Anderson (RVB) states, is constructed using local unitary transformations. Formation of valence bonds causes attraction between holes in the *d*-channel and corresponding superconductivity compatible with antiferromagnetic spin order. Our calculations indicate that there is a fairly wide range of doping with antiferromagnetic order in isolated CuO_2 planes. The shape of the Fermi surface and phase transition curve are sensitive to the value and sign of the hopping interaction t' between diagonal neighboring sites. In underdoped samples, the dielectrization of various sections of the Fermi boundary, depending on the sign of t', gives rise to a pseudogap detected in photoemission spectra for various quasi-momentum directions. In particular, in bismuth- and yttrium-based ceramics ($t'>0$), the transition from the normal state of overdoped samples to the pseudogap state of underdoped samples corresponds to the onset of dielectrization on the Brillouin zone boundary near $\mathbf{k}=(0,\pi)$ and transition from ''large'' to ''small'' Fermi surfaces. The hypothesis about *s*-wave superconductivity of La- and Nd-based ceramics has been revised: a situation is predicted when, notwithstanding the *d*-wave symmetry of the superconducting order parameter, the excitation energy on the Fermi surface does not vanish at all points of the phase space owing to the dielectrization of the Fermi boundary at $k_x=\pm k_y$. The model with orthorhombic distortions and two peaks on the curve of T_c versus doping is discussed in connection with experimental data for the yttrium-based ceramic.
[S1063-7761(99)01902-2]

1. INTRODUCTION

In recent years, important results concerning electronic structure of HTSC[1–6] have been reported. They include the proof of the *d*-wave symmetry of the superconducting order parameter in a number of cuprates,[7,8] detection of such effects as a small Fermi surface,[9] a pseudogap in the normal state[10,11] of underdoped samples, and characteristic resonances in spectra of inelastic neutron scattering. The task of theory is to develop a self-consistent description of these phenomena and their relation to the phase diagram and magnetic properties of HTSCs.

Theoretical studies of strongly correlated systems (localized approaches, calculations for finite clusters, and band calculations) have identified the most important types of correlations,[1,2] namely, the antiferromagnetic alternation of spins and short-range correlations of type of the valence bond formation. The former is fairly adequately described by both the localized approach and mean-field band calculations with double magnetic elementary cells. The important role of antiferromagnetic correlations is also recognized in recent publications.[12,13] The idea of a correlation nature of superconducting pairing, which was set forth and justified in some publications,[14,15] has remained quite attractive. According to Refs. 16 and 17, the valence bond formation causes attraction between holes in the *d*-channel and *d*-wave superconductivity. One test of the theory should be its ability to account for the anisotropic pseudogap in the normal state of underdoped samples. In preliminary interpretations,[10,11] the pseudogap anisotropy was identified with the anisotropy of the *d*-wave superconducting gap. For this reason, the pseudogap was considered as a precursor of superconductivity in underdoped samples. Our previous publications[16,17] expressed a different view on the pseudogap nature and related it to the structure of the lower Hubbard subband.

The present work is a continuation of our pseudogap studies. Here we set forth a clear and quantitatively accurate interpretation of the phase diagram and some spectral properties of cuprates in terms of a variational correlated state, which takes into account both antiferromagnetic correlations and correlations of the valence bond type. A classification of cuprates in accordance with the Fermi surface topology and pseudogap anisotropy in the normal state of underdoped samples is proposed. Experiments that could test the proposed scheme are discussed. In particular, we propose arguments in favor of the hypothesis of *d*-wave superconductivity in all cuprates, including $La_{2-x}Sr_xCuO_4$ (LSCO) and $Nd_{2-x}Ce_xCuO_{4-y}$ (NCCO), which have been traditionally treated as *s*-wave superconductors.

The term resonating valence bonds (RVB) first suggested by Anderson[18] means that system configurations are composed of singlet pairs of particles that form the bonds.

Recently[16,19] we constructed variational correlated functions—the band analogues of RVB states—for a Hubbard model with arbitrary doping. The method of unitary local transformations used in Refs. 16 and 19 (in contrast to representations like Gutzwiller's ansatz[20]) allows one to construct not only the correlated function, but also an effective Hamiltonian. The latter (unlike the Hamiltonian of the t–J model) is not subject to additional constraints, so it can be analyzed in the mean-field approximation. This allowed us to calculate constants of superconducting pairing variationally without empirical parameters, and calculations did not violate the additional local constraints.

The specific goal of the present work is to study (on the base of the generalized Hubbard model) the role of weak hopping interaction t' between neighboring diagonal sites in formation of the pseudogap, the effect of t' on the phase diagram, the Fermi surface topology, and behavior of the superconducting gap. Concurrently, we propose a possible classification of specific cuprates in accordance with the types of their electron bands and Fermi surfaces. As background, Sec. 2 gives a brief description of the calculation techniques and main results of the previous work.[16]

2. CORRELATED STATE OF VALENCE BONDS IN THE BAND MODEL

Consider the generalized Hubbard model providing the single band mapping[21–23] of the CuO_2 plane of HTSC:

$$H=H(U,t)+\Delta H(V,t'),$$

$$H(U,t)=-t\sum_{\langle nm\rangle,\sigma}(c^{\dagger}_{n\sigma}c_{m\sigma}+\text{H.c.})+\sum_{n}Un_{n\uparrow}n_{n\downarrow}, \quad (1)$$

$$\Delta H(V,t')=V\sum_{\langle nm\rangle}n_n n_m+t'\sum_{\langle\langle nm\rangle\rangle}\sum_{\sigma}(c^{\dagger}_{n\sigma}c_{m\sigma}+\text{H.c.}). \quad (2)$$

As will be shown below, the low-temperature properties of a system described by Eq. (1) are very sensitive to the weak hopping interaction t' between diagonal neighboring sites $\langle\langle nm\rangle\rangle$. For this reason, the interaction t' and Coulomb-like interaction V between neighboring sites are added to the Hamiltonian $\tilde{H}(U,t)$ of the basic Hubbard model (1).

The variational function Ψ of the correlated state of valence bonds is constructed by applying a unitary transformation to the uncorrelated state Φ:

$$\Psi=\hat{W}(\alpha)\Phi, \quad \hat{W}(\alpha)=\exp(\alpha Z), \quad Z=\sum_{\langle nm\rangle}Z_{nm}. \quad (3)$$

Here Φ is an uncorrelated function of the most general form, namely, a BCS wave function with a double magnetic cell for testing the feasibility of anticipated antiferromagnetic and superconducting pairing. The unitary operator $W(\alpha)$ responsible for valence bond formation depends on the variational parameter α. The local anti-Hermitian operator Z_{nm} acting on the bond $\langle nm\rangle$ between neighboring sites is

$$Z_{nm}=-\frac{1}{2}\sum_{\sigma}j_{nm\sigma}\Delta_{nm,-\sigma},$$
$$j_{nm\sigma}=(c^{\dagger}_{n\sigma}c_{m\sigma}-\text{H.c.}), \quad \Delta_{nm,-\sigma}=n_{n-\sigma}-n_{m-\sigma}. \quad (4)$$

The sense of this transformation is illustrated by a two-site system $\{a,b\}$ with molecular orbitals $g_\sigma(u_\sigma)=(a_\sigma \pm b_\sigma)/\sqrt{2}$. In this case, operator $W=\exp(\alpha Z_{ab}) \equiv\exp[\alpha(g^{\dagger}_{\uparrow}g^{\dagger}_{\downarrow}u_{\downarrow}u_{\uparrow}-\text{H.c.})]$ transforms the uncorrelated two-particle dimer state $\Phi(a,b)=|g^{\dagger}_{\uparrow}g^{\dagger}_{\downarrow}\rangle$ to the exact singlet dimer state when α assumes the optimal value.

Thus, the unitary transformation W allows one to adjust the hole localization degree and optimize the bond charge state. The unitarity of W enables one to calculate the effective (transformed) Hamiltonian

$$\tilde{H}(\alpha)=W^{\dagger}(\alpha)HW(\alpha), \quad (5)$$

acting in the space of functions Φ, and to express the average energy

$$\overline{H}=\langle\Psi H\Psi\rangle=\langle\Phi\tilde{H}\Phi\rangle \quad (6)$$

in terms of one-electron averages. Hence it is possible to use the self-consistent procedure of energy minimization with respect to Φ and then with respect to the variational parameter α.

Previously two types of valence bond structures were studied in detail. One of them is an alternating structure of nonoverlapping dimers of a particular orientation.[19] In its description, $\alpha=\alpha(nm)$ was assumed to be nonzero only for intradimer bonds, and an exact formula for the effective Hamiltonian (5) was derived. The pesent study, like that in Ref. 16, is concerned with another type of state, namely, a homogeneous state of valence bonds with parameter α common for all bonds in the system. In this case, the effective Hamiltonian was calculated and analyzed by taking terms of the two lowest orders in α:

$$\tilde{H}(\alpha)\approx H+\alpha[H,Z]+\frac{\alpha^2}{2}[[H,Z],Z]. \quad (7)$$

Thus, expansion (7) is applicable only for $U/t\leqslant 9$, which corresponds to small values of the variational parameter $\alpha\leqslant 0.23$. An explicit expression for the effective Hamiltonian $\tilde{H}(U,t)$ in terms of fermion operators allowed us to obtain self-consistent solutions with antiferromagnetic and superconducting order. The calculation procedure was described in detail.[16] In reality, all calculation in the previous[16] and present work were performed with the effective Hamiltonian

$$\tilde{H}=\tilde{H}(\alpha,U,t)+\Delta H(V,t'), \quad (8)$$

in which the main part $H(U,t)$ was acted upon by the unitary transformation, according to Eq. (7), whereas in the weak interaction $\Delta H(V,t')$ only the terms of zeroth order in α were retained.

The basic results of Ref. 16 are the following.

1. Most of the gain in energy is due to the antiferromagnetic ordering of spins accompanied by formation of valence bonds (the optimal α is nonzero). Valence bond correlations

narrow the doping range $|1-n|<\delta_c$ in which antiferromagnetic order occurs: $\delta_c \sim 0.3$ at $U/t \sim 8$, whereas the simple mean-field approximation predicts $\delta_c \sim 0.45$. But both estimates of the region of long-range two-dimensional antiferromagnetism are considerably wider than the experimentally determined range of bulk antiferromagnetism $\delta_c^{\exp} \sim 0.05$. It is feasible that in reality the range of antiferromagnetic correlations R_{AF} is finite, although fairly large. A solution of the effective problem in the mean-field approximation cannot determine this radius. But for calculating the energy and short-range effects of valence bond formation, it suffices that R_{AF} is much larger than the bond length, unless we investigate diffusion or spin-wave processes of antiferromagnetic order breaking at long distances. Arguments in favor of this two-dimensional antiferromagnetism in CuO_2 planes in a wide range of doping are given in Ref. 16 and in this paper.

2. The attraction between holes in the *d*-channel is largely due to terms like

$$\sim \alpha U[c_{n\sigma}^{\dagger} c_{m\sigma} n_{n,-\sigma} n_{m,-\sigma} + \text{H.c.}],$$

which appear during formation of valence bonds. Unlike Hirsch's empirical coupling constants of correlated hopping interaction,[15] here it is determined variationally through the parameter α. Note that a similar correlated hopping interaction occurs also in the t–J Hamiltonian, but, unlike the effective Hamiltonian $\tilde{H}(\alpha)$, it cannot be analyzed in the mean-field approximation since double occupation of sites is strictly forbidden.

3. *d*-wave superconductivity exists only in the doping range of two-dimensional antiferromagnetism, i.e., where $\delta<\delta_c$ because the density of states on the Fermi boundary increases when the original band is split into two.

To conclude this section, let us recall the sense of quasiparticle bands for the correlated state (3) based on a solution of the effective problem $\tilde{H}$ in the mean-field approximation. The uncorrelated state Φ, which minimizes the average energy (6), is characterized by a set of one-electron energies $E_{k\lambda}$ and the one-electron function $\chi_{k\lambda}^{\dagger}$, which are eigenfunctions of linearized Hamiltonian $(\tilde{H})_L$ (for details see the Appendix and Ref. 16):

$$(\tilde{H})_L \chi_{k\lambda}^{\dagger} = E_{k\lambda}^{\dagger} \chi_{k\lambda}^{\dagger}. \tag{9}$$

For the most general type of states $\Psi = \hat{W}(\alpha)\Phi$ with both antiferromagnetic and superconducting order, $E_{k\lambda}$ and $\chi_{k\lambda}^{\dagger}$ are derived from the equations

$$\chi_{k\lambda}^{\dagger} = \{c_{k\uparrow}^{\dagger}, c_{\tilde{k}\uparrow}^{\dagger}, c_{\_k\downarrow}, c_{\_\tilde{k}\downarrow}\}_i S_{i\lambda}, \quad i,j,\lambda = 1,\ldots,4,$$

$$\tilde{k} = (\pi,\pi) + k, \tag{10}$$

$$h_{ij}(k) S_{j\lambda} = S_{i\lambda} E_{k\lambda}, \quad i,j,\lambda = 1,\ldots,4, \tag{11}$$

$$E_{k\lambda} = \pm \sqrt{(E_{\nu}^{\mathrm{AF}} - \mu)^2 + W_{\nu}^2}, \quad \nu = 1,2. \tag{12}$$

The matrix h_{ij} and expressions for E_{ν}^{AF} and W_{ν}^{AF} are given by Eqs. (33), (38), and (39) in the Appendix (see also Ref. 16). In the absence of superconductivity, when the function Φ does not describe pairing of electrons with different spin projections, Fermi excitations are determined by the upper and lower Hubbard subbands $E_{1(2)}^{\mathrm{AF}}$.

Finally, note that any time-dependent perturbation of the form

$$\hat{V}(t) = \sum_k [v_k(t) c_{k\sigma} + \text{H.c.}]$$

acting on the system in the correlated ground state $\Psi = W\Phi$ generates both one-particle and multiparticle excitations. Specifically,

$$\hat{V}(t)\Psi = W \sum_k [v_k(t) \tilde{c}_{k\sigma} + \text{H.c.}] |\Phi\rangle. \tag{13}$$

Here the ''dressed'' operator $\tilde{c}_{k\sigma} = W^{\dagger} c_{k\sigma} W$ contains both one-fermion and multifermion operators:

$$\tilde{c}_{k\sigma} = \frac{1}{\sqrt{N}} \sum_n \left\{ c_{n\sigma} - \frac{\alpha}{2} \sum_{m \in \langle nm \rangle} (c_{n\sigma} j_{nm,-\sigma} + c_{m\sigma} \Delta_{nm,-\sigma}) + \frac{\alpha^2}{2} [\ldots] \right\}. \tag{14}$$

The operators j and Δ in Eq. (14) are determined by Eq. (4).

3. PHASE DIAGRAM AND STRUCTURES OF BANDS AND FERMI SURFACE

Let us discuss in detail the phase diagram of the model (1), (2). The calculation techniques and procedures were described in Ref. 16.

One unexpected result was that the phase boundary, i.e., the curve of the superconducting transition temperature T_c as a function of doping $\delta = |1-n|$, had a broad plateau in the basic Hubbard model $H(U,t)$, and the critical temperature remained almost constant up to $\delta \sim 0.02$ at $U/t \sim 8$. The calculations for the model $H(U,t)$ did not show a sharp drop of $T_c(\delta)$ on both sides of the optimal doping δ_{opt} corresponding to maximum T_c. In an attempt to eliminate this discrepancy, interactions defined by Eq. (2) were introduced. It was shown that the interaction $V>0$ changes the shape of $T_c(\delta)$ in the desired manner, but it also suppresses superconductivity and reduces T_c considerably. In the present work, we have used $V/t = 0.1$ in all calculations. Given this value of V and $t'>0$, the maximum critical temperature is $T_c^{\max} \sim 0.014t$, which gives $T_c^{\max} \sim 80$ K for $t \sim 0.5$ eV.[22] For $V/t = 0$ and $t'>0$, we have $T_c^{\max} = 0.023t \sim 135$ K.

The most interesting question is the effect of the hopping interaction t' between nearest diagonal lattice sites $\langle\langle nm \rangle\rangle$ with $|n-m| = \sqrt{2}$ on the function $T_c(\delta)$. The interest in interaction t' is connected with three reasons.

1. This interaction has considerable effect on the Fermi surface structure and its image in angle-resolved photoemission spectra (ARPES). This effect is connected with the position of flat regions of lower electron band near the Fermi surface.

2. In the one-band projection of the Emery model, the magnitude and even the sign of constant t' (unlike other parameters U, t, and V of the Hubbard model) are sensitive to the input parameters ϵ_d, ϵ_p, t_{pd}, and t_{pp} of the CuO_2 plane owing to the competition between two diagonal-hopping channels, namely, the direct channel via t_{pp} and the

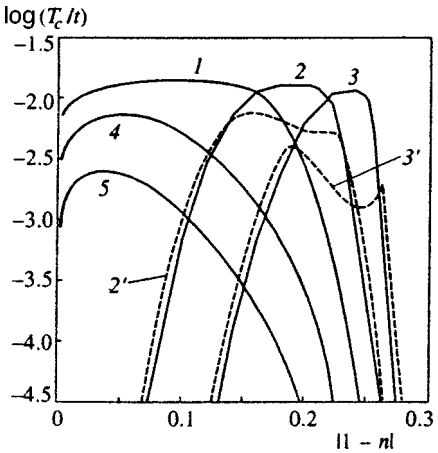

FIG. 1. Logarithm of the superconductivity transition temperature T_c versus doping $\delta=|1-n|$ for models with $U=8t$, $V=0.1t$, and various t'. Curves *1*, *2*, *3*, *4*, and *5* are characterized by values $t'/t=0$, 0.05, 0.1, -0.05, and -0.1, respectively. The dashed lines plot the results for the model (24) with the orthorhombic perturbation $\tau=0.05t$ and $t'=0.05t$ (curve *2′*) and $t'=0.1t$ (curve *3′*) at the same U and V.

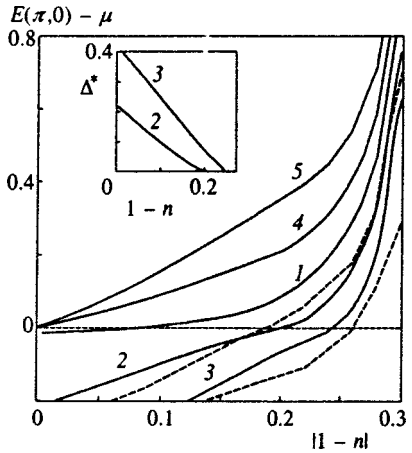

FIG. 2. Energy $E(0,\pi)-\mu$ of the lower Hubbard subband measured with respect to the chemical potential, versus doping for the same model parameters as in Fig. 1. The curves *1*, *2*, *3*, *4*, and *5* are characterized by the values $t'/t=0$, 0.05, 0.1, -0.05, and -0.1, respectively (solid lines). The dashed curves plot the maximum energies $E^{\max}_{y(x)}-\mu$ on sections of the lower subband along k_x- or k_y-axes for the models with orthorhombic perturbation. The inset shows the pseudogap Δ^* as a function of doping in the normal state at $t'/t=0.05$ and 0.1 in the regions where a pseudogap exists.

second-order process $\sim t_{pd}^2/(\epsilon_d-\epsilon_p)$.[21,24] Thus, parameter t' depends on the material and its measurements in various cuprates are of fundamental importance.[24–26]

3. As will be shown below, in the presence of antiferromagnetic splitting, energy variations $\delta E=E(\pi,0)-E(\pi/2,\pi/2)\sim -4t'$ along the nesting line are responsible for the pseudogap effects detected in various experiments. Parameter $4t'$ determines the energy scale of these effects. [Hereafter $E(k)=E_1^{\rm AF}(k)$ is the energy of the lower Hubbard subband in the case of hole doping.]

Figure 1 shows the logarithm of the critical temperature T_c as a function of doping for systems with $U=8t$, $V=0.1t$, $t'/t=0,\pm 0.05,\pm 0.1$. The phase curve shape is strongly affected by the magnitude and sign of t'. For $t'>0$ the shape of the $T_c(\delta)$ curve is similar to that measured in the bismuth-based ceramic,[2] in contrast to the curve with a broad plateau at $t'=0$ and $V=0$.[16] The maximum transition temperature $T_c^{\max}$ is almost constant with t' for $t'>0$, but the optimal doping shifts with t' to higher values, while remaining inside the region of two-dimensional (or latent) antiferromagnetism.[16] On the contrary, at $t'<0$ curves of $T_c(\delta)$ have wider peaks, and $T_c^{\max}$ notably decreases with growing $|t'|$. The difference between the $T_c(\delta)$ shapes at $t'<0$ and $t'>0$ is caused by the different behavior of the energies $E(k)$ of the Hubbard subbands (the lower subband in the case of hole doping and the upper in the case of electron doping). Recall that $E(k)$ is one of the eigenvalues of the linearized Hamiltonian $(\tilde{H})_L$, i.e., the one-particle energy obtained by diagonalizing effective Hamiltonian (5) in the mean-field approximation. Figure 2 shows the energy of the lower Hubbard subband with respect to the chemical potential, $E(\pi,0)-\mu$, at the point $M=(\pi,0)$. In systems with $t'>0$ this point corresponds to the most flat region of the band spectrum and to a Van Hove singularity in the density of states $n(E)$. By comparing Figs. 1 and 2, one can see that the doping δ at which the Fermi boundary passes through the point M, i.e., when $E(\pi,0)-\mu=0$, coincides with the optimal doping $\delta_{\rm opt}$ corresponding to maximum T_c (curves *2* and *3* in Figs. 1 and 2).

In systems with $t'<0$ (curves *4* and *5* in Figs. 1 and 2) one has $E(\pi,0)-\mu>0$ for all $\delta=|1-n|$. The density of states on the Fermi boundary is lower than in systems with $t'>0$ and decreases with doping. This leads to a lower $T_c^{\max}$ than in the case of $t'>0$, a wider peak on the phase curve $T_c(\delta)$, and the superconductivity range shifted to lower δ.

The difference between the two different types of bands for $t'>0$ and $t'<0$, and the corresponding Fermi surfaces is illustrated by Fig. 3.

Since in the interesting range of doping $\delta<\delta_c$ two-dimensional antiferromagnetism persists, let us recall some familiar concepts concerning evolution of Hubbard subbands with the strong on-site interaction U. It is very important for interpreting both ARPES data and the occurrence of a pseudogap.

For $U=0$ and $t>0$ there is only one unperturbed band

$$E_0(k)=-2t(\cos k_x+\cos k_y)+4t'\cos k_x\cos k_y \qquad (15)$$

with a minimum at the point $\Gamma(0,0)$ and a maximum at the point $Y(\pi,\pi)$ in the Brillouin zone $-\pi<k_{x(y)}<\pi$ of the original crystal lattice. For $\delta<\delta_c$, the lowest in energy self-consistent solution of the effective Hamiltonian (5) yields nonzero spin densities of opposite signs in the two sublattices:

$$d_l=\frac{1}{2}\sum_\sigma \frac{\sigma}{|\sigma|}\langle c^\dagger_{n\sigma}c_{n+l,\sigma}\rangle_\Phi(-1)^{n_x+n_y},$$

$$|l|=0,\sqrt{2},2,\ldots, \qquad (16)$$

and the band splits into two subbands $E^{\rm AF}_{1(2)}(k)$ separated by the antiferromagnetic gap $\Delta_{\rm AF}\sim Ud_0$. Figure 3 shows the shapes of the lower subbands $E(k_x,k_y)$ in the magnetic Brillouin zone $|k_x\pm k_r|\leqslant\pi$. The corresponding Fermi surfaces of two types of underdoped systems with $t'>0$ and $t'<0$

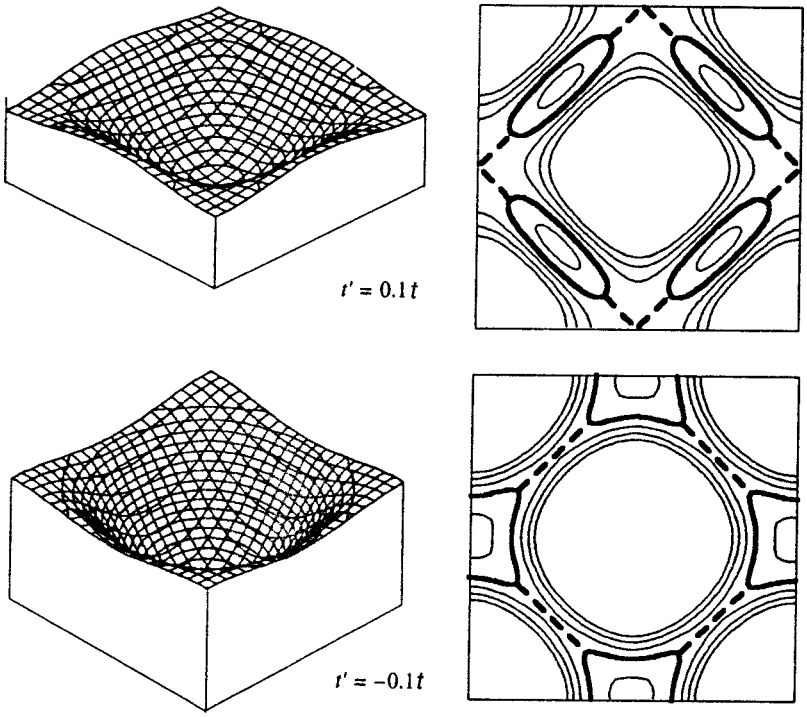

FIG. 3. Left: energy $E(k_x,k_y)$ of the lower Hubbard subband as a function of components k_x and k_y within the magnetic Brillouin zone, $|k_x \pm k_y| \leq \pi$, for an underdoped system ($\delta = 0.15$) with $t'/t = 0.1$ (top) and $t'/t = -0.1$ (bottom); the other parameters are standard: $U/t = 8$ and $V = 0.1t$. Right: Fermi boundaries (solid curves) and nearest energy levels (with a step of $0.02t$) are shown in the full $|k_{x(y)}| \leq \pi$ Brillouin zone. Dielectric sections of the generalized Fermi boundary are shown by bold dashes (see the text).

are shown in the full Brillouin zone. In the absence of superconducting order, the shape of the lower Hubbard subband $E = E_1^{\mathrm{AF}}$ is determined by Eq. (38) in the Appendix, which can be approximated by

$$E(k_x,k_y) = [\epsilon_0 + 2t' \cos k_x \cos k_y + \dots] - \sqrt{[Ud_0 + \dots]^2 + \{2t(\cos k_x + \cos k_y) + \dots\}^2}. \tag{17}$$

The dots in brackets $[\dots]$ and $\{\dots\}$ denote omitted contributions of either even or odd (l_x,l_y) harmonics in (k_x,k_y).

The band spectrum (17) is very different from the standard representation of band energy in terms of a sum of harmonics (so-called tight-binding approximation[27]):

$$E(\mathbf{k}) = \sum_{(l_x,l_y)} t_l \cos k_x l_x \cos k_y l_y,$$
$$l = (0,1),(1,0),(1,1), \dots \tag{18}$$

It is precisely this approximation that is usually used for fitting a band spectrum to ARPES data. The main difference between representations (17) and (18) is that the lower Hubbard subband (17) is periodic with the period of the magnetic Brillouin zone, whereas the spectrum (18) is periodic with the full Brillouin zone. This means that $E(k)$ defined by Eq. (17) is the same at the points $\Gamma = (0,0)$ and $Y = (\pi,\pi)$, which is not the case for the fitting formula (18). There are several reasons why susch different band shapes can be fitted to the same ARPES data.

1. Direct photoemission measures only a part of an electron spectrum below the chemical potential, $E(k) < \mu$, since one can extract an electron only from a populated level. In the case of a single unsplit band and hole doping, this means that photoemission is possible only from states with quasimomenta k near the point $\Gamma = (0,0)$ and impossible from the region near $Y = (\pi,\pi)$. Thus, only the spectrum around the point Γ is scaled and fitted to ARPES data.

2. In the case of antiferromagnetically split bands, both the lower and upper Hubbard subbands are periodic with the magnetic Brillouin zone. But the photoemission intensity has no such period.[28] Roughly speaking, the lower subband is divided into the main part in the first magnetic Brillouin zone (the solid curves in Fig. 4) and the shadow section outside it (the dashed line in Fig. 4). In reality, the transition matrix element drops off gradually in the transition from the main to the shadow part of the band.

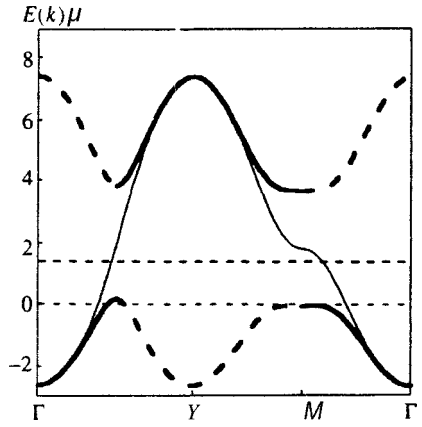

FIG. 4. Profiles of band energy $E(k) - \mu$ as a function of quasimomentum varied along the contour $\Gamma(0,0) \to Y(\pi,\pi) \to M(\pi,0) \to \Gamma$ for a system with $U/t = 8$, $V = 0.1t$, $t' = 0.05t$, and $\delta = 0.2$ in the normal state. The solid and dashed lines correspond to the main (nonshadow) and shadow parts of the lower and upper Hubbard subbands. The thin line shows the spectrum calculated without antiferromagnetic splitting. The horizontal dashed lines are levels of chemical potential in the split and unsplit band.

Figure 4 shows that the shape of the unsplit band below the chemical potential and the nonshadow part of the lower Hubbard subband are similar. This is why both these shapes can be fitted to APRES data, but the two representations yield different ratios t'/t of the hopping interactions.

Direct photoemission, in particular, ARPES, cannot measure the band levels $E_k > \mu$ and thus cannot detect directly the dielectric gap in the hole-doped materials. This gap, however, should reveal itself in electron-doped Nd-based materials with a partial population of the upper Hubbard subband, which is seen, apparently, in the ARPES spectra of these materials at $E-\mu < -300$ meV.[3]

Let us reconsider the effect of diagonal hopping t' on the shape of the spectrum and Fermi surface. The ratios $t'/t = \pm 0.05, \pm 0.1$ used in our calculations are smaller than empirical values $t'/t \sim 0.2$–0.4.[24] The most consistent determination[25] of the ratio t'/t in the t–t'–J model or in t–t'–U Hubbard model is based on fitting the energies of lowest excitations in a Cu_5 cluster with four, five, and six holes to a three-band model whose parameters were calculated in the local density-functional approach (LDA). This procedure yielded $t'/t = 0.17$ (0.13) for the hole (electron) doped materials.[25] Given that t' is very sensitive to model parameters[21] and changes its sign as t_{pp} is varied, the values used in our calculations, namely $\pm 0.05, \pm 0.1$, do not seem anomalous. Even such small variations in t' may radically change the phase diagram and low-energy characteristics of the models. Calculations based on the tight-binding approximation (18) and fitting of the Fermi surface shape to LDA calculations[26] or to bands derived from ARPES data[27] overestimate t'/t, since the approximation based on Eq. (18) ignores the antiferromagnetic band splitting, even though the necessary renormalizations of unperturbed LDA bands are taken into account.

Consider the first type of Fermi surface, which corresponds to $t' > 0$ in a hole-doped material. For small $\delta < \delta_{\rm opt}$, the Fermi surface (shown by solid lines in Fig. 3) has the form of hole pockets around the points $(\pi/2, \pi/2)$. This shape of the Fermi surface has been discussed in several publications.[1,29,30] The heavy dashes in Fig. 3 show the boundaries of the magnetic Brillouin zone on which $E(k) < \mu$ holds. When k changes along any line $\mathbf{k}(l)$ crossing this boundary, the work function $|E(k)-\mu|$ remains finite, and the minimum work function on this line is value of the pseudogap derived from ARPES measurements with $\mathbf{k}$ varied along this line. Note that, if the band were unsplit, like that described by Eq. (18), there would be no trajectory connecting Γ and Y not crossing the real Fermi boundary. On this trajectory, there would always be a point where $E_k - \mu = 0$ and no pseudogap would be detected.

The behavior of the phase curve $T_c(\delta)$ also becomes understandable. The density of states at the Fermi level is maximal at $\delta = \delta_{\rm opt}$ (Fig. 5a). In the range $\delta > \delta_{\rm opt}$ it drops rapidly, so $T_c \to 0$. In the range $\delta < \delta_{\rm opt}$ the density of states also drops, although not so fast. Note that the density of states $n(E)$ in the lower Hubbard subband shown in Fig. 5a reproduces the density of state with the peak due to the Van Hove singularity below the chemical potential, which was

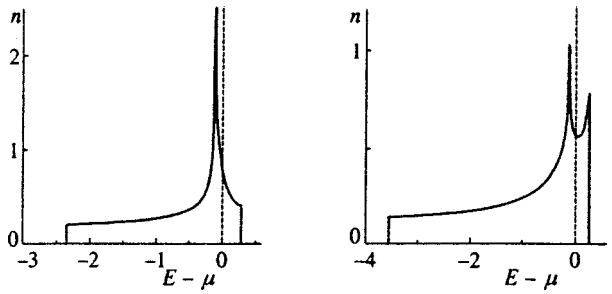

FIG. 5. Density of states $n(E)$ for the same systems as in Fig. 3 with $t'/t = 0.1$ (left) and $t'/t = -0.1$ (right). For $t' > 0$ the Van Hove singularity corresponds to the pseudogap energy $E - \mu = -\Delta^*(\delta)$. For $t' < 0$, two Van Hove singularities correspond to energies $E(\pi,0) - \mu > 0$ and $E(\pi/2, \pi/2) - \mu < 0$.

introduced for interpretation of transport properties of cuprates in the normal state.[31]

Another type of Fermi surface corresponds to $t' < 0$ in hole-doped materials (or to $t' > 0$ in electron-doped systems). Our attention is still focused on the region of two-dimensional antiferromagnetism $\delta < \delta_c$, since only inside this region are the solutions with superconductivity obtained. Two Van Hove singularities in the density of states of the lower Hubbard subband (Fig. 5b) peak at energies $\epsilon_1 = E(\pi,0)$ and $\epsilon_2 = E(\pi/2, \pi/2) < \epsilon_1$. For low doping ($\epsilon_2 < \mu$) the small Fermi surfaces that are shown in Fig. 3 are formed around the points $M = (\pm\pi, 0)$, $(0, \pm\pi)$. The dashed lines in this graph show sections of the boundary of the magnetic Brillouin zone. When the quasimomentum $\mathbf{k}$ varies along a trajectory connecting the points $\Gamma(0,0)$ and $Y(\pi, \pi)$ and crossing one of these boundary sections, the electron work function $\Delta\omega(k)$ measured in the ARPES experiments as a lower edge of the electron distribution function (in brief, as a photoemission edge) nowhere vanishes. The minimum shift $\min|\Delta\omega(k)|$ along this trajectory takes place at the point $k_x + k_y = \pi$ where it crosses the zone boundary.

Thus, for small doping and $t' < 0$ a pseudogap should be detected in the ARPES spectra in the direction $\mathbf{k} \sim (1,1)$. At the same time, the absorption edge should spread, like the broadened photoemission edge on the boundary of the magnetic Brillouin zone in the case of an underdoped dielectric. For high doping, when $\epsilon_2 = E(\pi/2, \pi/2) > \mu$, the Fermi surface transforms to a large connected Fermi surface around the point Γ or Y. Moreover, for antiferromagnetic states, there are both main and shadow Fermi boundaries inside and outside the magnetic Brillouin zone in hole-doped materials. In the case of an electron-doped material, the main and shadow Fermi boundaries interchange.

As in the basic two-parameter Hubbard model $H(U,t)$, in the extended model described by Eqs. (1) and (2) for $t' < 0$ the Fermi level never crosses the energy $\epsilon_1 = E(\pi,0)$ at the point M corresponding to the Van Hove singularity at the dielectric gap edge (Fig. 5). As a result, the maximum critical temperature $T_c^{\max}$ is lowered, the peak on the phase curve $T_c(\delta)$ is broadened, and the entire curve shifts to lower doping.

The main feature of models of this type with $t' < 0$ is the

absence of fermion excitations with zero energy in the superconducting state, regardless of the presence of nodes in the *d*-wave superconducting order parameter. This means that measurements of specific heat and the NMR Knight shift, in other words, all experiments measuring the absolute value of the excitation gap, should display characteristics typical of *s*-wave superconductors, namely an exponential temperature dependence with a finite minimum gap. Such behavior has always been considered to be a proof of *s*-wave symmetry of the order parameter. Calculations based on models with $t'<0$ indicate that *d*-wave superconductors can also have such properties.

4. MANIFESTATIONS OF PSEUDOGAP AND SUPERCONDUCTING GAP IN PHOTOEMISSION SPECTRA

Since one of the major sources of information about HTSC band structure is APRES, let us try to calculate photoemission spectra based on our models. The simplest scheme assumes that the interaction with electric field $E(t)=E_0e^{i\omega t}$ removing a high-energy electron with momentum $\mathbf{q}$ and its projection q_{ab} on the ab plane is described by the operator

$$V(t)\sim \mathbf{E}_0\mathbf{f}(\mathbf{q},\mathbf{k})[e^{i\omega t}a^{\dagger}_{q\sigma}c_{k\sigma}+\text{H.c.}]\delta(q_{ab}-k\pm 2\pi n). \tag{19}$$

Here we do not specify the structure of on-site operators $c_{n\sigma}$ and polarization form-factor $\mathbf{f}(\mathbf{q},\mathbf{k})$ and suppose that it is a smooth function of q and k. Then the direct angle-resolved photoemission (ARPES) signal is proportional to the spectral function

$$A(k\omega)=\frac{1}{Z}\sum_{i,f}|\langle\Psi_f^{N_e-1}|c_{k,\sigma}|\Psi_i^{N_e}\rangle|^2e^{-\beta E_i}\delta \times(\omega-\mu+E_f^{N_e-1}-E_i^{N_e}). \tag{20}$$

Here Z is the partition function and $\beta=1/kT$. In the basis of one-determinant functions Φ, the matrix element in Eq. (20) can be expressed as

$$M(k)=\langle\Phi_f^{N_e-1}|\tilde{c}_{k,\sigma}|\Phi_i^{N_e}\rangle,\quad \tilde{c}_{k\sigma}=W^{\dagger}(\alpha)c_{k\sigma}W(\alpha). \tag{21}$$

Unlike the simple operator $c_{k\sigma}$, the dressed quasiparticle operator $\tilde{c}_{k\sigma}$ including terms of up to second order in α generates both one-particle and multiparticle excitations against the background of the uncorrelated mean-field state Φ. Let us restrict our analysis to one-particle excitations. Then the contribution to $\tilde{c}_{k\sigma}$ can be represented in the form of an expansion in the eigenfunctions $\chi_{k\lambda}$ of the linearized Hamiltonian: $(\tilde{H})_L$

$$\tilde{c}_{k\sigma}=\varphi(k)c_{k\sigma}+\frac{\sigma}{|\sigma|}\eta(k)c_{\tilde{k}\sigma} =\sum_{\lambda}\left\{\left(\frac{1}{2}+\sigma\right)R_{\lambda}\chi_{k\lambda}+\left(\frac{1}{2}-\sigma\right)Q_{\lambda}\chi^{\dagger}_{k\lambda}\right\}. \tag{22}$$

Here $\tilde{k}=k-(\pi,\pi)$, and functions φ, η, R, and Q including terms of up to the second order in α are given by Eqs. (41), (42), and (44) in the Appendix. The final result was calculated by the formula

$$A(k\omega)=\sum_{\lambda}|R_{\lambda}|^2f(E_{\lambda})g(E_{\lambda}-\omega) \equiv\sum_{\lambda}|Q_{\lambda}|^2[1-f(E_{\lambda})]g(E_{\lambda}+\omega), \tag{23}$$

where $f(E)$ is the Fermi distribution function, and $\delta(x)$ is replaced by the intentionally broadened Gaussian function $g(x)=\exp[-(x/\sigma)^2]$ with $\sigma=0.02t$ for more visible presentation of the results and to simulate a finite spectral resolution.

Figure 6 shows calculations of photoemission intensity $A(k,\omega)$ (in arbitrary units) calculated by Eq. (23) as a function of ω (in units of t) for several values of k uniformly distributed over the segments $[C–D]$ or $[A–M–B]$ in the first quadrant of the phase plane (the lower left section of Fig. 6). In an underdoped system with $\delta=0.15$, the shift of the photoemission edge $\Delta\omega(M)$ at the point M occurs in both the superconducting and normal states and it is almost entirely due to the dielectric pseudogap. At the optimal doping $\delta=0.2$, the shift $\Delta\omega(M)$ vanishes in the normal state. This means that at $T<T_c$ this shift is entirely determined by the superconducting gap.

Figure 7 shows the shift $\Delta\omega(k)$ on the generalized Fermi boundary (M,M'), which consists of the dielectric sections of the magnetic Brillouin zone boundary and of the nonshadow parts of the Fermi boundary. Calculations for $\Delta\omega(k)$ were carried out for the superconducting and normal states of an underdoped system with $\delta=0.15$, $t'=0.05t$, $U=8t$, and $T_c=0.0091t$. Figure 7 also shows similar $\Delta\omega(k)$ curves for k varied along the real Fermi boundary (S,S') of the optimally doped system at $\delta=0.2$, other parameter being the same. For this type of band $(t'>0)$, there is always a point on the Fermi surface, $k_x=k_y$, where the superconducting gap vanishes, in accordance with the conventional wisdom about *d*-wave superconductivity.

The situation is different in underdoped systems with $t'<0$. The shift of the photoemission edge, $\Delta\omega(k)$ on the generalized Fermi surface (see the curve for $\delta=0.1$, $t'=-0.05t$, and $U/t=8$ in Fig. 8) does not vanish anywhere for $T<T_c$, although the sign of the superconducting order parameter changes. This means that the features of all physical parameters sensitive to the minimum energy of fermion excitation (such as specific heat, Knight shift in NMR, etc.) should be similar to those of conventional *s*-wave superconductors. Nonetheless, phase-sensitive experiments should still detect a change in the sign of the order parameter.

This section can be summarized differently if we follow the system evolution in the reversed direction from a higher doping $\delta>\delta_c\sim 0.3$ to $\delta=0$, i.e., to the undoped state of the antiferromagnetic dielectric. For $\delta>\delta_c$ we have a metallic paramagnetic state with an unsplit energy band. For $\delta<\delta_c$ there is a long-range, two-dimensional antiferromagnetic order with the energy band split into two and an increase in the

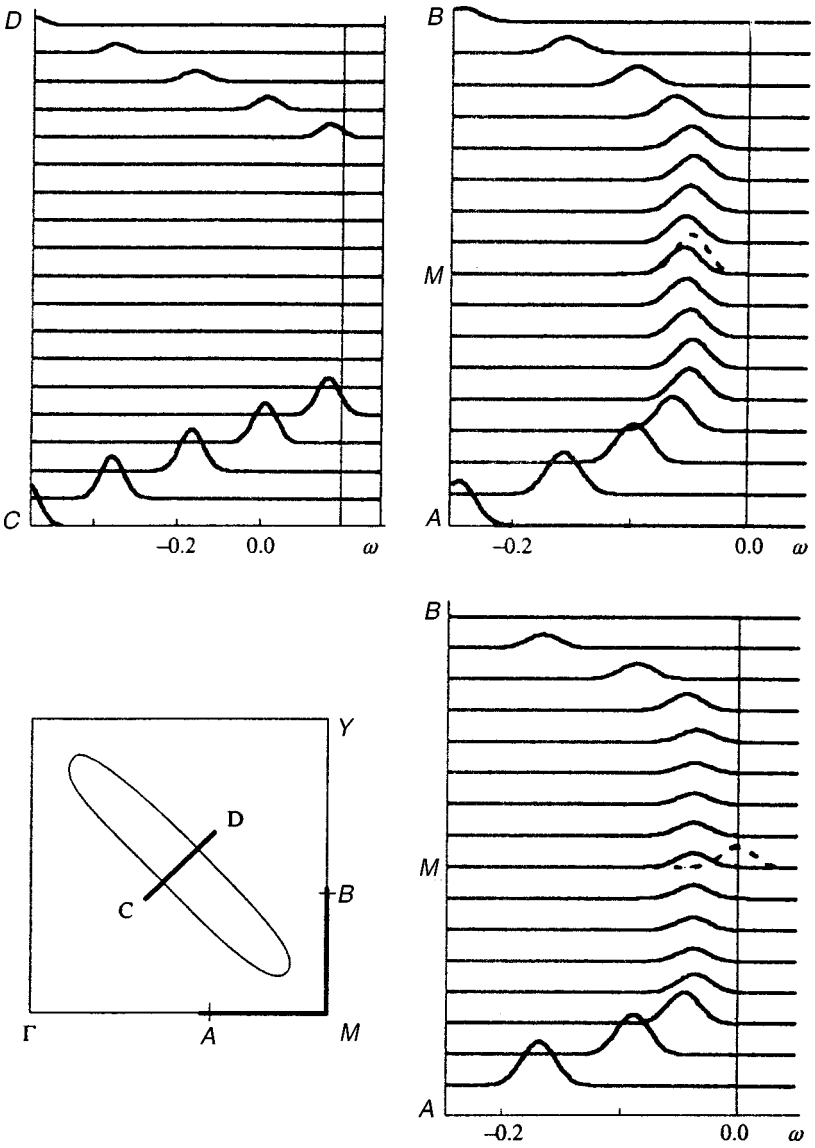

FIG. 6. Top: the spectral functions $A(k\omega)$ plotted against ω and calculated by Eq. (23) for an underdoped system with $\delta=0.15$, $U=8t$, $V=0.1t$, and $t'/t=0.05$ in the superconducting state for a set of quasimomenta uniformly distributed over the segments $[C-D]$ and $[A-M-B]$ in the first quadrant of the full Brillouin zone (left section at the bottom). The bottom right section shows spectral functions for an optimally doped system with $\delta=0.2$ and $k\in[A-M-B]$. The dashed curve corresponding to the point M with $k=(0,\pi)$ shows $A(k\omega)\times 1.5$ for the same system in the normal state for $T=0.02t>T_c$ (the factor 1.5 is introduced for better visibility).

density of states. But in this material, which is still overdoped, the Fermi surface remains large down to $\delta=\delta_{\rm opt}$ at $t'>0$ or $\delta=\delta_2$ at $t'<0$. In the latter case $E(\pi/2,\pi/2)-\mu=0$ at $\delta=\delta_2$. Only below $\delta_{\rm opt}$ (or $\delta<\delta_2$ at $t'<0$) do some sections of the previously large Fermi surface become dielectric, which appears as a pseudogap in specific directions of the quasimomentum in the normal state. In this state, the large Fermi surface decomposes into small hole pockets around the points $(\pm\pi/2,\pm\pi/2)$ at $t'>0$ or around $(0,\pi)$, $(\pi,0)$ at $t'<0$.

5. SUGGESTED CLASSIFICATION OF CUPRATES IN ACCORDANCE WITH BAND CONFIGURATIONS

It seems worthwhile to discuss which of cuprates can be described in terms of which type of electron band ($t'>0$ or $t'<0$).

Usually[32] cuprates are divided into two groups in accordance with their properties. Typical representatives of group I are $Bi_2Sr_2CaCu_2O_{8+\delta}$ (BSCCO) and $YBa_2Cu_3O_{7-\delta}$ (YBCO). Group II includes $La_{2-x}Sr_xCuO_4$ (LSCO) and the only electron-doped material $Nd_{2-x}Ce_xCuO_{4-y}$ (NCCO). The features on which this classification is based are as follows.

1. Different $T_c^{\max}$, which varies between 90–120 K for the cuprates of group I and is about 36 and 25 K for LSCO and NCCO, respectively.[32,33]

2. Vanishing (for group I) and finite (for group II) minimum energy of fermion excitations in the superconducting state. This parameter is derived from measurements of specific heat, the Knight shift of NMR, and other physical properties. These measurements are usually considered as tests of whether the superconducting order parameter has the *d*-wave (group I) or *s*-wave symmetry (LSCO and NCCO of group II).

3. Direct detection of *d*-wave symmetry of the superconducting order parameter in BSCCO and YBCO of group I.[7,8] No such measurements of group II materials are available. Recently[34,35] a small *s*-wave component added to the main *d*-wave order parameter due to orthorhombic distortions has been detected.

4. In ARPES spectra of underdoped BSCCO materials in

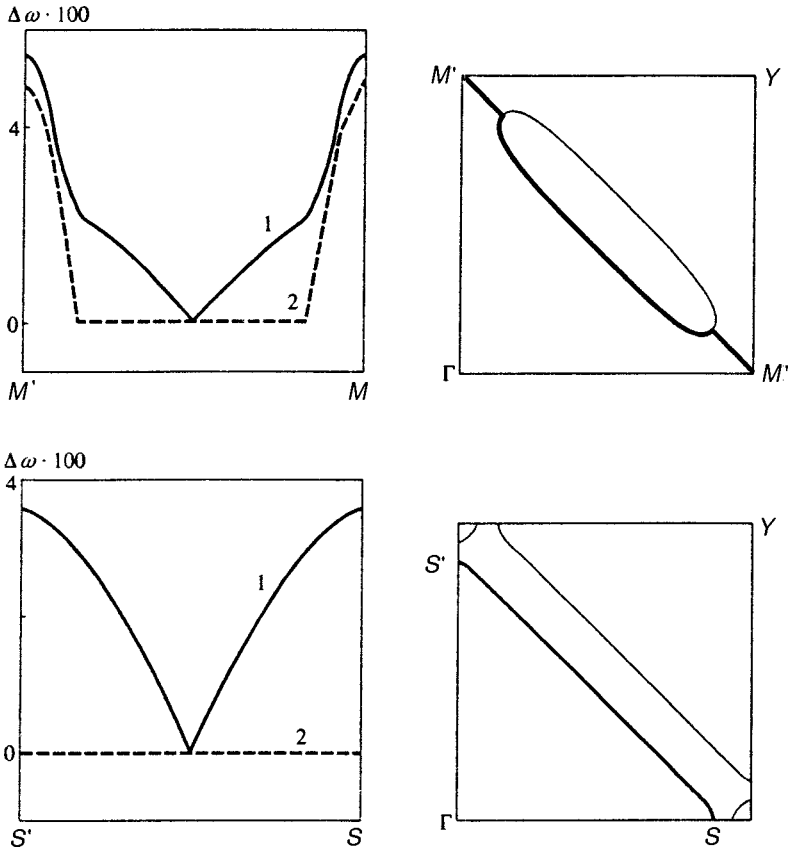

FIG. 7. Top left: minimal energy $\Delta\omega(k)$ (in units of t) of fermion excitations versus quasimomentum k varied along the generalized Fermi boundary $[M',M]$ shown on the right and composed of dielectric parts and nonshadow part of the Fermi boundary around the hole pocket centered at the point $(\pi/2,\pi/2)$. The system parameters are $U=8$, $V=0.1$ and $t'=0.05$ in units of t. Curves *1* and *2* correspond to the superconducting ($T=0.0017t<T_c$) and normal ($T=0.02t>T_c$) states. Bottom: the same for a system with optimal doping $\delta=0.2$ with the quasimomentum varied along the Fermi boundary $[S',S]$ (right).

the normal state, an anisotropic pseudogap (a shift of the electron energy distribution edge) was detected around the point $M=(\pi,0),(0,\pi)$. No such measurements for materials of group II are available.

5. A small Fermi surface in the form of a hole pocket around $k\sim(\pi/2,\pi/2)$ in underdoped BSCCO (group I) was detected.[10,11]

6. The Hall coefficient in LSCO and NCCO (group II) changes its sign at a certain doping.[36–38]

7. Two materials from different groups, namely YBCO and LSCO, have different peaks in coherent inelastic neutron scattering at excitation energy $\hbar\omega\to 0$. In YBCO there is one peak at $Q=(\pi,\pi)$. In LSCO there are four peaks at $Q=(\pi\pm\Delta,\pi)$, $(\pi,\pi\pm\Delta)$.[39,40] Theoretical analysis[41,42] unambiguously ascribes this difference to differences in parameters and shapes of Fermi surfaces of corresponding uncorrelated Emery models, although in an intermediate stage these calculations include renormalizations of the paramagnetic susceptibility, which tent to be based more on empirical data.

8. For the two cuprate groups, the main difference in their Fermi surface shapes derived from quantum-chemical calculations of uncorrelated bands[43] by the LDA method is in the role of interactions between non-nearest neighbors. For cuprates of group I the Fermi surface displays a nesting with $Q\sim(\pi,\pi)$.[43] It is shaped as a square with rounded angles and sides along the lines $|k_x\pm k_y|=\pi$, and is centered at the point Y. In group II the calculated Fermi surface looks more like a turned rounded square with the sides parallel to k_x- and k_y-axes. In terms of the one-band unperturbed model, this form corresponds to $t'/t<0$. Note that unperturbed bands of the tight-binding model cannot be used directly

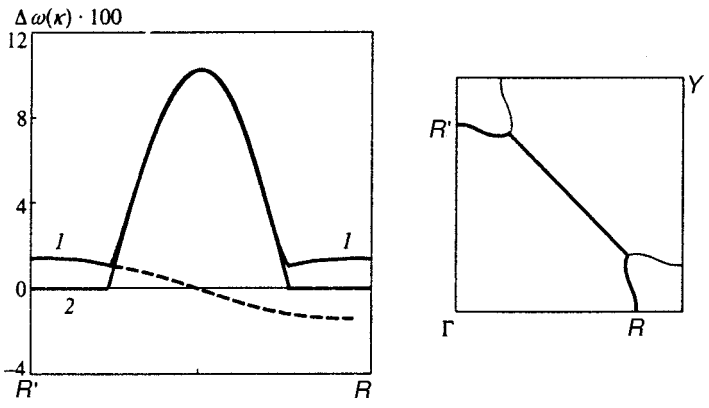

FIG. 8. The same as in Fig. 7 for an underdoped model of type II with $t'=-0.05t$ and $\delta=0.1$ versus k varied along the generalized Fermi boundary composed of sections around hole pockets centered at the points $(\pi,0)$ and $(0,\pi)$, and the section of the dielectric part of the magnetic Brillouin zone boundary. Curves *1* and *2* correspond to the superconducting ($T=0.0017t$) and normal ($T=0.02t$) states. The superconducting gap $W_1^{\mathrm{AF}}(k)$ plotted by the dashed line [Eq. (39) in the Appendix] changes sign.

(without renormalization) to analyze the low-energy processes.

Now let us briefly discuss the direct comparison between uncorrelated bands deriving from LDA calculations and those constructed using ARPES data in accessible regions of the phase space (see the comments on Fig. 4 in the text). On average, the calculations with the predetermined resolution adequately describe the Fermi surface shape, but they overestimate band dispersions in comparison with measurements. This discrepancy is usually eliminated by introducing renormalizations, often of empirical character.[43] The calculated and measured bands can differ because the former do not take into account alternations of the spin density and the resulting valence band splitting. As was mentioned above, the antiferromagnetic gap can be directly observed in photoemission spectra only in electron-doped materials. Sakisaka[44] and King *et al.*[45] in fact detected such a gap at $\hbar\omega = E_k - \mu < -300$ meV, which contradicts LDA calculations. Moreover, the finite resolution in both energy and k may distort the averaged interpretation of ARPES spectra. The uncertainty in the Fermi boundary derived from poorly resolved spectra was analyzed in Ref. 24. It seems necessary to reprocess ARPES spectra on the base of Eqs. (17) and (23) for band energies and intensities, which take account of the antiferromagnetic band splitting and partial dielectrization of the zone boundary. Yet the only material in which the presence of small Fermi surfaces and pseudogap has been experimentally proven is the BSCCO ceramic.[9–11]

All this allows us to associate cuprates of groups I and II with two types of correlated Hubbard models with $t'/t > 0$ and $t'/t < 0$, respectively. Specifically, self-consistent calculations taking account of antiferromagnetic correlations and correlations like valence bonds have revealed the following:

1. The maximum critical temperatures in models with $t'/t = 0.05$ and 0.1 are $T_c^{\max} = 0.0129t$ and $0.0114t$, respectively, whereas in models with $t'/t = -0.05$ and -0.1 only $T_c^{\max} = 0.0072t$ and $0.00247t$. These results apply to models with $V/t = 0.1$. At $V = 0$ the critical temperature is a factor of ~ 1.7 higher. The ratio between $T_c^{\max}$ for the two types of the model is in fair agreement with this ratio for two cuprate groups.

2. In models with $t'/t < 0$, unlike the case $t'/t > 0$, the minimum energy of Fermi excitations in the superconducting state is nonvanishing (see Fig. 8) almost over the entire range of superconductivity ($\delta < \delta_2$), although the superconducting order parameter has the *d*-wave symmetry. This allows us to abandon the hypothesis of the *s*-wave superconductivity in LSCO and NCCO and suggest that *d*-wave superconductivity occurs in all cuprates, including LSCO and NCCO. Owing to the dielectrization of Brillouin zone boundaries in the direction $k_x = k_y$, the latter hypothesis is in accord with measurements that yield a finite minimum energy of excitations. Note also that in electron-doped systems the second type of Fermi surfaces with electron pockets around the points $(\pm\pi, 0)$ and $(0, \pm\pi)$ corresponds to the positive sign of t' which is opposite to the sign of t' for hole-doped system with the same shape of Fermi surface. The reason is that for $t' > 0$ the shape of the upper Hubbard subband looks like the inverted lower subband for $t' < 0$ (Fig. 3).

3. In checking the hypothesis of the *d*-wave superconductivity in all cuprates, phase-sensitive experiments with LSCO and NCCO are very important. From the theoretical viewpoint, the *d*-wave symmetry is preferable for CuO_2 planes in all cuprates, since the on-site repulsion does not suppress the *d*-wave order parameter, unlike the *s*-wave superconductivity.

4. The model with $t'/t > 0$ provides a reasonable interpretation of the anisotropic pseudogap manifesting itself in the shift of the photoemission edge for $\mathbf{k}$ near the directions $(0, \pi)$ and $(\pi, 0)$ in underdoped BSCCO samples. The pseudogap results from the dielectrization of zone boundaries near the point $M = (0, \pi)$ under conditions of antiferromagnetic band splitting. Some authors supposed[10,11] that the pseudogap has the *d*-wave symmetry, implying that a certain parameter characterizing the pseudogap changes sign. The pseudogap was regarded as a precursor of the *d*-wave superconducting gap. Our models makes such hypotheses redundant. Our interpretation of properties of group II cuprates might be verified by experiment on detection of the pseudogap in LSCO and NCCO along directions $k_x = \pm k_y$.

5. The detection of the small Fermi surface (hole pockets around the point $(\pi/2, \pi/2)$ in BSCCO)[9] supports our assumption that this material is described by the model with $t'/t > 0$. There is hope that, given a higher ARPES resolution, it will be possible to distinguish between the smooth dielectric and sharp metallic parts of the generalized Fermi boundary.

6. Changes in the sign of the Hall constant in LSCO and NCCO may be due to transitions from small to large Fermi surfaces.

7. An important task of the theory is to check out whether features in the magnetic susceptibility $\chi(Q, \omega)$ detected in neutron scattering could be described in terms of antiferromagnetically split bands without arbitrary renormalizations. (One has to use various renormalizations when considering an unsplit, unperturbed band.) The affirmative answer to this question would support the hypothesis that the main mechanism forming the lower and upper Hubbard subband is antiferromagnetic ordering of spins.

6. MODEL OF ORTHORHOMBIC PERTURBATIONS IN YBCO

Of special note is the effect of orthorhombic distortions on the band structure, Fermi surface, and shape of the phase curve $T_c(\delta)$. Such distortions occur in YBCO owing to the layers composed of CuO chains. In order to produce a simple description of orthorhombic effects, let us introduce different hoppings t_x and t_y for bonds aligned with the *x*- and *y*-axes: $t_{x(y)} = t \mp \tau$, i.e., we will analyze the model

$$H = H(U, t, V, t') + \Delta H(\tau),$$

$$\Delta H(\tau) = \tau \sum_{n,\sigma} \left[(c_{n\sigma}^{\dagger} c_{n+e_x,\sigma} - c_{n\sigma}^{\dagger} c_{n+e_y,\sigma}) + \text{H.c.} \right]. \quad (24)$$

To construct a correlated state (3) with correlations of valence bond type, generally speaking, requires using a unitary transformation with different parameters α_x and α_y for bonds aligned with *x*- and *y*-axes. Our rough calculations,

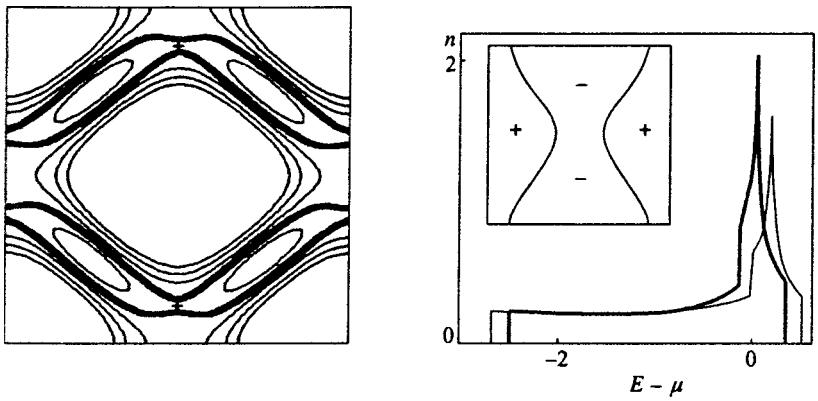

FIG. 9. Left: Fermi boundary (bold line) and nearest zone energy levels (with a step of $0.02t$) in the region $-\pi < k_{x(y)} < \pi$ calculated by Eq. (25) with $U=8$, $V=0.1$, $t'=0.1$, and $\tau=0.05$ (in units of t) at a doping $\delta=0.15$. Right: density of states (bold curve). The thin curve corresponds to the doping $\delta=0.2$ of the second peak on the phase curve. The inset shows lines of superconducting gap nodes defined by Eq. (29) with the *s*-wave contribution at doping corresponding to the first peak on the curve of $T_c(\delta)$.

however, were based on a variational function with one parameter $\alpha_x=\alpha_y=\alpha$. We have used the following approximate effective Hamiltonian

$$\tilde{H}=\tilde{H}(\alpha,U,t)+\Delta H(V,t')+\Delta\tilde{H}(\tau), \tag{25}$$

and calculated the first main term to two orders of α, the weak interaction $\Delta H(V,t')$ in zeroth order, and the third term in the zeroth and first orders in α. By applying the model to YBCO, where $t_b>t_a>0$ for the *a*- and *b*-axes, and determining the signs of t and τ in Eqs. (1) and (24), we obtain $\tau>0$.

The phase curves $T_c(\delta)$ for the model described by Eqs. (24) and (25) with $\tau=0.05$ and parameters $U=8$, $V=0.1$, $t'=0.1$ and 0.05 (in units of t) are plotted in Fig. 1 (dashed lines). They have two peaks due to the more complex band structure. For the model with $t'=0.1$ with doping $\delta=0.15$, the Fermi surface and density of states in the lower Hubbard subband are shown in Fig. 9. Now the two energy parameters $4t'$ and 2τ characterize features of the lower Hubbard subband $E(k)$ near the Fermi level. The parameter τ determines the difference between maximum energies $E_x^{\max}$ and $E_y^{\max}$ in profiles of the band energy for k varying along the path $\Gamma(0,0)\to\Gamma(2\pi,0)$ on the k_x-axis and along the path $\Gamma(0,0)\to\Gamma(0,2\pi)$ on the k_y-axis. On the k_x-axis the curve $E(k)$ has one peak $E_x^{\max}$ at $k=(\pi,0)$. On the k_y-axis the profile of $E(k)$ has two peaks $E_y^{\max}$ of equal energies at $k=(0,\pi\pm\Delta_y)$ (the crosses in Fig. 9). Since the band energy is a periodic function with the period of the magnetic Brillouin zone, $E(0,\pi)=E(\pi,0)=E_x^{\max}<E_y^{\max}$. Curves of $E_{x(y)}^{\max}(\delta)$ for the model (24) with $\tau=\text{const}=0.05t$ and $t'=0.1t$ are given in Fig. 2. The corresponding density of states is shown in Fig. 9, where the main Van Hove singularity corresponds to energy $E_x^{\max}$. This singularity is associated with the first peak on the phase curve $T_c(\delta)$ (the dashed line $3'$ in Fig. 1). It corresponds to the doping δ_1, at which the chemical potential coincides with the energy of the Van Hove singularity: $E_y^{\max}-\mu=0$. The doping δ_2, for which $E_x^{\max}-\mu=0$ holds, approximately corresponds to the second peak on the phase curve $T_c(\delta)$.

The measured phase diagram of $YBa_2Cu_3O_{6+y}$, in fact, has two regions with $T_c^{\max}\sim 45$ K and $T_c^{\max}\sim 95$ K at excess oxygen concentrations $y\sim 0.65$ and $y\sim 0.96$. The parameter y is related (but not identical) to the doping δ of the CuO_2 plane. In contrast to results obtained for model (24) with $\tau=0.05t$, the region of higher doping on the experimental phase curve has a higher critical temperature T_c than the region of lower doping. This discrepancy may be due to the crude approximation used in our calculations. Undoubtedly, the correlation function with different parameters α_x and α_y for bonds aligned with *x*- and *y*-axes should correspond to a lower energy. More accurate calculations, together with investigations of superconductivity in alternating dimer structures of valence bonds,[19] might clarify the situation.

Note that, throughout the studied doping range $\delta_1<\delta<\delta_2$, we have different connectivities of regions in the phase space on k_x- and k_y-axes occupied by electrons. Unlike the line $\Gamma(0,0)\to\Gamma(0,2\pi)$, the line $\Gamma(0,0)\to\Gamma(2\pi,0)$ does not cross the Fermi boundary. Such is the structure of the lower Hubbard subband. There is an unanswered question of whether this can be a reason of the observed high anisotropy of resistivity, $\alpha_a/\alpha_b=2.2$, even at a small anisotropy of the hopping interaction, $t_a/t_b=0.9$ ($\tau=0.05t$). The resistivity anisotropy prescribes $t_a/t_b=0.6$ in the tight binding approximation with an unsplit valence band, i.e., $\tau=0.25t$.[26] The difference between these two values may be attributed to different susceptibilities of the Fermi surface to weak interactions t' and τ with and without band splitting in Eqs. (17) and (18), respectively.

Finally, let us estimate the admixture of the *s*-wave symmetry to the superconducting order parameter due to orthorhombic distortions in our model. For simplicity, we consider only three anomalous averages $w_i=\{w_1^d,w_0,w_1^s\}_i$, $i=1,2,3$. Here we retain only the main order parameter $w_1^d\equiv w_1$ of all *d*-wave anomalous averages w_l and add two *s*-wave parameters. Thus, the following anomalous averages are included:

$$w_1^{d(s)}=\frac{1}{8N}\sum_n \langle[c_{n\downarrow}c_{n+e_x\uparrow}+c_{n\downarrow}c_{n-e_x\uparrow}] \mp[c_{n\downarrow}c_{n+e_y\uparrow}+c_{n\downarrow}c_{n-e_y\uparrow}]+\text{H.c.}\rangle, \tag{26}$$

$$w_0^s=\frac{1}{2N}\sum_n \langle c_{n\downarrow}c_{n\uparrow}+\text{H.c.}\rangle. \tag{27}$$

The transition temperature T_c is calculated by equating to zero the determinant of the uniform linear equation system for the anomalous averages w_i:

$$w_i - \sum_j D_{ij} w_j = 0. \tag{28}$$

The matrix D_{ij} for the given set of anomalous averages is determined by Eqs. (45)–(48) in the Appendix. The eigenvector $\{w_i\}$ of equation system (28) corresponding to the zero eigenvalue allows us to express the anisotropic superconducting gap [paring interaction in the lower Hubbard subband determined by Eqs. (39) and (49)] as follows:

$$W_1^{\mathrm{AF}}(k) = A_1(c_x - c_y) + A_2\beta(k) + A_3(c_x + c_y),$$
$$c_{x(y)} = \cos k_{x(y)}. \tag{29}$$

The coefficients A_i and function $\beta(k) = \cos\gamma_k$ are given by Eqs. (40) and (50) of the Appendix. One can easily check out that function $\beta(k)$ has the same symmetry, $\beta(\tilde{k}) = -\beta(k)$, as the third term on the right of Eq. (29). The superconducting gap described by Eq. (29) changes sign, as in a pure *d*-wave superconductor. At a doping $\delta \sim \delta_1$, corresponding to the first peak of $T_c(\delta)$, the curve of nodes of the superconducting gap (29) is that shown in Fig. 9. The deviation of the nodal line of function (29) from lines $k_x = \pm k_y$ prescribed for the pure *d*-wave symmetry rises with increasing δ in the model with constant τ. The Fermi surface asymmetry is, probably, more important for the observed orthorhombic distortions in YBCO[34,35] than the small admixture of the *s*-wave symmetry to the main *d*-wave contribution in pairing interaction (29). In order to clarify the situation, more detailed investigations are needed, including studies of the role of interaction between two neighboring CuO_2 planes in YBCO.

7. CONCLUSION

Let us summarize the main results of the present work.

A description of superconductivity and other low-energy effects in cuprates in the range of doping where CuO_2 planes are characterized by long-range antiferromagnetic correlations can be based on correlated states of the one-band Hubbard model with correlations of valence bond type. The short-range attraction between holes in the *d*-channel is due to formation of valence bonds, and the constants of this attraction can be calculated variationally.

The phase curve $T_c(\delta)$ for the superconducting transition and other properties are very sensitive to the interaction between second nearest neighbors. The sign of this interaction t'/t determines the band structure and Fermi surface topology at low doping. Models of the first type ($t'>0$) for underdoped systems are characterized by the presence of small Fermi surfaces, namely, hole pockets around the point $(\pm\pi/2, \pm\pi/2)$. The latter account for the presence of the anisotropic pseudogap in ARPES spectra and its doping dependence. The optimal doping is uniquely determined by the coincidence between the chemical potential and the energy of Van Hove singularity connected with the flat region of the lower Hubbard subband around the point $M = (\pi, 0)$. Arguments in favor of applicability of this model to BSCCO and YBCO materials have been presented.

The models of the second type correspond to $t'<0$ for hole-doped systems (or $t'>0$ for electron-doped systems). They show lower transition temperatures and broader peaks on the phase curve $T_c(\delta)$ shifted in the region of lower doping. These properties are determined by the different band structures and Fermi surface shapes in these models. In underdoped systems of this sort, hole (electron) pockets are located around the points $M = (0, \pm\pi), (\pm\pi, 0)$, and there are dielectric sections of the magnetic Brillouin zone boundary near the nodes of the superconducting gap. As a result, the minimal energy of fermion excitations is finite everywhere. The applicability of these models to LSCO and NCCO has been argued. Thus, we have purt forward the hypothesis that all the cuprates can be described in terms of the *d*-wave order parameter. For NCCO, this hypothesis is supported by the coincidence between the sign of calculated $t'/t > 0$[25–27] and the sign of $t'/t > 0$ required for shaping of type II Fermi surfaces in the case of electron doping. This hypothesis, however, seems more questionable in the case of LSCO, where the calculated and empirical t'/t have opposite signs. The role of the lattice superstructure in LSCO has also remained unclear.[46] Phase-sensitive experiments for LSCO and NCCO, and detection of pseudogaps with asymmetry different from that of BSCCO might help in discovering the truth.

Models taking account of orthorhombic distortions yield two peaks on the phase curve $T_c(\delta)$ due to new features in the lower Hubbard subband. The approximate calculation, however, yields an inverse ratio between two maximum critical temperatures $T_c^{\max}$ in comparison with that measured in YBCO. The contribution of an *s*-wave component added to the basic *d*-wave order parameter has also been estimated.

The present work was supported by the Russian Fund for Fundamental Research (Projects 97-03-33727A and 96-15-97492) and ISTC (Grant No. 872) which we gratefully acknowledge here. The authors are indebted to V. Ya. Krivnov for helpful discussions.

APPENDIX

For the one-determinant BCS function Φ with a double magnetic elementary cell, the mean energy $\overline{H}(y_i) = \langle \psi H \Psi \rangle = \langle \Phi \tilde{H} \Phi \rangle$ depends on the set of averages $y_i = \{r_l, d_l, w_l\}_i$ of one-electron operators

$$\hat{r}_l = \frac{1}{2n_l} \sum_{l,\sigma} c^\dagger_{n,\sigma} c_{n+l,\sigma},$$

$$\hat{d}_l = \frac{1}{2n_l} \sum_{l,\sigma} (-1)^n \frac{\sigma}{|\sigma|} c^\dagger_{n,\sigma} c_{n+l,\sigma}, \tag{30}$$

$$\hat{w}_l = \frac{1}{4n_l} \sum_{l,\sigma} \mathrm{sign}(l_x^2 - l_y^2)[c^\dagger_{n,\sigma} c^\dagger_{n+l,-\sigma} + \mathrm{H.c.}] \tag{31}$$

calculated with wave function Φ. Here $l = 0, 1, \sqrt{2}, 2, \sqrt{5}, 3,$ and n_l is the number of all vectors $\mathbf{l}$ of length $l = |\mathbf{l}|$ over which the sums are performed. The values d_l are nonzero only when $(l_x + l_y)$ is even, and anomalous averages w_l of the *d*-wave symmetry only $l_x \neq l_y$ are included. The one-determinant function minimizing $\overline{H}$ is an eigenfunction of the linearized Hamiltonian, which is expressed in the basis of fermion operators $b_{ki} = \{c^\dagger_{k\uparrow}, c^\dagger_{\tilde{k}\uparrow}, c_{-k\downarrow}, c_{-\tilde{k}\downarrow}\}_i$ in the form

$$(\tilde{H})_L=\sum_k{}^F \sum_{i,j=1}^4 h_{ij}(k)b_{ki}^\dagger b_{kj}+\text{const.} \tag{32}$$

The matrix h_{ij} is given by the formula

$$h_{ij}=\begin{pmatrix} \epsilon_k-\mu & \Delta_k & W_k & 0 \\ \Delta_k & \epsilon_{\tilde{k}}-\mu & 0 & W_{\tilde{k}} \\ W_k & 0 & -(\epsilon_k-\mu) & -\Delta_k \\ 0 & W_{\tilde{k}} & -\Delta_k & -(\epsilon_{\tilde{k}}-\mu) \end{pmatrix}, \tag{33}$$

$$\epsilon_k=\frac{1}{2}\sum_l \frac{\partial\overline{H}}{\partial r_l}g_l(k), \quad \Delta_k=\frac{1}{2}\sum_l \frac{\partial\overline{H}}{\partial d_l}g_l(k),$$

$$W_k=\frac{1}{2}\sum_l \frac{\partial\overline{H}}{\partial w_l}q_l(k). \tag{34}$$

The functions $g_l(k)$ and $q_l(k)$ are determined by the equations

$$g_l(k)=\frac{1}{2}[\cos k_x l_x \cos k_y l_y+\cos k_y l_x \cos k_x l_y],$$

$$q_l(k)=\frac{1}{2}\text{sign}(l_x^2-l_y^2)[\cos k_x l_x \cos k_y l_y - \cos k_y l_x \cos k_x l_y]. \tag{35}$$

Here $\tilde{k}=k+(\pi,\pi)$. In sums over l in the expression for ϵ_k, Δ_k, and W_k, the index l runs through all r_l, or d_l, or w_l from the complete set of one-electron averages in $\overline{H}$. One-electron eigenfunctions $\chi_{k\lambda}$ and spectrum $E_\lambda(k)$ of the linearized Hamiltonian are calculated by diagonalizing the matrix (33):

$$\chi_{k\lambda}^\dagger=\sum_j b_{kj}^\dagger S_{j,\lambda}, \quad h_{ij}S_{j,\lambda}=S_{j,\lambda}E_\lambda \tag{36}$$

for all $k\in F$ within the magnetic Brillouin zone. The approximate numerical diagonalization yields fairly accurate results thanks to the smallness $W_k \ll \Delta_k$ of the superconducting gap in comparison with the antiferromagnetic one. The approximate eigenvalues h_{ij} are

$$E_\lambda=\mp\sqrt{(E_\nu^{\text{AF}}-\mu)^2+(W_\nu^{\text{AF}})^2}, \quad \lambda=1,\ldots,4, \quad \nu=1,2. \tag{37}$$

Here the energies $E_{1(2)}^{\text{AF}}$ of the upper and lower Hubbard subbands are

$$E_{1(2)}^{\text{AF}}=\frac{1}{2}(\epsilon_k+\epsilon_{\tilde{k}})\mp\sqrt{\frac{1}{4}(\epsilon_k-\epsilon_{\tilde{k}})^2+\Delta_k^2}, \tag{38}$$

and interaction W_ν^{AF} responsible for the superconducting pairing in one of the subbands in the case of hole ($\nu=1$) or electron ($\nu=2$) doping is given by

$$W_\nu^{\text{AF}}=\frac{1}{2}[W_k-W_{\tilde{k}}\pm\cos\gamma_k(W_k+W_{\tilde{k}})], \quad \nu=1,2, \tag{39}$$

$$\tan\gamma_k=2\Delta_k/(\epsilon_k-\epsilon_{\tilde{k}}). \tag{40}$$

Given the eigenfunctions and spectrum, one can calculate the desired averages y_i and so obtain a self-consistent solution. Details of the calculation procedure were described elsewhere.[16]

The one-particle contribution to annihilation operators $\tilde{c}_{k\sigma}$ is determined by Eq. (22), where the functions φ and η are determined by the expressions

$$\varphi(k)=1-\alpha^2(c_x+c_y)[(1-2r_0)r_1 + r_0(1-r_0)+d_0^2+2r_1^2], \tag{41}$$

$$\eta(k)=-2(c_x+c_y)d_0(1+\alpha^2 r_1^2), \quad c_{x(y)}=\cos k_{x(y)}. \tag{42}$$

Let us expand $\tilde{c}_{k\sigma}$ in (22) in terms of one-particle operators $\chi_{k\sigma}$, which determine the eigenfunction Φ of the linearized Hamiltonian $\tilde{H}_L$:

$$c_{k\uparrow}=\Sigma_\lambda R_\lambda \chi_{k\lambda}, \quad c_{k\downarrow}=\Sigma_\lambda Q_\lambda \chi_{k\lambda}^\dagger. \tag{43}$$

The expressions for R_λ and Q_λ are different for quasimomenta inside ($k\in F$) and outside ($k\notin F$) the first magnetic Brillouin zone:

$$R_\lambda(k)=Q_{\lambda'}(k)=\begin{cases} \varphi(k)S_{1\lambda}+\eta(k)S_{2\lambda}, & k\in F, \\ -\eta(k)S_{1\lambda}+\varphi(k)S_{2\lambda}, & k\notin F. \end{cases} \tag{44}$$

Here $\lambda'=\lambda'(\lambda)$ corresponds to the eigenvalue $E_{\lambda'}=-E_\lambda$. By virtue of symmetry relations (44) between Q_λ and R_λ and the formula $f(E_\lambda)=1-f(-E_\lambda)$ for the Fermi distribution function, we have final expression (23) for the spectral function $A(k\omega)$.

In analyzing the model described by Eqs. (24) and (25) and including the perturbation of the orthorhombic symmetry proportional to τ, we consider only the principal average $w_1^d\equiv w_1$ of the d-wave anomalous averages (31) and add two s-wave anomalous averages w_0 and w_1^s given by Eqs. (26) and (27). Thus, we treat only three superconducting order parameter $w_i=\{w_1^d,w_0,w_1^s\}$. Then the critical temperature of a hole-doped system is determined by Eq. (28) with matrix $D_{ij}=\partial w_i/\partial w_j$ at $w_l=0$ equal to

$$D_{ij}=-N^{-1}\sum_k{}^F R_{il}(k)a_{lj}\frac{1-2f(E_1)}{2E_1}, \tag{45}$$

$$R_{11}=\frac{1}{4}(c_x-c_y)^2, \quad R_{22}=\cos\gamma_k^2, \quad R_{33}=\frac{1}{4}(c_x+c_y)^2, \tag{46}$$

$$R_{12}=\frac{1}{2}(c_x-c_y)\cos\gamma_k, \quad R_{13}=\frac{1}{4}(c_x^2-c_y^2),$$

$$R_{23}=\frac{1}{2}(c_x+c_y)\cos\gamma_k. \tag{47}$$

Here $c_{x(y)}=\cos k_{x(y)}$, γ_k derives from Eq. (40), f is the Fermi distribution function, $E_1(k)=E_1^{\text{AF}}(k)-\mu$ is the energy of the lower Hubbard subband with respect to the chemical potential, and coefficients a_{ij} are given by

$$a_{ij}=\left.\frac{\partial^2\overline{H}}{\partial w_i\partial w_j}\right|_{w_i=0}. \tag{48}$$

In a hole-doped system with antiferromagnetic band splitting, the role of the superconducting pairing (superconducting gap) is played by quantity (39). If the anomalous averages (26) and (27) are retained, it includes *d*-wave and *s*-wave contributions,

$$W_k=A_1(c_x-c_y)+2A_2\cos\gamma_k+A_3(c_x+c_y), \tag{49}$$

with coefficients

$$A_i=\frac{1}{2}[a_{i1}w_1^d+a_{i2}w_0+a_{i3}w_1^s]. \tag{50}$$

Here γ_k and a_{ij} are given by Eqs. (40) and (48).

Note added in proof (18 December 1998). Some of our results are close to those obtained in Ref. 1 (below), where the same problems were studied in the different (probably equivalent) language of the $t-J$ model and spin-polaron pairing. In both cases, superconductivity is described without any empirical parameters and occurs inside the doping region where 2D antiferromagnetic spin order exists. In both cases the value and sign of diagonal hopping greatly influence the phase curve $T_c(\delta)$ and the type of Fermi surface. Reference 2 should also be mentioned.

1. N. M. Plakida, V. S. Oudovenko, P. Horsh, and A. L. Liechtenstein, Phys. Rev. B **55**, R11997 (1997).

2. R. S. Markiewicz, Phys. Rev. B **56**, 9091 (1997).

*)E-mail: movchin@center.chph.ras.ru

[1] E. Dagotto, Rev. Mod. Phys. **66**, 763 (1994).
[2] D. J. Scalapino, Phys. Rep. **250**, 329 (1995).
[3] Z.-X. Shen and D. S. Dessau, Phys. Rep. **253**, 1 (1995).
[4] J. W. Allen, R. Claessen, R. O. Anderson *et al.*, in *The Physics of the Hubbard Model*, ed. by D. K. Campbel, JMP Carmelo and F. Guinea, Plenum Press, New York (1994).
[5] E. Dagotto and T. M. Rice, Science **271**, 618 (1996).
[6] Yu. A. Izyumov, Usp. Fiz. Nauk **167**, 465 (1997).
[7] J. R. Kirtley, C. C. Tsuei, J. Z. Sun *et al.*, Nature (London) **373**, 225 (1995).
[8] D. A. Brawner, C. Mancer, and H. R. Ott, Phys. Rev. B **55**, 2788 (1997).
[9] D. S. Marshall, D. S. Dessau, A. G. Loeser *et al.*, Phys. Rev. Lett. **76**, 4841 (1996).
[10] A. G. Loeser, Z.-X. Shen, D. S. Dessau *et al.*, Science **273**, 325 (1997).
[11] H. Ding, T. Yokoya, J. C. Campuzano *et al.*, Nature (London) **382**, 51 (1996).
[12] N. Nagaosa, Science **275**, 1078 (1997).
[13] S. C. Zhang, Science **275**, 1089 (1997).
[14] R. O. Zaitsev, JETP Lett. **55**, 135 (1992); **56**, 339 (1992).
[15] J. E. Hirsch, Phys. Rev. Lett. **54**, 1317 (1985).
[16] A. A. Ovchinnikov, M. Ya. Ovchinnikova, and E. A. Plekhanov, Zh. Éksp. Teor. Fiz. (1998) [JETP **87**, 534 (1998).
[17] A. A. Ovchinnikov, M. Ya. Ovchinnikova, and E. A. Plekhanov, JETP Lett. **67**, 369 (1998).
[18] P. W. Anderson, Science **235**, 1196 (1987).
[19] A. A. Ovchinnikov and M. Ya. Ovchinnikova, Zh. Éksp. Teor. Fiz. **110**, 342 (1996); **112**, 1409 (1997) [JETP **83**, 184 (1996); **85**, 767 (1997)].
[20] M. C. Gutzwiller, Phys. Rev. A **137**, 1726 (1965).
[21] J. H. Jefferson, H. Eskes, and L. F. Feiner, Phys. Rev. B **45**, 7959 (1992).
[22] H. B. Schuttler and A. J. Fedro, Phys. Rev. B **45**, 7588 (1992).
[23] A. A. Ovchinnikov and M. Ya. Ovchinnikova, J. Phys.: Condens. Matter **6**, 10317 (1994).
[24] D. Duffy and A. Moreo, Phys. Rev. B **52**, 15607 (1995).
[25] M. S. Hybertsen, E. B. Stechel, M. Schluter, and D. R. Jennison, Phys. Rev. B **41**, 11068 (1990).
[26] J. Yu and A. J. Freeman, J. Electron Spectrosc. Relat. Phenom. **66**, 387 (1994).
[27] R. J. Radke and M. R. Norman, Phys. Rev. B **50**, 9554 (1994).
[28] A. R. Kamp and J. R. Schriefer, Phys. Rev. B **42**, 7967 (1990).
[29] E. Dagotto, A. Nazarenko, and A. Boninsegni, Phys. Rev. Lett. **73**, 728 (1994).
[30] S. Haas, Phys. Rev. B **51**, 11748 (1995).
[31] J. H. Kim, K. Levin, and A. Auerbach, Phys. Rev. B **39**, 11633 (1989).
[32] J. Makert *et al.*, in *Physical Properties of High Temperature Superconductors*, Vol. I, ed. by D. M. Ginzberg, World Scientific, Singapore (1989).
[33] J. B. Torrance, A. Y. Tokura, A. I. Nazzal *et al.*, Phys. Rev. Lett. **61**, 1127 (1988).
[34] A. G. Sun, A. Truscoff, A. S. Katz *et al.*, Phys. Rev. B **54**, 6734 (1996).
[35] K. A. Kuznetzov, A. G. Sun, B. Chen *et al.*, Phys. Rev. Lett. **79**, 3050 (1997).
[36] H. Takagi, T. Ito, S. Ishabashi *et al.*, Phys. Rev. B **40**, 2254 (1989).
[37] J. L. Peng, S. Y. Li, and R. L. Greene, Phys. Rev. B **43**, 13606 (1991).
[38] P. Benard, L. Chen, and A. H. Tremblay, Phys. Rev. B **47**, 15217 (1993).
[39] T. E. Mason, G. A. Aepli, S. M. Hayden *et al.*, Phys. Rev. Lett. **71**, 919 (1993).
[40] R. J. Birgenot and J. Chiran, in *Physical Properties of High Temperature Superconductors*, ed. by D. M. Ginzberg, World Scientific, Singapore (1989).
[41] Q. Si, Y. Zha, K. Levin, and J. P. Lu, Phys. Rev. B **47**, 9055 (1993).
[42] P. B. Littlewood, J. Zaanen, and G. Aepli, Phys. Rev. B **48**, 487 (1993).
[43] J. Yu and A. J. Freeman, J. Electron Spectrosc. Relat. Phenom. **66**, 281 (1994).
[44] Y. Sakisaka, J. Electron Spectrosc. Relat. Phenom. **66**, 387 (1994).
[45] D. H. King, Z.-X. Shen, D. S. Dessau *et al.*, Phys. Rev. Lett. **70**, 3159 (1993).
[46] A. Bianconi, N. L. Saini, M. Missori *et al.*, Phys. Rev. Lett. **76**, 3412 (1997).

Translation provided by the Russian Editorial office.

Formation of periodic surface structures by ultrashort laser pulses

M. B. Agranat,*) S. I. Ashitkov, and V. E. Fortov

Research Center for the Thermophysics of Pulsed Processes, Joint High-Temperature Institute, Russian Academy of Sciences, 127412 Moscow, Russia

S. I. Anisimov

L. D. Landau Institute of Theoretical Physics, Russian Academy of Sciences, 142432 Chernogolovka, Moscow Region, Russia

A. M. Dykhne

Troitsk Institute of Innovational and Thermonuclear Studies, 142092 Troitsk, Moscow Region, Russia

P. S. Kondratenko

Institute of Safe Development of Nuclear Power, Russian Academy of Sciences, 113191 Moscow, Russia
(Submitted 19 June 1998)
Zh. Éksp. Teor. Fiz. **115**, 675–688 (February 1999)

The formation of periodic surface structures by ultrashort laser pulses was observed experimentally and explained theoretically. The experiments were performed on graphite with picosecond laser pulses. The spatial period of the structures is of the order of the wavelength of the incident radiation, and the orientation of the structures is correlated with the direction of polarization of the light. The key point of the theoretical model proposed is resonance excitation of surface electromagnetic waves, which under conditions such that the temperature of the electronic subsystem is decoupled from the temperature of the crystal lattice causes a ''temperature grating'' to be written on the flat solid surface of the sample while the laser pulse is being applied on account of the temperature dependence of the surface impedance. The formation of a periodic surface profile from the temperature grating occurs by the volume expansion of a melted layer near the surface of the material. For typical values of the surface tension and viscosity for metals, there is not enough time for the periodic profile to be resorbed before the liquid layer solidifies. The formation of periodic surface structures is delayed in time relative to the laser pulse.
[S1063-7761(99)02002-8]

1. INTRODUCTION

The formation of periodic surface structures, which are responsible for spatially modulated surface profiles with a period on the order of the wavelength of light, is one of the most common phenomena occurring under the action of laser radiation on condensed media.[1] The mechanism of this phenomenon for metals and semiconductors is based on resonance excitation of surface electromagnetic waves,[2] whose interference with the incident wave results in spatial modulation of the energy release, which in the presence of positive feedback gives rise to a periodic surface profile by means of an appropriate thermophysical mechanism (for example, evaporation or thermal expansion).

As far as we know, all investigations performed thus far on periodic surface structures concern the action of comparatively short laser pulses, so that the physical processes accompanying the development of the phenomenon can be treated as quasistationary. At the same time, for ultrashort laser pulses (picosecond and shorter duration), not only can the assumption of quasistationarity break down, but it may even become impossible for some of these processes to occur during the laser pulse. Under such conditions, the formation of periodic surface structures, if it happens, should occur after the laser action ceases; this is what distinguishes fundamentally the ultrashort laser pulse regime from the conventional case. In addition, there are other important features of the ultrashort laser action regime: decoupling of the conduction-electron temperature from the crystal lattice temperature[3] and smallness of the modulated-heating depth. In our opinion, all this gives grounds for believing that the discovery of the formation of periodic surface structures under given conditions in itself can be exceedingly important for understanding the physics of the interaction of ultrashort laser pulses with condensed media.

In the present paper we report experimental data that confirm the formation of periodic surface structures on graphite by picosecond laser pulses, and we propose a theory of the formation of such structures by ultrashort laser pulses.

This paper is organized as follows. In Sec. 1, the results of an experimental investigation of the formation of periodic surface structures on graphite are reported. In Sec. 2, the physical processes that result in the formation of such structures by ultrashort laser pulses interacting with a metal are analyzed theoretically. Laser absorption, instability with re-

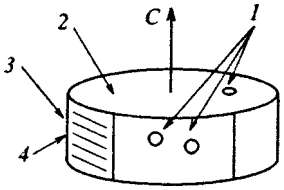

FIG. 1. Arrangement of the focusing spots (*1*) of the heating pulse on the graphite sample: *2* — sample surface; *3* — side face of the sample; *4* — graphite layers.

spect to the formation of spatially modulated temperature fields over the application time of the laser pulse, and the formation of a periodic profile after the laser pulse ends are studied separately. The results are discussed in Sec. 3.

2. EXPERIMENT

A periodic surface structure was observed on the surface of microcrystalline pyrolytic graphite when ~ 1 ps laser pulses were applied to the surface of the sample. The graphite sample was prepared in the form of tablets (Fig. 1), whose flat surface is perpendicular to the symmetry axis C. Correspondingly, the side face of the tablet is parallel to the symmetry axis.

Investigation of the initial sample by transmission electron microscopy, x-ray diffractometry, and Raman scattering spectroscopy (see Figs. 2 and 3 below) showed that the initial sample consists of microcrystalline graphite with 10–30 nm grains. The diffraction pattern for the sample surface resembles that obtained when the crystallites have a distinguished predominant orientation. The graphite layers are arranged parallel to one another, but the distance between them is not fixed. A distinct texture is observed, i.e., microcrystals are slightly disoriented with respect to one another. On the basis of these investigations it can be concluded that the properties (specifically, the thermal conductivity) of this sample are strongly anisotropic.

The experimental investigations were conducted using the laser system based on multicascade SRMB and SRS compression of the initial nanosecond pulse from a Nd:YAG laser, generating simultaneously several pulses of different duration in the femto- and picosecond ranges with repetition frequencies up to 10 Hz. The parameters were monitored using the measurement part of the system, including photosensors for the pulse energy, an image converter with ~ 3 ps time resolution, and an autocorrelator with ~ 50 fs resolution. All measurements, including diagnostic, were fed into a multichannel digital information reading and processing unit. This unit consisted of several highly sensitive digital video cameras and analog signal inputs, an information reading synchronization unit, a controller for the the laser setup and the computer that made it possible to control the setup according to a prescribed program, extract and process simultaneously information from the outputs of the image converter, oscillograph, autocorrelator, energy sensors, photosensors, and spectroscopic instruments.

A heating laser pulse of length $\simeq 1$ ps with wavelength 780 nm was focused into a spot $\simeq 200\ \mu$m in diameter on the lateral face and the surface of the graphite at an angle close

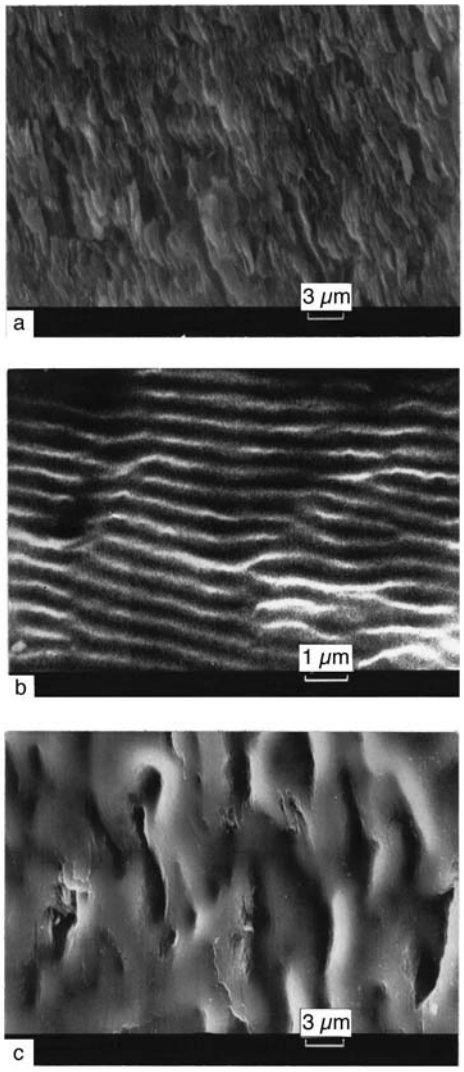

FIG. 2. Electron microscopy of microcrystalline pyrolytic graphite (in the side face plane) after laser action: initial surface before irradiation (a) and after irradiation with pulses with duration $\tau = 1$ ps (b) and 500 ps (c).

to normal incidence (Fig. 1). The energy density in the focusing spot was $\simeq 0.5$ J/cm^2. When laser pulses are applied repeatedly with repetition frequency 10 Hz, a strongly reflecting (reflection three to four times more intense than the initial reflection) microregion arises at the center of the focusing spot on the side face and increases in size up to the size of the focusing spot. The reflection increases very little in the flat surface of the sample. An electron microscope showed that a periodic surface relief with period $\simeq 0.6\ \mu$m on the spots on the side faces (Fig. 2b) and with period $\simeq 1\ \mu$m in the spot on the surface of the sample is present in all spots. The structure on the sample surface is not as sharp as on the side face.

Raman spectroscopy (Fig. 3) showed that for the crater located on the surface of the sample, the presence of an amorphous phase of carbon is identified at both the center and periphery. The fraction of the amorphous phase is higher at the center of the crater. Reflections corresponding to the microcrystalline phase of graphite are observed simultaneously with the amorphous phase at the same points. At the

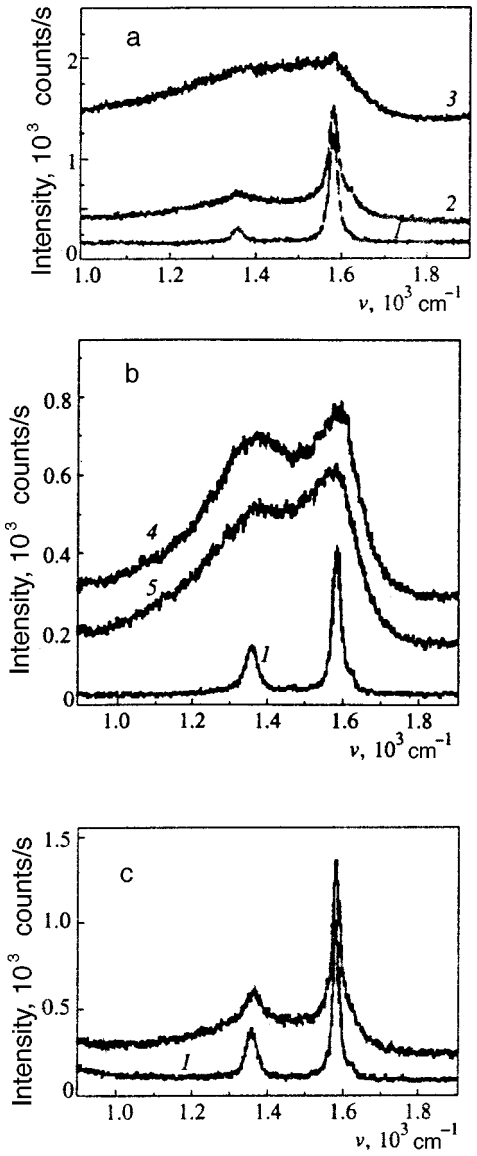

FIG. 3. Raman spectroscopy of pyrolytic graphite after laser action (curves *1* — initial surface): a — heating of the surface (curve *2*) and of the side face (curve *3*) of the sample by laser pulses with duration $\tau \simeq 1$ ps; b — heating of the side face of the sample with $\tau \simeq 500$ ps (curve *4*) and 1 ps (curve *5*, vacuum); c — heating of the side face of the single crystal for $\tau \simeq 1$ ps.

same time, for the spots located on the side face of the sample, the Raman measurements show the presence of only an amorphous carbon phase without any traces of microcrystalline graphite.

The following investigations were performed (in all experiments, the side face of the experimental sample of microcrystalline graphite was heated).

1. When the polarization plane of the heating pulse changed (rotated by 45°), the orientation of the lines of the periodic surface structure also changed by the amount $\varphi \approx 45°$.

2. Formation of periodic surface structures was not observed (Fig. 2c) when the duration of the heating laser pulse was increased to $\tau \approx 500$ ps (wavelength 530 nm). The Raman spectra (Fig. 3b, curve *4*) show the presence of a microcrystalline structure with traces of an amorphous phase.

3. The formation of periodic surface structures and an amorphous phase of graphite were likewise observed when the experimental sample was heated with laser pulses with duration $\tau \approx 1$ ps in a shallow vacuum (10^{-6} bar) (Fig. 3b, curve *5*).

4. Such investigations were performed on a different form of graphite, a pyrolytic graphite single crystal. Raman spectroscopy of this sample reveals a single-crystal structure on the surface and a microcrystalline structure on the side face (Fig. 3c). The formation of a periodic surface structures was not observed after repeated application of laser pulses ($\tau \approx 1$ ps). The Raman spectra show the presence of a microcrystalline structure with traces of an amorphous phase.

5. An accumulation effect is observed with the appearance of periodic surface structures on samples of microcrystalline graphite, i.e., if a periodic structure is observed at the center of a focusing spot with one application of a heating pulse with the appropriate energy density, then under repeated application of pulses with approximately the same parameters, the region of formation of such a structure is observed to increase in size until the entire area of the focusing spot is filled (the energy distribution in the focusing spot is close to Gaussian.

3. THEORETICAL ANALYSIS

3.1. Formulation of the problem

A characteristic feature of the action of picosecond (and shorter) laser pulses with moderate energy densities ($\leqslant 1-5$ J/cm^2) on strongly absorbing condensed media is that radiation absorption occurs over the duration τ of the laser pulse and the absorbed energy is redistributed between the subsystems in the material (conduction electrons and lattice), while all other processes, such as thermal expansion, vaporization, and possibly melting,[4] require a longer time and proceed after the laser pulse ends. Thus, elastic unloading of a heated layer of matter of thickness $\delta \sim 10^{-5}$ cm, which results in thermal expansion, occurs over a time $\sim \delta/s \simeq 3-10$ ps (s is the sound speed) appreciably longer than the duration τ of the laser pulse. Hence it follows that a stage when a spatially modulated temperature field (writing of a temperature grating) forms as a result of an unstable process with positive feedback[1)] over a time $t \leqslant \tau$, should precede the formation of a periodic surface structure. An important characteristic of this process is that under the action of picosecond and shorter pulses it proceeds with a large decoupling of the electron temperature from the crystal-lattice temperature.

The process leading to the formation of a periodic temperature profile by action on high-conductivity materials must be based on the resonance excitation of surface electromagnetic waves accompanying the diffraction of the incident laser wave by the flat surface with spatial modulation of the optical properties by the periodic modulation of the temperature. It is significant that the modulation depth can be both greater than and of the order of the skin-layer depth.

If periodic surface structures form for $t > \tau$ as a result of the thermal expansion of a melted layer of matter, stamping of a grating (second stage), then it is followed by the reverse

process (as the third stage): resorption of the grating as a result of surface tension. The fixation of a periodic surface profile ultimately depends on whether or not spatial modulation of the surface temperature remains before the melted surface layers of the material can solidify and whether or not the viscosity of the melt can be sufficiently effective to prevent resorption of the periodic profile by surface tension before solidification occurs.

A detailed examination of the above-enumerated elements of the formation of periodic surface structures by ultrashort laser pulses is the content of the analysis of this section.

3.2. Resonant diffraction by a flat surface with periodic modulation of the optical properties

To solve this problem it is convenient to use an impedance boundary condition in the Leontovich form:[6]

$$\mathbf{E}_t = \tilde{\zeta}\mathbf{n}\times\mathbf{H}, \tag{1}$$

where $\mathbf{E}_t$ is the part of the electric field intensity vector that is tangent to the surface, $\mathbf{H}$ is the magnetic field vector, $\mathbf{n}$ is the unit vector of the normal directed into the sample,

$$\tilde{\zeta} = \zeta + \delta\zeta \tag{2}$$

is the effective surface impedance, consisting of a uniform (averaged over the surface) part ζ and a spatially modulated part $\delta\zeta$, where $|\delta\zeta| \ll |\zeta|$. We shall assume below that the depth of uniform (in the plane of the surface) heating is greater than the skin-layer depth. Then ζ is an equilibrium function of the crystal-lattice temperature T_i averaged over the surface. Assuming that the temperature T_e to which the conduction electrons are heated is low compared with the Fermi temperature and the temperatures corresponding to interband transitions, ζ is independent of T_e. Here and below, the metal under study is assumed to be nonmagnetic.

To find the small spatially modulated correction $\delta\zeta$ to the surface impedance, we proceed in the same manner as in the derivation of the impedance boundary condition in the conventional case:[6] Using Maxwell's equations in matter we express the tangential components of the electric and magnetic fields at the boundary in terms of one another. Comparing the relation so obtained and the expressions (1) and (2), we find

$$\delta\zeta = \frac{\partial\zeta}{\partial T_i}\langle\delta T_i\rangle, \quad \langle\delta T_i\rangle \equiv \tilde{\mu}\int_0^\infty dz\ \delta T_i \exp(-\tilde{\mu}z), \tag{3}$$

where $\tilde{\mu} = -2i(\omega/c)(1/\zeta)$ and ω and c are, respectively, the circular frequency and the speed of light. We emphasize that in such a simple form the boundary condition (1) for nonuniform (as a function of depth) modulated (over the surface) heating of the metal is valid to first orer in δT_i.

We note that the spatially modulated parts of the surface impedance and temperatures depend on the coordinates in the boundary plane as

$$\delta\zeta, \delta T_i, \delta T_e \propto \exp(i\mathbf{g}\cdot\mathbf{r}), \tag{4}$$

where $\mathbf{g}$ is the periodicity vector of the surface structure, which satisfies the resonance condition for a surface electromagnetic wave,

$$|\mathbf{k}_t + \mathbf{g}| \simeq \omega/c, \tag{5}$$

$\mathbf{k}_t$ is the projection of the wave vector of the incident wave on the boundary plane of the sample, $|\mathbf{k}_t| = (\omega/c)\sin\theta$, and θ is the angle of incidence.

The solution of the problem of diffraction of a plane wave by a flat surface with spatially modulated optical properties, determined by the relations (2)–(5) consistent with the boundary condition (1) as done for a similar problem in Ref. 7 leads to expressions for the magnetic and electric field vectors of a surface wave that give the following relation for the interference spatially-modulated part of the density of the absorption of laser power:

$$\delta Q(z) = -\mu(1-R)IAB\cos\vartheta\exp(-\mu z + i\vartheta)\frac{\langle\delta T_i\rangle}{T_i},$$

$$B = \frac{T_i}{\zeta'}\frac{\partial\zeta}{\partial T_i}, \quad \zeta' = \mathrm{Re}\,\zeta. \tag{6}$$

Here R is the reflection coefficient corresponding to the lattice temperature T_i, I is the intensity of the incident radiation, $\mu = 2(\omega/c)\mathrm{Im}(1/\zeta)$,

$$A = \frac{|E_p\cos\psi + E_s\cos\theta\sin\psi|^2}{|E_p|^2 + |E_s|^2\cos^2\theta}, \tag{7}$$

E_p and E_s are the components of the field amplitude of the incident wave, respectively, in and perpendicular to the plane of incidence, ψ is the angle between the vectors $\mathbf{k}_t + \mathbf{g}$ and $\mathbf{k}_t$, and ϑ is an angle in the range

$$\frac{\pi}{2} < \vartheta \leqslant \pi + \tan^{-1}\frac{|\zeta''|}{\zeta'}, \quad \zeta'' = \mathrm{Im}\,\zeta, \tag{8}$$

and is determined by the detuning of the resonance of the incident laser wave relative to the surface wave.

We note that the expression (6) for the spatially modulated part of the power absorption density corresponds to resonant diffraction with excitation of one surface wave with the wave vector $\mathbf{k}_t + \mathbf{g}$. The situation corresponding to angles $\psi = \pm(\pi/2 - \theta)$ requires a special analysis. In this case, two surface electromagnetic waves with wave vectors $\mathbf{k}_t + \mathbf{g}$ and $\mathbf{k}_t - \mathbf{g}$ are excited simultaneously and are in resonance. The expression (6) for δQ is modified by adding to it an additional term obtained from the initial term by complex conjugation and by the substitutions $\mathbf{g} \to -\mathbf{g}$, $\psi \to -\psi$.

3.3. Formation of a periodic profile of the temperature field

The periodically modulated radiation absorption is sharply selective with respect to the periodicity vector $\mathbf{g}$ on account of resonance in the surface electromagnetic wave, as calculated in the preceding subsection. In the presence of positive feedback an instability against the formation of a periodically modulated temperature field therefore develops in the surface layer of the sample. To study this process we shall employ the two-temperature model developed in Refs.

3 and 8. In accordance with this model, the system of equations for the spatially modulated part of the conduction-electron temperature and the crystal-lattice temperature under the conditions $T_e \gg T_i$, $\delta T_e \gg \delta T_i$ has the form

$$\left(c_e \frac{\partial}{\partial t} + \tilde{\alpha} \right) \delta T_e = \lambda_e \frac{\partial^2}{\partial z^2} \delta T_e + \delta Q, \quad c_i \frac{\partial \delta T_i}{\partial t} = \alpha \delta T_e . \tag{9}$$

Here c_e and c_i are, respectively, the specific heats of the conduction electrons and crystal lattice,

$$\tilde{\alpha} = \alpha + \lambda_e g^2, \tag{10}$$

α is a parameter describing energy transfer between the conduction electrons and the crystal lattice, and λ_e is the electronic thermal conductivity.

We seek the unstable solution of the system (9), as usual, in the form

$$\delta T_e , \delta T_i \propto \exp(\gamma t), \quad \text{Re}\, \gamma > 0. \tag{11}$$

Two regimes of the behavior of the growth rate γ with respect to the intensity of the incident radiation can be distinguished, depending on the role of the electronic thermal conductivity. In the first regime, where the diffusion length of heat into the sample for spatially modulated heating is large compared with the skin-layer depth μ^{-1} (relatively long laser pulses), the last term on the right side of Eqs. (9) plays the role of a surface source. Then, taking account of Eq. (6) and substituting the expression (11) into Eq. (9), we arrive at the following equation for the growth rate:

$$\gamma \sqrt{\gamma + \frac{1}{\tau_e}} = \frac{\delta Q(0)}{c_i \langle \delta T_i \rangle} \frac{a}{\sqrt{\tau_e}} , \tag{12}$$

where

$$\tau_e = \frac{c_e}{\tilde{\alpha}}, \quad a = \sqrt{\alpha^2 / \lambda_e \tilde{\alpha} \mu^2} . \tag{13}$$

In the visible and near-infrared ranges, where the conduction-electron momentum relaxation time τ_m satisfies $\omega \tau_m \gg 1$, the quantity B appearing in the expression (6) for $\delta Q(z)$ (and defined there) at temperatures above the Debye temperature, where $\tau_m \propto T_i^{-1}$, is real, essentially independent of the lattice temperature, and close to unity:

$$B \approx 1. \tag{14}$$

Under these conditions, taking account of the expression (6) for the source and the condition (8) on its phase, it follows from Eq. (12) that in the regime under study, for radiation intensities below the threshold determined in order of magnitude by the expression

$$I_{\text{th}} \sim [(1-R) a \mu \tau_e]^{-1} c_i T_i , \tag{15}$$

there is no solution of the system (9) that increases with time, and so an aperiodic surface structure does not form.

In what follows, we shall assume that the condition $\omega \tau_m \gg 1$ is not necessary. The expression (14), however, to within a factor of order unity, remains valid over the entire range of applicability of the impedance approximation, and for this reason, we shall employ it in what follows to make estimates without any restrictions on the frequency of the incident radiation.

For intensities appreciably above threshold, the growth rate of the temperature grating in the regime of deep spatial modulation is given, according to Eq. (12), by the expression

$$\gamma \simeq \left(\frac{\delta Q(0)}{c_i \langle \delta T_i \rangle} \frac{a}{\sqrt{\tau_e}} \right)^{2/3} . \tag{16}$$

This expression and the analogous subsequent formula for the growth rate have two free parameters: the angles ϑ and ψ appearing in the expression for δQ [see the relations (6)–(8)]. Here and below, these quantities must be determined by maximizing the real part of the growth rate. Specifically, the wave vector of the periodic spatial structure is established in this manner.

On the basis of this and the relations (6)–(8), it follows from Eq. (16) that by the moment the laser pulse ends in the regime under study, the real part of the exponent in the relations (11), $\Gamma = \tau \,\text{Re}\, \gamma$, is approximately

$$\Gamma \approx \left(a^2 \tilde{Q}^2 \frac{\tau}{\tau_e} \right)^{1/3}, \quad \tilde{Q} = \frac{\mu (1-R) I \tau}{c_i T_i} . \tag{17}$$

In accordance with the role of the thermal conductivity described above, the applicability condition of the expressions (16) and (17) for the first regime is

$$\lambda_e \mu^2 / c_e |\gamma| \gg 1. \tag{18}$$

We shall now consdier the second regime of formation of a temperature grating, where the heat diffusion length into the sample for spatially-modulated heating is small compared with the skin-layer depth,

$$\lambda_e \mu^2 / c_e |\gamma| \ll 1. \tag{19}$$

In this case the thermal conductivity in the system of equations (9) can be neglected, and the expression

$$\gamma = \sqrt{\frac{\tilde{\mu} \alpha}{(\mu + \tilde{\mu}) \tilde{\alpha}} \frac{\delta Q(0)}{c_i \langle \delta T_i \rangle \tau_e}} . \tag{20}$$

is obtained for the growth rate of the instability with respect to formation of a temperature grating. Hence it follows, taking account of the relations (6)–(8), that by the moment the laser pulse ends the real part of the exponent in the relations (11) in the regime under study is given approximately by

$$\Gamma \approx \sqrt{\tilde{Q} \frac{\tau}{\tau_e}} . \tag{21}$$

3.4. Evolution of periodic surface structures at the stage of a melted near-surface layer

We shall now discuss the evolution of periodic surface structures for the case in which the structures form from a temperature grating by thermal expansion of a melted layer near the surface of the material. After the laser pulse ends, the electron and lattice temperatures equalize ($T_e = T_i = T$, $\delta T_e = \delta T_i = \delta T$), and a melted layer forms near the surface, a periodic surface profile with depth

$$h=\frac{1}{3}\beta\delta T\Delta, \tag{22}$$

where Δ is the effective depth of the spatially modulated heating and β is the thermal expansion coefficient of the melt, forms as a result of elastic relaxation accompanied by isotropic thermal expansion. At this stage, surface tension, which tends to flatten the boundary and thereby cause the the periodic structure to be resorbed, comes into play. The solution of the equations of hydrodynamics taking account of the surface tension and viscosity gives an exponential time dependence of the depth of the surface profile:

$$h(t)=h(0)\exp(-t/t^*). \tag{23}$$

The relaxation time appearing here is determined, in the shallow-water approximation[9] $g\Delta\ll 1$, $g=|\mathbf{g}|$, by

$$t^*=\frac{3\eta}{\sigma g(g\Delta)^3}, \tag{24}$$

where η is the dynamic viscosity and σ is the surface tension.

4. DISCUSSION

The writing of a spatially modulated temperature field (temperature grating) is a characteristic, very important, and inevitable (under the conditions of ultrashort laser pulses) preliminary stage in the formation of periodic spatial structures. As shown in the preceding section, this stage and therefore the entire process leading to the formation of periodic structures under the conditions considered are threshold processes. We shall estimate the threshold intensity determined by the expression (15). Setting $\alpha\approx 10^{12}$ W/cm$^3\cdot$K, $\mu\approx 3\times 10^5$ cm^{-1}, $1-R\approx 0.7$, $\lambda_e\approx 1$ W/cm$\cdot$K, $c_i\approx 1.6$ J/cm$^3\cdot$K, $\tau_e\approx 10^{-3}$ s, $T_i\approx 10^3$ K, we obtain from Eq. (13) $a\approx 3$ and from Eq. (15) $I_{\text{th}}\approx 2\times 10^{10}$ W/cm^2.

Noting that the real part of the exponent in Eq. (11) should be $\Gamma\approx 10$ when the laser pulse ends, for the values of μ and λ_e adopted above we conclude that for laser pulse durations $\tau\approx 10^{-12}$ s the second of the two regimes considered in the preceding section is realized in the experiment performed. Since the energy flux in the experiment was 0.1–0.5 J/cm^2, we find on the basis of the relations (21) and (17) that the required values of Γ are attained.

We note that since according to Eqs. (9) and (11) $\delta T_i=(\alpha/\gamma c_i)\delta T_e$ and in view of the fact that the threshold (15) corresponds to the condition $|\gamma|\tau_e\sim 1$, for laser pulse intensities above threshold the periodically modulated parts of the electron and lattice temperatures satisfy

$$|\delta T_i|<\frac{\alpha c_e}{\alpha c_i}|\delta T_e|.$$

Hence it follows, since $c_e\ll c_i$, that a temperature grating can form only if the modulated part of the electron temperature is substantially decoupled from the lattice temperature. At the same time, satisfaction of the inequality $|\gamma|\tau_e>1$ above threshold, taking account of the estimate $|\gamma|\tau\approx 10$, implies that under these conditions decoupling of the electronic temperature, which is uniform in the interface, should also occur (though the decoupling is not as large).

To interpret the experimental data concerning the subsequent stages of the formation of periodic surface structures, we proceed from the fact that after the temperature grating is fixed a periodic relief forms by the thermal expansion of the melted near-surface layer of the material. An indirect argument in favor of this is that both the periodic structure and amorphous graphite[10] appear in much sharper form on the side face section of the sample boundary than on the surface of the sample. This is easily understood on the basis of the fact that the near-surface melted layer plays a dominant role compared with vaporization followed by deposition on a surface, if the following circumstance is taken into account. On account of the layered structure of the sample, the thermal conductivity in a direction normal to the boundary is appreciably higher on the side face than on the surface of the sample. As a result, the melted layer cools more rapidly on the side face and correspondingly the conditions for a transition of the liquid layer into the amorphous state are more favorable. The same is true for the formation of periodic surface structures, since, as noted above, the mechanism based on spatially modulated thermal expansion of the liquid layer requires rapid solidification (due to the diffusion of heat into the sample) and slow relaxation of the spatially modulated temperature field due to the diffusion of heat in the boundary plane. We shall compare the characteristic times of the corresponding processes.

The surface tension and viscosity for liquid graphite that are required in order to estimate the relaxation time of the periodic profile on a liquid surface [Eq. (24)] are unknown. However, the ratio of these two characteristics for liquid metals falls into a comparatively narrow range. For liquid Al, Bi, Pb, Cu, and Fe, this ratio lies in the range $\eta/\sigma=(1-4)\times 10^{-5}$ s/cm. Assuming that for liquid graphite this quantity does not fall far outside these limits, and setting $\Delta\approx 2\times 10^{-6}$ cm and $g\approx 10^5$ cm^{-1}, we obtain the estimate $t^*>1$ ns.

The relaxation time of the spatially modulated thermal profile due to diffusion of heat in the boundary plane of the sample is determined by the expression $t_r\approx(\chi g^2)^{-1}$, where χ is the thermal diffusivity. Setting $\chi\approx 0.1$ cm^2/s with $g\approx 10^5$ cm^{-1} we have $t_r\approx 1$ ns.

Finally, according to Ref. 4, solidification of the surface for the laser action regime under consideration occurs over a time of order $t_s^*\approx 100$ ps.

Comparing the times t^*, t_r, and t_s we conclude that on the one hand the spatial modulation of the surface temperature remains before the melted layer near the surface solidifies, while on the other hand there is not enough time for the periodic surface profile to be resorbed on account of surface tension. All this suggests that the proposed mechanism of formation of periodic surface structures due to the thermal expansion of a melted layer near the surface of the graphite corresponds to the experiment performed.

We note that for sufficiently long laser pulses, such that decoupling of the electronic temperature does not occur, the system of two equations (9) becomes a single equation for the modulated part of the equilibrium temperature δT_e

$=\delta T_i=\delta T$. The solution of the equation in the form (11), taking account of the relations (6), (8), and (14), corresponds to Re $\gamma<0$.

To conclude this section, we wish to make two additional remarks. First, the character and quality of the periodic surface modes which are formed can depend strongly on the initial state of the surface. Thus, in Refs. 11–14 it is shown that a specially produced nonuniformity of the surface, for example, in the form of a straight line, greatly facilitates the formation of periodic surface modes and improves their quality. The second remark concerns the high laser power densities used in our experiments. As shown in Ref. 13, for intensities $\sim 10^{12}$ W/cm^2 the field corrections to the permittivity can become appreciable, which can cause the period of the surface structures to depend on the laser power density. Both of these aspects have a direct bearing on the physics of the formation of periodic surface modes by ultrashort laser pulses and merit special investigations.

5. CONCLUSIONS

The main results of this work are as follows.

The formation of light-induced periodic structures on a graphite surface by picosecond laser pulses was detected experimentally. The closeness of the spatial period of the structure to the wavelength of the incident radiation and the correlation of their orientation and the radiation polarization made it possible to link the mechanism leading to the formation of such structures to resonant excitation of surface electromagnetic waves.

A theoretical model was proposed for the process leading to the formation of periodic surface structures by ultrashort laser pulses. In this model the process proceeds in three stages. In the first stage, spatial modulation of the temperature and pressure along the surface is produced over the application time of the pulse under conditions such that the electronic temperature is decoupled from the lattice temperature. In the second stage, elastic unloading of a layer near the surface occurs and, on account of thermal expansion, a periodic surface relief with preservation of its temperature modulation is formed. In the third stage, two processes occur simultaneously: resorption of the periodic relief due to surface tension and cooling of the surface with solidification. The periodic surface structure becomes fixed if the solidification process is more rapid than the relaxation of the surface profile on the liquid surface.

The estimates made above confirm that the proposed model adequately describes the experiment with formation of periodic surface structures on graphite by picosecond laser pulses.

This work was supported by the Russian Fund for Fundamental Research (Grants 96/52-18494a and 96/52-18495a).

[a)]E-mail: agranat@tiv.phys.msu.su

[1)]We note that an instability induced in the periodic surface structure by the spatial modulation of the temperature under the action of relatively long laser pulses and therefore differing fundamentally from that studied in the present paper was investigated in Ref. 5.

[1] A. E. Siegman and P. M. Fauchet, IEEE J. Quantum Electron. **QE-22**, 1384 (1986).
[2] A. M. Bonch-Bruevich, M. K. Kochengina, M. N. Libenson, V. S. Makin, S. D. Pudkov, and V. V. Turbaev, Izv. Akad. Nauk SSSR, Ser. Fiz. **46**, 1186 (1982).
[3] S. I. Anisimov, B. L. Kapeliovich, and T. L. Perel'man, Zh. Éksp. Teor. Fiz. **66**, 776 (1974) [Sov. Phys. JETP **39**, 375 (1974)].
[4] M. B. Agranat, S. I. Anisimov, S. I. Ashitkov, A. V. Kirillin, P. S. Kondratenko, A. V. Kostanovskiĭ, and V. E. Fortov, Zh. Éksp. Teor. Fiz. **113**, 2162 (1998) [JETP **86**, 1184 (1998)].
[5] A. A. Kovalev, P. S. Kondratenko, and Yu. N. Orlov, Izv. Akad. Nauk SSSR, Ser. Fiz. **53**, 572 (1989).
[6] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Pergamon Press, N. Y., 1984) [Russian original, Nauka, Moscow (1982)].
[7] G. M. Gandel'man and P. S. Kondratenko, Zh. Éksp. Teor. Fiz. **88**, 1470 (1985) [Sov. Phys. JETP **61**, 880 (1985)].
[8] S. I. Anisimov and B. Rethfeld, Proc. SPIE **3093**, 192 (1997).
[9] L. D. Landau and E. M. Lifzhitz, *Fluid Mechanics* (Pergamon Press, N. Y.) [Russian original, Nauka, Moscow (1986)].
[10] M. B. Agranat, S. I. Anisimov, S. I. Ashitkov, A. V. Kirillin, P. S. Kondratenko, A. V. Kostanovskiĭ, and V. E. Fortov, JETP Lett. **66**, 699 (1997).
[11] B. I. Makshantsev and N. F. Pilipetskiĭ, Kvant. Élektron. **12**, 860 (1985) [Sov. J. Quantum Electron. **15**, 563 (1985)].
[12] Yu. V. Vigant, A. A. Kovalev, O. L. Kulikov, B. I. Makshantsev, N. F. Pilipetskiĭ, and E. A. Sukhareva, Zh. Éksp. Teor. Fiz. **91**, 213 (1986) [Sov. Phys. JETP **64**, 308 (1986)].
[13] B. I. Makshantsev and É. A. Manykin, Zh. Éksp. Teor. Fiz. **93**, 437 (1987) [Sov. Phys. JETP **66**, 251 (1987)].
[14] Yu, V. Vigant, A. A. Kovalev, O. L. Kulikov, B. I. Makshantsev, and N. F. Pilipetskii, Appl. Phys. B **44**, 61 (1987).

Translated by M. E. Alferieff

Magnetic anisotropy of epitaxial iron films on single-crystal MgO(001) and $Al_2O_3(11\bar{2}0)$ substrates

Yu. V. Goryunov and I. A. Garifullin

Kazan Physicotechnical Institute, Kazan Science Center, Russian Academy of Sciences, 420029 Kazan, Russia

Th. Mühge and H. Zabel

Ruhr-Universität Bochum, D-44780 Bochum, Germany
(Submitted 19 June 1998)
Zh. Éksp. Teor. Fiz. **115**, 689–703 (February 1999)

The ferromagnetic resonance and magnetization of single-crystal thin (27–100 Å films grown in the (110) direction are measured in the temperature range 20–400 K. The films are prepared by molecular-beam epitaxy on single-crystal sapphire $Al_2O_3(11\bar{2}0)$ substrates with a Nb(110)buffer layer. The angular dependence of the parameters of the ferromagnetic resonance spectrum is observed to have a 180° character when the static magnetic field is rotated in the plane of the sample. It is established that this angular dependence can be described on the assumption that the lattice distortions are essentially trigonal. A comparative analysis of previous data for Fe(001) films with the data for Fe(110) films shows that the source of the corrections to the cubic anisotropy constant is the characteristic distribution of the strains along the thickness of the film. © *1999 American Institute of Physics.* [S1063-7761(99)02102-2]

1. INTRODUCTION

The last decade has witnessed enormous interest in multilayer systems made up of alternating thin layers of ferromagnetic and nonmagnetic metals. The reason is twofold: an encouraging outlook for practical applications of such systems and the discovery of extraordinary phenomena in them—exchange interaction of the ferromagnetic layers that oscillates as a function of the thickness of the nonmagnetic metal layer,[1] and giant magnetoresistance.[2] Giant magnetoresistance is observed for antiferromagnetic mutual orientation of the magnetic moments of adjacent ferromagnetic layers. This phenomenon can be utilized in the design of exceptionally sensitive magnetic sensors or magnetic read heads. To achieve any kind of mutual orientation of the magnetizations, it is necessary that the interaction energy of the layers exceed the magnetic anisotropy energy in each ferromagnetic layer. Consequently, to understand the magnetic and transport properties of multilayer magnetic systems, apart from studying the interaction of the ferromagnetic layers, it is of utmost importance to know the sources of magnetic anisotropy of ferromagnetic thin films.

At the present time we see abundant examples of the powerful influence of surface and interface effects on the magnetic anisotropy of films. The classic manifestation of their influence is the so-called Néel anisotropy,[3] which is a consequence of symmetry breaking of the internal crystal field on the surface of a film or at an interface. As a result of this effect the magnetization of films having a thickness on the order of a few atomic layers tends to order in the direction normal to the film surface. Interface anisotropy is also influenced by the perturbation of the crystal field and band hybridization at the interface. Roughness of the film surface also contributes to the magnetic anisotropy of a film by diminishing the demagnetizing factor when the magnetization is oriented perpendicular to the plane of the film. Among the various mechanisms that contribute to magnetic anisotropy, interface effects induced by a difference in the periods of the crystal lattices of contiguous layers of the substrate and film (the epitaxial misfit is $\eta=(a_d-a_s)/a_s$, where a_d and a_s are the lattice constants of the film and the substrate, respectively) are decisive in many instances.[4–6] For example, in ultrathin Fe films on single-crystal Ag, Au, Cu, and Pd substrates with thicknesses up to 15 Å grown in the (001) orientation epitaxial misfit induces an homogeneous lattice strain, adding fourfold corrections to the cubic anisotropy constants and creating twofold anisotropy.[4]

Experimental data on the magnetic anisotropy of an Au/Co/Au sandwich with Co film thicknesses from 12 Å to 80 Å have been analyzed with allowance for epitaxial misfit.[5] In the analysis it was assumed that a film of thickness smaller than a critical value L_c completely bonds with the substrate, i.e., the lattice strain $\varepsilon=-\eta$. For large thicknesses L the ideal epitaxy states becomes metastable, and epitaxial dislocations appear. In this case the strains relax according to the law $\varepsilon\simeq-\eta L_c/L$. Estimates have shown that the magnetic anisotropy proportional to L^{-1} in Au/Co/Au and Cu/Co/Cu sandwiches can be explained by taking into account the stated strain relaxation law, which causes the magnetic anisotropy to change by virtue of spontaneous magnetostriction. For thicker ferromagnetic films, on the other hand, a qualitative analysis of experimental data on the magnetic properties of epitaxial Fe films with thicknesses up to 200 Å

on GaAs substrates suggests that the magnetic anisotropy contribution due to epitaxial misfit is dominant for such films as well.[6] Despite a fair number of papers on the role of epitaxial misfit in the magnetic properties of ferromagnetic films, we still lack a clear understanding of how epitaxial misfit affects magnetic anisotropy.

The objective of the present investigations is to clarify the nature of the influence of growth distortions induced by epitaxial misfit of the film and substrate on the magnetic anisotropy of ferromagnetic films. Iron films grown on MgO(001) and $Al_2O_3(1\bar{1}20)$ substrates have been chosen for the investigations. This choice is dictated by the abundant use of iron in the preparation of superlattices. It is therefore of utmost importance to understand the nature of the magnetic anisotropy of iron thin films. The choice of MgO(001) as the substrate material stems from the results of a small-angle x-ray scattering (SAXS) study of the strain distribution using synchrotron radiation in application to Fe(001) films grown by rf sputtering in argon.[7] In contrast with iron films on MgO(001) substrates, whose plane coincides with the crystallographic (001) plane, the plane of iron films grown on sapphire $Al_2O_3(1\bar{1}20)$ substrates with a Nb(110) buffer layer is oriented parallel to the crystallographic (110) plane.

In the first stage of our investigations[8,9] we have measured the ferromagnetic resonance (FMR) and magnetization of Fe(001) films of thickness from 25 Å to 500 Å on MgO(001) substrates. An analysis of the data has shown that the corrections to the cubic anisotropy constant and the perpendicular uniaxial anisotropy constant depend strongly on the film thickness L, varying as $1/L$ for $L \geqslant 50$ Å and tending to saturation for $L < 50$ Å. The results are interpreted on the basis of the notion that the epitaxial misfit of the film and the substrate provide the dominant contribution to the magnetic anisotropy of the investigated films.

In this paper we give the experimental results of investigations of the FMR and magnetization of Fe(110) films grown on sapphire substrates with a Nb(110) buffer layer. A comparative analysis of the data for Fe(001) and Fe(110) films has shown that the corrections to the cubic anisotropy constant in Fe(001) films and the anisotropy constant induced by trigonal distortions in Fe(110) films are attributable to the same mechanism: a characteristic (nonideal) distribution of the epitaxial misfit-induced strains along the thickness of the film.

The article is organized as follows. The preparation and characteristics of the samples are described in Sec. 2. The experimental FMR results are given in Sec. 3, and they are analyzed in Sec. 4. The results for the anisotropy constants are discussed in Sec. 5.

2. SAMPLES

2.1. Film preparation

The Fe films were grown by molecular epitaxy on sapphire $Al_2O_3(1\bar{1}20)$ substrates with a Nb buffer layer. The working vacuum in the chamber was $\sim 10^{-10}$ mbar. The single-crystal $Al_2O_3(1\bar{1}20)$ substrates were degassed and were annealed at 1000 °C for one hour immediately prior to deposition. The niobium was evaporated by electron-beam heating from a crucible of volume 14 cm^3. The substrate temperature was maintained close to 900 °C during niobium deposition. The niobium cooling rate was measured during evaporation by an optical method; a rate of 0.5 Å/s was found to be optimal from the standpoint of obtaining high-quality Nb(110) single crystals. The thickness of the Nb films in our samples ranged from 100 Å to 600 Å.

Upon completion of the deposition process the niobium was subjected to a short-time (15-min) anneal at 950 °C. Iron was deposited from a evaporator unit producing a highly stable flow. The iron film was grown at a rate of 0.1 Å/s; its thickness was estimated from the growth time and was subsequently refined on the basis of SAXS data. The substrate temperature was lowered to 100 °C for iron deposition. Since the iron evaporator unit was tilted relative to the substrate holder, the latter was rotated to ensure a uniform sample thickness. Protective layers of niobium and palladium of thickness 30 Å each were deposited on top of the iron layer. During the growth of each layer its thickness was monitored *in situ* by recording the small-angle diffraction of electrons.

2.2. Small-angle x-ray scattering

Small-angle x-ray scattering experiments were conducted using an 18-kV x-ray source with a copper anode. These measurements were performed to determine the thicknesses of the layers in the sample and the roughness parameters of the surface and interfaces. Typical angular plots of the SAXS intensity exhibit distinct oscillations associated with the interference of x-rays reflected from the film surface and the film-substrate interface. Fitting of the resulting angular scattering diagrams by the Parratt formalism[10] yields a interface roughness height < 6 Å for films prepared by both methods. The film thicknesses determined from this fitting of the x-ray spectra were subsequently used as the final thicknesses of the iron layers.

2.3. Bragg x-ray diffraction

Bragg x-ray scattering measurements in the direction perpendicular to the plane of the film have shown that iron on $Al_2O_3(1\bar{1}20)$ substrates with a Nb(110) buffer layer grows in the (110) orientation. The structural coherence length is comparable with the film thickness. Similar in-plane measurements have enabled us to determine the epitaxial ratios of the components in our samples. We have established that an iron film grows as a continuation of the niobium structure, i.e., the Nb and Fe [001] axes are parallel. The Nb and Fe [$\bar{1}$11] axes are oriented parallel to the c axis of the $Al_2O_3(1\bar{1}20)$ substrate in this case.

The indicated sample preparation procedures and their characteristics are discussed in greater detail in Refs. 11 and 12.

2.4. Magnetization measurements

The magnetization was measured by means of a SQUID magnetometer in the temperature range from 4.2 K to 300 K. To preclude the influence of the demagnetizing field, the

TABLE I. Saturation magnetization at $T=20$ K for the investigated samples.

Sample	L, Å	$4\pi M$, kG
Fe(110)	100	20.1
Fe(110)	48	18.8
Fe(110)	27	17.6
Bulk sample[13]		21.9

field dependence of the magnetic moment was measured in a magnetic field of strength up to 5 kOe, which was oriented in the longitudinal direction relative to the plane of the film. At temperatures below 100 K the paramagnetic contribution of uncontrollable impurities in the substrate material to the magnetic moment of the sample became comparable with the ferromagnetic moment of the film. For this reason the saturation magnetic moments at low temperatures were obtained by linear extrapolation of the field dependence of the magnetic moment of the sample from high fields to zero fields. For more accurate accounting of the substrate contribution to the magnetic moment of the sample the magnetic moments of the substrates were measured separately after the iron films had been removed. Estimates show that such a correction can impart at most 10% error to the resulting magnetic moment. Another error source in the determination of the magnetic moments of the films is roughness of the film surface. As mentioned, the degree of surface roughness of our films was less than 6 Å. For a film of thickness 100 Å this height can produce an additional error of order 3%. Consequently, the total error of determination of the saturation magnetic moments of the investigated films is expected to be less than 13%.

Table I gives the results of measurements of the magnetization of Fe(110) films on sapphire substrates with a Nb buffer layer at 20 K. The observed slight thickness dependence of the magnetization has previously[14] been attributed to the presence a nonmagnetic iron layer having a thickness of order 5 Å at the iron-niobium interface. This layer is the result of cross diffusion of Fe and Nb at the Fe/Nb interface.

3. EXPERIMENTAL RESULTS ON FERROMAGNETIC RESONANCE

The FMR measurements were performed on a Bruker Instruments BE-R 418^S spectrometer in a rectangular TE_{102} cavity at a frequency of 9400 MHz in the temperature range from 4.2 K to 400 K. The angular dependence of the parameters of the FMR spectrum was measured relative to the direction of the static magnetic field with the static and microwave magnetic fields oriented parallel to the plane of the film. The direction of the static magnetic field for the Fe(110) films was characterized by the angle θ_H relative to the in-plane [001] axis of the iron film. The detected FMR signal corresponded to the derivative of the absorbed microwave power with respect to the magnetic field. The resonance fields H_0 were chosen half way between the fields corresponding to extrema of the derivative of the absorption. The peak-to-peak uncertainty ΔH_{pp} of the linewidth was 10%, so that the error of determination of the resonance field was of order $0.1\Delta H_{pp}$. The error in determining the magnetic field orientation was less than 2°. The magnetic field strength was determined by means of an NMR magnetometer.

Figure 1 shows FMR spectra for iron films of thickness 100 Å in the direction of a static magnetic field corresponding to various angles θ_H at two temperatures, and Fig. 2 shows the angular dependence of the resonance field at the same temperatures. At high temperatures $T>340$ K and for any direction of the static magnetic field a solitary line is observed, and at low temperatures two resonance lines are observed in the vicinity of the [110] axis of the iron film. These lines merge and vanish when the direction of the static magnetic field deviates approximately 25° from the [110] axis. For thinner samples the angular dependence of H_0 is observed to form a loop over the entire investigated temperature range.

The linewidth as a function of θ_H, as in the case of the Fe(001) films, becomes a maximum at angles corresponding to the maximum values of $|\partial H_0/\partial\theta_H|$. This behavior is observed for both types of angular dependence of H_0.

4. ANALYSIS OF THE EXPERIMENTAL RESULTS

To analyze the FMR results, we use Suhl's equation[15] specifying the FMR condition:

$$\left(\frac{\omega}{\gamma}\right)^2=\frac{1}{(M\sin\theta)^2}\left[\frac{\partial^2F}{\partial\theta^2}\,\frac{\partial^2F}{\partial\phi^2}-\left(\frac{\partial^2F}{\partial\theta\partial\phi}\right)^2\right]. \tag{1}$$

Here $\gamma=g\mu_B/\hbar$, and g is the g factor, F is the magnetic part of the free energy density, and θ and ϕ are the polar and

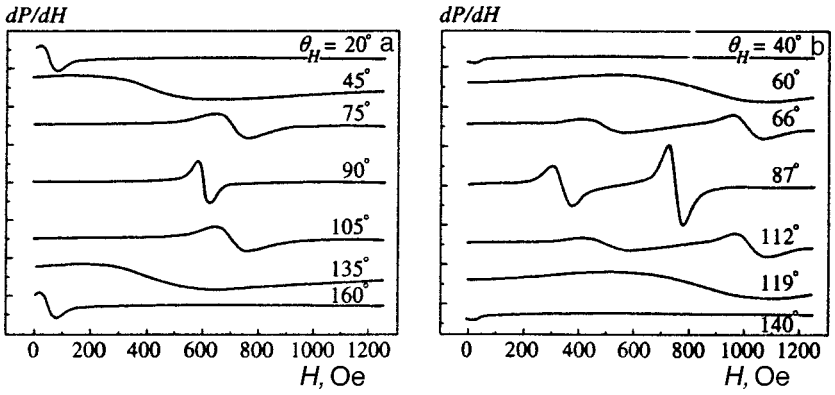

FIG. 1. Ferromagnetic resonance spectra for an Fe(110) film of thickness $L=100$ Å with the static magnetic field in directions corresponding to various angles θ_H at two temperatures: (a) $T=365$ K; (b) $T=295$ K.

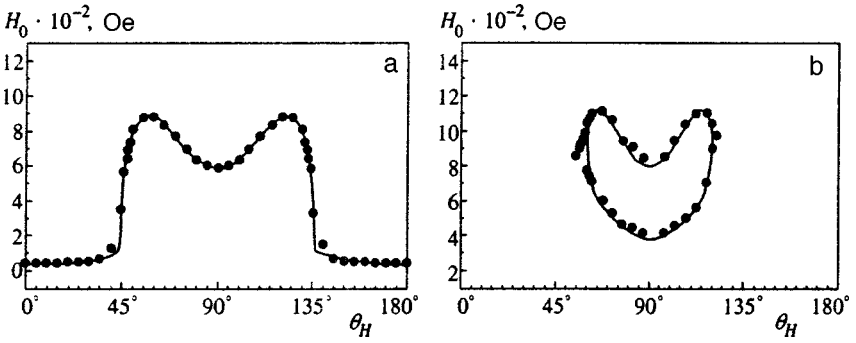

FIG. 2. Angular dependence of the resonance field for an Fe(110) sample of thickness $L=100$ Å: (a) $T=365$ K; (b) $T=295$ K.

azimuth angles characterizing the direction of the magnetization **M**. The equilibrium direction of the magnetization, characterized by the angles θ_0 and ϕ_0, is given by the zeros of the first derivative of the free energy.

The magnetic part of the free energy density consists of the Zeeman energy, the demagnetization energy, and the magnetic anisotropy energy.

An external magnetic field augments the free energy of the film with the Zeeman energy

$$F_Z = -\mathbf{M}\cdot\mathbf{H}. \tag{2}$$

For thin films it is also necessary to take into account the anisotropy associated with the shape of the sample and governed by the demagnetization energy:

$$F_D = 2\pi M_\perp^2, \tag{3}$$

where $M_\perp$ is the component of the saturation moment perpendicular to the film surface.

For an ideal cubic crystal the crystal anisotropy contribution to the free energy density is written in the form

$$F_A = -\frac{1}{2}K_1(\alpha_x^4+\alpha_y^4+\alpha_z^4). \tag{4}$$

Here K_1 is the cubic anisotropy constant and α_i denotes the direction cosines of the magnetization.

For thin films it is also necessary to include the twofold contribution to the anisotropy energy from symmetry breaking of the crystal field on the surface of the film or at an interface (the so-called Néel contribution[3]). This contribution has the same angular dependence as the contribution from the demagnetization energy. For this reason it is customary to include it in the FMR equations in combination with the demagnetization term by introducing the effective magnetization M_{eff} (Ref. 16):

$$4\pi M_{\text{eff}} = 4\pi M - 2K_n/M, \tag{5}$$

where K_n is the uniaxial anisotropy constant, which varies with the thickness L as $1/L$ when this contribution is a pure surface effect, and M is the magnitude of the equilibrium magnetization.

For Fe(110) films it is more convenient to analyze the FMR data in a coordinate system where θ is the polar angle characterizing the deviation of the magnetization vector **M** from the [001] axis, and ϕ is the azimuth angle measured from the iron [110] axis in the xy plane. The external magnetic field is applied at the angles θ_H relative to the [001] axis and ϕ_H relative to the [110] axis. In our experiments $\phi_H = 0$.

These free energy contributions due to magnetic anisotropy are valid only for films having an ideal cubic structure. In reality, however, the crystal lattice of an epitaxial film is distorted. The epitaxial misfit of the film and substrate lattices induces in-plane deformations, which, in turn, create Poisson deformations perpendicular to the plane of the film.

As a rule (see, e.g., Ref. 16), films with a cubic structure grown along the crystallographic [001] direction have tetragonal distortions.

The growth distortions induced by epitaxial misfit in Fe(110) films mainly have a trigonal character, because the iron [111] axis coincides with the **c** axis of the sapphire substrate, whose lattice has axial symmetry. We have used crystal field theory (see, e.g., Ref. 17) to analyze the free energy contributions induced by in-plane trigonal distortions of an iron film. These contributions have a complex general form and contain a whole set of parameters. However, numerical estimates of these parameters indicate small terms that can be disregarded. Combining various terms with identical angular dependence, we find that the main factors contributing to the crystal anisotropy energy in our case, apart from the pure cubic contribution (4), can be written in the form $F_{\text{trig}} = K_u \sin^2\theta + K_u' \sin^4\theta$. This conclusion is consistent with the assumption set forth in Ref. 18.

In the adopted coordinate system the total free energy density is written in the form

$$\begin{aligned} F = &-MH(\cos\theta\cos\theta_H + \sin\theta\sin\theta_H\cos\phi) \\ &+ 2\pi M_{\text{eff}}^2 \sin^2\theta\sin^2\phi + \frac{1}{4}K_1(\sin^2 2\theta \\ &+ \sin^4\theta\cos^2 2\phi) + K_u\sin^2\theta + K_u'\sin^4\theta. \end{aligned} \tag{6}$$

Suhl's general condition for ferromagnetic resonance (1) gives

$$\begin{aligned} \left(\frac{\omega}{\gamma}\right)^2 = &\left[H\cos(\theta_0-\theta_H) + \frac{1}{2}\frac{K_1}{M}(\cos 2\theta_0 + 3\cos 4\theta_0)\right. \\ &\left. + 2\frac{K_u}{M}\cos 2\theta_0 + 2\frac{K_u'}{M}(\cos 2\theta_0 - \cos 4\theta_0)\right] \\ &\times\left[H\cos(\theta_0-\theta_H) + 4\pi M_{\text{eff}}\right. \end{aligned}$$

$$+\frac{1}{8}\frac{K_1}{M}(3\cos 4\theta_0+16\cos 2\theta_0-3)\Big]. \qquad (7)$$

Equation (7) together with the equilibrium condition

$$H\sin(\theta_0-\theta_H)=-\frac{1}{4}\frac{K_1}{M}\sin 2\theta_0(1+3\cos 2\theta_0)$$
$$-\frac{K_u}{M}\sin 2\theta_0-\frac{1}{2}\frac{K_u'}{M}(2\sin 2\theta_0-\sin 4\theta_0) \qquad (8)$$

determines H_0 as a function of the angle θ_H, the effective magnetization $4\pi M_{\text{eff}}$, and the anisotropy constants K_u and K_u'. The constant K_1 is set equal to its value for bulk Fe (Ref. 13).

An analysis has shown that two kinds of angular dependence are possible for certain relations between the effective magnetization $4\pi M_{\text{eff}}$ and the anisotropy fields, one an open curve and the other in the form of a loop, as observed experimentally (Fig. 2). The influence of the parameters $4\pi M_{\text{eff}}$ and K_u differs in these two cases. In the first case, when FMR is observed for all orientations of the external magnetic field, the influence of $4\pi M_{\text{eff}}$ reduces to a change in the average position of the line above the horizontal axis. The parameter K_u characterizes the amplitude of the variation of H_0. In contrast, in the second case (a closed loop), when FMR is observed only near a hard magnetization axis, the parameter K_u gives the position of the center of the loop above the horizontal axis, whereas the parameter $4\pi M_{\text{eff}}$ gives the amplitude of the loop. The parameter K_u', on the other hand, in both cases characterizes the depth of the trough near the hard axis in the angular dependence of H_0.

Typical results of fitting the angular diagrams of the resonance fields H_0 for $g=2.09$ (Ref. 13) are shown in Fig. 2. They exhibit good agreement with the experimental data. These calculations give us the values of the anisotropy constants K_u and K_u' and the effective magnetization $4\pi M_{\text{eff}}$, which are shown in Figs. 3 and 4.

For comparison Fig. 5 shows the cubic anisotropy corrections $K_t=K_1-K_1^b$ for Fe(001) films on MgO substrates as functions of the film thickness. These corrections are attributable to tetragonal distortions of the films along the [001] axis and have been calculated from data given in Ref. 9, Fig. 4. The quantity K_1^b is interpreted as the cubic anisotropy constant for bulk iron samples.[13]

5. DISCUSSION OF THE RESULTS

To achieve our stated goal in the investigations, i.e., to ascertain how epitaxial misfit of the film and substrate influences the magnetic anisotropy of ferromagnetic films, we need to discuss the totality of our experimental data on iron films, including the data reported in our earlier work.[9] Such a discussion can lead to conclusions regarding the interrelationship between the nature of the growth of ferromagnetic films in general and their magnetic anisotropy.

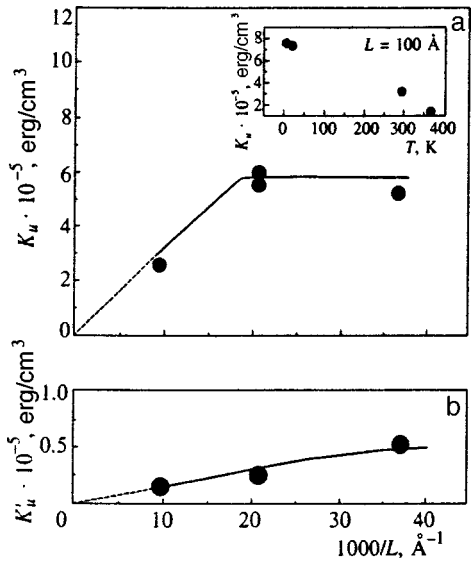

FIG. 3. Thickness dependence of the anisotropy constants K_u and K_u' for Fe(110) films at $T=300$ K.

5.1. Correction K_t to the cubic anisotropy constant

The experimentally determined correction K_t to the cubic anisotropy constant in the Fe(001)/MgO samples depends strongly on the sample thickness (Fig. 5). At $T=300$ K in the range of thicknesses from 80 Å to 500 Å this dependence can be written

$$\frac{K_t}{M}=-\frac{7.3}{d}\,\text{kOe}, \qquad (9)$$

where d is the film thickness expressed in monolayers ($d=L/1.433$ for bcc Fe). This dependence is qualitatively similar to one obtained previously[4] for Fe(001) thin (<15 Å) films on single-crystal Ag(001)) substrates, for which the correction to the cubic anisotropy constant $K_t=-1.25/d$ is associated with a fourfold surface correction.

The negative sign of this contribution for Fe(001)/Ag(001) samples indicates that the surface anisotropy energy

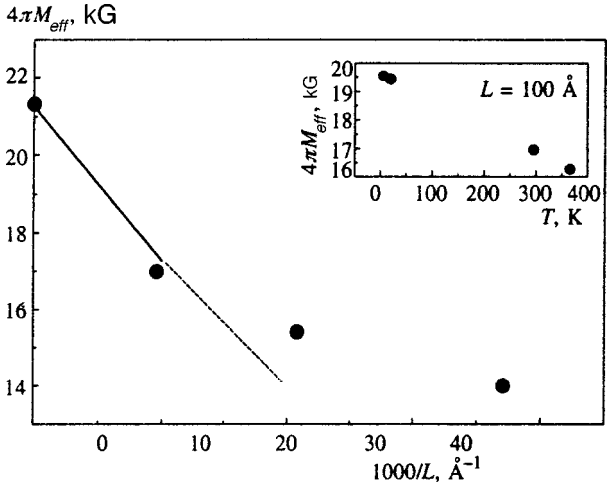

FIG. 4. Thickness dependence of $4\pi M_{\text{eff}}$ for Fe(110) films at $T=300$ K. Inset: temperature dependence of $4\pi M_{\text{eff}}$ for a sample with $L=100$ Å.

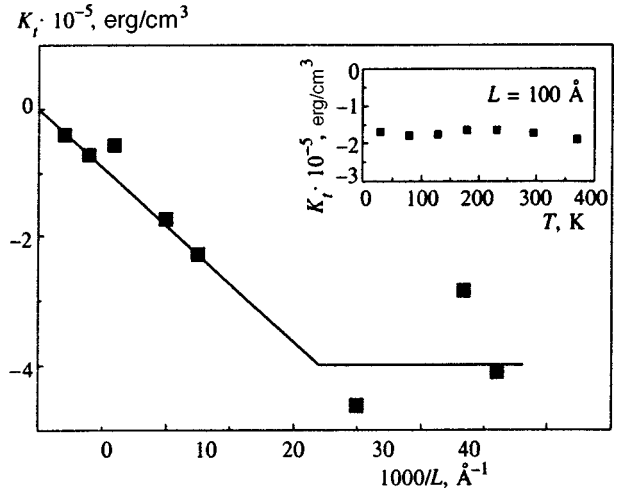

FIG. 5. Thickness dependence of the correction to the cubic anisotropy constant K_t for Fe(001) films at $T=300$ K. Inset: temperature dependence of K_t for a sample with $L=100$ Å.

has an easy magnetization axis parallel to the {110} direction, implying the existence of a certain thickness at which the observed anisotropy constant $K_{1\|}=K_1+K_t$ changes sign, i.e., the easy magnetization axis switches from the {100} to the {110} direction. If we assume that the thickness dependence of K_t in our situation has a form consistent with Eq. (9) down to the smallest thicknesses, the change of sign of the film anisotropy constant should take place for $L=36$ Å ($K_1/M=0.9$ kOe at $T=300$ K, Ref. 13).

However, it is evident from Fig. 5 that this is not the case. It is highly probable that in our case the observed K_t contribution does not come solely from the surface. We can assume that our observed contribution K_t is attributable to epitaxial misfit of the film and the substrate. We know from SAXS experiments[7] that Fe films grow on MgO(001) substrates to thicknesses on the order of ten atomic layers in the form of islands. The islands, which initially harbor small strains, merge as the film thickness increases. The maximum strains are observed at this time. The strain then begins to decline rapidly. Strains localized near an interfacial zone can alter the magnetic anisotropy constant in these boundary layers through magnetostriction. The variation of the anisotropy constant in the vicinity of the interface can exert the following influence the magnetic anisotropy of the film as whole.

It is well known[16] that for films having a thickness smaller than the so-called exchange length $\lambda_{ex}=(A/2\pi M^2)^{1/2}$ the anisotropy field measured by the FMR method can be written as a combination of the bulk anisotropy contribution and the surface anisotropy contribution, which is inversely proportional to the film thickness L. Since the room-temperature spin stiffness coefficient of iron is $A\simeq 1\times 10^{-6}$ erg/cm and the saturation magnetization is $M\simeq 1700$ G, we obtain $\lambda_{ex}\simeq 50$ Å. It has been assumed previously[19] that the anisotropy constant of films with $L\geqslant\lambda_{ex}$ exhibits a more complex dependence on the surface anisotropy contribution than the L^{-1} law. However, Tagirov's calculations in Ref. 9 have shown that a ferromagnetic film subjected to an in-plane external magnetic field can be regarded as a ''dynamically thin'' film up to thicknesses of order 10^3 Å. The FMR-measured contribution of surface anisotropy for such films is therefore averaged over the entire thickness, exactly as in the case of films with thicknesses $L<\lambda_{ex}$ (i.e., the contribution is proportional to L^{-1}). This conclusion does not apply exclusively to a pure surface contribution to the magnetic anisotropy. Owing to the presence of long-wavelength magnetization transfer processes, an interface anisotropy source localized in a boundary layer of thickness $\delta L\ll L$ near the interface will appear to be a surface anisotropy contribution, depending on the film thickness as L^{-1}. It is obvious that as soon as the film thickness L attains the boundary layer thickness δL, the thickness dependence of the anisotropy constant must reach saturation.

We assume, then, that the strains induced by epitaxial misfit at the film-substrate interface have the effect of altering the magnetic anisotropy constant of the boundary layer. It has been established by SAXS in the system Fe(001)/MgO(001) that these lattice strains are localized in a layer of thickness on the order of 30 monolayers near the interface. However, it is impossible to construct the detailed form of the strain distribution on the basis of existing experimental data. For a crude analysis of our experimental magnetic anisotropy data we assume an elementary distribution of the strains in the form of a step function, which yields the following distribution of the corresponding contribution to the anisotropy constant:

$$K_t(z)=\begin{cases} 0 & \text{for} \quad 0\leqslant z\leqslant L-\delta L, \\ K_t^0 & \text{for} \quad L-\delta L<z\leqslant L. \end{cases} \tag{10}$$

The magnetoelastic contribution averaged over the film thickness can then be written as

$$K_t=\frac{1}{L}\int_0^L K_t(z)\,dz=\begin{cases} (\delta L/L)K_t^0 & \text{for} \quad L>\delta L, \\ K_t^0 & \text{for} \quad L\leqslant\delta L. \end{cases} \tag{11}$$

As long as $L<\delta L$ holds, Eq. (11) provides a magnetic anisotropy contribution that depends on the thickness as L^{-1}. Once the thickness L becomes smaller than δL, K_t saturates, eliciting deviations from the law (9) and ruling out a change of sign of the anisotropy field. It is obvious that a certain smooth distribution of K_t exist in reality, accounting for the observed behavior of K_t with a continuous transition from the L^{-1} law to saturation at small thicknesses. Fitting the experimental data on K_t to the thickness dependence given by Eq. (11), we obtain the parameters $K_t^0\sim -4\times 10^5$ erg/cm^3 and $\delta L\sim 45$ Å. The resulting estimate of the thickness of the strained layer is in excellent agreement with SAXS data,[7] whence it follows that the lattice strains of the film drop abruptly at thicknesses on the order of 30 monolayers. Unfortunately, the lack of a microscopic theory of the magnetic anisotropy of ferromagnets makes it impossible to form valid estimates of the correction K_t itself.

The validity of the proposed model has been further corroborated in FMR measurements on specially prepared Fe(001) films having thicknesses of 120 Å, 250 Å, and 500 Å grown on MgO(001) substrates at a temperature of 470 K. The thickness dependence of K_t for this set of samples can be described by the empirical relation

$$\frac{K_t}{M} = -\frac{4.1}{d}\ \text{kOe}. \tag{12}$$

A comparison of K_t for samples grown at different substrate temperatures shows that the product $K_t^0 \delta L$ [see Eq. (11)] for samples grown at 470 K is smaller than for samples prepared at room temperature. This disparity can be attributed to the fact that in films grown at a higher substrate temperature the epitaxial strains are smaller, and they relax at shorter distances from the film-substrate interface. Our data therefore show that the fourfold correction K_t to the cubic anisotropy constant in our investigated films is attributable to strains induced by epitaxial misfit.

5.2. Trigonal anisotropy constants K_u and K_u'

For MgO/Fe(001) the in-plane symmetry of the substrate-induced strains coincides with the crystal symmetry of the iron film. The additional free magnetic energy term has fourfold symmetry. Films of Nb(110) and Fe(110) are oriented so that the [111] axes are parallel to the c axis of an $Al_2O_3(11\bar{2}0)$ substrate. This epitaxial relation with the Al_2O_3 substrate induces trigonal distortions of the Fe and Nb lattices. These distortions require the introduction of corrections to the magnetic anisotropy of an Fe film, which can be effectively taken into account in the form of the twofold and fourfold constants K_u and K_u' (see Sec. 4). Clearly, the constants K_u and K_u' must vanish in the transition to the bulk material, as is indeed observed experimentally (Fig. 3). It is also evident from this figure that at the threshold of ~ 50 Å the anisotropy constants no longer increase as the film thickness decreases, i.e., again we have the same characteristic thickness of the strained layer as in the case of Fe(001)/MgO(001).

5.3. Effective magnetization

The results of our magnetization measurements using a SQUID magnetometer (see Fig. 3 in Ref. 9) show that the saturation moment M of Fe(001)/MgO(001) films does not depend on the thickness of the iron film, whereas the saturation moment measured in the present study (see Table I) for $Fe(110)/Nb(110)/Al_2O_3(11\bar{2}0)$ films exhibits a slight thickness dependence, which, as mentioned above, is associated with the formation of a nonmagnetic iron layer at the Fe/Nb interface. This means that for the magnetically active part of an Fe(110) film the real saturation moment again does not depend on the thickness of the iron film. In addition, the effective magnetization of Fe(001) films (Fig. 5 in Ref. 9) and Fe(110) films (Fig. 4) are essentially identical as functions of thickness. With high probability the resulting magnetization curves are determined by the uniaxial twofold anisotropy constant K_n [see Eq. (5)]. This type of anisotropy can be induced by magnetostriction or by perpendicular Néel anisotropy.[3] A previous analysis[9] has shown that the magnetostriction contribution causes M_{eff} to increase, contradicting our experimental results. Estimates of the Néel contribution have shown that the sign of this contribution is correct, but its value is too small to account for the observed reduction of the magnetization.

An analysis of the data in Figs. 3–5 brings us to the conclusion that K_t, K_u, K_u', and K_n depend on thickness similarly; they all change considerably as L decreases, and they stop changing below a certain thickness. This behavior is most likely indicative of a common origin of the thickness dependence of all these parameters. It has been shown previously[9] that strains produce a combination of opposite sign from that observed experimentally. Note, however, that the relaxation of strains induced by epitaxial misfit at film thicknesses greater than the critical value $L_c \sim 10$ Å occurs because of the onset of epitaxial dislocations. Symmetry is broken on a dislocation line in exactly the same way as on the surface. It is difficult to estimate this contribution to the perpendicular uniaxial anisotropy. On the other hand, it is obvious that the density of dislocations, like the number of atomic planes with broken symmetry, is extremely high and fosters rapid relaxation of the lattice misfit at distances of order 40–50 Å. We can therefore conclude that the Néel mechanism, which gives the correct sign of K_n, can be magnified considerably by epitaxial dislocations, and the results obtained for Fe(110) are qualitatively the same as the results for Fe(001).

In summary, we have shown that epitaxial misfit of the film and the substrate exerts a powerful influence on the magnetic anisotropy of iron films. The epitaxial misfit corrections to the in-plane fourfold cubic anisotropy constants and the perpendicular twofold anisotropy constant vary as L^{-1} as the film thickness L decreases, obeying this law down to a certain thickness of order 50 Å. When the thickness is further decreased, these values reach saturation. The observed behavior of the magnetic anisotropy constants is attributable to the presence in the films of regions having a thickness of order 50 Å near the interface, where the strains induced by epitaxial misfit are localized for the most part. These strains alter the magnetic anisotropy constants in the given region by virtue of magnetostriction and the presence of epitaxial dislocations. Owing to the existence of long-wavelength magnetization transfer processes, this source of anisotropy is manifested as a surface contribution to the anisotropy of the film, depending on the film thickness as L^{-1} for large thicknesses.

A comparison of the results of our investigations with existing data shows that the nature of the magnetic anisotropy of thin films is complex and multifaceted. It is governed by the growth characteristics of the films. Depending on the material chosen for the substrate and the conditions of deposition of the ferromagnetic layer, the growth of a film can proceed one atomic monolayer after another, or it can begin with island growth followed by overlapping of the islands when certain thicknesses are attained. A model strain distribution in the vicinity of the interface is applicable in the first case, the films bonding completely with the substrate at thicknesses smaller than a critical value L_c, i.e., the lattice strain $\varepsilon = -\eta$, and then the ideal epitaxy state becoming metastable at greater thicknesses, with the onset of epitaxial dislocations. Strain relaxation obeys the law $\varepsilon \simeq -\eta L_c/L$. In the second case, as established in our investigations, the source of ''interface'' anisotropy is an expansive region situated near the interface. Thus, the magnetic anisotropy of thin

films requires detailed study in each specific instance. Accordingly, in testing various theoretical models of the interaction of ferromagnetic layers in multilayer systems, it is imperative to bear in mind the real values of the magnetic anisotropy of the ferromagnetic layers, as they can exceed the interaction constants and obstruct the formation of particular magnetic structures.

This work has received financial support from the Russian Fund for Fundamental Research (Project 96-02-16332a).

[1] P. Grünberg, R. Schreiber, Y. Pang, M. B. Brodsky, and C. H. Sowers, Phys. Rev. Lett. **57**, 2442 (1986).
[2] M. N. Baibich, J. M. Broto, A. Fert, F. Nguen Van Dau, F. Petroff, P. Etienne, G. Creuzet, A. Friederich, and J. Chazelas, Phys. Rev. Lett. **61**, 2472 (1988).
[3] L. Néel, C.R. Acad. Sci. **237**, 1623 (1953).
[4] B. Heinrich, Z. Celinski, J. F. Cochran, A. S. Arrott, and K. Myrtle, J. Appl. Phys. **70**, 5769 (1991).
[5] C. Chappert and P. Bruno, J. Appl. Phys. **64**, 5736 (1988).
[6] J. J. Krebs, F. J. Rachford, P. Lubitz, and G. A. Prinz, J. Appl. Phys. **53**, 8058 (1982).
[7] B. M. Lairson, A. P. Payne, S. Brennan, N. M. Rensing, B. J. Daniels, and B. M. Clemens, J. Appl. Phys. **78**, 4449 (1995).
[8] Yu. V. Goryunov, G. G. Khaliullin, I. A. Garifullin, L. R. Tagirov, F. Schreiber, P. Bödeker, K. Bröhl, Ch. Morawe, Th. Mühge, and H. Zabel, J. Appl. Phys. **76**, 6096 (1994).
[9] Yu. Goryunov, N. N. Garif'yanov, G. G. Khaliullin, I. A. Garifullin, L. R. Tagirov, F. Schreiber, Th. Mühge, H. Zabel, Phys. Rev. B **52**, 13450 (1995).
[10] L. G. Parratt, Phys. Rev. **95**, 359 (1954).
[11] Th. Mühge, A. Stierle, N. Metoki, U. Pietsch, and H. Zabel, Appl. Phys. A **59**, 659 (1994).
[12] K. Theis-Bröhl, R. Scheid, Th. Zeidler, F. Schreiber, H. Zabel, Th. Mathieu, C. Mathieu, and B. Hillebrands, Phys. Rev. B **53**, 11613 (1996).
[13] Landolt-Börnstein, *Numerical Data and Functional Relationships in Science and Technology, New Series*, Vol. III/19a, Springer-Verlag, Heidelberg (1986).
[14] Th. Mühge, N. N. Garif'yanov, Yu. V. Goryunov, K. Theis-Bröhl, K. Westerholt, I. A. Garifullin, and H. Zabel, Physica C **296**, 325 (1998).
[15] H. Suhl, Phys. Rev. **97**, 555 (1955).
[16] B. Heinrich and J. F. Cochran, Adv. Phys. **42**, 523 (1993).
[17] S. A. Altshuler and B. M. Kozyrev, *Electronic Paramagnetic Resonance in Compounds of Transition Elements*, Halstead, New York (1975).
[18] G. A. Prinz, G. T. Rado, and J. J. Krebs, J. Appl. Phys. **53**, 2087 (1982).
[19] G. T. Rado, Phys. Rev. B **26**, 295 (1982); Phys. Rev. B **32**, 6061 (1985); J. Appl. Phys. **61**, 4262 (1987).

Translated by James S. Wood

Echo spectroscopy of two-level systems in a Y_2SiO_5: Pr^{3+} crystal

R. S. Borisov, B. V. Grinev, Yu. V. Malyukin, and B. I. Minkov

Institute of Single Crystals, Ukrainian National Academy of Sciences 310001 Khar'kov, Ukraine

N. V. Znamenskiĭ, É. A. Manykin, D. V. Marchenko, and E. A. Petrenko

Russian Science Center ''Kurchatov Institute'', 123182 Moscow, Russia

(Submitted 25 June 1998)
Zh. Éksp. Teor. Fiz. **115**, 704–715 (February 1999)

The results of an experimental investigation of low-temperature optical spectra and phase relaxation of electronic excitations of Pr^{3+} impurity ions in a Y_2SiO_5 crystal are reported. It is established that at low temperatures spectral lines are broadened by a mechanism that is uncharacteristic for crystals and is due to the interaction of impurity ions with two-level systems. The constants characterizing the interaction of Pr^{3+} impurity ions with phonons and two-level systems are found. © *1999 American Institute of Physics.* [S1063-7761(99)02202-7]

1. INTRODUCTION

The numerous investigations of the dynamics of electronic excitations of crystals have been concerned mainly with determining the general mechanisms for the homogeneous broadening of the spectral lines of impurity centers.[1–5] Universal mechanisms that determine spectral line broadening as a result of elastic, magnetic, and electric interactions of an impurity center with its crystal environment were established.[1–4] The universality of these mechanisms is such that any change in the crystal matrix or impurity is reflected only in the value of the parameters that characterize them.

An ''impurity-center–crystal-environment system'' can have, together with the universal interactions of the impurity center with the crystal, a specific local interaction that often cannot be specified at a microscopic level. As a result of interactions of this type, a quasicontinuous or discrete energy spectrum of interatomic origin can appear at an impurity center. An example is the inhomogeneous broadening of spectral lines[6,7] or Shpol'skiĭ multiplets.[8] It is believed[6–10] that the inhomogeneous broadening of the optical spectra of crystals is determined by static disorder. However, by analogy to glasses,[11–14] static disorder in crystals can also depend on the time scale of the experiment.[4,15,16] For glasses at low temperatures, the time dependence of the disorder is determined by tunneling transitions in two-level systems[17,18] which are associated with the multiwell adiabatic potential of the nuclei.[11–14,17,18] Such excitations are called tunnelons.[13,14]

The interaction of an impurity center with two-level systems reduces to transitions of the impurity center between nonequivalent states with different energies. This mechanism also contributes to the homogeneous broadening of the spectral lines of an impurity center.[11–14] For crystals, the multiwell adiabatic potential remains exotic,[19–24] and the existence of two-level systems and their manifestation in relaxation processes are especially unclear. For this reason, cases where a multilevel adiabatic potential with a low activation energy of two-level systems, which ensures efficient quantum tunneling at low and ultralow temperatures, are of special interest, since the contributions of phonons and two-level systems to the homogeneous width of spectral lines can be distinguished in this case. Since disorder is weaker in crystals than in glasses, the number of different two-level systems can be limited, and it may be possible to identify a generalized coordinate of a multilevel adiabatic potential and to study the quantum motion of impurity centers on a microscopic level. As an example, the variety of two-level systems in glasses is so large (the relaxation times can vary over 10 orders of magnitude,[11–14] starting with 10^{-15} s), that this problem is seems to be impossible to solve.

Thus, the crystal YSO : Pr^{3+} is of great interest. In this crystal a hierarchy of nonequivalent states with different energies can exist for impurity ions. These states can be associated primarily with the substitution of two nonequivalent cationic sites in the YSO lattice.[25] Even though this feature is observed in YSO : Nd^{3+} (Ref. 26) and YSO : Eu^{3+} (Ref. 27) crystals, as is confirmed by the presence of two optical centers, it is not been observed in the YSO : Pr^{3+} crystal.[28] But in contradistinction to this, in Ref. 29 spectral lines identified as lines belonging to impurity ions Pr^{3+} occupying nonequivalent cationic sites were observed in a limited spectral range in an investigation of the optical spectra of YSO : Pr^{3+}. A detailed investigation of the contour of a spectral line[30] corresponding to the transition $^3H_4(0)-{}^3P_0$ of impurity ions in a YSO : Pr^{3+} crystal as well as the external Stark effect[29] in the transition $^3H_4(0)-{}^1D_2(0)$ revealed nonequivalent impurity-ion states of a different type, but within the localization site. In the latter case, there is a specific local interaction of the impurity ion with the crystalline environment. This interaction can vanish if either the crystal matrix or the impurity changes.

In the present work an investigation of the low-temperature optical spectra of a YSO : Pr^{3+} crystal unequivocally established two Pr^{3+} optical centers, which correspond to localization of impurity ions in two nonequivalent cationic sites. Analysis of the temperature dependence of the

amplitude of a two-pulse photon echo in the transition $^3H_4(0)-{}^3P_0$ of Pr^{3+} impurity ions established a new (for crystals) dephasing mechanism due to thermally stimulated transitions of the Pr^{3+} ion between nonequivalent states within the localization site. The constants characterizing the interaction of impurity ions with phonons and two-level systems are obtained in a multiwell adiabatic potential model.

2. EXPERIMENTAL TECHNIQUE

The optical absorption and luminescence spectra of a YSO : Pr^{3+} crystal were investigated using an automated spectrofluorimeter, which was based on a MDR-23 grating monochromator. The spectra were detected with a FEU-100 photomultiplier operating in the photon-counting mode. Electronic modules, assembled in a CAMAC standard, controlled the stepping motor of the monochromator and recorded the single-photon counts. The CAMAC crate was coupled through an interface with a personal computer running on an Intel 286 processor.

An R-118 helium optical cryostat was used to obtain low temperatures. In the crysostat the samples were placed in helium vapor.

To eliminate any influence due to the optical anisotropy of the experimental crystal, the two-pulse photon echo was observed in a collinear geometry. The experimental equipment used to excite and detect the echo signal is described in Refs. 30 and 31.

The YSO : Pr^{3+} crystals were grown by the Czochralski method. The concentration of praseodymium in the charge was 0.1 at.%. The experimental samples consisted of plates 1–5 mm thick.

3. SPECTROSCOPY OF THE YSO : PR^{3+} CRYSTAL

A Pr^{3+} impurity ion embedded in different crystal matrices is probably one of the most thoroughly studied ions.[4] In the Russell–Saunders basis,[32] the $4f^2$ electronic configuration of the Pr^{3+} ion gives rise to four singlet terms, $^1S_0, {}^1D_2, {}^1G_4$, and $^1I_6$, and nine triplet terms, $^3H_4, {}^3H_5, {}^3H_6, {}^3F_2, {}^3F_3, {}^3F_4, {}^3P_0, {}^3P_1, {}^3P_2$. Their relative arrangement and therefore the structure of the energy levels of the Pr^{3+} ion are determined primarily by the Coulomb interaction between the $4f$ electrons and by the spin-orbit interaction.[32,33] The fine structure of the energy levels of a rare-earth ion in a crystal is determined by the Stark splitting of the terms, which depends on the point symmetry group of the localization site of the impurity ion.[32,33] Since both nonequivalent cationic sites in a YSO crystal possess C_1 point symmetry,[25] the degeneracy of the terms of Pr^{3+} should be completely lifted. The energy scale of the Stark splitting of the terms of rare-earth ions can vary from tens to hundreds of cm^{-1}.[32,33] In the first-order perturbation theory[32,33] the crystal field does not shift the center of gravity of the multiplets. We shall employ this fact to interpret the optical spectra of a YSO : Pr^{3+} crystal on the basis of spectroscopic data for different crystals activated by Pr^{3+} ions.[4,34–38]

The term $^3H_4$ corresponds to the ground state of the impurity ion. This follows from Hund's rule[32,33] and the Pauli principle[32,33] and is confirmed by existing experimental data.[4,34–38] The next terms in order of increasing energy are $^3H_5, {}^3H_6, {}^3F_2, {}^3F_4$, and $^1G_4$. This group of terms determines the energy spectrum of the Pr^{3+} impurity ion, including the IR range.[34–38] The optical spectrum of the Pr^{3+} impurity ion forms as a result of quantum transitions between Stark components of the terms $^3H_4, {}^2D_1, {}^3P_0, {}^3P_1, {}^1I_6$, and $^3P_6$.[4,34–38] The energy splitting between the centers of gravity of the terms $^1G_4$ and $^1D_2$ is $7000\ cm^{-1}$ on the average.[4,34–38] The frequencies of the electronic transitions with participation of the $^1S_0$ term lie in the far-UV range.[4,34–38] Since the experimentally most interesting[4,29,30] terms $^2D_1$ and $^3P_0$ are well separated in energy from other terms, the spectrum of optical transitions between the Stark components of the terms $^3H_4, {}^1D_2$, and $^3P_0$ is simple and easy to interpret. We shall assume that the qualitatively described energy level structure of the Pr^{3+} ion remains valid in the YSO crystal. For this reason we shall employ a simplified scheme (Fig. 1) to interpret the optical spectra of a YSO : Pr^{3+} crystal and to find the exact arrangement of the terms $^3H_4, {}^1D_2$, and $^3P_0$ and their Stark splitting parameters.

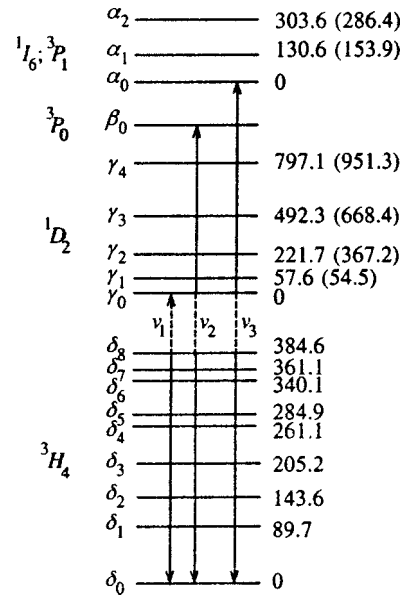

FIG. 1. Simplified scheme of energy levels of a Pr^{3+} ion in a Y_2SiO_5 crystal: $\nu_1 = 16538.2$ (16483.7) cm^{-1}, $\nu_2 = 20540.2$ (20742.6) cm^{-1}, $\nu_3 = 20981.1$ (20867.2) cm^{-1} (the numbers indicate the position of the Stark components in cm^{-1}; the data for the second type of optical centers of Pr^{3+} are shown in parentheses).

The absorption spectrum of the YSO : Pr^{3+} crystal (Fig. 2) consisted of two groups of spectral lines which could be distinguished according to their narrowness. The ratio of the widths at half-height of the corresponding spectral lines in each group was 3:1. We start the analysis and interpretation of the absorption spectrum (Fig. 2) with the group of widest spectral lines, marked by the letters. According to the energy scale of the Stark splitting of the terms of rare-earth ions,[4,34–38], it should be expected that at $T = 6.0$ K only the lowest of the nine Stark components of the term $^3H_4$ (Fig. 1) will be populated. Taking account of this and the results of the investigation of the optical spectra of the Pr^{3+} ion in other crystals,[4,34–38] we shall compare the spectral lines $\gamma_0, \gamma_1, \ldots, \gamma_4$ and β_0 (Fig. 2) with the optical transitions

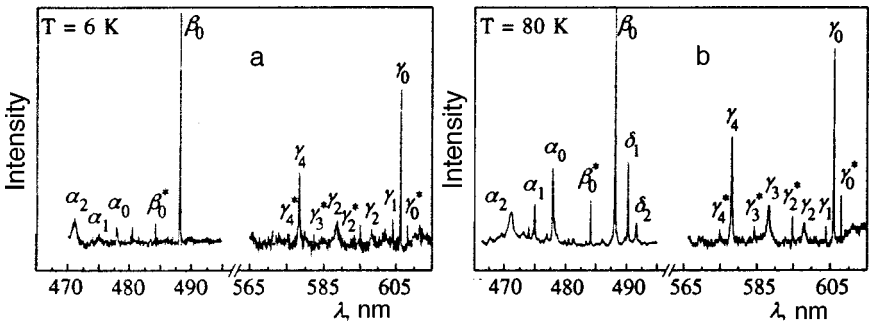

FIG. 2. Fragments of the absorption spectra of a Y_2SiO_5: Pr^{3+} crystal at different temperatures (two groups of lines belonging to different optical centers Pr^{3+} are marked by letters with and without an asterisk).

from the lowest Stark component δ_0 of the term $^3H_4$ to the five Stark components of the $^1D_2$ term and the $^3P_0$ term (Fig. 1).

The validity of this interpretation of the spectral lines $\gamma_0, \gamma_1, \ldots, \gamma_4$ and β_0 (Fig. 2) is additionally confirmed by the following experimental facts. The excitation of luminescence of Pr^{3+} impurity ions to any of the spectral lines $\beta_0, \alpha_0, \alpha_1$, and α_2 (the spectral lines α_0, α_1, and α_2 correspond to optical transitions with the participation of the Stark components of the terms $^3P_1$ and $^1I_6$) is accompanied by the formation of a spectrum consisting of two identical sections (only one section of the spectrum is displayed in Fig. 3), originating from the two spectral lines β_0 and γ_0 which are in resonance with the same lines in the absorption spectrum (Fig. 2). This agrees with the well-known fact[4,34–38] that the emission spectrum of the Pr^{3+} impurity ions forms as a result of optical transitions from two metastable levels β_0 and γ_0 (Fig. 1).

As the temperature of the experimental crystal increases, temperature-dependent spectral satellites, located in the low-frequency range with respect to each of the analyzed lines $\gamma_0, \gamma_1, \ldots, \gamma_4$ and β_0, appear in the absorption spectrum (Fig. 2b). The most intense satellites are observed near the β_0 line (in the scale of Fig. 2b the spectral satellites near the other lines are hard to see). At $T=80$ K two such spectral satellites, δ_1 and δ_2, can be indicated near the β_0 line (Fig. 2b). The frequency intervals determining their position with respect to the analyzed spectral lines coincide exactly with the frequency intervals determining the position of the same spectral lines in the luminescence spectrum (Fig. 3). Therefore the spectral satellites δ_1 and δ_2 appear at $T=80$ K (Fig. 2b) as a result of absorption from thermally occupied Stark components δ_1 and δ_2 of the term $^3H_4$ (Fig. 1). Aside from δ_1 and δ_2, the seven remaining spectral lines, corresponding to transitions from the metastable levels β_0 and γ_0 to the nine Stark components $\delta_0, \ldots, \delta_8$ of the term $^3H_4$ (Fig. 1), can be seen in the luminescence spectrum (Fig. 3). Therefore one group of spectral lines (Fig. 2) can be ''tied'' to the adopted energy level scheme (Fig. 1), and the exact energy parameters, which determine the relative arrangement of the terms $^3H_4, ^1D_2$, and $^3P_0$ and their splitting by the crystal field, can be found from the experimental spectra (Figs. 2 and 3). The energy parameters found in this manner are shown in Fig. 1.

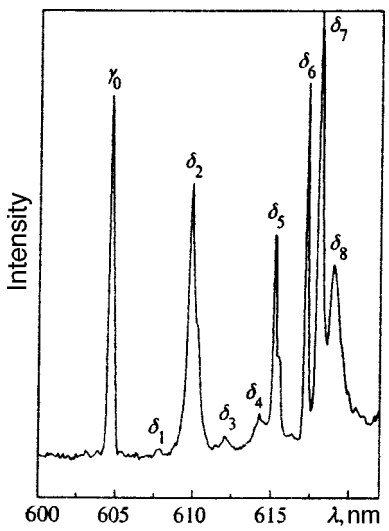

FIG. 3. Fragment of the luminescence spectrum of a Y_2SiO_5: Pr^{3+} crystal at temperature $T=6.0$ K (the spectrum does not depend on the excitation on any of the spectral lines $\beta_0, \alpha_0, \alpha_1$, and α_2).

The weaker but narrower spectral lines present in the absorption spectrum of a YSO : Pr^{3+} crystal (in Fig. 2 they are marked by a letter with an asterisk), do not fit, according to their number and spectral position, into the scheme of levels of the Pr^{3+} ion with these energy parameters (Fig. 1). In Ref. 29 the spectral lines γ_0 and γ_0^* were attributed to absorption in the transition $^3H_4(0)-^1D_2(0)$ of Pr^{3+} impurity ions occupying nonequivalent cationic sites in a YSO crystal.[25] Considering the larger set of ''narrow'' spectral lines $\beta_0^*, \gamma_0^*, \gamma_2^*, \ldots, \gamma_4^*$ (apparently, the line γ_1^* accidentally coincided with the intense line γ_0, Fig. 2), we assume that they can be ''tied'' to the energy level scheme of the Pr^{3+} ion (Fig. 1) but with different parameters. Therefore a second type of Pr^{3+} optical centers can be identified. In this interpretation, the spectral lines $\beta_0^*, \gamma_0^*, \ldots, \gamma_4^*$ correspond to the absorption of Pr^{3+} optical centers of the second type. This absorption is due to transitions from the lowest Stark component of the term $^3H_4$ to the Stark components of the terms $^1D_2$ and $^3P_0$. ''Narrow'' spectral lines are also present in the region of optical transitions involving the Stark components of the terms $^3P_1$ and $^1I_6$ (Fig. 2). However, since the Stark components of the terms $^3P_1$ and $^1I_6$ intermix,[34–38] it is difficult to interpret them unequivocally and additional investigations are required.

On the microscopic level, one cationic site is coordinated with six oxygen atoms and the other with eight.[25,26] In addition, the cationic sites have substantially different average ion–ligand distances and therefore different average volumes. For this reason, the Pr^{3+} ions occupy predominantly a

cationic position with a large volume, because their ionic radius (1.06 Å) is greater than the ionic radius of Y^{3+} (0.92 Å).[39] Therefore, since the ''narrow'' spectral lines are weaker (Fig. 2) and the splitting of the terms by the crystal is somewhat greater (Fig. 2), these lines belong to Pr^{3+} ions which are localized in more dense cationic sites. Selective excitation of Pr^{3+} optical centers of one type, for example, on the β_0 absorption line (Fig. 2) at temperature $T=6.0$ K led only to their luminescence (Fig. 3). Therefore Pr^{3+} impurity ions occupy nonequivalent cationic sites in different unit cells and do not interact with one another (the average distance between different cationic sites of Y^{3+} within the same unit cell is 3 Å in order of magnitude[25,26]).

4. MICROSCOPIC MECHANISM OF THE PHASE RELAXATION OF THE ELECTRONIC EXCITATIONS OF IMPURITY IONS IN A YSO : PR^{3+} CRYSTAL

A photon echo in a YSO : Pr^{3+} crystal was invstigated for one type of Pr^{3+} optical centers which have wider absorption lines (Fig. 2). To obtain the maximum amplitude of the photon echo signal, the laser line ($0.2\ cm^{-1}$) was scanned inside the spectral line β_0 (Fig. 2). It was noted that the intensity of the photon echo signal varies nonmonotonically along the spectral contour of the line β_0. According to Ref. 29, this signifies that a restructuring occurred with respect to the spectral contours which correspond to nonequivalent positions of the impurity praseodymium ion within cationic sites of a given type during the scanning process.

Since the amplitude of the echo signal, determined by the relation $I_{\text{echo}}\sim\exp(-2\gamma(T)\delta)$ (where $\gamma(T)$ is the homogeneous, temperature-dependent width of the spectral line, δ is the time interval between the exciting laser pulses, and T is temperature), makes it possible to obtain directly the law $\gamma(T)=\pi/T_2(T)$ ($T_2(T)$ is the phase relaxation time of electronic excitations), a two-pulse photon echo[5] was used to investigate the temperature mechanisms of the phase relaxation of the electronic excitations of impurity centers in a YSO : Pr^{3+} crystal.

To identify the characteristic features of the phase relaxation of the electronic excitations in a YSO : Pr^{3+} crystal, we shall make a comparative analysis of the results with the analogous results for an LaF_3: Pr^{3+} crystal.[40] To verify the correctness of our experiment,[30] we completely reproduced the results of Ref. 40.

The temperature-dependence of the amplitude of a photon echo was presented in Ref. 40 in a plot in which the double logarithm of the intensity of the echo signal was plotted along the ordinate and the reciprocal of the temperature was plotted along the abscissa. Then the experimental points fall on a straight line, and this temperature behavior was explained in Ref. 40 by direct absorption and emission of phonons with the participation of the Stark components closest in energy to the resonant optical transition. Specifically, for an LaF_3: Pr^{3+} crystal the Stark component $^3H_4(1)$, lying $57\ cm^{-1}$ above $^3H_4(0)$, will determine the phase relaxation of electronic excitations in the transition $^3H_4(0)-{}^3P_0$.[4,34] Since the energy levels of the terms $^3P_1$ and $^1I_6$ closest to $^3P_0$ lie at quite high energies,[4,34–38] they can be neglected at helium temperature. For this reason, the temperature contribution to the homogeneous width of the spectral line of the transition $^3H_4(0)-{}^3P_0$ will have the form[40]

$$\gamma_{\text{ph}}(T)=\alpha\exp\left(-\frac{\Delta\varepsilon}{kT}\right), \tag{1}$$

where α is the width of the Stark component $^3H_4(1)$ at half-height and $\Delta\varepsilon$ is the energy gap between the Stark components $^3H_4(0)$ and $^3H_4(1)$.

If we take account of the exponential dependence of the amplitude of the photon echo on $\gamma(T)$[5,40] and the relation (1), it becomes clear that the slope of the experimental curve presented in the plot in Ref. 40 gives directly the energy gap between the Stark components $^3H_4(0)$ and $^3H_4(1)$. The parameter $\Delta\varepsilon$ determined in this manner was $49\ cm^{-1}$, different from the value $57\ cm^{-1}$ found directly from the optical spectra of the crystal LaF_3: Pr^{3+}.[4,34] Despite this discrepancy, the temperature dependence of the amplitude of the photon echo in an LaF_3: Pr^{3+} crystal[40] was described satisfactorily by the relation (1) with $\alpha=2.4\times10^{11}\ s^{-1}$.[40]

In contrast to LaF_3: Pr^{3+},[40] the temperature dependence of the amplitude of the photon-echo signal in a YSO : Pr^{3+} crystal in the same electronic transition $^3H_4(0)-{}^3P_0$ of the impurity ions (the spectral line β_0 in Fig. 2a) had two characteristic sections where the slopes of the experimental curves,[30] presented in a plot similar to that in Ref. 40, were different. The slope of the experimental curves on the high-temperature section ($T=10-19$ K) was virtually identical to the Stark splitting of the components $^3H_4(0)$ and $^3H_4(1)$ of the term $^3H_4$ in the YSO crystal (Fig. 1). The behavior of the experimental curves within the entire temperature interval of $\gamma(T)$ (Fig. 4) could not be described satisfactorily using the relation (1) with a variable parameter α and $\Delta\varepsilon$

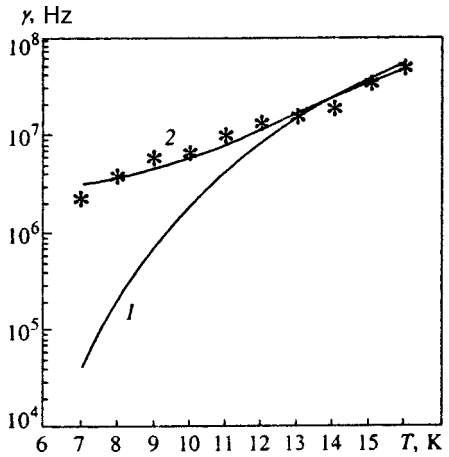

FIG. 4. Temperature dependence of the homogeneous width of the spectral line β_0 corresponding to the transition $^3H_4(0)-{}^3P_0$ of Pr^{3+} impurity ions in a Y_2SiO_5 crystal. The asterisks are experimental points; curve 1 is a least-squares fit of the experimental points on the basis of the relation (1) with $\Delta\varepsilon=89\ cm^{-1}$ and variable parameter α; curve 2 is a least-squares fit of the experimental points on the basis of the simultaneous application of the relations (1) and (3) with $\Delta\varepsilon=89\ cm^{-1}$, $\varepsilon=0.25\ cm^{-1}$, and variable parameters α and β.

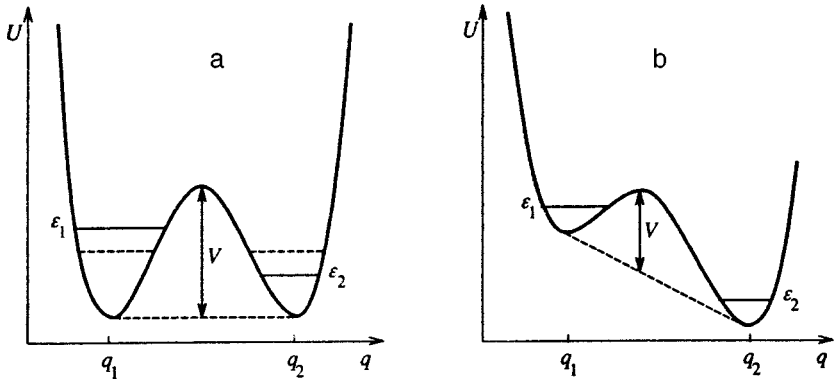

FIG. 5. Models of the symmetric (a) and asymmetric (b) two-well adiabatic potentials ($\Delta=\varepsilon_1-\varepsilon_2$ is the activation energy of two-level systems, V is the barrier height, and q is the generalized anharmonic coordinate).

$=89\ \mathrm{cm}^{-1}$. For $\alpha=1.4\times10^{10}\ \mathrm{s}^{-1}$, a satisfactory description was obtained only in the temperature interval 10–19 K (Fig. 4). The description of the experimental results did not improve even when higher energy Stark components of the term $^3H_4$ (Fig. 1) were included in the analysis. We considered also the Raman mechanism of scattering of phonons by an impurity center of the type[1–4,13] that makes a contribution $\sim T^7$ to the homogeneous width of the spectral line $\sim T^7$, though it is known[4,30,40] to be unimportant in the temperature range 6–10 K. As expected, this likewise did not improve the description of the experimental results (Fig. 4). In other words, the known temperature mechanisms[1–4,13] determining the homogeneous broadening of the spectral lines of impurity centers in crystals could describe satisfactorily the variation of $\gamma(T)$ for the YSO : Pr^{3+} crystal over the entire temperature range (Fig. 4).

Analysis of the structure of the YSO : Pr^{3+} crystal and its rare-earth analogs[25,26] and the way in which rare-earth ion impurities enter these crystals[25,26,29] enabled us to infer that a mechanism, previously not encountered in crystals but known for glasses,[11–14] of temperature broadening of spectral lines due to the interaction of an impurity center with the two-level systems of its multiwell adiabatic potential can operate in YSO : Pr^{3+}. The key point in this supposition was the presence of a number of nonequivalent positions of impurity ions in the YSO : Pr^{3+} crystal. The positions associated with the substitution of nonequivalent cationic sites[25–27] are of no interest, since an impurity ion cannot pass from one position to another. For this reason, the nonequivalent states of the Pr^{3+} impurity ions within a localization site[29,30] that can be represented in a multiwell adiabatic potential model are of interest, though at the microscopic level it remains unclear which generalized nuclear coordinate gives rise to this adiabatic potential. Specifically, this potential will have four minima[29] for cationic sites of one type and two minima[29] for cationic sites of a different type. The populations of the minima of the adiabatic potential in the ground state of the impurity ion are close, since γ_0 is formed by four and γ_0^* by two spectral lines of essentially the same intensity.[29] Using the widths of the spectral lines γ_0 and γ_0^*, the energy splitting between the lowest levels (the activation energy of two-level systems) can be estimated in each minimum of the adiabatic potential. For this, the total width at half-height of the lines γ_0 and γ_0^* must be divided by the number of nonequivalent states of the impurity ion corresponding to them. Since according to Ref. 29 the total width of the line γ_0 at half-height is $1\ \mathrm{cm}^{-1}$ and that of the line γ_0^* is $0.3\ \mathrm{cm}^{-1}$, the desired energy differences are $0.25\ \mathrm{cm}^{-1}$ and $0.15\ \mathrm{cm}^{-1}$, respectively. Such a small energy difference can also arise as a result of tunneling splitting of the levels[20,22,33] and asymmetry of the multiwell adiabatic potential.[20,22,33] If only two minima are considered, as a simplification, then the adiabatic potentials displayed in Fig. 5 can be realized for a Pr^{3+} impurity ion in a YSO crystal. In a symmetric two-well adiabatic potential, temperature-independent quantum tunneling can occur with the frequency

$$\omega=\varpi_0\exp\left(-\frac{|q_1-q_2|}{\hbar}\sqrt{2mV}\right)$$

(where m is the mass of the tunneling system and $\bar{\omega}$ is the average vibrational frequency near one of the minima of the adiabatic potential)[17,18,20,22] as well as temperature-dependent tunneling with the participation of phonons.[17,18,20,22] For an asymmetric potential (Fig. 5b) both processes will become weaker with increasing asymmetry.[17,18,20,22]

The nonequivalent states, which we are discussing, of a Pr^{3+} ion in a YSO crystal have no relation with the Jahn–Teller effect[32,41] in the ''slow rotation'' limit[41], since the degeneracy of the terms of the Pr^{3+} ion with respect to the total angular momentum is lifted. This is unequivocally confirmed by the optical spectra (Fig. 2), and the energy scale of the Stark splitting of the terms (Fig. 1) likewise precludes the manifestation of the pseudo Jahn–Teller effect.[41]

The effect of two-level systems on the homogeneous width of the spectral lines of impurity centers which interact with these systems has been analyzed systematically for glasses.[13,14] In our case the problem simplifies substantially, since according to our assumption a Pr^{3+} impurity ion interacts only with one two-level system. Therefore, to describe the experimental results presented in Fig. 4 we must sum the additive contributions to the homogeneous width of a spectral line which form as a result of the interaction of an impurity center with phonons and with a two-level system. By analogy with phonons,[1,3,13] the contribution to the homogeneous width of a spectral line in the case of an interaction of an impurity center with one two-level system is determined by the change in the structure of the adiabatic potential. This change is expressed as the difference of the Hamiltonians of the ground H_g and excited H_e states[13,14]

$$H_e - H_g = \varepsilon a^+ a + \beta(a^+ + a)q - \frac{\varepsilon}{2}, \tag{2}$$

where $\varepsilon = \Delta_e - \Delta_g$ is the difference of the activation energies of the two-level systems in the excited and ground states of the impurity center, $\beta = (\beta_e - \beta_g)/2$ is the difference of the coupling constants with phonons in the excited and ground states of the impurity center, a^+ and a are tunnelon creation and annihilation operators, and q is the generalized coordinate.

In the dynamic approximation[13,14] the first term in Eq. (2) gives the contribution to the homogeneous spectral line width that exhibits saturation[14] with increasing temperature. We shall not consider this contribution, since it does not improve the description of the experimental results. The second term in the relation (2), associated with the phonon stimulation of transitions between the minima of the adiabatic potential, gives the contribution to the homogeneous width with a temperature dependence of the form[13,14]

$$\gamma_{TLS}(T) = \lambda \sinh^{-1}\left(\frac{\Delta}{kT}\right), \tag{3}$$

where λ is a parameter that depends on β and the spectral function of the phonons[13,14] and $\Delta = \Delta_e \approx \Delta_g$.

The relation (3) has the property that we require. Specifically, for $kT \geqslant \Delta$ it gives an almost linear temperature dependence for the homogeneous width of the spectral line, and for this reason it will be used together with Eq. (1) to describe $\gamma(T)$ (Fig. 4). To obtain the best least-squares fit of the experimental points, the parameters α and λ in the relations (1) and (3) were varied independently. In the relation (1) the value $\Delta\varepsilon = 89\ \mathrm{cm}^{-1}$ (Fig. 1) was used, and Δ in the relation (3) could assume two values. One value is $0.25\ \mathrm{cm}^{-1}$, which follows from the estimate according to the width of the spectral line γ_0. The other value equals the energy gap found in Eq. (30) between the spectral line β_0 and its high-frequency satellite, which is $8.1\ \mathrm{cm}^{-1}$. The latter value was considered in a preliminary analysis[30] to be the activation energy between nonequivalent states of the impurity praseodymium ion. In both cases, the relations (1) and (3) describe the experimental points (Fig. 4) very well, and the values of the varied parameters are $\alpha = 1.04 \cdot 10^{10}\ \mathrm{s}^{-1}$ and $\lambda = 1.1 \cdot 10^5\ \mathrm{s}^{-1}$ for $\Delta = 0.25\ \mathrm{cm}^{-1}$ and $\alpha = 1.01 \cdot 10^{10}\ \mathrm{s}^{-1}$ and $\lambda = 3.9 \cdot 10^6\ \mathrm{s}^{-1}$ for $\Delta = 8.1\ \mathrm{cm}^{-1}$.

5. CONCLUSIONS

In summary, it has been shown in this work that the familiar temperature-dependent mechanisms in crystals for broadening lines in the optical spectrum of impurity centers do not permit describing systematically the temperature variation of the homogeneous width $\gamma(T)$ of a spectral line corresponding to the transition ${}^3H_4(0) - {}^3P_0$ of impurity ions in a Y_2SiO_5: Pr^{3+} crystal. The low-temperature spectra and ultrahigh resolution spectroscopic data made it possible to observe in a Y_2SiO_5: Pr^{3+} crystal two types of nonequivalent states, having different energies, of impurity ions. The states of one type are associated with the substitution of impurity ions at two nonequivalent cationic sites in the Y_2SiO_5 crystal lattice. The energy difference between the same Stark components of impurity ions in nonequivalent cationic sites depends on the type of term and fluctuates from tens to hundreds of cm^{-1}. The structural crystallographic nature of the differences of two nonequivalent cationic sites in the YSO lattice signifies that the impurity ions substituting for them at the sites behave as independent ensembles of impurity particles. Nonequivalent states of a different type in the YSO crystal lattice are realized within the localization site of an impurity ion. The energy difference between such states is small, and according to the experimental data it can range from tenths to several cm^{-1}. In this paper it was suggested that at low temperatures the impurity ions in a Y_2SiO_5: Pr^{3+} crystal can undergo phonon-induced transitions from one nonequivalent state to another within a localization site. Such transitions of an impurity ion were represented formally in a multilevel adiabatic potential model. The temperature broadening of the spectral line of the transition ${}^3H_4(0) - {}^3P_0$ of impurity ions in a YSO crystal was described very well by using an additional mechanism whereby an impurity ion interacts with a two-level system of its multilevel adiabatic potential. Determining more accurately the reasons for the appearance of a multilevel adiabatic potential for a Pr^{3+} impurity ion in a YSO crystal will make it possible in the future to perform an experiment that is unthinkable for glasses: investigation of the quantum motions of impurity centers in crystals at the microscopic level.

[1] V. M. Agranovich and R. M. Hochstrasser [Eds.], *Spectroscopy and Excitation Dynamics of Condensed Molecular Systems* (Elsevier, N. Y., 1983) [Nauka, Moscow, 1987].
[2] P. De Bree and D. A. Wiersma, J. Chem. Phys. **70**, 790 (1979).
[3] I. S. Osad'ko, Usp. Fiz. Nauk **128**, 31 (1979) [Sov. Phys. Usp. **22**, 311 (1979)].
[4] A. A. Kaplyanskii and R. M. Macfarlane [Eds.], *Spectroscopy of Solids Containing Rare Earth Ions* (North-Holland, Amsterdam, 1987).
[5] É. A. Manykin and V. V. Samartsev, *Optical Echo Spectroscopy* (Nauka, Moscow, 1984).
[6] K. M. Macfarlane, J. Lumin. **45**, 1 (1990).
[7] J. L. Skinner, B. B. Laird, and L. Koot, J. Lumin. **45**, 6 (1990).
[8] É. V. Shpol'skiĭ, Usp. Fiz. Nauk **71**, 215 (1960) [Sov. Phys. Usp. **3**, 372 (1960)].
[9] R. I. Personov, B. I. Al'shits, and L. A. Bykovskaya, Zh. Éksp. Teor. Fiz. **65**, 1825 (1973) [Sov. Phys. JETP **38**, 912 (1974)].
[10] M. Stoneham, Rev. Mod. Phys. **41**, 82 (1969).
[11] R. Jankowiak and G. J. Small, Science **237**, 618 (1987).
[12] H. C. Meijers and D. A. Wiersma, J. Chem. Phys. **101**, 6927 (1994).
[13] I. S. Osad'ko, Phys. Rep. **206**, 43 (1991).
[14] I. S. Osad'ko, Adv. Polym. Sci. **114**, 123 (1994).
[15] R. Flach, D. S. Hamilton, P. S. Selzer *et al.*, Phys. Rev. Lett. **35**, 1034 (1975).
[16] P. S. Selzer, D. S. Hamilton, and W. M. Yen, Phys. Rev. Lett. **38**, 858 (1977).
[17] P. W. Anderson, B. I. Halperin, and C. M. Varma, Philos. Mag. **25**, 1 (1972).
[18] W. A. Phillips, J. Low Temp. Phys. **7**, 351 (1972).
[19] A. J. Horsewill, P. J. McDonald, and D. Vijayaraghavan, J. Chem. Phys. **100**, 1889 (1994).
[20] J. L. Skinner and H. P. Trommsdorff, J. Chem. Phys. **89**, 897 (1988).
[21] A. M. Alsanoosi and A. J. Horsewill, Chem. Phys. **160**, 25 (1992).
[22] J. P. Sethna, Phys. Rev. **24**, 698 (1981).
[23] C. P. Flynn and A. M. Stoneham, Phys. Rev. **1**, 3966 (1970).
[24] M. Neumann, M. R. Johnson, and H. P. Trommsdorff, J. Lumin. **66–67**, 146 (1996).
[25] B. A. Maksimov, V. V. Ilyukhin, Yu. A. Kharitonov *et al.*, Kristallografiya **15**, 926 (1970) [Sov. Phys. Crystallogr. **15**, 806 (1970)].
[26] A. M. Tkachuk, A. K. Przhevusskiĭ, L. G. Morozova *et al.*, Opt. spek-

trosk. **60**, 288 (1986) [Opt. Spectrosc. **60**, 176 (1986)].
[27] M. Mitsunaga, T. Takagahara, K. Yano *et al.*, Phys. Rev. Lett. **68**, 3216 (1992).
[28] N. V. Kuleshov, V. P. Mikhaĭlov, S. A. Radkevich *et al.*, Opt. spektrosk. **77**, 244 (1994) [Opt. Spectrosc. **77**, 220 (1994)].
[29] K. Holliday, M. Croci, E. Vanthey *et al.*, Phys. Rev. **47**, 14741 (1993).
[30] Yu. V. Malyukin, N. L. Pogrebnyak, V. P. Seminozhenko *et al.*, Zh. Éksp. Teor. Fiz. **108**, 485 (1995) [JETP **81**, 260 (1995)].
[31] P. N. Zhmurin, N. L. Pogrebnyak, V. P. Seminozhenko *et al.*, Ukr. Fiz. Zh. **39**, 791 (1994).
[32] I. B. Bersuker, *Electronic Structure and Properties of Coordination Compounds* (Khimiya, Leningrad, 1986).
[33] W. H. Flygare, *Molecular Structure and Dynamics* (Prentice-Hall, Englewood Cliffs, N. J., 1978) [Russian translation, Mir, Moscow, 1982, Vol. 1].
[34] W. N. Yen, W. C. Scot, and A. L. Schawlow, Phys. Rev. **136**, 271 (1964).
[35] C. A. Morrison and R. P. Leavit, J. Chem. Phys. **71**(6), 2366 (1979).
[36] H. H. Caspers, H. E. Rast, and R. A. Buchanan, J. Chem. Phys. **43**(6), 2124 (1965).
[37] S. Matthies and D. Welsch, Phys. Status Solidi B **6S**, 125 (1975).
[38] B. R. Reddy and L. E. Erickson, Phys. Rev. **27**(9), 5217 (1983).
[39] E. S. Makarov, *Isomorphism of Atoms in Crystals* (Atomizdat, Moscow, 1973).
[40] N. Takeuchi, J. Lumin. **12/13**, 743 (1978).
[41] A. M. Stoneham, *The Theory of Defects in Solids: The Electronic Structure of Defects in Insulators and Semiconductors* (Clarendon Press, Oxford, 1975) [Russian translation, Mir, Moscow, 1978, Vol. 1).

Translated by M. E. Alferieff

Skyrmions in electron gas with nonlocal exchange in a strong magnetic field

S. V. Iordanskiĭ and S. G. Plyasunov

L. D. Landau Institute of Theoretical Physics, Russian Academy of Sciences, 142432 Chernogolovka, Moscow Region, Russia

V. I. Falko

School of Physics and Chemistry, Lancaster University LA1 44B UK; Department de Physique, University J. Fourier, Grenoble 1 France

(Submitted 19 August 1998)
Zh. Éksp. Teor. Fiz. **115**, 716–725 (February 1999)

The energy and action for skyrmions in two-dimensional electron gas with nonlocal exchange have been calculated. The energy of positively charged skyrmions is considerably lower than the energy of negatively charged skyrmions and does not contain an exchange contribution. The action has been calculated taking into account collective skyrmion null modes.
 [S1063-7761(99)02302-1]

After the pioneering work by Sondhi *et al.*,[1] a number of publications were dedicated to calculating the energy of a specific spin texture—a skyrmion having an odd filling factor of Landau levels in a 2D electronic structure.[2–4] Such textures are characterized by rotation of average spin with a nonzero degree of the 2D plane mapping on the unit sphere of spin directions. It was found that generation of a pair of skyrmions with opposite charges yields a lower energy than formation of a spin exciton.

Calculations of the energy of an isolated skyrmion in subsequent publications were not fully self-consistent. The authors assumed that the required textures could be obtained via rotation of spins in the uniform state using a coordinate-dependent rotation matrix $U(\mathbf{r})$, but they assumed that the electron spinors belonged to states of the same Landau level both before and after rotation. Moreover, they used a certain reduced rotation matrix[4] $U_{\rm red}$, which is not unitary: $U_{\rm red}^{\dagger}U_{\rm red}\neq 1$. In addition, the unreduced rotation matrices initially used were considered to be functions of two Euler angles, which necessarily gives risesingularities for nontrivial degrees of mapping, whereas the effective formulas did not take into account this fact.

These inconsistencies can be eliminated using a self-consistent approach based on smooth rotation matrices $U(r)$ which are functions of the three Euler angles (or another equivalent set of parameters), taking into consideration the full representation of electron spinors in terms of Landau wave functions, and using an expansion in terms of derivatives of the rotation matrix $U(\mathbf{r})$.[5,6]

A flaw of previous calculations was the assumption that the exchange term was uniform, which applies only to the case of a short-range interaction potential. Moreover, the final expressions for the action contained a number of serious errors. The present work is aimed at eliminating these flaws, which is essential for calculating the energy of an isolated skyrmion.

1. CALCULATION OF ACTION AND ENERGY

Let us briefly discuss the procedure for calculating the action.[6] The action is calculated in the Hartree–Fock approximation, which is valid for the case of one fully filled Landau level. One can also obtain this approximation via the Hubbard–Stratonovich transform. In this case, the term in the Hamiltonian describing the interaction can be expressed as

$$-\int V(\mathbf{r}-\mathbf{r}_1)\langle\psi_\alpha^{\dagger}(\mathbf{r})\psi_\beta(\mathbf{r}')\rangle\psi_\beta^{\dagger}(\mathbf{r}')\psi_\alpha(\mathbf{r})d^2r\,d^2r'.$$

Here α and β are spin indices, $X_{\alpha\beta}(\mathbf{r},\mathbf{r}')=\langle\psi_\alpha^{\dagger}(\mathbf{r})\psi_\beta(\mathbf{r}')\rangle$ is the Hubbard–Stratonovich exchange field corresponding to the softest degree of freedom, namely, the spin rotation in the exchange approximation (the angular brackets denote quantum-mechanical averaging). The direct interaction can be included using the well-known expression for the average charge density.[1–6] In the uniform case (the lower spin sublevel of the lower Landau level is fully populated), the quantum-mechanical average has the form

$$X_{\alpha\beta}^{0}(\mathbf{r},\mathbf{r}')=\sum\ \delta_{\alpha\beta}\delta_{\alpha 0}\Phi_{p0}^{*}(\mathbf{r})\Phi_{p0}(\mathbf{r}'). \tag{1}$$

We introduce instead of $X_{\alpha\beta}^{0}$ the transformed matrix

$$X_{\mu\nu}(\mathbf{r},\mathbf{r}_1)=U_{\alpha\mu}^{\dagger}(\mathbf{r})U_{\nu\beta}(\mathbf{r}_1)X_{\alpha\beta}^{0}(\mathbf{r},\mathbf{r}_1),$$

where $U(\mathbf{r})$ is a matrix of spinor rotation ($U^{\dagger}U=1$). One can get rid of the rotation operator in the interaction term by introducing new spinors χ and $\chi^{\dagger}$ via the unitary transform $\psi(\mathbf{r})=U(\mathbf{r})\chi(\mathbf{r})$, so that the interaction term regains its initial form

$$-\int V(\mathbf{r}-\mathbf{r}_1)X_{\alpha\beta}^{0}(\mathbf{r},\mathbf{r}_1)\chi_\beta^{\dagger}(\mathbf{r}_1)\chi_\alpha(\mathbf{r})d^2r\,d^2r_1$$

for the transformed spinors, since their potential interaction does not change under this unitary transform. In the case of a

coordinate-dependent rotation matrix, $U(\mathbf{r})$ must be differentiated in the expression for the kinetic energy in the full Hamiltonian. As a result, we have

$$H=\frac{1}{2m}\int \chi^{\dagger}(-i\nabla+\mathbf{A}_0+\mathbf{\Omega}^l\sigma_l)^2\chi d^2r + \int \chi^{\dagger}\Omega_t^l\sigma_l\chi d^2r-\frac{1}{2} \times\int V(\mathbf{r}-\mathbf{r}_1)\chi_\alpha^{\dagger}(\mathbf{r})\chi_\beta(\mathbf{r}_1)\chi_\beta^{\dagger}(\mathbf{r}_1) \times\chi_\alpha(\mathbf{r})d^2rd^2r_1.$$

We have written out the full second-quantization Hamiltonian, where $\Omega_\nu^l\sigma_l=-iU^{\dagger}\partial_\nu U$, σ_l are Pauli matrices and A_0 is the vector potential of applied magnetic field. Here we use a system of units such that $\hbar=1$, $l_H^2=c\hbar/eH=1$, and $H=1$, so the reciprocal electron mass $1/m$ is measured in units of $\hbar\omega_c$. Parameters Ω_ν^l expressed in terms of Euler's angles are given by

$$\Omega_\nu^z=\frac{1}{2}(\partial_\nu\alpha+\cos\beta\partial_\gamma),$$

$$\Omega_\nu^x=\frac{1}{2}(\sin\beta\cos\alpha\partial_\nu\gamma-\sin\alpha\partial_\nu\beta),$$

$$\Omega_\nu^y=\frac{1}{2}(\sin\beta\sin\alpha\partial_\nu\gamma+\cos\alpha\partial_\nu\beta).$$

The assumed small value of gradients of U allows us to develop a perturbation theory in terms of Ω (gradient expansion) for calculating the action as a functional of U. To this end, we decompose the Hamiltonian into two parts:

$$H_0=\int d^2r\left\{\frac{1}{2m}\chi^{\dagger}(-i\nabla+\mathbf{A}_0+\mathbf{\Omega}^l\sigma_l)^2\chi+\chi^{\dagger}\Omega_t^l\sigma_l\chi\right\} -\int V(\mathbf{r}-\mathbf{r}_1)X_{\beta\alpha}^0(\mathbf{r}_1,\mathbf{r})\chi_\alpha^{\dagger}(\mathbf{r})\chi_\beta(\mathbf{r}_1)d^2rd^2r_1 \quad (2)$$

and

$$H'=-\frac{1}{2}\int V(\mathbf{r}-\mathbf{r}_1)\chi_\alpha^{\dagger}(\mathbf{r})\chi_\beta(\mathbf{r}_1)[\chi_\beta^{\dagger}(\mathbf{r})\chi_\alpha(\mathbf{r}_1) -2X_{\beta\alpha}^0(\mathbf{r}_1,\mathbf{r})]d^2rd^2r_1, \quad (3)$$

where $X_{\alpha\beta}^0$ is averaged over the fully filled spin-up Landau sublevel.

The action, which is a function of the rotation matrix U, can be calculated in the Hartree–Fock approximation and is expressed by diagrams in Fig. 1, where δG denotes the correction to the Green's function of the full Landau level due to terms with Ω_ν in Hamiltonian H_0. The quantity S_0 is the action calculated with the Hamiltonian H_0:

$$S_0=i\,\mathrm{Tr}\ln\frac{G}{G_0},\quad G=G_0+\delta G.$$

The corrections linear in δG are included in S_0. The quantity S_0 was calculated previously to lowest order in $1/m$ for uniform spin exchange,[6] but that calculation contained an algebraic error. This paper presents calculations including terms through zeroth order in $1/m$ without assuming that the spin exchange is uniform, unlike the previous publication.[6]

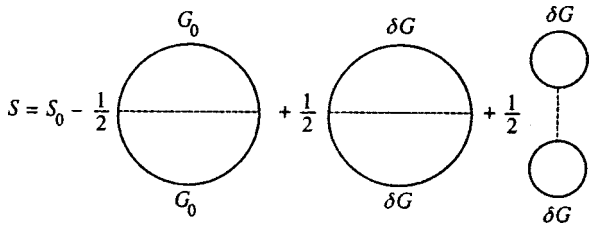

FIG. 1. Hartree–Fock diagrams for action. The dashed lines denote the interaction potential.

The Hamiltonian H_0 can be decomposed into the main part

$$H_{00}=\frac{1}{2m}\int \chi^{\dagger}(-i\nabla+\mathbf{A}_0)^2\chi d^2r -\int\int V(\mathbf{r}-\mathbf{r}_1)X_{\beta\alpha}^0(\mathbf{r}_1,\mathbf{r})\chi_\alpha^{\dagger}(\mathbf{r})\chi_\beta(\mathbf{r}_1)d^2rd^2r_1,$$

plus small corrections of first order in Ω,

$$H_1=\int \Omega_t^l\sigma_l d^2r+\frac{1}{m}\int \chi^{\dagger}\Omega_\mu^l\sigma_l(-i\partial_\mu+A_{0\mu})\chi d^2r$$

and second order in Ω,

$$H_2=\frac{1}{2m}\int \chi^{\dagger}[(\Omega_\mu^l\sigma_l)^2-i\partial_\mu\Omega_\mu^l\sigma_l]\chi d^2r.$$

As was shown in the previous publication,[6] the contribution of H_2 is offset by the contribution of H_1 in the second order. In addition, the contribution of $\Omega_t^l\sigma_l$ can be trivially calculated[6] and is not discussed here.

The Green's function G_0 for the Hamiltonian H_{00} corresponding to the fully filled lower Landau spin sublevel is calculated very easily and has the form

$$G_0(\mathbf{r},\mathbf{r}',\omega)=\sum_{p,s}\hat{g}_s(\omega)\Phi_{sp}(\mathbf{r})\Phi_{sp}^*(\mathbf{r}'). \quad (4)$$

Here we have performed the Fourier transform to the frequency representation; Φ_{sp} are the electron wave functions of the sth level in the Landau gauge. The levels needed for our calculations are $s=0$:

$$\hat{g}_0(\omega)=\frac{1+\sigma_z}{2}\frac{1}{\omega+E_0/2-i\delta}+\frac{1-\sigma_z}{2}\frac{1}{\omega-E_0/2+i\delta},\quad \delta\to+0, \quad (5)$$

and $s=1$:

$$\hat{g}_1(\omega)=\frac{1+\sigma_z}{2}\frac{1}{\omega-1/m+(E_1-E_0/2)+i\delta} +\frac{1-\sigma_z}{2}\frac{1}{\omega-1/m-E_0/2+i\delta}. \quad (6)$$

Higher levels s are essential only for calculation of contributions higher than second order in Ω. The Green's function is calculated for the chemical potential $\mu=1/2m-1/2E_0$, which

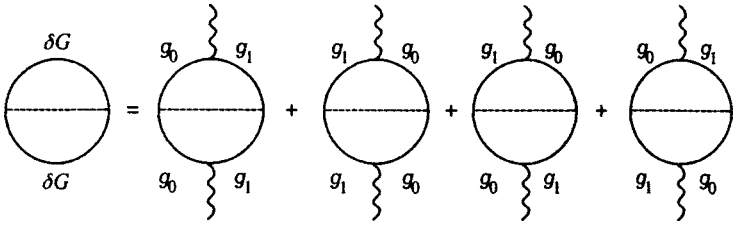

FIG. 2. Diagrams illustrating different ways of inserting into the exchange diagram the first-order corrections containing H_1, shown by a wavy line.

corresponds to the full population of the lower spin sublevel. The exchange energies for the fully filled Landau levels are calculated easily:

$$E_0=\int \frac{d^2q}{(2\pi)^2}e^{-q^2/2}V(q),$$

$$E_1=\int \frac{d^2q}{(2\pi)^2}\frac{q^2}{2}e^{-q^2/2}V(q),$$

where $V(q)$ is the Fourier transform of the interaction potential. For the Coulomb interaction

$$E_1=\frac{1}{2E_0}=\sqrt{\frac{\pi}{2}}\frac{e^2}{2\kappa l_H}, \tag{7}$$

where κ is the dielectric constant.

The expansion of the action S_0 to second order in H_1 has the form

$$\delta S_0=i\,\mathrm{Tr}\,H_1G_0+\frac{i}{2}\mathrm{Tr}\,H_1G_0H_1G_0+i\,\mathrm{Tr}\,H_2G_0. \tag{8}$$

The calculation procedure uses the representation of H_1 in the form

$$H_1=\frac{1}{m}\int \chi^\dagger(\Omega_+^l\hat{\Pi}_-+\Omega_-^l\hat{\Pi}_+)\sigma_l\chi d^2r,$$

where $\Omega_\pm^l=(\Omega_y^l\mp i\Omega_x^l)/2$ and the operators $\hat{\Pi}_\pm$ raise and lower the Landau level index:

$$\hat{\Pi}_-\Phi_{sp}=\sqrt{2s}\Phi_{s-1p},\quad \hat{\Pi}_+\Phi_{sp}=\sqrt{2(s+1)}\Phi_{s+1p}. \tag{9}$$

After omitting the terms compensated for by H_2, we have in the first order[6]

$$S_0^{(1)}=-\frac{1}{2m}\int \mathrm{curl}\,\mathbf{\Omega}^z\frac{d^2rdt}{2\pi}. \tag{10}$$

The terms of second order are expressed as[6]

$$S_0^{(2)}=-\frac{2}{m^2}\,\mathrm{Tr}\left\{\sigma_l\frac{1+\sigma_z}{2}\sigma_{l_1}\hat{g}_1\left(-\frac{E_0}{2}\right)\right\}\times\int \Omega_-^l\Omega_+^{l_1}\frac{d^2rdt}{2\pi}. \tag{11}$$

By expanding $\hat{g}_1(-E_0/2)$ to second order in m and using the identity $\mathrm{curl}\,\mathbf{\Omega}^z=(i/2)\mathrm{Tr}\{\sigma_z\sigma_{l'}\sigma_l\}(\Omega_x^l\Omega_y^{l'}-\Omega_y^l\Omega_x^{l'})$, we have

$$\delta S_0=\delta S_0^{(1)}+\delta S_0^{(2)}=-2E_0\int\frac{d^2rdt}{2\pi}\Omega_-^l\Omega_+^l+2E_1\times\int\frac{d^2rdt}{2\pi}\Omega_-^z\Omega_+^z-2E_0\mathrm{Tr}\frac{\sigma_{l'}\sigma_l\sigma_z}{2}\times\int\Omega_+^{l'}\Omega_-^l\frac{d^2rdt}{2\pi}. \tag{12}$$

The contributions of order $1/m$ cancel out in the sum $S_0^{(1)}+S_0^{(2)}$ [this result differs from that given in Ref. 6 because the sign of the first term on the right of Eq. (12) was wrong].

Thus the contribution (10) of the mean field of order $1/m$ (all the spins are locally directed upwards) is fully cancelled out by fluctuations in the spin alignment in the second-order component of the action. The remaining quantities are of order the exchange energy $e^2/\kappa l_H$, and there are no terms proportional to $\hbar\omega_c$. In order to complete the calculation of the action in the second order in terms of gradients of rotation matrix U, we must calculate the contribution of the second graph in Fig. 1. This diagram is decomposed in Fig. 2, where the the different ways of inserting the Green's function components due to the ground and first Landau levels are shown. The contribution of the last two diagrams in Fig. 2 is equal to zero by virtue of isotropy of $V(\mathbf{r})$. The first two diagrams on the right yield equal contributions:

$$S_{HF}=\frac{(-2)}{m^2}\,2\mathrm{Tr}\left\{\frac{1+\sigma_z}{2}\sigma_l\hat{g}_1^2\left(-\frac{E_0}{2}\right)\sigma_{l'}\right\}\times\int\Omega_+^l\left(\frac{\mathbf{r}+\mathbf{r}'}{2}\right)\Phi_{0p}(\mathbf{r})\Phi_{1p_1}^*(\mathbf{r}')\Phi_{0p_1}^*(\mathbf{r})\Phi_{1p}(\mathbf{r}')\Omega_-^{l'}\times\left(\frac{\mathbf{r}+\mathbf{r}'}{2}\right)V(\mathbf{r}-\mathbf{r}')d^2rd^2r'. \tag{13}$$

Integration with respect to the coordinates is performed fairly easily, and we obtain

$$S_{HF}=2(E_0-E_1)\int\Omega_-^z\Omega_+^z\frac{d^2rdt}{2\pi}+2(E_0-E_1)\times\sum_{l\neq z}\int\Omega_-^l\Omega_+^l\frac{d^2rdt}{2\pi}+2(E_0-E_1)\times\int\mathrm{curl}\,\mathbf{\Omega}^z\frac{d^2rdt}{2\pi}. \tag{14}$$

As a result, we have the action to second order with respect to the gradients of $U(\mathbf{r},t)$:

$$\delta S=-\int \Omega_t^z \frac{d^2rdt}{2\pi}-2E_1\sum_{l\neq z}\int \Omega_-^l\Omega_+^l\frac{d^2rdt}{2\pi}$$

$$-2E_1\mathrm{Tr}\frac{\sigma_{l_1}\sigma_l\sigma_z}{2}\int \Omega_+^{l_1}\Omega_-^l\frac{d^2rdt}{2\pi}. \quad (15)$$

If Ω^l is expressed in terms of Euler's angles, the action for the skyrmion is given by the equation

$$\delta S=-\int \Omega_t^z\frac{d^2rdt}{2\pi}-\frac{E_1}{2}\int\frac{1}{4}\left(\frac{\partial n_i}{\partial x_k}\right)^2\frac{d^2rdt}{2\pi}-\frac{E_1}{2}$$

$$\times\int \mathrm{curl}\,\mathbf{\Omega}^z\frac{d^2rdt}{2\pi}+|g|\mu_B\int(\mathbf{H}\cdot\mathbf{n}-H)\frac{d^2rdt}{2\pi}$$

$$-\frac{1}{2}\int\frac{e^2}{|\mathbf{r}-\mathbf{r}'|}\mathrm{curl}\,\mathbf{\Omega}^z(r)\mathrm{curl}\,\mathbf{\Omega}^z(r')\frac{d^2rd^2r'}{(2\pi)^2}dt, \quad (16)$$

which includes the small contributions of the Zeeman and Coulomb energies and contains the unit vector aligned with the averaged spin:

$$\mathbf{n}=(\sin\beta\cos\alpha,\ \sin\beta\sin\alpha,\ \cos\beta).$$

In calculating the Coulomb energy, we have used the expression for density $\rho=(1+\mathrm{curl}\,\mathbf{\Omega}^z)/2\pi$ from Refs. 1 and 6.

As was shown by Belavin and Polyakov,[7] the minimum gradient energy is expressed in terms of the topological invariant

$$Q=\frac{1}{2\pi}\int \mathrm{curl}\,\mathbf{\Omega}^z d^2r,$$

which corresponds to the degree of mapping of the 2D plane on the unit sphere of spin directions, by the formula

$$\frac{1}{8\pi}\int\left(\frac{\partial n_i}{\partial x_k}\right)^2 d^2r=|Q|.$$

Thus, the energy of a skyrmion with charge Q is given by

$$E=\frac{E_1}{2}(Q+|Q|)+|g|\mu_B\int(\mathbf{H}\cdot\mathbf{n}-H)\frac{d^2r}{2\pi}+\frac{1}{2}$$

$$\times\int\frac{e^2}{|\mathbf{r}-\mathbf{r}'|}\mathrm{curl}\,\mathbf{\Omega}^z(\mathbf{r})\mathrm{curl}\,\mathbf{\Omega}^z(\mathbf{r}')\frac{d^2rd^2r'}{(2\pi)^2}. \quad (17)$$

For skyrmions with $Q<0$ the exchange contribution is zero since the gradient energy is cancelled by the topologically invariant term, so the energy contains only the Coulomb and Zeeman components. As for the skyrmions with $Q>0$, their energy contains an exchange component.

According to Ref. 7, the solution that yields the minimum energy decays with the distance as $r^{-|Q|}$, as a result, the Zeeman energy contains the large logarithmic factor $\ln(|g|\mu_B H/E_1)$ at $|Q|=1$, whereas the energy of skyrmions with $|Q|>1$ has no such factor. For this reason, the skyrmion with $Q=-2$ has a lower energy than that with $Q=-1$. States with lower Q $(Q<-2)$, however, have higher energies owing to the increase in the Coulomb energy with $|Q|$. This fact was pointed out by Nazarov and Khaetski.[8] Note that the suggested expression yields the skyrmion classical energy without quantum corrections.

2. NULL COLLECTIVE MODES

The skyrmion energy is independent of the rotation angle of the spin reference frame with respect to the orbital frame, since adding an arbitrary constant γ to angle α does not change energy (13) (the magnetic field is assumed to be aligned with the z-axis). This degeneracy can be lifted only by a spin–orbit coupling. Thus, $\gamma=\mathrm{const}$ corresponds to a null collective mode.[9] The corresponding motion must have the form of free rotation. In order to analyze this motion, let us use the standard technique applied to null modes,[9] namely, introduce a collective coordinate $\gamma(t)$ and calculate the action through expansion in terms of the angular velocity $\dot{\gamma}$.

By considering only terms that are not total derivatives, we obtain in the linear approximation

$$S_1=-\int\Omega_t^z\frac{d^2rdt}{2\pi}=\frac{1}{2}\int\dot{\gamma}\beta_1\sin\beta_0\frac{d^2rdt}{2\pi},$$

where β_0 is the minimizing stationary solution,[7] and β_1 is the small correction to the stationary solution. The term quadratic in $\dot{\gamma}$ in the action is easily derived from the second-order term $S_2=(i/2)\mathrm{Tr}\,\Omega_l^t\sigma_l G_0\Omega_t^k\sigma_k G_0$. In the lowest order in m, we can restrict our analysis to the lowest Landau level and obtain

$$S_2=\int\frac{I\dot{\gamma}^2}{2}dt,$$

where the ''inertia'' moment is given by

$$I=\frac{1}{2E_0}\int\sin^2\beta_0\frac{d^2r}{2\pi},$$

and we have neglected in the calculation the deviation of β from its time-independent value. The total action has the form

$$S=S(\alpha_1,\beta_1)+\int\frac{I\dot{\gamma}^2}{2}dt+\frac{1}{2}\int\dot{\gamma}\beta_1\sin\beta_0\frac{d^2rdt}{2\pi},$$

where the first term is the action as a quadratic functional of deviations α_1 and β_1 from the time-independent values of α_0 and β_0 without taking into account null modes. This action is independent of angle γ and is expressed in the mixed representation, which is the Hamiltonian in α_1 and β_1, and Lagrangian in γ. The transition to the fully Hamiltonian representation is performed by introducing the conserved angular moment

$$M=I\dot{\gamma}+\frac{1}{2}\int\frac{\beta_1}{2\pi}\sin\beta_0 d^2r,$$

which yields the Hamiltonian action

$$S=S(\alpha_1,\beta_1)+\int M\dot{\gamma}dt-\frac{1}{2I}\times\int\left(M-\frac{1}{2}\int\beta_1\sin\beta_0\right)^2dt. \tag{18}$$

The last term describes interaction between the null rotational mode and collective modes of higher orders.

In the presence of spin–orbit coupling, the action depends on the rotation angles of the spin reference frame and the orbital frame, so the degeneracy is lifted. We do not calculate the spin–orbit term in the action in terms of the microscopic theory, but restrict our analysis to the phenomenological approach, which yields an expression in the lowest order in the gradients:

$$S_{\rm so}=-\lambda_{\rm so}\int\mathbf{n}\cdot\boldsymbol{\nu}\,\mathrm{div}\,\mathbf{n}\frac{d^2rdt}{2\pi}, \tag{19}$$

where $\mathbf{n}$ is the vector of the average spin direction, $\boldsymbol{\nu}$ is the normal to the 2D plane. We assume that there is a certain asymmetry that defines a unique direction of $\boldsymbol{\nu}$. This expression is T-invariant and isotropic in the 2D plane. It can be shown that the constant $\lambda_{\rm so}$ is related to the bare spin–orbit coupling constant: $\lambda_{\rm so}=\lambda_0E_1/\hbar\omega_c$. The bare constant λ_0 is included in the spin–orbit Hamiltonian[10] $H_{\rm so}=\lambda_0\epsilon^{zjk}p_j\sigma_k$, where p_j is the electron momentum and ϵ^{zjk} is the absolutely antisymmetrical tensor. The parameter $S_{\rm so}$ can be expressed in terms of Euler angles:

$$S_{\rm so}=-\lambda_{\rm so}\int\frac{d^2rdt}{2\pi}\sin^2\beta(\cos\alpha\partial_x\beta+\sin\alpha\partial_y\beta). \tag{20}$$

Owing to the smallness of the spin–orbit coupling, we can substitute in the first approximation the steady-state solution[7] $\beta=\beta_0(\mathbf{r})$, $\alpha=Q\varphi+\gamma$, where Q is integer and φ is the polar angle) in the expression for the spin–orbit energy, which yields

$$E_{\rm so}=\lambda_{\rm so}\int\sin^2\beta_0\cos[(Q-1)\varphi+\gamma]\frac{d^2r}{2\pi}.$$

For $Q\neq1$ this energy vanishes after integration with respect to the angle, and the system remains degenerate in the first order in $\lambda_{\rm so}$. Let us consider the case of $Q=1$, when the energy has a minimum at $\gamma=\pi$ if $\lambda_{\rm so}d\beta_0/dr>0$ holds or at $\gamma=0$ if $\lambda_{\rm so}d\beta_0/dr<0$ holds. In this case, a specific mode is generated with oscillating γ and variable skyrmion core dimension.

In order to determine characteristic frequencies of such a radial-angular mode, let us use equations of motion obtained by varying the action and linearized in the neighborhood of the minimum solution:[7]

$$-i\omega\alpha_1\sin\beta_0+J[\Delta\beta_1+\sin^2\beta_0\nabla\alpha_0\nabla\alpha_1+\sin2\beta_0(\nabla\alpha_0)^2\beta_1]+\frac{\delta^2E_{ZC}}{(\delta\beta)^2}\beta_1=0, \tag{21}$$

$$i\omega\beta_1\sin\beta_0+J\nabla[\sin^2\beta_0\nabla\alpha_1+\beta_1\sin2\beta_0\nabla\alpha_0]+\lambda_{\rm so}\sin^2\beta_0\frac{d\beta_0}{dr}\alpha_1=0. \tag{22}$$

The factor $\delta^2E_{ZC}/(\delta\beta)^2$ denotes the second derivative with respect to β of the sum of the Zeeman and Coulomb energies at the steady-state value. These equations contain the large parameter $J=E_1/4$, but the operators proportional to J have nontrivial solution that cause them to go to zero: $\alpha_1=\gamma=\mathrm{const}$, $\beta_1=r\partial_r\beta_0$. Therefore, low-frequency oscillations are determined by solutions of this form, and one must consider, as usual, projections of the equations on the corresponding eigenfunctions. The first equation is to be be multiplied by unity, the second by $rd\beta_0/dr$, and each is integrated over the coordinates. As a result, the terms containing J vanish, and we obtain a system of linear equations

$$i\omega c_1\int\sin\beta_0\left(r\frac{\partial\beta_0}{\partial r}\right)d^2r=\lambda_{\rm so}\int\sin^2\beta_0\frac{\partial\beta_0}{\partial r}d^2r\gamma, \tag{23}$$

and

$$-i\omega\gamma\int\sin\beta_0\left(r\frac{\partial\beta_0}{\partial r}\right)d^2r=c_1\int\left(r\frac{\partial\beta_0}{\partial r}\right)^2\frac{\delta^2E_{cZ}}{(\delta\beta)^2}d^2r. \tag{24}$$

By equating to zero the determinant of this equation system, we obtain the characteristic frequency:

$$\hbar^2\omega^2=\frac{|\lambda_{\rm so}|\int\sin^2\beta_0\frac{\partial\beta_0}{\partial r}\frac{d^2r}{2\pi}\int\frac{d^2r}{2\pi}\left(r\frac{\partial\beta_0}{\partial r}\right)^2\frac{\delta^2E_{ZC}}{(\delta\beta)^2}}{\left(\int\sin\beta_0r\frac{\partial\beta_0}{\partial r}dr\right)^2}\sim\frac{\lambda_{\rm so}g\mu_BH}{L_c\ln\frac{L^*}{L_c}}. \tag{25}$$

Here $L_*=\sqrt{|g|\mu_BH/J}$ is the distance at which the Zeeman and gradient energies of the skyrmion are equal, and the skyrmion size is

$$L_c=\left(\frac{|g|\mu_BH}{J}\ln\frac{L_*}{L_c}\right)^{1/3}$$

(see, for example, Ref. 12). Here the lengths are measured in units of magnetic length l_H. Such oscillations change concurrently the skyrmion size L_c and angle γ between the spin and orbital axes.

There is also a collective mode described earlier[12] in which a skyrmion is translated as a whole, when the rotation matrix $U(\mathbf{r}-\mathbf{X})$ is a function of the skyrmion center position $\mathbf{X}(\mathbf{t})$. As in the zero rotational mode, the action contains a term linear in velocity $\dot{X}$ and describing interaction with other collective oscillations, so the action is expressed by

$$\delta S=\int\dot{\mathbf{X}}(\cos\beta_0\nabla\alpha_1-\beta_1\sin\beta_0\nabla\alpha_0)\frac{d^2rdt}{2\pi}+\int\frac{m_s(\dot{\mathbf{X}})^2}{2}dt+S(\alpha_1,\beta_1).$$

The skyrmion mass $m_s=|Q|\hbar^2/E_0l_H^2$ was calculated earlier.[12] Since the skyrmion has charge eQ, there is an additional term in the action due to the vector potential of the

external magnetic field. After going over to the Hamiltonian expression for the action, as in the derivation of Eq. (20), and taking into account the other collective modes, we obtain

$$S=\int M\dot{\gamma}dt+\int \mathbf{P}\dot{\mathbf{X}}dt-\frac{1}{2I}\int (M-l)^2dt$$
$$-\frac{1}{2m_s}\int\left(\mathbf{P}-\frac{eQ}{c}\mathbf{A}_0-\mathbf{a}\right)^2dt+S(\alpha_1,\beta_1),$$

where

$$l=\hbar\int \beta_1\sin\beta_0\frac{d^2r}{2\pi l_H^2},$$

$$\mathbf{a}=\hbar\int(\cos\beta_0\nabla\alpha_1-\beta_1\sin\beta_0\nabla\alpha_0)\frac{d^2r}{2\pi l_H^2},$$

$$m_s=|Q|\,\hbar/E_0l_H^2,\quad I=\frac{\hbar^2}{2E_0}\int\sin^2\beta_0\frac{d^2r}{2\pi l_H^2}.$$

The work was supported by the USA CRDF (Grant 452, RPI-273) and the Russian Fund for Fundamental Research (Project 98-02-16245), and RFBR–INTAS (Grant 950675).

[1] S. Sondhi, A. Kahlrede, S. Kivelson, and E. Rezai, Phys. Rev. B **47**, 16418 (1993).
[2] H. Fertig, L. Brey, R. Cote, and A. H. MacDonald, Phys. Rev. B **50**, 11018 (1994).
[3] K. Moon, H. Mori, Kun Yang, S. Girvin, and A. MacDonald, Phys. Rev. B **51**, 5138 (1995).
[4] Yu. Bychkov, T. Maniv, and I. Vagner, Phys. Rev. B **53**, 10148 (1995).
[5] S. Iordanskiĭ and S. Plyasunov, JETP Lett. **65**, 259 (1997).
[6] S. Iordanskiĭ and S. Plyasunov, Zh. Éksp. Teor. Fiz. **112**, 1899[JETP **85**, 1039 (1997)]; cond-mat/9706236.
[7] A. A. Belavin and A. M. Polyakov, JETP Lett. **22**, 245 (1975).
[8] Yu. Nazarov and A. Khaetski, Phys. Rev. Lett. **80**, 576 (1998).
[9] R. Radjaraman, *Solitons and Instantons. An Introduction to Solitons and Instantons in Quantum Field Theory*, North Holland, Amsterdam (1985).
[10] Yu. Bychkov and É. Rashba, JETP Lett. **39**, 78 (1984).
[11] S. Iordanski, J. Phys.: Condens. Matter **10**, L247 (1998).
[12] S. Iordanskiĭ, JETP Lett. **66**, 459 (1997).

Translation provided by the Russian Editorial office.

Generalized Ginzburg–Landau equation and the properties of superconductors with Ginzburg–Landau parameter κ close to 1

Yu. N. Ovchinnikov*)

L. D. Landau Institute of Theoretical Physics, Russian Academy of Sciences, 117940 Moscow, Russia
(Submitted 10 September 1998)
Zh. Éksp. Teor. Fiz. **115**, 726–739 (February 1999)

An expression is derived for the free energy of a superconductor near the critical temperature, taking account of the terms of next highest order in the parameter $1-T/T_c$. These terms become important for Ginzburg–Landau parameter values $|\kappa-1|\ll 1$, and in this case, in an external magnetic field H_0 close to H_{c2}, the structure of the order parameter is determined by the relative values of the three small parameters $|\kappa-1|$, $1-T/T_c$, and $(H_{c2}-H_0)/H_{c2}$. Three types of lattices are investigated: triangular with one and two flux quanta per cell and square with one flux quantum per cell. © *1999 American Institute of Physics.* [S1063-7761(99)02402-6]

1. INTRODUCTION

The Ginzburg–Landau equations[1] have the important property that for Ginzburg–Landau parameter $\kappa=1$ there exists in a cylindrical geometry a N-vortex solution with an arbitrary position of the zeros of the order parameter in a plane perpendicular to the axis of the cylinder.[2,3] As a result of this degeneracy, all coefficients (starting with the second one) in the Taylor series expansion of the Ginzburg–Landau free energy in powers of $H_{c2}-B$, where B is the magnetic field induction and H_{c2} is the critical field of the superconductor, vanish as $\kappa\to 1$.[1]

Naturally, even a small perturbation of the Ginzburg–Landau equations will completely destroy this enormous degeneracy. In real superconductors, the perturbing terms are proportional to $1-T/T_c$, where T_c is the superconducting transition temperature. As a result, in magnetic fields H_0 close to H_{c2}, $(H_{c2}-H_0)/H_{c2}\ll 1$, three small parameters arise: $|\kappa-1|$,$1-T/T_c$, and $(H_{c2}-B)/H_{c2}$. Depending on the relative value of the parameters $|\kappa-1|$ and $1-T/T_c$, the free energy $(F_S-F_N)/V$ of the mixed state as a function of the parameter $(H_{c2}-B)/H_{c2}$ can have the form of one of the three plots shown in Fig. 1 ($H_0<H_{c2}$) or Fig. 2 ($H_0>H_{c2}$).

In magnetic fields, $H_0<H_{c2}$, in the (case a) vortex-free state or a state with $B\sim H_{c2}$ forms. The case b corresponds to a classical type-II superconductor, and in the case c metastable vortex lattices can form. As will be shown below, all three cases can be realized. The free energy as a function of the parameter $(H_{c2}-B)/H_{c2}$ in magnetic fields $H_0>H_{c2}$ is displayed in Fig. 2, and all three possibilities (a,b,c) can obtain. It is obvious that the free energy depends on the form of the vortex lattice. We shall examine below three types of vortex lattices: triangular with one or two flux quanta per unit cell and square with one flux quantum per unit cell.

2. GINZBURG–LANDAU EQUATION WITH CORRECTION TERMS

To obtain expressions for the free energy of a superconductor taking account of the terms of next higher order in the parameter $1-T/T_c$, we shall make use of the system of equations for the Green's functions $\hat{G}$ integrated over the energy variable ξ.[4–6] These equations have the form

$$\left(\mathbf{v}\cdot\frac{\partial}{\partial\mathbf{r}}\right)\hat{G}+\hat{\omega}\hat{G}-\hat{G}\omega=0,$$

$$\hat{G}=\begin{pmatrix}\alpha, & -i\beta_{\mathbf{p}}\\ i\tilde{\beta}_{\mathbf{p}}, & -\alpha\end{pmatrix},\quad \hat{G}^2=\hat{1}, \tag{1}$$

where

$$\hat{\omega}=\omega\tau_z-ie(\mathbf{v}\cdot\mathbf{A})\tau_z-i\hat{\Delta}-in\hat{\Sigma}_{\mathbf{pp}},$$

$$\hat{\Delta}=\begin{pmatrix}0 & \Delta\\ -\Delta^* & 0\end{pmatrix},\quad \hat{\Sigma}_{\mathbf{pp}}=-\frac{iv}{2}\int\sigma_{\mathbf{pp_1}}\hat{G}_{\mathbf{p_1}}d\Omega_{\mathbf{p_1}}. \tag{2}$$

In Eq. (2) Δ is the order parameter of the superconductor, $\sigma_{\mathbf{pp}'}$ is the electron scattering cross section of an impurity, and n is the impurity density. The order parameter Δ is related to the Green's function $\beta_{\mathbf{p}}$ by

$$\Delta\ln\left(\frac{T_c}{T}\right)=2\pi T\sum_{\omega>0}\left(\frac{\Delta}{\omega}-\int\frac{d\Omega_{\mathbf{p}}}{4\pi}\beta_{\mathbf{p}}\right), \tag{3}$$

where T_c is the superconducting transition temperature. We seek only the first few corrections to the Ginzburg–Landau equations in the parameter $1-T/T_c$. The equation $\hat{G}^2=1$ implies

$$\alpha^2+\beta_{\mathbf{p}}\tilde{\beta}_{\mathbf{p}}=1. \tag{4}$$

Solving this equation up to fourth-order terms in Δ we find

$$\alpha_{\mathbf{p}}=\mathrm{sing}\,\omega\left[1-\frac{\beta_{\mathbf{p}}\tilde{\beta}_{\mathbf{p}}}{2}-\frac{(\beta_{\mathbf{p}}\tilde{\beta}_{\mathbf{p}})^2}{8}\right]. \tag{5}$$

We find from the system of equations (1)

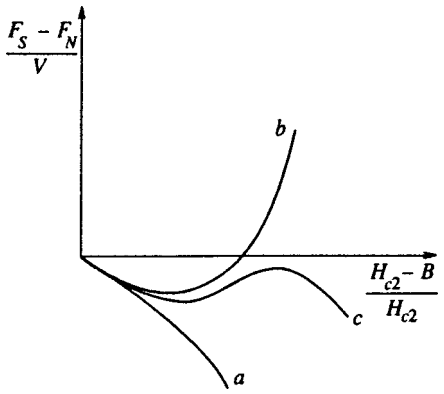

FIG. 1. Free energy $(F_S-F_N)/V$ versus the parameter $(H_{c2}-B)/H_{c2}$ in the ranges $x\ll 1$ and $H_0<H_{c2}$.

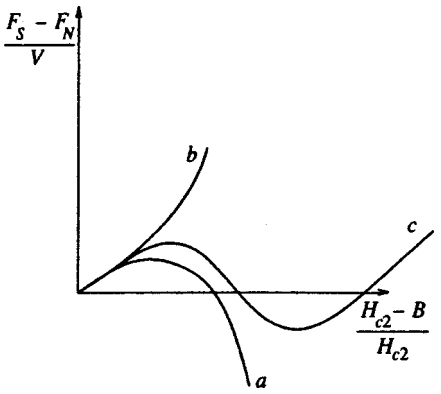

FIG. 2. Free energy $(F_S-F_N)/V$ versus the parameter $(H_{c2}-B)/H_{c2}$ in the ranges $x\ll 1$ and $H_0>H_{c2}$.

$$\frac{1}{2}(\mathbf{v}\cdot\boldsymbol{\partial}_-)\beta_{\mathbf{p}}+\omega\beta_{\mathbf{p}}-\Delta\alpha_{\mathbf{p}}+\frac{nv}{2}\int\sigma_{\mathbf{pp_1}}(\beta_{\mathbf{p}}\alpha_{\mathbf{p_1}}-\alpha_{\mathbf{p}}\beta_{\mathbf{p_1}})d\Omega_{\mathbf{p_1}}=0,$$

$$-\frac{1}{2}(\mathbf{v}\cdot\boldsymbol{\partial}_+)\tilde{\beta}_{\mathbf{p}}+\omega\tilde{\beta}_{\mathbf{p}}-\Delta^*\alpha_{\mathbf{p}}+\frac{nv}{2}\int\sigma_{\mathbf{pp_1}}(\tilde{\beta}_{\mathbf{p}}\alpha_{\mathbf{p_1}}-\alpha_{\mathbf{p}}\tilde{\beta}_{\mathbf{p_1}})d\Omega_{\mathbf{p_1}}=0, \tag{6}$$

where

$$\boldsymbol{\partial}_\pm=\frac{\partial}{\partial\mathbf{r}}\pm 2ie\mathbf{A}.$$

The system of equations (5) and (6) can be solved by expanding the quantities $\beta_{\mathbf{p}}$ and $\tilde{\beta}_{\mathbf{p}}$ in spherical harmonics with arbitrary electron mean-free path. Simple calculations using Eqs. (5) and (6) reduce Eq. (3) for the order parameter to the form

$$-\Delta\ln\left(\frac{T_c}{T}\right)-2\pi T\times\sum_{\omega>0}\left\{-\frac{\Delta|\Delta|^2}{2\omega^3}+\frac{3\Delta|\Delta|^4}{8\omega^5}+\frac{v^2}{12}\frac{\partial_-^2\Delta}{\omega^2(\omega+1/2\tau_{\rm tr})}+\frac{v^4(1/\tau-1/\tau_2)(\partial_-^2)^2\Delta}{288\omega^3(\omega+1/2\tau_{\rm tr})^2(\omega+(1/2)(1/\tau-1/\tau_2))}\right.$$
$$+\frac{v^4((\partial_-^2)^2+4e^2H^2-(4ie/3)(\boldsymbol{\nabla}\times\mathbf{H}\cdot\boldsymbol{\partial}_-))\Delta}{80\omega^2(\omega+1/2\tau_{\rm tr})^2(\omega+(1/2)(1/\tau-1/\tau_2))}-\frac{v^2\partial_-^2(\Delta|\Delta|^2)}{24\omega^4(\omega+1/2\tau_{\rm tr})}+v^2\Delta((\boldsymbol{\partial}_-\Delta)\cdot\boldsymbol{\partial}_+\Delta^*)$$
$$\left.+(\partial/\partial\mathbf{r})\cdot(\Delta\boldsymbol{\partial}_+\Delta^*)\right)48\omega^4\tau_{\rm tr}(\omega+1/2\tau_{\rm tr})^2+\frac{v^2\Delta((\partial_-\Delta)\cdot(\partial_+\Delta^*))}{24\omega^3(\omega+1/2\tau_{\rm tr})^2}-\frac{v^2(|\Delta|^2\partial_-^2\Delta+\Delta^2\partial_+^2\Delta^*)}{24\omega^4(\omega+1/2\tau_{\rm tr})}$$
$$+\frac{v^2\partial_-\cdot(\Delta(\Delta\partial_+\Delta^*-\Delta^*\partial_-\Delta))}{24\omega^3(\omega+1/2\tau_{\rm tr})^2}=0. \tag{7}$$

The scattering ''time'' τ_N in Eq. (6) is determined in the standard manner as

$$\frac{1}{\tau_N}Y_N\frac{\mathbf{P}}{|P|}=nv\int\sigma_{\mathbf{pp_1}}Y_N\frac{\mathbf{P}_1}{|P|}d\Omega_{\mathbf{p_1}},\quad\frac{1}{\tau_{\rm tr}}=\frac{1}{\tau}-\frac{1}{\tau_1}. \tag{8}$$

In accordance with the general assumptions, Eq. (6) is a variational derivative of the free energy with respect to the parameter Δ^*. Direct calculations give the following expression for the free energy F_S:

$$F_S-F_N=\nu\int d^3r\left\{-\ln\left(\frac{T_c}{T}\right)|\Delta|^2-2\pi T\sum_{\omega>0}\left[-\frac{|\Delta|^4}{4\omega^3}+\frac{|\Delta|^6}{8\omega^5}-\frac{v^2}{12\omega^2}\frac{|\partial_-\Delta|^2}{(\omega+1/2\tau_{\rm tr})}+\frac{v^4(1/\tau-1/\tau_2)|\partial_-^2\Delta|^2}{288\omega^3(\omega+1/2\tau_{\rm tr})^2(\omega+(1/2)(1/\tau-1/\tau_2))}\right.\right.$$
$$+\frac{v^4(|\partial_-^2\Delta|^2+4e^2H^2|\Delta|^2-(2ie/3)\boldsymbol{\nabla}\times\mathbf{H}\cdot(\Delta^*\partial_-\Delta-\Delta\partial_+\Delta^*))}{80\omega^2(\omega+1/2\tau_{\rm tr})^2(\omega+(1/2)(1/\tau-1/\tau_2))}+\frac{v^2}{24\omega^3(\omega+1/2\tau_{\rm tr})^2}(|\Delta|^2(\partial_-\Delta)\cdot(\partial_+\Delta^*))$$
$$\left.\left.+\frac{v^2}{24\omega^4(\omega+1/2\tau_{\rm tr})}\left(\frac{1}{2}\left(\frac{\partial|\Delta|^2}{\partial r}\right)^2+2|\Delta|^2(\partial_-\Delta)\cdot(\partial_+\Delta^*)\right)\right]\right\}+\frac{1}{8\pi}\int d^3r(\boldsymbol{\nabla}\times\mathbf{A}-\mathbf{H}_0)^2. \tag{9}$$

The last term in Eq. (9) is the magnetic-field energy and $\nu = mp_0/2\pi^2$ is the density of states at the Fermi surface. This coefficient is found by setting the standard expression for the current density equal to the corresponding expression obtained from Eq. (9) by variation with respect to the vector potential $\mathbf{A}$. Performing the summation over the frequency ω in Eq. (9), we obtain finally

$$
\begin{aligned}
F_S - F_N = \nu \int d^3r \Bigg\{ & -\ln\left(\frac{T_c}{T}\right)|\Delta|^2 + \frac{7\zeta(3)}{16\pi^2 T^2}|\Delta|^4 \\
& + \frac{\pi D}{8T}|\partial_-\Delta|^2 - \frac{31\zeta(5)|\Delta|^6}{128\pi^4 T^4} - \frac{v^4(1/\tau - 1/\tau_2)}{288} \\
& \times a_1|\partial_-^2\Delta|^2 - \frac{v^2}{24}a_2|\Delta|^2|\partial_-\Delta|^2 - \frac{v^2}{24}b_2 \\
& \times\left(\frac{1}{2}\left(\frac{\partial|\Delta|^2}{\partial\mathbf{r}}\right)^2 + 2|\Delta|^2|\partial_-\Delta|^2\right) - \frac{v^4}{80}b_1 \\
& \times\Bigg(|\partial_-^2\Delta|^2 + 4e^2H^2|\Delta|^2 - \frac{2ie}{3}\boldsymbol{\nabla}\times\mathbf{H}\cdot(\Delta^*\partial_-\Delta \\
& - \Delta\partial_+\Delta^*)\Bigg\} + \frac{1}{8\pi}\int d^3r(\boldsymbol{\nabla}\times\mathbf{A} - \mathbf{H}_0)^2. \qquad (10)
\end{aligned}
$$

In Eq. (10) H_0 is the external magnetic field and $\zeta(x)$ is the Riemann zeta function. Expressions for the coefficients $a_{1,2}, b_{1,2}$, and D are given in Appendix A.

We note that the expression (10) for the free energy of a superconductor holds near T_c taking account of the first corrections with respect to the parameter $1 - T/T_c$ and for arbitrary values of all other parameters: the magnetic field, the mean-free path, the Ginzburg–Landau parameter κ, and so on.

3. SUPERCONDUCTOR IN A MAGNETIC FIELD H_0 CLOSE TO H_{c2}

The structure of the superconducting state of a superconductor in an external magnetic field H_0 close to H_{c2} was first obtained by Abrikosov.[7] However, the situation arising for values of the parameter κ in the limit $|\kappa - 1| \ll 1$ is much more interesting and diverse (see also Refs. 8 and 9).

We seek the order parameter Δ and the vector potential $\mathbf{A}$ in the range of fields $|H_{c2} - H_0| \ll H_{c2}$ in the form of a power series in $(H_{c2} - B)$ (Refs. 8 and 9):

$$
\begin{aligned}
& \Delta = \Delta_0 + \Delta_1 + \Delta_2 + \ldots, \\
& \mathbf{A} = (0, Bx, 0) + \mathbf{A}_1 + \mathbf{A}_2. \qquad (11)
\end{aligned}
$$

In the gauge $\boldsymbol{\nabla}\cdot\mathbf{A} = 0$ the quantities $\mathbf{A}_{1,2,\ldots}(x,y)$ are bounded functions in the (x,y) plane, and we represent the order parameter $\Delta_{0,1,\ldots}$ in the form[7–9]

$$
\begin{aligned}
& \Delta_0 = \sum_N C_N \exp\{2ieBNx_1 y - eB(x - Nx_1)^2\}, \\
& \Delta_1 = \sum_{M=1}^{\infty} \alpha_M \Delta_1^M, \quad \langle \Delta_0 \Delta_M^* \rangle_{M=1,2,\ldots} = 0, \qquad (12)
\end{aligned}
$$

$$
\Delta_1^M = \sum_{N=-\infty}^{\infty} C_N \exp(-2ieBNx_1 y) D_M(2\sqrt{eB}(x - Nx_1)).
$$

In Eq. (12) the $D_M(x)$ are parabolic cylinder functions.

Using Eqs. (10) and (11) we find the critical field H_{c2} and the equation for the first correction $\mathbf{A}_1$ to the vector potential $\mathbf{A}$:

$$
\begin{aligned}
& \ln\left(\frac{T_c}{T}\right) - \frac{\pi eDH_{c2}}{4T} + \frac{e^2H_{c2}^2v^4}{8}\left(\frac{a_1}{9}\left(\frac{1}{\tau} - \frac{1}{\tau_2}\right) + \frac{4b_1}{5}\right) = 0, \\
& \frac{1}{4\pi}\frac{\partial^2\mathbf{A}_1}{\partial\mathbf{r}^2} = \nu\Bigg\{\frac{e\pi D}{4T}\boldsymbol{\nabla}\times(0,0,|\Delta_0|^2) - \frac{ev^4}{120}b_1\boldsymbol{\nabla} \\
& \qquad \times\left(0,0,\frac{\partial^2|\Delta_0|^2}{\partial\mathbf{r}^2}\right) - e^2v^4B\Bigg(\frac{1}{36}\left(\frac{1}{\tau} - \frac{1}{\tau_2}\right)a_1 \\
& \qquad + \frac{1}{5}b_1\Bigg)\boldsymbol{\nabla}\times(0,0,|\Delta_0|^2)\Bigg\}. \qquad (13)
\end{aligned}
$$

To obtain Eq. (13) for the vector potential $\mathbf{A}_1$, we employed the relation

$$
\Delta_0^*\partial_-^0\Delta_0 - \Delta_0\partial_+^0\Delta_0^* = -i\boldsymbol{\nabla}\times(0,0,|\Delta_0|^2), \qquad (14)
$$

where $\boldsymbol{\partial}_-^0 = \partial/\partial\mathbf{r} - 2ie\mathbf{A}_0$.

The equation (14) follows from the expression (12) for Δ_0. Using the gauge $\boldsymbol{\nabla}\cdot\mathbf{A}_1 = 0$, we find from the second equation (13) the following expression for the magnetic field $\mathbf{H}_1$:

$$
\begin{aligned}
& \mathbf{H}_i = \boldsymbol{\nabla}\times\mathbf{A}_i, \quad i = 1,2,\ldots \\
& \mathbf{H}_1 = -\pi\nu(0,0,1)\Bigg\{\frac{\pi eD}{T} - e^2v^4B\left(\frac{a_1}{9}\left(\frac{1}{\tau} - \frac{1}{\tau_2}\right) + \frac{4}{5}b_1\right) \\
& \qquad \times(|\Delta_0|^2 - \langle|\Delta_0|^2\rangle) - \frac{ev^4}{30}b_1\frac{\partial^2}{\partial\mathbf{r}^2}|\Delta_0|^2\Bigg\}. \qquad (15)
\end{aligned}
$$

In the expression (10) for the free energy, the terms containing H_2 cancel up to terms of order $|\Delta_0|^6$, inclusively. To check this assertion, we employ the following relations:

$$
\begin{aligned}
\langle|\partial_-\Delta|^2\rangle = {} & 2eB\langle|\Delta_0|^2\rangle + 2e\langle(H_1 + H_2)|\Delta_0|^2\rangle \\
& + 4e^2\langle A_1^2|\Delta_0|^2\rangle + 4ie\langle\mathbf{A}_1\cdot(\Delta_0^*\partial_-^0\Delta_1 - \Delta_0\partial_+^0\Delta_1^*)\rangle \\
& - \langle\Delta_1(\partial_+^0)^2\Delta_1^*\rangle, \\
\langle|\partial_-^2\Delta|^2\rangle = {} & 4e^2B^2\langle|\Delta_0|^2\rangle + 8ieB\langle\mathbf{A}_1\cdot(\Delta_0^*\partial_-^0\Delta_0 \\
& - \Delta_0\partial_+^0\Delta_0^*)\rangle + 8ie^2B\langle\mathbf{A}_1\cdot(\Delta_0^*\partial_-^0\Delta_1 - \Delta_0\partial_+^0\Delta_1^*) \\
& + 8e^2B\langle H_2|\Delta_0|^2\rangle + 16e^3B\langle\mathbf{A}_1^2|\Delta_0|^2\rangle \\
& + \langle((\partial_-^0)^2\Delta_1)((\partial_+^0)^2\Delta_1^*)\rangle + 4ie\langle(\mathbf{A}_1\cdot\partial_+^0\Delta_0^*) \\
& \times((\partial_-^0)^2\Delta_1) - (\mathbf{A}_1\cdot\partial_-^0\Delta_0)((\partial_+^0)^2\Delta_1^*)\rangle \\
& + 16e^2\langle(\mathbf{A}_1\cdot\partial_-^0\Delta_0)(\mathbf{A}_1\cdot\boldsymbol{\partial}_+^0\Delta_0^*)\rangle, \qquad (16)
\end{aligned}
$$

$$\langle H^2|\Delta|^2\rangle = B^2(\langle|\Delta_0|^2\rangle+\langle|\Delta_1|^2\rangle)+\langle H_1^2|\Delta_0|^2\rangle+ +2B\langle H_1(|\Delta_0|^2+(\Delta_0^*\Delta_1+\Delta_0\Delta_1^*))\rangle +2B\langle H_2|\Delta_0|^2\rangle,\langle\nabla\times\mathbf{H}\cdot(\Delta^*\partial_-\Delta-\Delta\partial_+\Delta^*)\rangle =i\left\langle \mathbf{H}_1\frac{\partial^2|\Delta_0|^2}{\partial\mathbf{r}^2}\right\rangle+i\left\langle \mathbf{H}_2\frac{\partial^2|\Delta_0|^2}{\partial\mathbf{r}^2}\right\rangle+\langle\nabla\times\mathbf{H}_1\cdot (-ie\mathbf{A}_1|\Delta_0|^2+2(\Delta_0^*\partial_-^0\Delta_1-\Delta_0\partial_+^0\Delta_1^*))\rangle.$$

Subtracting from Eq. (10) the equation (13) for H_{c2} multiplied by $\nu\int d^3r(|\Delta_0|^2+|\Delta_1|^2)$ and using Eqs. (15) and (16), we reduce the expression (10) for the free energy to the form

$$F_S-F_N=\frac{1}{8\pi}\int d^3r((B-H_0)^2-H_1^2)+\nu\int d^3r\Bigg\{-\frac{\pi eD}{4T}(H_{c2}-B)|\Delta_0|^2+e^2v^4(H_{c2}^2-B^2)|\Delta_0|^2\left(\frac{a_1}{72}\left(\frac{1}{\tau}-\frac{1}{\tau_2}\right)+\frac{b_1}{10}\right)$$
$$+\frac{7\zeta(3)}{16\pi^2T^2}(|\Delta_0|^4+2|\Delta_0|^2(\Delta_0^*\Delta_1+\Delta_1^*\Delta_0))-\frac{31\zeta(5)}{128\pi^4T^4}|\Delta_0|^6+\frac{\pi D}{8T}[4e^2\mathbf{A}_1^2|\Delta_0|^2+4ie\mathbf{A}_1\cdot(\Delta_0^*\partial_-^0\Delta_1-\Delta_0\partial_+^0\Delta_1^*)$$
$$-\Delta_1^*(\partial_-^0)^2\Delta_1-2eH_{c2}|\Delta_1|^2]-v^4\left(\frac{a_1}{288}\left(\frac{1}{\tau}-\frac{1}{\tau_2}\right)+\frac{b_1}{80}\right)[8ie^2B\mathbf{A}_1\cdot(\Delta_0^*\partial_-^0\Delta_1-\Delta_0\partial_+^0\Delta_1^*)+16e^3B\mathbf{A}_1^2|\Delta_0|^2$$
$$+((\partial_-^0)^2\Delta_1)((\partial_+^0)^2\Delta_1^*)+4ie((\mathbf{A}_1\cdot\partial_+^0\Delta_0^*)(\partial_-^0)^2\Delta_1-(\mathbf{A}_1\cdot\partial_-^0\Delta_0)(\partial_+^0)^2\Delta_1^*)+16e^2(\mathbf{A}_1\cdot\partial_-^0\Delta_0)(\mathbf{A}_1\cdot\partial_+^0\Delta_0^*)]$$
$$-\frac{v^2}{48}b_2\left(\left(\frac{\partial|\Delta_0|^2}{\partial\mathbf{r}}\right)^2+2\frac{\partial|\Delta_0|^2}{\partial\mathbf{r}}\cdot\frac{\partial}{\partial\mathbf{r}}(\Delta_0\Delta_1^*+\Delta_0^*\Delta_1)\right)-\frac{v^2}{24}(a_2+2b_2)[|\Delta_0|^2((\partial_-^0\Delta_0)\cdot(\partial_+^0\Delta_0^*))+2e\mathbf{A}_1\cdot|\Delta_0|^2\nabla$$
$$\times(0,0,|\Delta_0|^2)+2eB|\Delta_0|^2(\Delta_0^*\Delta_1+\Delta_0\Delta_1^*)-(\Delta_0^*\Delta_1^*(\partial_-^0\Delta_0)^2+\Delta_0\Delta_1(\partial_+\Delta_0^*)^2)]+e^2v^4H_{c2}^2\left(\frac{a_1}{72}\left(\frac{1}{\tau}-\frac{1}{\tau_2}\right)+\frac{b_1}{10}\right)|\Delta_1|^2$$
$$-\frac{e^2v^4}{20}b_1[B^2|\Delta_1|^2+H_1^2|\Delta_0|^2+2BH_1(\Delta_0\Delta_1^*+\Delta_0^*\Delta_1)]+\frac{ev^4}{60}b_1\nabla\times\mathbf{H}_1\cdot(2e\mathbf{A}_1|\Delta_0|^2+i(\Delta_0^*\partial_-^0\Delta_1-\Delta_0\partial_+^0\Delta_1^*))\Bigg\}. \tag{17}$$

In what follows we shall study solutions of the form (11), which form in the (x,y) plane for $|\Delta|^2$ lattices with unit-cell vectors $\mathbf{a}_{1,2}$.[8,9] Let $\mathbf{K}_{1,2}$ be the elementary reciprocal-lattice vectors. We shall represent the order parameter $|\Delta_0|^2$ in this case in the form of a Fourier series as

$$|\Delta_0|^2=\sum_{N,M=-\infty}^{\infty}C_{NM}\exp(i(N\mathbf{K}_1+M\mathbf{K}_2)\cdot\mathbf{r}). \tag{18}$$

The function $\Delta_0^*\Delta_1^M$ is also a periodic function of the coordinates and therefore can be expressed in a Fourier series as

$$\Delta_0^*\Delta_1^M=\sum_{N,K}C_{NK}^M\exp(i\mathbf{K}_{NK}\cdot\mathbf{r}) \tag{19}$$

where $\mathbf{K}_{NK}=N\mathbf{K}_1+K\mathbf{K}_2$.

The values of $\langle|\Delta_0|^2\rangle$ and α_M can be found from the conditions of an extremum of the free energy (17) with respect to these parameters, i.e.,

$$\frac{\partial(F_S-F_N)}{\partial\langle|\Delta_0|^2\rangle}=0,\quad \frac{\partial(F_S-F_N)}{\partial\alpha_M}=0. \tag{20}$$

On the basis of Eqs. (20), it is sufficient to find the values of $\langle|\Delta_0|^2\rangle$ and α_M to leading order in $1-T/T_c$. Using Eqs. (17)–(20) we find

$$\langle|\Delta_0|^2\rangle=\frac{2\pi^3eDT}{7\zeta(3)}\frac{H_{c2}-B}{\beta_A-(\beta_A-1)/\kappa^2},$$

$$\alpha_M=-\frac{1}{M\,M!B\langle|\Delta_0|^2\rangle}\frac{7\zeta(3)}{4\pi^3eDT}\Bigg\{\sum_{N,L}C_{NL}C_{NL}^{M*} -\frac{1}{\kappa^2}\sum_{\mathbf{K}_{N,L}\neq0}C_{NL}C_{NL}^{M*}\left(1-\frac{4eBM}{\mathbf{K}_{NL}^2}\right)\Bigg\}. \tag{21}$$

In Eqs. (21), the quantity β_A is the Abrikosov structural parameter,

$$\beta_A=\langle|\Delta_0|^4\rangle/\langle|\Delta_0|^2\rangle^2, \tag{22}$$

and the Ginzburg–Landau parameter κ is determined by the equation

$$\kappa=\frac{1}{\pi^2eD}\left(\frac{7\zeta(3)}{2\pi\nu}\right)^{1/2}. \tag{23}$$

We employed the following relations[9] to obtain the expressions (21) for α_M:

$$\langle \Delta_1^M \Delta_1^{M*} \rangle = M! \langle |\Delta_0|^2 \rangle,$$

$$\Delta_1^M \partial_+^0 \Delta_0^* - \Delta_0^* \partial_-^0 \Delta_1^M = i\left(\frac{\partial}{\partial y}; -\frac{\partial}{\partial x} \right)(\Delta_0^* \Delta_1^M) - 2\sqrt{eB} M \Delta_0^* \Delta_1^{M-1}(1;-i). \tag{24}$$

The equations (13), (15), (21), and (24) make it possible to reduce Eq. (17) for the free energy to an expression containing only sums of various combinations of the quantities C_{NK} and C_{NK}^M. For arbitrary κ this expression for $F_S - F_N$ is given in Appendix B. For $|\kappa^2 - 1| \ll 1$, we find from Eq. (B.1)

$$\begin{aligned}
\frac{F_S - F_N}{V} = {} & \frac{1}{8\pi}(H_{c2} - H_0)^2 - \frac{1}{4\pi}(H_{c2} - B)(H_{c2} - H_0) + \frac{\beta_A (H_{c2} - B)^2}{8\pi}\Bigg[(\kappa^2 - 1) + \frac{288\pi^4 T^4 \ln(T_c/T)}{(7\zeta(3))^2 \kappa_{\rm cl}^2}\left(\frac{a_1}{9}\left(\frac{1}{\tau} - \frac{1}{\tau_2} \right) + \frac{11 b_1}{15} \right) \\
& - \frac{16\pi^4 T^4 \ln(T_c/T)}{(7\zeta(3))^2 \kappa_{\rm cl}}(a_2 + 3b_2) \Bigg] - \frac{(H_{c2} - B)^3}{\pi H_{c2}}(\kappa^2 - 1) \sum_M \frac{1}{M!} R_1^M R_2^M - \frac{31\zeta(5)\beta_{A1}(H_{c2} - B)^3 \ln(T_c/T)}{8\pi (7\zeta(3))^2 H_{c2}} \\
& - \frac{\pi^3 (H_{c2} - B)^3 T^4 \ln(T_c/T)}{(7\zeta(3))^2 \kappa_{\rm cl}^2 H_{c2}} \Bigg\{ 18\Bigg[(2\beta_A - 1)\left(\frac{a_1}{9}\left(\frac{1}{\tau} - \frac{1}{\tau_2} \right) + \frac{4 b_1}{5} \right) - \frac{2\beta_A b_1}{15} \Bigg] + 72\Bigg[\frac{G_2 + G_3 - 8G_1}{2}\left(\frac{a_1}{9}\left(\frac{1}{\tau} - \frac{1}{\tau_2} \right) \right. \\
& \left. + \frac{2 b_1}{5} \right) + \frac{b_1}{15}(G_2 - 12 G_1) + \frac{b_1}{10}\left(1 + \beta_{A1} - 2\beta_A + \frac{2}{3} G_1 \right) \Bigg] + 144 \sum_M \frac{M+1}{M!} R_2^M (R_1^M - 2M R_2^M)\left(\frac{a_1}{9}\left(\frac{1}{\tau} - \frac{1}{\tau_2} \right) + \frac{2 b_1}{5} \right) \\
& + 144 \sum_M \frac{1}{M!} R_2^M (R_1^M - 4M R_2^M)\left(\frac{a_1}{9}\left(\frac{1}{\tau} - \frac{1}{\tau_2} \right) + \frac{4 b_1}{5} \right) - \frac{48}{5} b_1 \sum_M \frac{1}{M!} R_2^M ((R_3^M - 4M R_1^M) - 6 R_1^M) \Bigg\} \\
& + \frac{2\pi^3 (H_{c2} - B)^3 T^4 \ln(T_c/T)}{(7\zeta(3))^2 \kappa_{\rm cl} H_{c2}} \Bigg\{ b_2\left(\beta_A + 4 \sum_M \frac{R_3^M R_2^M}{M!} \right) + (a_2 + 2b_2)\left(\beta_A + 8 \sum_M \frac{1}{M!} R_1^M R_2^M - 2 G_2 \right) \Bigg\}.
\end{aligned} \tag{25}$$

For the three types of lattices—triangular with one or two flux quanta per cell and square with one flux quantum per cell—the coefficients C_{NL} and C_{NL}^M were found in Ref. 9. For this reason, we shall not present them here. A numerical calculation of the sums in Eq. (25) gives the following values of the free energy as a function of the parameter B and type of lattice.

1. Triangular lattice with one flux quantum per cell:

$\beta_A = 1.159595, \quad \beta_{A_1} = 1.423012,$

$$\begin{aligned}
\frac{F_S - F_N}{V} = {} & \frac{1}{8\pi}(H_{c2} - H_0)^2 - \frac{1}{4\pi}(H_{c2} - B)(H_{c2} - H_0) + \frac{\beta_A (H_{c2} - B)^2}{8\pi}\Bigg[(\kappa^2 - 1) + \frac{16\pi^4 T^4 \ln(T_c/T)}{(7\zeta(3))^2}\Bigg(\frac{18}{\kappa_{\rm cl}^2}\Bigg(\frac{a_1}{9}\left(\frac{1}{\tau} - \frac{1}{\tau_2} \right) \\
& + \frac{11 b_1}{15} \Bigg) - \frac{1}{\kappa_{\rm cl}}(a_2 + 3b_2) \Bigg) \Bigg] - \frac{(H_{c2} - B)^3}{8\pi H_{c2}} \Bigg\{ 4.361 \cdot 10^{-2}(\kappa^2 - 1) + 1.423 \cdot \frac{31\zeta(5)}{(7\zeta(3))^2} \ln(T_c/T) \\
& + \frac{18\pi^4 T^4 \ln(T_c/T)}{(7\zeta(3))^2 \kappa_{\rm cl}^2}\left(1.3424 a_1\left(\frac{1}{\tau} - \frac{1}{\tau_2} \right) + 8.36257 b_1 \right) - \frac{16\pi^4 T^4 \ln(T_c/T)}{(7\zeta(3))^2 \kappa_{\rm cl}} 1.4666(a_2 + 3b_2) \Bigg\}.
\end{aligned} \tag{26}$$

2. Triangular lattice with two flux quanta per cell:

$\beta_A = 1.33898, \quad \beta_{A_1} = 2.02588,$

$$\begin{aligned}
\frac{F_S - F_N}{V} = {} & \frac{1}{8\pi}(H_{c2} - H_0)^2 - \frac{1}{4\pi}(H_{c2} - B)(H_{c2} - H_0) + \frac{\beta_A (H_{c2} - B)^2}{8\pi}\Bigg[(\kappa^2 - 1) + \frac{16\pi^4 T^4 \ln(T_c/T)}{(7\zeta(3))^2}\Bigg(\frac{18}{\kappa_{\rm cl}^2}\Bigg(\frac{a_1}{9}\left(\frac{1}{\tau} - \frac{1}{\tau_2} \right) \\
& + \frac{11 b_1}{15} \Bigg) - \frac{1}{\kappa_{\rm cl}}(a_2 + 3b_2) \Bigg) \Bigg] - \frac{(H_{c2} - B)^3}{8\pi H_{c2}} \Bigg\{ 0.252(\kappa^2 - 1) + 2.0258 \frac{31\zeta(5)}{(7\zeta(3))^2} \ln \frac{T_c}{T} \\
& + \frac{18\pi^4 T^4 \ln(T_c/T)}{(7\zeta(3))^2 \kappa_{\rm cl}^2}\left(2.2488 a_1\left(\frac{1}{\tau} - \frac{1}{\tau_2} \right) + 14.3985 b_1 \right) - \frac{16\pi^4 T^4 \ln(T_c/T)}{(7\zeta(3))^2 \kappa_{\rm cl}} 2.27793(a_2 + 3b_2) \Bigg\}.
\end{aligned} \tag{27}$$

3. Square lattice with one flux quantum per cell:

$\beta_A = 1.18034, \quad \beta_{A_1} = 1.4971,$

$$\frac{F_S-F_N}{V}=\frac{1}{8\pi}(H_{c2}-H_0)^2-\frac{1}{4\pi}(H_{c2}-B)(H_{c2}-H_0)+\frac{\beta_A(H_{c2}-B)^2}{8\pi}\Bigg[(\kappa^2-1)+\frac{16\pi^4T^4\ln(T_c/T)}{(7\zeta(3))^2}\Bigg(\frac{18}{\kappa_{\rm cl}^2}\Bigg(\frac{a_1}{9}\Bigg(\frac{1}{\tau}-\frac{1}{\tau_2}\Bigg)+\frac{11b_1}{15}\Bigg)-\frac{1}{\kappa_{\rm cl}}(a_2+3b_2)\Bigg)\Bigg]-\frac{(H_{c2}-B)^3}{8\pi H_{c2}}\Bigg\{0.068798(\kappa^2-1)+1.4971\cdot\frac{31\zeta(5)}{(7\zeta(3))^2}\ln\frac{T_c}{T}+\frac{18\pi^4T^4\ln(T_c/T)}{(7\zeta(3))^2\kappa_{\rm cl}^2}(1.45306a_1+9.102163b_1)-\frac{16\pi^4T^4\ln(T_c/T)}{(7\zeta(3))^2\kappa_{\rm cl}}\cdot 1.56589(a_2+3b_2)\Bigg\}. \tag{28}$$

Aside from Eq. (B.5) the relation

$$2\sum_M\frac{R_2^MR_3^M}{M!}=4\sum_M\frac{R_1^MR_2^M}{M!}-G_2. \tag{29}$$

probably also holds.

The ratio of the Ginzburg–Landau parameters in dirty ($\kappa_{\rm dir}$) and clean ($\kappa_{\rm cl}$) superconductors can be easily found with the aid of Appendix A and is

$$\frac{\kappa_{\rm dir}}{\kappa_{\rm cl}}=\frac{7\zeta(3)}{2\pi^3T\tau_{\rm tr}}. \tag{30}$$

Let us assume that in a clean superconductor we have $\kappa_{\rm cl}\ll 1$. Then the condition $\kappa\approx 1$ signifies that $2\pi T\tau_{\rm tr}\ll 1$. In the limit of a short electron mean-free path, in Eqs. (26)–(28) the contributions of the terms proportional to a_2 and b_1 are small. In this limit, we find for the coefficients a_1 and b_2 from Appendix A:

$$a_{1\rm dir}=\frac{14\zeta(3)\tau_{\rm tr}^2}{\pi^2T^2(1/\tau-1/\tau_2)},$$

$$b_{2\rm dir}=\frac{\pi\tau_{\rm tr}}{24T^3}. \tag{31}$$

From Eqs. (30) and (31) it is easy to see that the coefficient of $(H_{c2}-B)^3$ in Eqs. (26)–(28) can be made positive, changing the sign of the coefficient of $(H_{c2}-B)^2$ by changing the value of κ. Clean superconductors satisfy

$$b_1=b_2=a_2. \tag{32}$$

The effect of Eqs. (32) is that if in a clean superconductor the parameter $\kappa_{\rm cl}$ is close to 1, then the coefficient of the term $(H_{c2}-B)^3$ in Eqs. (26)–(28) can be easily made negative. We have thus shown that all cases shown in Figs. 1 and 2 can be realized by an appropriate choice of the parameters of the superconductor. We recall also that the field induction B is determined from the condition[8]

$$\frac{\partial(F_S-F_N)}{\partial B}=0. \tag{33}$$

Using the condition (33) it can be shown that there exists a parameter range where a triangular lattice with two flux quanta per cell has a lower energy than a triangular or square lattice with one flux quantum per cell.

4. CONCLUSIONS

The strong degeneracy arising in the Ginzburg–Landau equation with $\kappa=1$ can be removed by correction terms proportional to $1-T/T_c$. Near the transition temperature, the type of solution is determined by the ratio of the small parameters $|\kappa^2-1|$ and $1-T/T_c$. It was shown that in an external magnetic field H_0 close to H_{c2} all possible types of dependences of the free energy on the parameter B can be realized with an appropriate choice of the quantities $\kappa_{\rm cl}$ and $1-T/T_c$ and electron mean-free path. The sign of the coefficient of the quadratic term $(H_{c2}-B)^2$ can be changed (with κ close to 1) by a small change in the impurity density. Since all types of dependence of the free energy on the parameter B shown in Figs. 1 and 2 can be realized, the transition into the superconducting state can be either first- or second-order. We also note that the minimum of the free energy can be reached, for example, on a triangular lattice with two flux quanta per cell.

This work was supported by the CRDF (Grant RP1-194).

5. APPENDIX A

From Eqs. (9) and (10) we find the coefficients D, $b_{1,2}$, and $a_{1,2}$ appearing in the expression for the free energy:

$$D=\frac{vl_{\rm tr}}{3}\eta_{(T)},\quad \eta_{(T)}=1-\frac{8T\tau_{\rm tr}}{\pi}\Bigg(\psi\Bigg(\frac{1}{2}+\frac{1}{4\pi T\tau_{\rm tr}}\Bigg)-\psi\Bigg(\frac{1}{2}\Bigg)\Bigg),$$

$$b_2=2\pi T\sum_{\omega>0}\frac{1}{\omega^4(\omega+1/2\tau_{\rm tr})}=\frac{\tau_{\rm tr}}{4\pi^3T^3}\Bigg\{\frac{\pi^4}{6}-4\pi T\tau_{\rm tr}[7\zeta(3)-2\pi^3T\tau_{\rm tr}\eta(T)]\Bigg\},$$

$$a_2=2\pi T\sum_{\omega>0}\frac{1}{\omega^3(\omega+1/2\tau_{\rm tr})^2}=\frac{\tau_{\rm tr}^2}{\pi^2T^2}\Bigg\{7\zeta(3)-4\pi^3T\tau_{\rm tr}-4\pi T\tau_{\rm tr}\psi'\left(\frac{1}{2}+\frac{1}{4\pi T\tau_{\rm tr}}\right)$$
$$+3(4\pi T\tau_{\rm tr})^2\left(\psi\left(\frac{1}{2}+\frac{1}{4\pi T\tau_{\rm tr}}\right)-\psi\left(\frac{1}{2}\right)\right)\Bigg\},$$

$$b_1=2\pi T\sum_{\omega>0}\frac{1}{\omega^2(\omega+1/2\tau_{\rm tr})^2(\omega+(1/2)(1/\tau-1/\tau_2))}=\frac{1}{16\pi^4T^4}\Bigg\{\frac{2\pi^3T(4\pi T\tau_{\rm tr})^2}{1/\tau-1/\tau_2}+\frac{(4\pi T)^3\tau_{\rm tr}^2}{1/\tau_1-1/\tau_2}\psi'\left(\frac{1}{2}+\frac{1}{4\pi T\tau_{\rm tr}}\right)$$
$$-\frac{(4\pi T)^4}{(1/\tau-1/\tau_2)^2(1/\tau_1-1/\tau_2)^2}\left(\psi\left(\frac{1}{2}+\frac{1}{4\pi T}\left(\frac{1}{\tau}-\frac{1}{\tau_2}\right)\right)-\psi\left(\frac{1}{2}\right)\right)-(4\pi T)^3\tau_{\rm tr}^2\left(\frac{8\pi T\tau_{\rm tr}}{1/\tau_1-1/\tau_2}-\frac{4\pi T}{(1/\tau_1-1/\tau_2)^2}\right)$$
$$\times\left(\psi\left(\frac{1}{2}+\frac{1}{4\pi T\tau_{\rm tr}}\right)-\psi\left(\frac{1}{2}\right)\right)\Bigg\},$$

$$a_1=2\pi T\sum_{\omega>0}\frac{1}{\omega^3(\omega+1/2\tau_{\rm tr})^2(\omega+(1/2)(1/\tau-1/\tau_2))}=\frac{1}{32\pi^5T^5}\Bigg\{7\zeta(3)\frac{(4\pi T)^3\tau_{\rm tr}^2}{1/\tau-1/\tau_2}-\frac{(4\pi T)^4\tau_{\rm tr}^3}{1/\tau_1-1/\tau_2}\psi'\left(\frac{1}{2}+\frac{1}{4\pi T\tau_{\rm tr}}\right)$$
$$+\frac{(4\pi T)^5}{(1/\tau-1/\tau_2)^3(1/\tau_1-1/\tau_2)^2}\left(\psi\left(\frac{1}{2}+\frac{1}{4\pi T}\left(\frac{1}{\tau}-\frac{1}{\tau_2}\right)\right)-\psi\left(\frac{1}{2}\right)\right)+(4\pi T)^5\tau_{\rm tr}^3\left(\frac{3\tau_{\rm tr}}{1/\tau_1-1/\tau_2}-\frac{1}{(1/\tau_1-1/\tau_2)^2}\right)$$
$$\times\left(\psi\left(\frac{1}{2}+\frac{1}{4\pi T\tau_{\rm tr}}\right)-\psi\left(\frac{1}{2}\right)\right)+\frac{\pi^2}{2}(4\pi T)^4\left(\tau_{\rm tr}^2\left(\frac{1}{(1/\tau_1-1/\tau_2)^2}-\frac{2\tau_{\rm tr}}{1/\tau_1-1/\tau_2}\right)-\frac{1}{(1/\tau_1-1/\tau_2)^2(1/\tau-1/\tau_2)^2}\right)\Bigg\}.$$

6. APPENDIX B

For arbitrary κ the free energy (17) can be put into the following form using Eqs. (13), (15), (21), and (24):

$$\frac{F_S-F_N}{V}=\frac{1}{8\pi}(H_{c2}-H_0)^2-\frac{1}{4\pi}(H_{c2}-B)(H_{c2}-H_0)+\frac{\beta_A(\kappa^2-1)(H_{c2}-B)^2}{8\pi\kappa^2(\beta_A-(\beta_A-1)/\kappa^2)}$$
$$+\frac{36\pi^3\beta_A(H_{c2}-B)^2T^4\ln(T_c/T)}{(7\zeta(3))^2\kappa_{\rm cl}^2(\beta_A-(\beta_A-1)/\kappa^2)^2}\left[\frac{a_1}{9}\left(\frac{1}{\tau}-\frac{1}{\tau_2}\right)+\frac{4b_1}{5}-\frac{b_1}{15\kappa^2}\right]-\frac{2\pi^3\beta_A(H_{c2}-B)^2T^4\ln(T_c/T)}{(7\zeta(3))^2\kappa\kappa_{\rm cl}(\beta_A-(\beta_A-1)/\kappa^2)^2}$$
$$\times(a_2+3b_2)-\frac{(H_{c2}-B)^3}{8\pi\kappa^2H_{c2}(\beta_A-(\beta_A-1)/\kappa^2)^3}\left[\frac{4G_1}{\kappa^4}+\sum_M\frac{1}{MM!}\left(\left(1-\frac{1}{\kappa^2}\right)R_1^M+\frac{4M}{\kappa^2}R_2^M\right)^2\right]$$
$$-\frac{31\zeta(5)\beta_{A1}(H_{c2}-B)^3\ln(T_c/T)}{8\pi(7\zeta(3))^2\kappa^2H_{c2}(\beta_A-(\beta_A-1)/\kappa^2)^3}-\frac{\pi^3(H_{c2}-B)^3T^4\ln(T_c/T)}{(7\zeta(3))^2\kappa_{\rm cl}^2H_{c2}(\beta_A-(\beta_A-1)/\kappa^2)^3}\Bigg\{18\Bigg[\left(\beta_A+\frac{\beta_A-1}{\kappa^2}\right)$$
$$\times\left(\frac{a_1}{9}\left(\frac{1}{\tau}-\frac{1}{\tau_2}\right)+\frac{4b_1}{5}\right)-\frac{2\beta_Ab_1}{15\kappa^2}\Bigg]+\frac{72}{\kappa^4}\Bigg[\frac{G_2+G_3-8G_1}{2}\left(\frac{a_1}{9}\left(\frac{1}{\tau}-\frac{1}{\tau_2}\right)+\frac{2b_1}{5}\right)+\frac{b_1(G_2-12G_1)}{15}+\frac{b_1}{10}$$
$$\times\left(1+\beta_{A1}-2\beta_A+\frac{2}{3}G_1\right)\Bigg]+18\sum_M\frac{M+1}{MM!}\left(\left(1-\frac{1}{\kappa^2}\right)R_1^M+\frac{4M}{\kappa^2}R_2^M\right)\left(\left(1+\frac{1}{\kappa^2}\right)R_1^M-\frac{4M}{\kappa^2}R_2^M\right)\left(\frac{a_1}{9}\left(\frac{1}{\tau}-\frac{1}{\tau_2}\right)\right.$$
$$\left.+\frac{2b_1}{5}\right)+\frac{36}{\kappa^2}\sum_M\frac{1}{MM!}\left(\left(1-\frac{1}{\kappa^2}\right)R_1^M+\frac{4M}{\kappa^2}R_2^M\right)(R_1^M-4MR_2^M)\left(\frac{a_1}{9}\left(\frac{1}{\tau}-\frac{1}{\tau_2}\right)+\frac{4b_1}{5}\right)-\frac{12b_1}{5\kappa^2}$$
$$\times\sum_M\frac{1}{MM!}\left(\left(1-\frac{1}{\kappa^2}\right)R_1^M+\frac{4M}{\kappa^2}R_2^M\right)((R_3^M-4MR_1^M)-6R_1^M)\Bigg\}+\frac{2\pi^3(H_{c2}-B)^3T^4\ln(T_c/T)}{(7\zeta(3))^2\kappa\kappa_{\rm cl}H_{c2}(\beta_A-(\beta_A-1)/\kappa^2)^3}$$
$$\times\Bigg\{.b_2\Bigg[\beta_A\left(\beta_A-\frac{\beta_A-1}{\kappa^2}\right)+\sum_M\frac{R_3^M}{MM!}\left(\left(1-\frac{1}{\kappa^2}\right)R_1^M+\frac{4M}{\kappa^2}R_2^M\right)\Bigg]+(a_2+2b_2)\Bigg[\beta_A\left(\beta_A-\frac{\beta_A-1}{\kappa^2}\right)$$

$$+2\sum_M \frac{R_1^M}{MM!}\left(\left(1-\frac{1}{\kappa^2}\right)R_1^M+\frac{4M}{\kappa^2}R_2^M\right)-\frac{2}{\kappa^2}G_2\Bigg]\Bigg\}. \tag{B.1}$$

In Eq. (B.1), $\kappa_{\rm cl}$ is the Ginzburg–Landau parameter of a clean superconductor

$$\kappa_{\rm cl}^2=\frac{18\pi T^2}{7\zeta(3)\nu e^2 v^4}, \tag{B.2}$$

We have written $\beta_{A1}=\langle|\Delta_0|^6\rangle/\langle|\Delta_0|^2\rangle^3$, and V is the volume of the superconductor. We also introduced the following notation:

$$\sum_{\mathbf{K}_{NL}\neq 0} C_{NL}C_{NL}^{M*}=R_1^M\langle|\Delta_0|^2\rangle^2,$$

$$\sum_{\mathbf{K}_{NL}\neq 0} C_{NL}C_{NL}^{M*}/\mathbf{K}_{NL}^2=R_2^M\langle|\Delta_0|^2\rangle^2/eB,$$

$$\sum_{\mathbf{K}_{NL}\neq 0} C_{NL}C_{NL}^{M*}\mathbf{K}_{NL}^2=R_3^M\langle|\Delta_0|^2\rangle^2 eB,$$

$$\sum_{\mathbf{K}_{NL}\neq 0}\sum_{\mathbf{K}_{N_1L_1}\neq 0} C_{NL}C_{N_1L_1}C_{N+N_1,L+L_1}^{*}(\mathbf{K}_{NL}\cdot\mathbf{K}_{N_1L_1})/(\mathbf{K}_{NL}^2\mathbf{K}_{N_1L_1}^2)=G_1\langle|\Delta_0|^2\rangle^3/eB, \tag{B.3}$$

$$\sum_{\mathbf{K}_{NL}\neq 0}\sum_{\mathbf{K}_{N_1L_1}\neq 0} C_{NL}C_{N_1L_1}C_{N+N_1,L+L_1}^{*}(\mathbf{K}_{NL}\cdot\mathbf{K}_{N_1L_1})/\mathbf{K}_{N_1L_1}^2=G_2\langle|\Delta_0|^2\rangle^3,$$

$$\sum_{\mathbf{K}_{NL}\neq 0}\sum_{\mathbf{K}_{N_1L_1}\neq 0} C_{NL}C_{N_1L_1}C_{N+N_1,L+L_1}^{*}(\mathbf{K}_{NL}\cdot\mathbf{K}_{N_1L_1})^2/(\mathbf{K}_{NL}^2\mathbf{K}_{N_1L_1}^2)=G_3\langle|\Delta_0|^2\rangle^3.$$

To obtain the expression (B.1) for the free energy, we used the relation

$$\sum_{N,L} C_{NL}C_{NL}^{*}\mathbf{K}_{NL}^2=2eB\beta_A\langle|\Delta_0|^2\rangle^2. \tag{B.4}$$

In accordance with the general assertions, the following equation should hold:

$$G_1+4\sum_M \frac{M}{M!}(R_2^M)^2=0. \tag{B.5}$$

*)E-mail: ovc@itp.ac.ru

[1] V. L. Ginzburg and L. D. Landau, Zh. Éksp. Teor. Fiz. **20**, 1064 (1950).
[2] E. B. Bogomol'nyi, Yad. Fiz. **24**, 861 (1976) [Sov. J. Nucl. Phys. **24**, 449 (1976).
[3] A. V. Efanov, Phys. Rev. B **56**, 7839 (1997).
[4] E. Eilenberger, Z. Phys. **214**, 195 (1968).
[5] A. I. Larkin and Yu. N. Ovchinnikov, Zh. Éksp. Teor. Fiz. **55**, 2262 (1968) [Sov. Phys. JETP **28**, 1200 (1969)].
[6] Yu. N. Ovchinnikov and E. H. Brandt, Phys. Status Solidi B **67**, 301 (1975).
[7] A. A. Abrikosov, Zh. Éksp. Teor. Fiz. **32**, 1442 (1957) [Sov. Phys. JETP **5**, 1174 (1957)].
[8] Yu. N. Ovchinnikov, JETP Lett. **65**, 632 (1997).
[9] Yu. N. Ovchinnikov, Zh. Éksp. Teor. Fiz. **112**, 1499 (1997) [JETP **85**, 818 (1997)].

Translated by M. E. Alferieff

Parametric instability and Hamiltonian chaos in cavity semiclassical electrodynamics

S. V. Prants*) and L. E. Kon'kov

Pacific Oceanographic Institute, Far East Branch of the Russian Academy of Sciences, 690041 Vladivostok, Russia

(Submitted 9 July 1998)
Zh. Éksp. Teor. Fiz. **115**, 740–753 (February 1999)

We study the nonlinear dynamics of the interaction of two-level atoms and a selected mode of a high-Q cavity with frequency modulation analytically and numerically. In the absence of modulation, the corresponding semiclassical Heisenberg equations for the expectation values of the collective atomic observables and the field-mode amplitudes allow, in the rotating wave approximation and in the strong-coupling limit, an exact solution with arbitrary detuning. Using this solution, we detect the coherent effect of trapping of the population of atomic levels and of trapping of the number of photons in the cavity. The explanation for this effect lies in the destructive interference of the atomic dipoles and the field mode. The integrable version of the system of equations exhibits a separatrix near which a stochastic layer is formed when modulation is introduced. The width of the layer is found to gradually increase with degree of modulation, and finally it fills the entire energy-permissible volume of the phase space. We show that the rotating wave approximation does not hinder the formation of Hamiltonian chaos in cavity semiclassical electrodynamics. The calculation of the maximum Lyapunov indices of nonlinear (in this approximation) equations of motion as functions of the modulation frequency δ and the frequency of natural Rabi oscillations of the atom–field system, Ω, suggests that Hamiltonian chaos appears first in the area of the fundamental parametric resonance, $\delta/2\Omega \simeq 1$. Parametric instability increases with increasing modulation and decreasing detuning from the atom–field resonance, generating at exact resonance new areas of chaos corresponding to multiple parametric resonances. The results of numerical experiments and estimates of the characteristic parameters show that Rydberg atoms placed in a high-Q microwave cavity are possible objects for observing parametric instability and dynamical chaos.
 [S1063-7761(99)02502-0]

1. INTRODUCTION

The basis for cavity quantum electrodynamics (for a review see, e.g., Ref. 1) is the micromaser,[2] a real device operating with Rydberg atoms in a high-Q superconducting microwave cavity, which in the single-mode regime and the strong-coupling limit is described by the Jaynes–Cummings model.[3] Among other things, an attractive feature of such systems is the possibility of experimentally studying the action of the atoms and the radiation field on the semiclassical, semiquantum, and fully quantum levels by varying the number of atoms in the cavity. The study of the quantum behavior of the atom–field system in regimes where the semiclassical description of the system leads to dynamical instability and chaos touches upon a new aspect of the old problem of quantum–classical correspondence, the relation between quantum mechanics and nonlinear classical mechanics.

This paper studies analytically and numerically Hamiltonian chaos in a parametric atom–field system whose equations of motion can be found from the operator Heisenberg equations via a semiclassical procedure of factorizing the quantum correlators. On the one hand, such an approach leads to a dynamical system that can be analyzed by methods of nonlinear classical mechanics. On the other, it allows one to take into account some quantum correlation effects by carrying out semiquantum factorization.

Semiclassical factorization of the standard Jaynes–Cummings model leads to a system of equations that are exactly integrable in the rotating wave approximation. As is well known, when this approximation is discarded, i.e., when nonresonant terms describing virtual transitions are taken into account in the Hamiltonian, the semiclassical model becomes chaotic.[4–6] However, these terms play an important role when the number density of the atoms is extremely high, $\simeq 10^{21}\,\mathrm{cm}^{-3}$, i.e., when the validity of the approximation of noninteracting atoms adopted in Refs. 4–6 is problematic.

In our recent papers[7,8] we showed that when two-level atoms travel through a single-mode cavity, Hamiltonian chaos appears even in the rotating wave approximation in the absence of external agents. Such a situation can be realized in micromaser experiments. Chaotic oscillations of atomic inversion and the number of photons may appear in the semiclassical model[7] and in the semiquantum model.[8] In the latter case, the field, interatomic, and first-order atom–field quantum correlations are taken into account when the equations of motion for the expectation values are derived.[9] As a result a new effect of cavity quantum electrodynamics appears, which became known as chaotic Rabi vacuum oscillations.[8] This effect reflects the dynamical instability and chaos in the interaction of atoms and the vacuum in a high-Q cavity. Note that reliable diagnostics of chaos in numerical experiments

by the values of the maximum Lyapunov index requires high velocities of the atoms, $v \simeq 10^8$ cm s^{-1}, and large values of the vacuum Rabi frequency, $\Omega_0 \simeq 10^7$ rad s^{-1}. For moving atoms, the mechanism of chaos formation is related to the modulation of the vacuum Rabi frequency $\Omega_0 f(r)$ due to the spatially inhomogeneous structure of the cavity mode described by a function $f(r)$ (see Ref. 10). Here the natural restrictions on the atomic velocities and hence on the modulation frequency limit parametric oscillations in the atom–field system to the adiabatic regime.

In the case of modulation of the energy of the active atomic transition and/or the selected cavity mode with a frequency ω_m, the atom–field system may demonstrate various regimes of parametric oscillations: the adiabatic regime ($\omega_m \ll \Omega_0 \sqrt{N}$), the nonadiabatic regime ($\omega_m \gg \Omega_0 \sqrt{N}$), and nonlinear parametric resonance ($n\omega_m \simeq 2\Omega_0 \sqrt{N}$, $n=1,2,\ldots$, where N is the number of atoms in the cavity). In the present research we use analytical solutions and numerical experiments to show how in the parametric resonance regime a system of immobile two-level atoms in a single-mode high-Q cavity becomes extremely sensitive to the slightest variations in the initial state of the atoms and the field mode. This becomes observable in chaotic oscillations of the atomic inversion and the number of photons.

2. MODEL

The interaction of N two-level atoms and a single mode in a high-Q cavity with parametric modulation is described in the rotating wave approximation by the Jaynes–Cummings N-atom Hamiltonian with time-dependent parameters:

$$H = \frac{1}{2}\hbar\omega_a \sum_{j=1}^{N} \sigma_z^j + \hbar\omega_f \left(a^\dagger a + \frac{1}{2} \right) + \hbar\Omega_0 \sum_{j=1}^{N} (a\sigma_+^j + a^\dagger \sigma_-^j). \quad (1)$$

For simplicity we assume that all the atoms are at rest and have the same electric-dipole transition frequency ω_a and the same vacuum Rabi frequency Ω_0, i.e., the linear dimensions of the atomic sample are assumed much smaller than the radiation wavelength. Modulations of the frequency of the atomic transition by an external field and of the frequency of the selected cavity mode are possible (at least in principle). The second type of modulation can be achieved in cavities with movable walls[11] or in a cavity filled with a homogeneous medium with a time-dependent dielectric constant.[12] Lately such systems have drawn much attention in connection with the possibility of producing photons from the vacuum[13] and of generating squeezed electromagnetic-field states.[12] Ignoring all quadratic effects, we model the selected cavity mode by a simple harmonic oscillator with a variable frequency ω_f. The results obtained in this paper can easily be generalized to the case of a modulated atomic frequency ω_a, since the ratio (6) of these frequencies is actually the independent variable of the problem.

We study the complexity of the Hamiltonian semiclassical dynamics generated by the operator (1). Hence our results are valid only for times much shorter than the field-mode decay time T_f and the atomic relaxation time T_a. Such a situation occurs, for instance, with Rydberg atoms in high-Q ($Q \simeq 10^{10}$) microwave cavities cooled to subkelvin temperatures. In these conditions, the lifetime of microwave photons in the cavity, T_f, is roughly 10^{-1}–10^{-2} s, and the lifetime of circular Rydberg states of atoms, $T_a \simeq 10^{-2}$ s (see Ref. 1), is several orders of magnitude longer than even the period of single-atom Rabi vacuum oscillations, $T_0 = 2\pi/\Omega_0 < 10^{-5}$ s. Note that the characteristic natural frequency of a multiatomic system is the collective vacuum Rabi frequency $\Omega_N = \Omega_0 \sqrt{N}$ at which the N-atom ensemble and the field mode coherently exchange energy. Thus, the Hamiltonian approach provides a meaningful description of the atom–field dynamics only in what is known as the strong-coupling regime, $\Omega_N \gg T_a^{-1}, T_f^{-1}$.

A classical dynamical system generated by the Hamiltonian (1) is obtained as a result of replacing the Heisenberg equations for the operators by equation for the expectation values of these operators. Of course, the result of such a transition is an infinite hierarchy of equations for the expectation values. In the semiclassical approximation this infinite-dimensional system of equations is truncated, and the result is a simple closed self-consistent system of equations in which the reciprocal effect of the radiation field is taken into account but the quantum correlations are ignored. The semiclassical approximation for N atoms can be shown (see the Appendix) to be accurate to $O(1/N)$. In the semiclassical limit, the natural quantum states over which averaging is performed are generalized coherent states, i.e., the direct product of a Glauber field coherent state and a Bloch atomic coherent state. As a result we arrive at the nonlinear dynamical system

$$\dot{x} = -y - \Omega z P, \quad \dot{y} = x - \Omega z E, \quad \dot{z} = \Omega(xP + yE), \quad (2)$$

$$\dot{E} = \omega P - \Omega y, \quad \dot{P} = -\omega E - \Omega x$$

for three atomic variables,

$$x = \frac{1}{N}\left\langle \sum_{j=1}^{N} \sigma_x^j \right\rangle, \quad y = \frac{1}{N}\left\langle \sum_{j=1}^{N} \sigma_y^j \right\rangle,$$

$$z = \frac{1}{N}\left\langle \sum_{j=1}^{N} \sigma_z^j \right\rangle, \quad (3)$$

and two field variables,

$$E = \frac{1}{\sqrt{N}}\langle a + a^\dagger \rangle, \quad P = \frac{i}{\sqrt{N}}\langle a^\dagger - a \rangle. \quad (4)$$

The dot in (2) indicates the derivative with respect to the dimensionless time $\tau = \omega_a t$, and the control parameters are the dimensionless collective vacuum Rabi frequency

$$\Omega = \frac{\Omega_0 \sqrt{N}}{\omega_a} \quad (5)$$

and the dimensionless variable detuning from resonance, which for the sake of definiteness has been chosen in the form

$$\omega=\omega_0(1+\alpha\sin\delta\tau), \tag{6}$$

where $\omega_0=\omega_f/\omega_a$, $\delta=\omega_m/\omega_a$ is the dimensionless harmonic-modulation frequency of the field mode, and α is the degree of modulation.

To be sure, the simple factorization of the correlators means that we project the motion of observables in an infinite-dimensional space of a completely quantum model on the five-dimensional space of the variables (3) and (4) of the semiclassical model. The presence of natural constants of motion,

$$x^2+y^2+z^2=1, \quad E^2+P^2+2z=W, \tag{7}$$

which reflect the unitarity of atomic evolution and the conservation of the total energy, limits the motion of the dynamical system (2) to three-dimensional space. By introducing new complex-valued variables,

$$\eta=x+iy, \quad \xi=P+iE, \tag{8}$$

we can reduce the system of five nonlinear equation to a single second-order complex-valued equation with variable coefficients,

$$\ddot{\xi}-i(\omega+1)\dot{\xi}-\left(\omega+i\dot{\omega}+\frac{1}{2}\,\Omega^2W\right)\xi+\frac{1}{2}\,\Omega^2\xi|\xi|^2=0. \tag{9}$$

This is a complex Duffing oscillator with parametric excitation, which in our case describes the semiclassical Hamiltonian dynamics of the interaction of two-level atoms and a single-mode frequency-modulated field. It is natural to suppose that such an oscillator may be chaotic, despite the zero right-hand side of its equation of motion (9).

3. EXACT SOLUTION OF AN INTEGRABLE SYSTEM

In the absence of modulation ($\alpha=0$), the initial system (2) acquires an additional constant of motion,

$$\Omega(xE-yP)-(\omega_0-1)z=C, \tag{10}$$

which reflects the conservation of the energy of interaction of the atoms and the field in the rotating wave approximation. By virtue of this constant of motion, the system of equations (2) yields a closed equation for atomic inversion,

$$\dot{z}=\pm\Omega\sqrt{(W-2z)(1-z^2)-\left[\frac{C+(\omega_0-1)z}{\Omega}\right]^2}. \tag{11}$$

This equation has an exact solution in terms of Jacobi's elliptic functions:

$$z=z_1+(z_2-z_1)\mathrm{sn}^2\left[\sqrt{\frac{z_3-z_1}{2}}\,\Omega(\tau-\tau_0);\frac{z_2-z_1}{z_3-z_1}\right], \tag{12}$$

where

$$\tau_0=\frac{1}{\sqrt{2}\,\Omega}\int_{z_0}^{z_1}\frac{dz}{\sqrt{(z-z_1)(z-z_2)(z-z_3)}}, \tag{13}$$

and $z_1<z_2<z_3$ are the roots of the characteristic equation

$$z^3-\left[\frac{W}{2}+\frac{(\omega_0-1)^2}{2\Omega^2}\right]z^2-\left[1+\frac{C(\omega_0-1)}{2\Omega^2}\right]z+\left[\frac{W}{2}-\frac{C^2}{2\Omega^2}\right]=0. \tag{14}$$

We seek the solution for the unknowns of the integrable version of the initial system (2) at $\alpha=0$ in the form

$$x=\frac{s}{v}\sin(\omega_0\tau+\theta)-u\cos(\omega_0\tau+\theta),$$

$$y=-\frac{s}{v}\cos(\omega_0\tau+\theta)-u\sin(\omega_0\tau+\theta),$$

$$E=v\sin(\omega_0\tau+\theta), \quad P=v\cos(\omega_0\tau+\theta). \tag{15}$$

Using the constants of motion, we can easily show that all the new variables θ, s, u, and v are functions of only one (old) variable z:

$$s=\frac{C+(\omega_0-1)z}{\Omega}, \quad v=\pm\sqrt{W-2z},$$

$$u=\pm\sqrt{1-z^2-\left(\frac{s}{v}\right)^2}, \quad \theta=\int_0^\tau\frac{C+(\omega_0-1)z}{W-2z}\,d\tau'. \tag{16}$$

Thus, in the absence of modulation the motion is limited to a two-dimensional surface and is fully characterized by the above general solution (12)–(16) describing the periodic exchange of energy between the atoms and the field generated by the atoms.

In terms of the new variables, the integrable version of (2) is

$$\dot{v}=\Omega u, \quad \dot{u}=\frac{1}{2}\Omega v(W-v^2)+\Omega\frac{s^2}{v^3}+(\omega_0-1)\frac{s}{v},$$

$$\dot{\theta}=\Omega\frac{s}{v^2}, \quad \dot{s}=(1-\omega_0)uv. \tag{17}$$

At exact resonance, $\omega_0=1$, for $C=0$ and $W>0$, the system of equation reduces to the equation of motion of the free Duffing oscillator, whose phase plane (u,v) contains a pair of homoclinic trajectories that converge at the singular point S_+: $(u_0=0,\ v_0=0)$. In terms of the old variables, this singular point has the coordinates $x_0=y_0=0$, $z_0=1$, and $E_0=P_0=0$. The point is always unstable at exact resonance and corresponds to the situation with fully inverted atoms in a cavity with the vacuum field. The solution on the separatrix ($W=2$) can be found from Eqs. (15) and (16) and has the following form:

$$x_s=\pm 2\sinh(\Omega\tau)\tanh(\Omega\tau)\cos(\tau+\theta_0),$$

$$y_s=\pm 2\sinh(\Omega\tau)\tanh(\Omega\tau)\sin(\tau+\theta_0),$$

$$z_s=1-2\sinh^2\Omega\tau, \tag{18}$$

$$P_s=\pm 2\sinh(\Omega\tau)\cos(\tau+\theta_0),$$

$$E_s = \pm 2\sinh(\Omega\tau)\sin(\tau+\theta_0),$$

where θ_0 is an arbitrary angle. Here and in what follows, to write the solution correctly, we select in the expressions with double signs either only the upper signs or only the lower signs.

4. TRAPPING OF THE POPULATION OF ATOMIC LEVELS

The general exact solution (15) of the integrable system contains an interesting particular solution that describes the coherent effect of trapping of the population of two-level atoms and trapping of the number of photons in the cavity. A direct check readily shows that a solution of the form

$$\begin{aligned}
x_{\rm tr} &= \pm\sqrt{1-z_0^2}\cos\left(\frac{\omega_0+1}{2}\tau+\varphi\right),\\
y_{\rm tr} &= \pm\sqrt{1-z_0^2}\sin\left(\frac{\omega_0+1}{2}\tau+\varphi\right),\\
z_0 &= \left(\frac{\omega_0-1}{2\Omega}\right)^2,\\
E_{\rm tr} &= \mp\frac{2\Omega}{\omega_0-1}\sqrt{1-z_0^2}\cos\left(\frac{\omega_0+1}{2}\tau+\varphi\right),\\
P_{\rm tr} &= \pm\frac{2\Omega}{\omega_0-1}\sqrt{1-z_0^2}\sin\left(\frac{\omega_0+1}{2}\tau+\varphi\right),
\end{aligned} \tag{19}$$

with φ an arbitrary phase, describes an interaction process in which the inversion density z and the photon density $n=(E^2+P^2)/4$ remain unchanged. If in (19) we put $\tau=0$, we arrive at the initial conditions that lead to population and radiation trapping. This phenomenon occurs when the atoms are prepared in a superposition state with a certain phase and the field is prepared in a coherent state with the same phase. Note that the initial inversion z_0 can take any value from the upper Bloch hemisphere, $0\leqslant z_0\leqslant 1$. The solutions (19) are valid for arbitrary detuning from the resonance between the mode frequency and the atomic transition frequency.

The limit $z_0=0$ is realized at exact resonance, or $\omega_0=1$. This corresponds to the initial atomic superposition $|x_0=\cos\varphi,\ y_0=\sin\varphi,\ z_0=0\rangle$ and the initial field with an infinite number of photons, which can be interpreted as the phase state of the field. The other limit, $z_0=1$, corresponds to the initial state $|x_0=y_0=E_0=P_0=0,\ z_0=1\rangle$, which is the equilibrium point S_+ in the semiclassical approximation, since in this approximation all sources of spontaneous emission are discarded. Allowance for interatomic quantum correlations[9] results in a situation in which the state with completely excited atoms decays in the vacuum, leading in the strong atom–field coupling regime to Rabi vacuum oscillations.

It is well known[14] that population trapping is possible in the Jaynes–Cummings model,[3] which in the rotating wave approximation describes the interaction of a two-level atom and a single mode of the quantized field in the case of zero detuning. Then the population inversion can be expressed analytically by a series in the occupation numbers, whose numerical integration shows[14] that when the phase of the initial states of the atoms is equal to the phase of the field, the amplitude of the oscillations of z is extremely small. The reason for population trapping is the destructive interference of the atomic dipoles and the field mode. In the semiclassical limit, one can find the explicit expressions for the corresponding initial conditions for N identical atoms and the field, which in the case of exact resonance, $\omega_0=1$, lead to population and radiation trapping:

$$\begin{aligned}
x_0 &= \pm\sqrt{1-z_0^2}\cos\psi,\\
y_0 &= \pm\sqrt{1-z_0^2}\sin\psi, \quad z_0 = -\beta^2,\\
E_0 &= \mp\sqrt{z_0-z_0^{-1}}\cos\psi, \quad P_0 = \pm\sqrt{z_0-z_0^{-1}}\sin\psi,
\end{aligned} \tag{20}$$

where β is an arbitrary number from the interval $0\leqslant\beta^2\leqslant 1$, and ψ is an arbitrary phase. Obviously, trapping is possible at resonance if the initial Bloch vector is in the lower Bloch hemisphere. The limit $z_0=-1$ corresponds to the initial state $|x_0=y_0=E_0=P_0=0,\ z_0=-1\rangle$, which is the trivial equilibrium point S_- of system (2).

5. PARAMETRIC RESONANCES AND HAMILTONIAN CHAOS: NUMERICAL RESULTS

Let us now go back to the system (2) with modulation of the frequency detuning (6). In Sec. 3 we established that the integrable version of this system has special homoclinic trajectories (18). On the phase portraits of the integrable system, these trajectories constitute a family of two-dimensional homoclinic tori converging to the saddle point S_+ and parametrized by the value of the total energy W. Each torus is the location of states of the atom–field system with completely excited atoms that emit and absorb light for an infinitely long time in the cavity with initial vacuum. It is to be expected that when a perturbation is introduced into the system (in our case the perturbation is the frequency modulation), the steady and unsteady manifolds of the saddle motion begin to cross transversely, thus generating chaotic oscillations of the atom–field system.

An illustration of the destruction of invariant manifolds and the onset of chaos is provided by Poincaré sections. Figure 1 depicts the results of successively mapping the states of the system in the plane of the field variables E and P for different degrees of modulation α and fixed values of the other control parameters ($\omega_0=\delta=1$ and $\Omega=0.5$). The initial state of the system, $|x_0=y_0=0,\ z_0=1,\ E_0=P_0=0.1\rangle$, corresponding to the values $W_0=2.02$ and $C=0$ of the constants of motion is near the singular point S_+. The trajectories of motion of the integrable system ($\alpha=0$) corresponds to a regular Poincaré section (Fig. 1a). When modulation is switched on (even if the modulation is very low), a stochastic layer appears near the separatrix. As the modulation α increases, the stochastic layer widens, and this manifests itself in the corresponding sections (Figs. 1b and 1c). Finally, when α gets large enough, the trajectory fills the entire energy-permissible part of the (E,P) plane (Fig. 1d). All these figures depict Poincaré sections for a single trajectory. Clearly, for the selected initial value of the total energy,

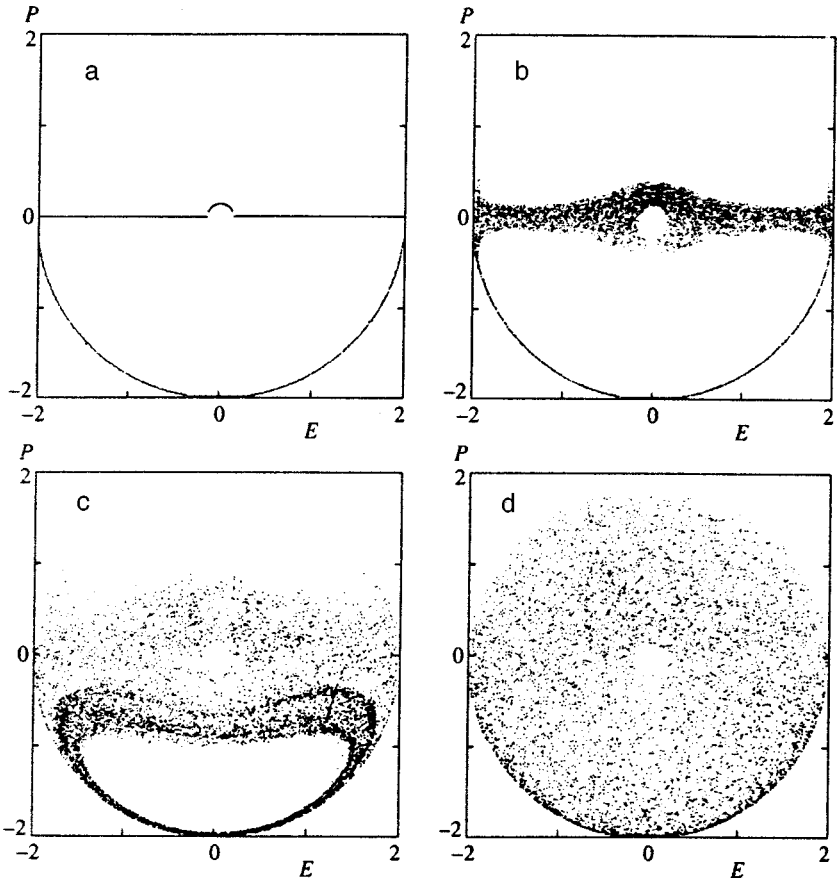

FIG. 1. Poincaré sections in the plane of the field variables E and P at $\delta=\omega_0=1$, $\Omega=0.5$, $W_0=2.02$, and $C=0$: (a) $\alpha=0$, (b) $\alpha=0.01$, (c) $\alpha=0.07$, and (d) $\alpha=0.2$.

$W=2.02$, the points of the mapping cannot land in the region near the center $E_0=P_0=0$ because of the restrictions imposed by the conservation laws (7).

Lyapunov indices serve as a quantitative characteristic of chaos, since they are the measure of local instability of a dynamical system. For systems with several control parameters, topographic λ-maps[6,7] usually provide a graphic picture of chaos. In these maps the values of the maximum Lyapunov index λ are depicted (usually by different shades of gray) in the form of a function of two control parameters at fixed values of the other parameters.

The results of calculating the maximum Lyapunov indices as functions of two characteristic frequencies of the atom–field system, the modulation frequency δ (on the log-linear scale) and the collective Rabi frequency Ω, are depicted in Figs. 2 and 3 at $\alpha=1$ and different values of the detuning amplitude, $\omega_0=0.9$ (Fig. 2) and $\omega_0=1$ (Fig. 3). The strength of Hamiltonian chaos is characterized by λ, whose scale is depicted in the right part of the figures by shades of gray. For the initial state we took $|x_0=y_0=0$, $z_0=1$, $E_0=P_0=1\rangle$ with completely excited atoms and a photon number density $n=1/2$. Areas of chaos are clearly visible on these maps, and the number of these areas and the chaos strength in them increase as the carrier frequency ω_f of the mode approaches the atomic transition frequency ω_a, i.e., as $\omega_0\to 1$. Under the same conditions the calculated (but not shown here) λ-map for $\omega_0=0.1$ suggests that the chaos is extremely weak (in the calculation range $0<\Omega\leqslant 5$, the maximum Lyapunov index is 0.006), its strength monotonically increases with Ω and is almost entirely independent of δ.

With moderate detuning from resonance, $\omega=0.9$, there are two areas of chaos (Fig. 2), whose centers in the (δ,Ω) plane correspond to a frequency ratio $\delta/2\Omega\simeq 1$. Here Ω is simply the frequency $\Omega_0/\sqrt{N}$ of natural oscillations of the unperturbed atom–field system divided by ω_a [see Eq. (5)]. Thus, $\delta/2\Omega\simeq 1$ is the condition for the fundamental parametric resonance. At exact atom–mode resonance, $\omega_0=1$, new areas of chaos corresponding to the fundamental resonance and the first subresonance with $\delta/2\Omega\sim 1/2$ appear (Fig. 3). The $\times$ in Figs. 2–4 indicate the contours of the parameter $\delta/2\Omega$, while the number in the parentheses indicates the value of this ratio, (1/2), (1), and (2), corresponding to a subresonance, the fundamental resonance, and a superresonance, respectively. It is quite natural that the stochastization of the atom–field system is the strongest when the modulation frequency and the Rabi frequency satisfy the condition for the fundamental parametric resonance. However, from the entire parametric instability area specified by the ratio $\delta/2\Omega\simeq 1$ we select values of δ that are integral multiples of the other characteristic frequencies of the system, ω_a and ω_f.

As expected, a decrease in the modulation narrows the areas of parametric instability of the system. Figure 4 depicts the λ-map in the (δ,Ω) plane for the same conditions as in Fig. 3 but for a lower modulation, $\alpha=0.2$. The broad area of chaos at $\alpha=1$ (Fig. 3) disintegrates into a chain of small

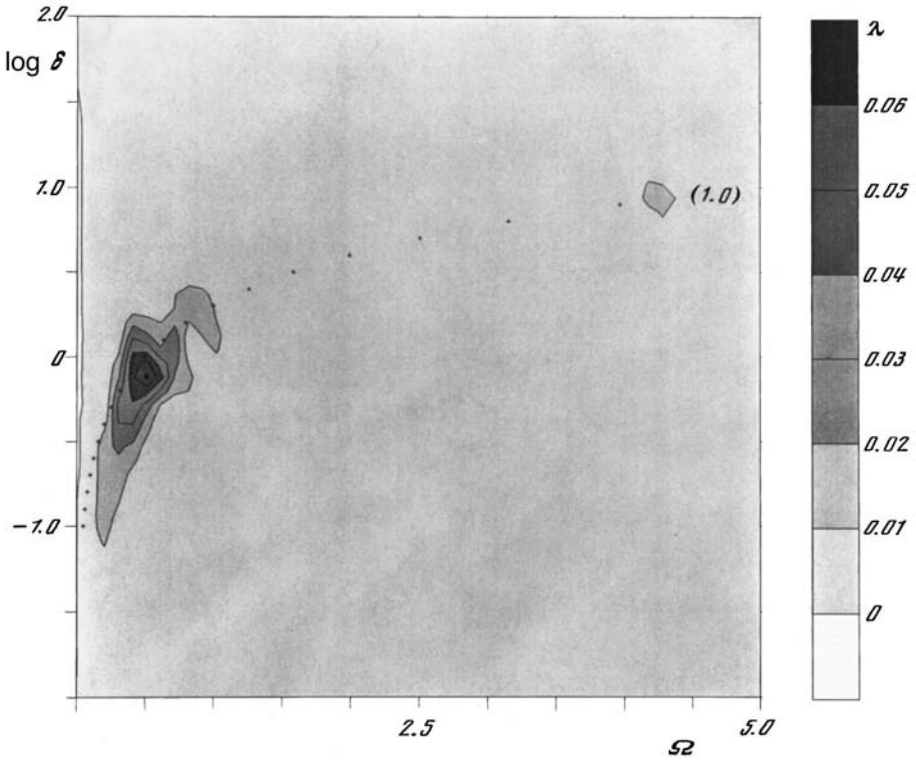

FIG. 2. Topographic map of maximum Lyapunov indices λ as functions of the modulation frequency δ (on the log-linear scale) and the Rabi oscillation frequency Ω at $\alpha=1$ and $\omega_0=0.9$.

''islets'' of chaos at $\alpha=0.2$ (Fig. 4). The chaos in these islets is much weaker than the chaos in the area: the corresponding maximum Lyapunov index is less than 0.01 in the islets.

As is well known,[15] Lyapunov indices are the measure of the rate of deformation of the phase volume along different directions. An m-dimensional system has m Lyapunov indices (λ_i, $i=1,2,\ldots,m$) ordered according to their values, $\lambda_1\leqslant\lambda_2\leqslant\cdots\leqslant\lambda_m\equiv\lambda$, and generally depending on the initial conditions. If $\lambda_j>0$ ($\lambda_j<0$), the volume element increases (decreases) in the respective direction as the system evolves. In a Hamiltonian system (which our semiclassical atom–field model is), the phase volume in an invariant measure, so that $\Sigma_{i=1}^{m}\lambda_i=0$.

Positive Lyapunov indices determine the Krylov–Kolmogorov–Sinaĭ metric entropy $h=\Sigma_{j=1}^{n<m}\lambda_j^{(+)}$ (see Ref.

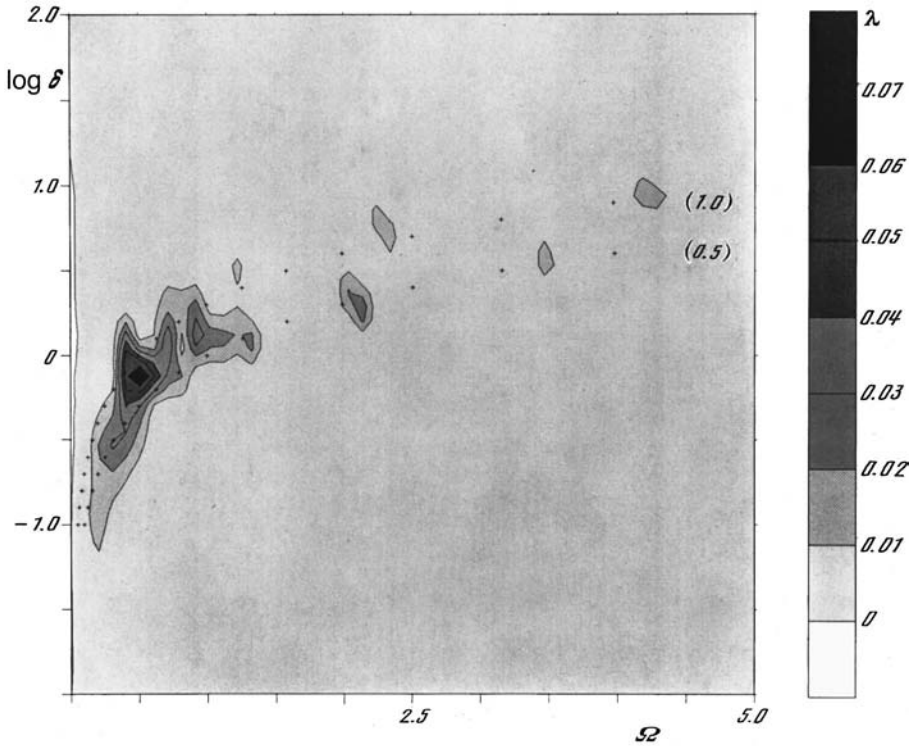

FIG. 3. The same as in Fig. 2 for $\omega_0=1$ and $\alpha=1$.

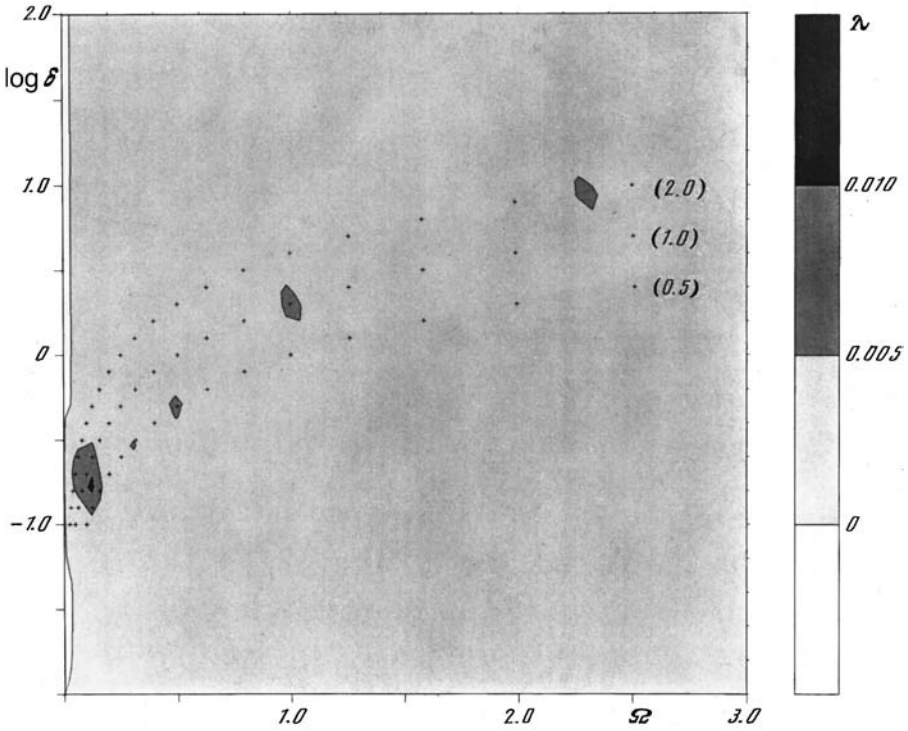

FIG. 4. The same as in Fig. 2 for $\omega_0=1$ and $\alpha=0.2$.

16). According to the algorithmic theory of dynamical systems,[17,18] the entropy h is formally equivalent to the algorithmic complexity K of a trajectory, a quantity that is the measure of the randomness of the sequence generated by this trajectory. On the other hand, the purely dynamical characteristic λ of the system is related to a statistical characteristic, the reciprocal correlation-decoupling time $1/\tau_c$, which is present in the expression for the correlation function $C(t)=C_0\exp(-t/\tau_c)$. As a result, the maximum Lyapunov index λ, the metric entropy h, the algorithmic complexity K, and the reciprocal correlation-decoupling time $1/\tau_c$ are quantities of the same order. Thus, statistical laws can emerge as result of local instability in the chaotic dynamics.[19,20]

6. CONCLUSION

We have studied the nonlinear dynamics of the interaction of an ensemble of identical two-level atoms and a selected cavity mode. In the rotating wave approximation, the semiclassical approximation, and the strong-coupling limit, we have derived explicitly the general exact solution of the respective Heisenberg equations for the expectation values of the atomic and field observables for an arbitrary detuning from resonance. Using this solution, we detected the coherent effect of trapping of the population of atomic levels and of trapping of the radiation in the cavity, an effect that arises because of destructive interferences of the atomic dipoles prepared in a superposition state with a certain phase and the field mode prepared in a coherent state with the same phase.

By using Poincaré sections we have found that the modulation of the frequency of the selected mode leads, even in the rotating wave approximation, to the formation of a stochastic layer near the separatrix of the integrable model, and the width of the layer increases with the modulation. The results of the numerical investigation of Hamiltonian chaos in the atom–field system are represented by topographic λ-maps in which the areas of equal values of the maximum Lyapunov index of the system λ are depicted (by different shades of gray) as functions of the values of two control parameters. The stochastization of the system is the strongest in the vicinity of the fundamental parametric resonance ($\delta/2\Omega\simeq 1$), when the modulation frequency is approximately two times higher than the frequency with which the atoms and the mode exchange energy. Parametric instability increases as the detuning from the resonance between the atomic and field frequencies decreases ($\omega\to 1$), and at exact resonance new areas of chaos corresponding to multiple parametric resonances arise. As the modulation α decreases, the system becomes more stable, which manifests itself in the disintegration of the broad area of chaos in the (δ,Ω) plane at $\alpha=1$ (Fig. 3) into a chain of small ''islets'' of chaos at $\alpha=0.2$ (Fig. 4) and in a sharp decrease in the values of λ.

Thus, in the semiclassical Jaynes–Cummings model with parametric modulation, Hamiltonian chaos appears even in the rotating wave approximation. While in a system without the rotating wave approximation chaos is a threshold effect, appearing for $\Omega\geqslant 0.5$–1 (see Refs. 4 and 6), in the parametric model, as shown by our numerical experiments, measurable chaos appears at much smaller values of the Rabi frequency, $\Omega\simeq 0.05$. This means that the dynamics of Rydberg atoms with a typical active transition frequency $\omega_a\simeq 10^{11}\,\mathrm{rad\,s^{-1}}$ that interact with frequency-modulated microwave radiation in the millimeter range becomes parametrically unstable, provided that $\Omega_0\sqrt{N}\geqslant 10^{10}$. Obviously, this must be reflected in the broadening of the Rabi oscillation spectrum. Thus, chaos appears at typical (in the strong-coupling limit) values of the one-photon vacuum Rabi frequency of Rydberg atoms, $\Omega_0\simeq 10^5\,\mathrm{rad\,s^{-1}}$ (see Ref. 1), when the number of these atoms is relatively small,

$N \simeq 10^{10}$. For such values of N, the rotating wave approximation, the semiclassical approximation, the approximation of noninteracting atoms, and the assumption that the linear dimensions of the atomic sample are small may be considered adequate.

What are the prospect for studying quantum–classical correspondence using Hamiltonian chaos in atom–field systems in cavities as an example? The Heisenberg approach makes it possible to take the quantum correlations and fluctuations into account (see the Appendix and the Refs. 8, 9, 21, and 22 cited therein). In a recent paper of ours,[8] we derive a closed system*o*equations that describes in the rotating wave approximation the dynamics of the interaction of two- level atoms and a cavity mode with allowance for the field, interatomic, and first-order atom–field quantum correlators.

In a future publication we plan to obtain the exact solution of the integrable version of this *semiquantum system*, study the onset of chaos in such a system in the presence of parametric modulation, and compare the results of the semiquantum and semiclassical analyses.

This work was supported by the Russian Fund for Fundamental Research (Projects 96-02-19827 and 96-02-18746)

APPENDIX

We introduce new operators normalized to the number N of atoms by the formulas

$$A=\frac{1}{\sqrt{N}}a, \quad A^{\dagger}=\frac{1}{\sqrt{N}}a^{\dagger},$$

$$S_x=\frac{1}{N}\left\langle \sum_{j=1}^{N} \sigma_x^j \right\rangle, \quad S_y=\frac{1}{N}\left\langle \sum_{j=1}^{N} \sigma_y^j \right\rangle, \tag{21}$$

$$S_z=\frac{1}{N}\left\langle \sum_{j=1}^{N} \sigma_z^j \right\rangle.$$

The commutators of these operators are

$$[A,A^{\dagger}]=\frac{1}{N}, \quad [S_i,S_j]=\frac{2i}{N}e_{ijk}S_k,$$

$$i,j,k=x,y,z. \tag{22}$$

In terms of the new operators and the collective vacuum Rabi frequency Ω_N, the Heisenberg equations are

$$\frac{dS_x}{dt}=-\omega_a S_y-\Omega_N S_z i(A^{\dagger}-A),$$

$$\frac{dS_y}{dt}=\omega_a S_x-\Omega_N S_z(A^{\dagger}+A),$$

$$\frac{dS_z}{dt}=i\Omega_N S_x(A^{\dagger}-A)+\Omega_N S_y(A^{\dagger}+A), \tag{23}$$

$$\frac{dA}{dt}=-i\omega_f A-\frac{i}{2}\Omega_N(S_x-iS_y),$$

$$\frac{dA^{\dagger}}{dt}=i\omega_f A^{\dagger}+\frac{i}{2}\Omega_N(S_x+iS_y).$$

In the semiclassical limit, the equations of motion for the corresponding expectation values of the operators,

$$\mathscr{X}=\langle S_x\rangle, \quad \mathscr{Y}=\langle S_y\rangle, \quad \mathscr{Z}=\langle S_z\rangle,$$
$$\mathscr{E}=\langle A^{\dagger}+A\rangle, \quad \mathscr{P}=i\langle A^{\dagger}-A\rangle, \tag{24}$$

in terms of the dimensionless time $\tau=\omega_a t$ have the form

$$\dot{\mathscr{X}}=-\mathscr{Y}-\Omega\mathscr{Z}\mathscr{P}, \quad \dot{\mathscr{Y}}=\mathscr{X}-\Omega\mathscr{Z}\mathscr{E},$$

$$\dot{\mathscr{Z}}=\Omega(\mathscr{X}\mathscr{P}+\mathscr{Y}\mathscr{E}), \quad \dot{\mathscr{E}}=\omega\mathscr{P}-\Omega\mathscr{Y},$$

$$\dot{\mathscr{P}}=-\omega\mathscr{E}-\Omega\mathscr{X}. \tag{25}$$

The commutators of the normalized operators vanish when N is large. Obviously, in this limit the semiclassical approximation is valid. For N finite, the relative error amounts to a quantity of order $1/N$. The equation of motion for these operators and the expectation values are independent of N. Using the normalization (21) and the semiclassical dynamical system, we can derive a hierarchy of dynamical systems in which the quantum corrections due to quantum correlations and fluctuations are taken into account explicitly. Indeed, when the commutators (22) do not vanish, the dependence on the small parameter $1/N$ manifests itself only in the quantum correlators. The quantum corrections can be taken into account by the $1/N$-expansion method by introducing cumulants of order $(1/N)^n$ (a description of how to use this approach in the Jaynes–Cummings model can be found in Ref. 21), a method based on averaging the operators over generalized coherent states at the initial moment (see Ref. 22 for a description of the Dicke model in an external coherent field), or by other methods[9] (see Ref. 8 for a description of the Dicke model with moving atoms).

*)E-mail: prants@medphys.marine.su

[1] J. M. Raimonds and S. Haroche, in *Confined Electrons and Photons*, E. Burstein and C. Weisbuch (Eds.), Plenum Press, New York (1995), p. 383.
[2] Y. Kaluzny, P. Goy, M. Gross, J. M. Raimonds, and S. Haroche, Phys. Rev. Lett. **51**, 1175 (1983); G. Rempe, H. Walter, and N. Klein, Phys. Rev. Lett. **58**, 3523 (1987).
[3] E. T. Jaynes and F. W. Cummings, Proc. IEEE **51**, 89 (1963).
[4] P. I. Belobrov, G. M. Zaslavskiĭ, and G. Kh. Tartakovskiĭ, Zh. Éksp. Teor. Fiz. **71**, 1799 (1976) [Sov. Phys. JETP **44**, 945 (1976)].
[5] R. F. Fox and J. C. Eidson, Phys. Rev. A **34**, 482 (1986).
[6] L. E. Kon'kov and S. V. Prants, J. Math. Phys. **37**, 1204 (1986).
[7] S. V. Prants and L. E. Kon'kov, Phys. Lett. A **225**, 33 (1997).
[8] L. E. Kon'kov and S. V. Prants, JETP Lett. **65**, 833 (1997).
[9] A. V. Andreev, V. I. Emel'yanov, and Yu. A. Il'inskiĭ, *Cooperative Phenomena in Optics: Superradiation, Bistability, and Phase Transitions* [in Russian], Nauka, Moscow (1988).
[10] I. L. Kirilyuk, L. E. Kon'kov, and S. V. Prants, submitted to Rep. Math. Phys.
[11] G. T. Moore, J. Math. Phys. **11**, 2679 (1970).
[12] V. V. Dodonov, A. B. Klimov, and V. I. Man'ko, P. N. Lebedev Physics Institute Preprint No. 41, Moscow (1991).
[13] B. S. Dewitt, Phys. Rep. **19**, 295 (1975); V. V. Hizhnyakov, Quantum Opt. **4**, 277 (1992).
[14] K. Zaheer and M. S. Zubairy, Phys. Rev. A **39**, 2000 (1989).
[15] V. I. Oseledec, Trudy Mosk. Matem. Obshch. **19**, 179 (1968) [English transl. in *Trans. Moscow Math. Soc.*, Vol. 19, American Mathematical Society, Providence, R.I. (1969)].

[16] Yu. B. Pesin, Dokl. Akad. Nauk SSSR **226**, 774 (1976) [*sic*].
[17] A. A. Brudno, Usp. Mat. Nauk **33**, 207 (1978).
[18] V. M. Alekseev and M. V. Yakobson, Phys. Rep. **75**, 287 (1981).
[19] D. V. Anosov and Ya. G. Sinaĭ, Usp. Mat. Nauk **22**, 107 (1967).
[20] G. M. Zaslavskiĭ and B. V. Chirikov, Usp. Fiz. Nauk **105**, 3 (1971) [Sov. Phys. Usp. **14**, 549 (1972)].
[21] A. Heidmann, J. M. Raimond, S. Reynaud, and N. Zagury, Opt. Commun. **54**, 189 (1985).
[22] G. P. Berman, E. N. Bulgakov, and D. D. Holm, Phys. Rev. A **49**, 4943 (1974).

Translated by Eugene Yankovsky

Superfluidity of ^{3}He in aerogel at $T=0$ in a magnetic field

G. Baramidze and G. Kharadze*)

Institute of Physics, Georgian Academy of Sciences, 380077, Tbitisi, Georgia
(Submitted 3 August 1998)
Zh. Éksp. Teor. Fiz. **115**, 754–762 (February 1999)

The transition of liquid ^{3}He to the superfluid B phase in aerogel at $T=0$ is considered. It is shown that in a magnetic field, the quantum phase transition with respect to pressure is split in two. The amount of splitting δP is estimated. The components of the superfluid density tensor are calculated near the critical pressures. © *1999 American Institute of Physics.* [S1063-7761(99)02602-5]

1. The behavior of superfluid ^{3}He in a silica aerogel environment is a subject of recent experimental investigations.[1–4] The scattering of quasiparticles on a random network of SiO_2 strands affects superfluid correlations, thus considerably modifying the phase diagram of liquid ^{3}He in the millikelvin temperature range. Interesting observations[4] were made concerning the behavior of ^{3}He at $T\simeq 0$ where, in contrast to bulk liquid ^{3}He, superfluidity shows up only above some critical density ρ_c (at pressures $P>P_c$).

The situation at $T=0$ for ^{3}He in aerogel was recently considered theoretically.[5] It is shown that according to a simple model with scattering effects characterized by the quasiparticle mean free path $l=v_F\tau$, the critical pressure P_c is given by the equation

$$T_{c0}(P_c)=\gamma_E/\pi\tau, \quad \ln\gamma_E=C\simeq 0.577, \tag{1}$$

where $T_{c0}(P)$ is the P-dependent critical temperature of the transition of bulk ^{3}He to the superfluid state. Near P_c the gap function (the order parameter) is

$$\Delta^2(P)=\frac{3}{\tau^2}\,\theta(P-P_c)\ln\frac{T_{c0}(P)}{T_{c0}(P_c)}. \tag{2}$$

The investigation carried out in Ref. 5 is based on the assumption that the superfluid state at $P>P_c$ is of the B-phase type. It should be remembered that the appearance of a B-phase-like state in aerogel at low pressures is expected when the magnetic contribution to quasiparticle scattering events is suppressed by ^{4}He layers covering the silica strands.[3] In what follows we extend the results of Ref. 5 to the B phase in a magnetic field. Our obvious motivation is to explore an expected magnetic splitting of quantum phase transition at $T=0$. The behavior of magnetically distorted ^{3}He-B in aerogel at $T\simeq T_c$ has been considered by us in Ref. 6.

2. In what follows we use quasiclassical Green's functions (the ξ-integrated Gorkov functions) in the Matsubara representation:

$$\hat g_\omega(\hat{\mathbf k})=\frac{1}{i\pi}\int \hat G_\omega(\hat{\mathbf k},\xi)\,d\xi=g_\omega\hat{\mathbf l}+\mathbf g_\omega\hat{\boldsymbol\sigma},$$

$$\hat f_\omega(\hat{\mathbf k})=\frac{1}{\pi}\int \hat F_\omega(\hat{\mathbf k},\xi)\,d\xi=(f_\omega\hat{\mathbf l}+\mathbf f_\omega\hat{\boldsymbol\sigma})i\hat\sigma_y, \tag{3}$$

where the unit vector $\hat{\mathbf k}$ specifies the location on the normal state Fermi sphere.

In an external magnetic field $\mathbf H=H_0\hat{\mathbf h}$, the functions $\hat g_\omega$ and $\hat f_\omega$ satisfy a set of equations ($\omega_0=\omega_0\hat{\mathbf h}$, $\omega_0=\gamma H_0$):

$$g_\omega\boldsymbol\Delta_t+\mathbf g_\omega\boldsymbol\Delta_s+\omega\mathbf f_\omega-\frac{i}{2}\,\omega_0 f_\omega=\mathbf 0,$$

$$g_\omega\boldsymbol\Delta_s+\mathbf g_\omega\boldsymbol\Delta_t+\omega f_\omega-\frac{i}{2}\,\omega_0\mathbf f_\omega=0. \tag{4}$$

Here pin-singlet and spin-triplet order parameters Δ_s and Δ_t, respectively, are the components of the matrix

$$\hat\Delta(\hat{\mathbf k})=(\Delta_s\hat{\mathbf l}+\boldsymbol\Delta_t\hat{\boldsymbol\sigma})i\hat\sigma_y, \tag{5}$$

which is found according to the self-consistency equation

$$\hat\Delta(\hat{\mathbf k})=\pi T\sum_\omega\langle V(\hat{\mathbf k},\hat{\mathbf k}')\hat f_\omega(\hat k')\rangle, \tag{6}$$

where $V(\hat{\mathbf k},\hat{\mathbf k}')$ is the Cooper pairing interaction and angle brackets denote averaging across the Fermi surface.

The set of equations (4) should be supplemented by the «boundary» conditions

$$g_\omega^2+\mathbf g_\omega^2+f_\omega^2+\mathbf f_\omega^2=1, \quad g_\omega\mathbf g_\omega+f_\omega\mathbf f_\omega=\mathbf 0. \tag{7}$$

The structure of Eqs. (4) and (7) implies that

$$f_\omega=a_\omega g_\omega, \quad \mathbf f_\omega=\mathbf a_\omega g_\omega, \quad \mathbf g_\omega=-a_\omega\mathbf a_\omega g_\omega, \tag{8}$$

and the solution is easily obtained:

$$g_\omega=-\frac{\operatorname{sign}\omega}{\sqrt{1+\mathbf a_\omega^2}\sqrt{1+\mathbf a_\omega^2}},$$

$$a_\omega=\frac{i}{2}\,\frac{2i\Delta_s+\omega_0\mathbf a_\omega}{\omega-\Delta_t\mathbf a_\omega}, \quad \mathbf a_\omega=\frac{i}{2}\,\frac{2i\Delta_t+\omega_0 a_\omega}{\omega-\Delta_s a_\omega}. \tag{9}$$

This is our starting point when considering the properties of the magnetized B phase. In order to take into account the quasiparticles scattering on aerogel spatial irregularities ($1/\tau\neq 0$), the ''impurity''-induced renormalization of the

frequencies ω and ω_0 and of the order parameter $\hat{\Delta}$ is to be performed according to a standard prescription:

$$\omega \to \tilde{\omega} = \omega + iM_{\tilde{\omega}}, \quad \omega_0 \to \tilde{\omega}_0 = \omega_0 - 2\hat{\mathbf{h}}M_{\tilde{\omega}},$$
$$\Delta(\hat{\mathbf{k}}) \to \hat{\Delta}_{\tilde{\omega}} = \hat{\Delta}(\hat{\mathbf{k}}) + \hat{m}_{\tilde{\omega}}, \tag{10}$$

where the ''impurity''-generated self-energies are given by

$$\hat{M}_\omega = M_\omega \hat{\mathbf{I}} + \mathbf{M}_\omega \hat{\boldsymbol{\sigma}} = \frac{i}{2\tau}\langle \hat{g}_\omega(\hat{\mathbf{k}})\rangle,$$
$$\hat{m}_\omega = (m_\omega \hat{\mathbf{I}} + \mathbf{m}_\omega \hat{\boldsymbol{\sigma}}) i\hat{\sigma}_y = \frac{1}{2\tau}\langle \hat{f}_\omega(\hat{\mathbf{k}})\rangle. \tag{11}$$

Before proceeding, an important comment should be made. Since we are going to consider a superfluid state with p-wave Cooper pairing for which $V(\hat{\mathbf{k}},\hat{\mathbf{k}}') = 3g(\hat{\mathbf{k}}\hat{\mathbf{k}}')$, from Eq. (6) it is certainly clear that $\Delta_s \equiv 0$ and $\boldsymbol{\Delta}_t$ satisfies the equation

$$\boldsymbol{\Delta}_t(\hat{\mathbf{k}}) = g\pi T \sum_\omega \langle 3\hat{\mathbf{k}}\hat{\mathbf{k}}' \mathbf{f}_{\tilde{\omega}}(\hat{\mathbf{k}}')\rangle. \tag{12}$$

The does not mean, however, that in our starting expressions for a_ω and $\mathbf{a}_\omega$ we have to forget about the presence of Δ_s. The point is that when considering the ''impurity'' renormalization $\hat{m}_\omega$, the contribution stemming from the spin-singlet part m_ω must be taken into account. This involves the calculation of the superfluid density tensor $\rho_{ij}^{(s)}$, which is a response of the system to the superfluid velocity field $\mathbf{v}_s$ and is contained in the expression for the supercurrent

$$\mathbf{j}_s = 2\pi i k_F N_F T \sum_\omega \langle \hat{\mathbf{k}} g_{\tilde{\omega}+iq}(\hat{\mathbf{k}})\rangle. \tag{13}$$

Here $\mathbf{v}_s$ is absorbed in $q(\hat{\mathbf{k}}) = k_F(\hat{\mathbf{k}}\mathbf{v}_s)$ and N_F denotes the quasiparticle density of states at the Fermi level. As we noted in Ref. 6, to first order in q, the spin-singlet part $m_{\tilde{\omega}+iq}$ is proportional to $(1/\tau)\omega_0 q$ and contributes to $\rho_{ij}^{(s)}$ in the magnetized B phase in the aerogel environment.

3. Now we turn to the calculation of equilibrium properties of magnetized ^{3}He-B in a quasiparticle scattering medium. Noticing that in the absence of superflow ($q=0$) $\hat{\Delta}(\hat{\mathbf{k}})$ is not renormalized in nonmagnetic scattering, and addressing Eqs. (9), it can be shown that to lowest order in the magnetic field strength,

$$\mathbf{f}_{\tilde{\omega}}(\hat{\mathbf{k}}) = \frac{1}{\sqrt{\tilde{\omega}^2 + \boldsymbol{\Delta}^2(\hat{\mathbf{k}})}} \times \left[\Delta - \frac{1}{4}\left(\boldsymbol{\Delta}_\parallel - \frac{3}{2}\frac{\boldsymbol{\Delta}_\parallel^2}{\tilde{\omega}^2+\boldsymbol{\Delta}^2}\boldsymbol{\Delta}\right)\frac{\tilde{\omega}_0^2}{\tilde{\omega}^2+\boldsymbol{\Delta}^2}\right]. \tag{14}$$

In this expression the longitudinal component of $\Delta = \Delta_\parallel + \Delta_\perp$ is given by

$$\boldsymbol{\Delta}_\parallel(\hat{\mathbf{k}}) = (\hat{\mathbf{h}}\boldsymbol{\Delta}(\hat{\mathbf{k}}))\hat{\mathbf{h}} = \Delta_\parallel(P,T)(\hat{\mathbf{l}}\hat{\mathbf{k}})\hat{\mathbf{h}}, \tag{15}$$

where the magnetic-field induced orbital anisotropy axis $\hat{\mathbf{l}}$ is defined as $\hat{l}_i = \hat{h}_\mu R_{\mu i}$ with $R_{\mu i}$ being the components of an orthogonal matrix of 3D relative spin-orbit rotations. We note also that

$$\Delta^2(\hat{\mathbf{k}}) = \Delta_\parallel^2(P,T)(\hat{\mathbf{l}}\hat{\mathbf{k}})^2 + \Delta_\perp^2(P,T)(\hat{\mathbf{l}}\times\hat{\mathbf{k}})^2, \tag{16}$$

and in zero magnetic field $\Delta_\parallel = \Delta_\perp = \Delta$.

Using Eqs. (12) and (14), and taking into account that

$$\tilde{\omega} = \omega - \frac{1}{2\tau}\langle g_{\tilde{\omega}}(\hat{\mathbf{k}})\rangle \tag{17}$$

with

$$g_{\tilde{\omega}} \simeq -\frac{\tilde{\omega}}{\sqrt{\tilde{\omega}^2+\boldsymbol{\Delta}^2}}\left[1 + \frac{3}{8}\frac{\tilde{\omega}_0^2}{\tilde{\omega}^2+\boldsymbol{\Delta}^2}\frac{\boldsymbol{\Delta}_\parallel^2}{\tilde{\omega}^2+\boldsymbol{\Delta}^2}\right], \tag{18}$$

equations for amplitudes $\Delta_\parallel(P,T)$ and $\Delta_\perp(P,T)$ can be readily derived.

Let us consider first the planar phase ($\Delta_\parallel = 0$). Near the transition to the normal state for $\Delta_\perp(\hat{\mathbf{k}})$ we obtain

$$\left[\ln\frac{T}{T_{c0}} + \psi\left(\frac{1}{2}+w\right) - \psi\left(\frac{1}{2}\right)\right]\boldsymbol{\Delta}_\perp(\hat{\mathbf{k}}) = \frac{3}{4}\frac{\psi^{(2)}(1/2+w)}{(2\pi T)^2}\langle \hat{\mathbf{k}}\hat{\mathbf{k}}'|\boldsymbol{\Delta}_\perp(\hat{\mathbf{k}}')|^2\boldsymbol{\Delta}_\perp(\hat{\mathbf{k}}')\rangle + \frac{w}{12}\frac{\psi^{(3)}(1/2+w)}{(2\pi T)^2}\langle \boldsymbol{\Delta}_\perp^2\rangle\boldsymbol{\Delta}_\perp(\hat{\mathbf{k}}), \tag{19}$$

where $w(T) = 1/4\pi T\tau$. After simple averaging we obtain an equation for $\Delta_\perp(P,T)$:

$$\Delta_\perp\left\{\ln\frac{T_{c0}}{T} + \psi\left(\frac{1}{2}\right) - \psi\left(\frac{1}{2}+w\right) + \left[\frac{1}{5}\psi^{(2)}\left(\frac{1}{2}+w\right) + \frac{\omega}{18}\psi^{(3)}\left(\frac{1}{2}+w\right)\right]\left(\frac{\Delta_\perp}{2\pi T}\right)^2\right\} = 0. \tag{20}$$

In the limit $T \to 0 (w \to \infty)$, it is found that for the planar phase

$$\Delta_\perp^2(P) = \frac{15}{16}\Delta^2(P), \tag{21}$$

where $\Delta^2(P)$ is given by the Mineev solution (at $\omega_0 = 0$), Eq. (2). The coefficient 15/16 in Eq. (21) is due to the averaging at $\Delta_\parallel = 0$ (which gives an answer analogous to the A phase).

As will be seen below, the solution (21) extends up to the pressure $P = P_\parallel$ where $\Delta_\parallel$ first appears. The new critical pressure $P_\parallel$ is given by

$$\ln\frac{T_{c0}(P_\parallel)}{T_{c0}(P_c)} = \frac{4}{9}(\omega_0\tau)^2. \tag{22}$$

To show this we turn back to Eqs. (12) and (14) and, after simple calculations, a set of equations for $\Delta_\perp$ and $\Delta_\parallel$ is obtained (again in the limit $T=0$):

$$\Delta_\perp\left[\Delta_\perp^2 + \frac{1}{15}(\Delta_\perp^2 - \Delta_\parallel^2) - \Delta^2\right] = 0,$$
$$\Delta_\parallel\left[\Delta_\parallel^2 + \frac{2}{15}(\Delta_\parallel^2 - \Delta_\perp^2) - \Delta^2 + \frac{3}{2}\omega_0^2\right] = 0. \tag{23}$$

The solution of Eqs. (23) for $\Delta_\parallel \neq 0$, $\Delta_\perp \neq 0$ is ($P \geqslant P_\parallel$)

$$\Delta_{\|}^2(P)=\Delta^2-\frac{4}{3}\omega_0^2=\frac{3}{\tau^2}\ln\frac{T_{c0}(P)}{T_{c0}(P_{\|})},$$

$$\Delta_{\perp}^2(P)=\Delta_{\|}^2(P)+\frac{5}{4}\omega_0^2. \tag{24}$$

It can be verified that $\Delta_{\perp}^2(P_{\|})$ matches the solution (24)

Finally we conclude that at $P<P_c$ the normal state is realized; in the pressure range $P_c<P<P_{\|}$ the planar phase is stabilized; and at $P>P_{\|}$ a magnetically distorted B phase appears.

At $P_{\|}$ a discontinuity of the magnetic susceptibility takes place (similar to a discontinuity of the compressibility at P_c). In order to demonstrate this property the superfluid contribution Φ_S to the thermodynamic potential density is to be constructed. In the Ginzburg–Landau region (which we consider)

$$\Phi_S=\frac{2}{3}N_F\left\{-\ln\frac{T_{c0}(P)}{T_{c0}(P_c)}\Delta_{\perp}^2-\frac{1}{2}\left[\ln\frac{T_{c0}(P)}{T_{c0}(P_c)}-\frac{1}{2}(\omega_0\tau)^2\right]\Delta_{\|}^2+\frac{\tau^2}{6}\left[\Delta_{\perp}^4+\frac{1}{2}\Delta_{\|}^4+\frac{1}{15}(\Delta_{\perp}^2-\Delta_{\|}^2)^2\right]\right\}. \tag{25}$$

It is easily verified that the solutions of Eqs. (23) realize the minima of Φ_S.

At zero magnetic field $(\omega_0=0)\Delta_{\perp}=\Delta_{\|}=\Delta$, and in this case (see Ref. 5)

$$\Phi_S=\Phi_{S0}=N_F\left[-\ln\frac{T_{c0}(P)}{T_{c0}(P_c)}\Delta^2+\frac{\tau^2}{6}\Delta^4\right]. \tag{26}$$

It can be shown that in equilibrium the magnetic field contribution $\Phi_{SM}=\Phi_S-\Phi_{S0}$ is

$$\Phi_{SM}^{(equ)}=\frac{1}{\tau^2}N_F\times\begin{cases}\dfrac{9}{16}\ln^2\dfrac{T_{c0}(P)}{T_{c0}(P_c)}, & P_c\leqslant P\leqslant P_{\|},\\ \dfrac{1}{2}\left[\ln\dfrac{T_{c0}(P)}{T_{c0}(P_c)}-\dfrac{2}{9}(\omega_0\tau)^2\right](\omega_0\tau)^2, & P\geqslant P_{\|}.\end{cases} \tag{27}$$

This expression is certainly continuous at $P=P_{\|}$. On the other hand,

$$\frac{\partial^2\Phi_{SM}^{(equ)}}{\partial H_0^2}=\begin{cases}0, & P=P_{\|}^-,\\ -\dfrac{4}{9}\gamma^2N_F(\omega_0\tau)^2, & P=P_{\|}^+,\end{cases} \tag{28}$$

which signals the discontinuity of magnetic susceptibility at $P_{\|}$.

Now we shall estimate the value of $\delta P=P_{\|}-P_c$, which characterizes the magnetic splitting of the superfluid phase transition of ^{3}He in aerogel at $T=0$. According to Eq. (22),

$$\frac{1}{T_{c0}(P_c)}\left(\frac{\partial T_{c0}}{\partial P}\right)_{P_c}\delta P\simeq\frac{4}{9}(\omega_0\tau)^2, \tag{29}$$

so that $\delta P=\alpha H_0^2$, where

$$\alpha=\left(\frac{2\gamma_E}{3\pi}\right)^2\frac{(\hbar\gamma/k_B)^2}{T_{c0}(P_c)(\partial T_{c0}/\partial P)_{P_c}}. \tag{30}$$

Using experimental data on $T_{c0}(P)$, we find that at $P_c=7$ bar the coefficient $\alpha\simeq2\cdot10^{-2}$ bar/(kG)^2 and increases gradually to $\alpha\simeq2.8\cdot10^{-2}$ bar/(kG)^2 at $P_c=10$ bar.

4. Now we turn to the calculation of the superfluid density tensor

$$\rho_{ij}^{(s)}=\rho_{\|}^{(s)}\hat{l}_i\hat{l}_j+\rho_{\perp}^{(s)}(\delta_{ij}-\hat{l}_i\hat{l}_j). \tag{31}$$

For this purpose Eq. (13) is to be used and $g_{\tilde{\omega}+iq}$ should be constructed. Returning to Eqs. (9) and performing ''impurity'' renormalizations according to Eqs. (10), it can be shown that

$$g_{\tilde{\omega}+iq}^{(1)}=-i\left\{\frac{\mathbf{\Delta}^2q+i\tilde{\omega}(\mathbf{\Delta}\mathbf{m}_{\tilde{\omega}}^{(1)})}{(\tilde{\omega}^2+\mathbf{\Delta}^2)^{3/2}}-\frac{3}{2}\frac{\tilde{\omega}_0^2}{(\tilde{\omega}^2+\Delta^2)^{5/2}}\times\left(1-\frac{5}{4}\frac{\Delta^2}{\tilde{\omega}^2+\Delta^2}\right)\Delta_{\|}^2q-\frac{3}{4}\frac{i\tilde{\omega}\tilde{\omega}_0^2}{(\tilde{\omega}^2+\Delta^2)^{5/2}}\times\left[\mathbf{\Delta}_{\|}\mathbf{m}_{\tilde{\omega}}^{(1)}-\frac{5}{2}\frac{\mathbf{\Delta}_{\|}^2}{\tilde{\omega}^2+\Delta^2}(\mathbf{\Delta}\mathbf{m}_{\tilde{\omega}}^{(1)})\right]-\frac{\tilde{\omega}_0}{(\tilde{\omega}^2+\Delta^2)^{3/2}}\left(1-\frac{3}{2}\frac{\Delta^2}{\tilde{\omega}^2+\Delta^2}\right)(\hat{\mathbf{h}}\mathbf{\Delta})m_{\tilde{\omega}}^{(1)}\right\}. \tag{32}$$

As will be seen shortly, $m_{\tilde{\omega}}^{(1)}$ and $\mathbf{m}_{\tilde{\omega}}^{(1)}$ are proportional to v_s.

The renormalized frequencies $\tilde{\omega}$ and $\tilde{\omega}_0$ obey

$$\omega=\omega+\frac{\tilde{\omega}}{2\tau}\left\langle\frac{1}{(\tilde{\omega}^2+\Delta^2)^{1/2}}+\frac{3}{8}\frac{\tilde{\omega}_0^2\mathbf{\Delta}_{\|}^2}{(\tilde{\omega}^2+\Delta^2)^{5/2}}\right\rangle, \tag{33}$$

$$\tilde{\omega}_0=\omega_0+\frac{\tilde{\omega}_0}{2\tau}\left\langle\frac{\mathbf{\Delta}_{\|}^2}{(\tilde{\omega}^2+\Delta^2)^{3/2}}\right\rangle. \tag{34}$$

In order to construct equations for $m_{\tilde{\omega}}^{(1)}$ and $\mathbf{m}_{\tilde{\omega}}^{(1)}$ we have to address the expressions for $f_{\tilde{\omega}+iq}$ and $\mathbf{f}_{\tilde{\omega}+iq}$. In the lowest order in ω_0 and Δ_s

$$f_{\omega}=\frac{i}{2}\frac{\omega\omega_0(\hat{\mathbf{h}}\mathbf{\Delta})-2i\omega^2\Delta_s}{(\omega^2+\Delta^2)^{3/2}}, \tag{35}$$

$$f_{\tilde{\omega}}^{(1)}\simeq\frac{\tilde{\omega}^2}{(\tilde{\omega}^2+\mathbf{\Delta}^2)^{3/2}}\left\{m_{\tilde{\omega}}^{(1)}+\frac{\tilde{\omega}_0}{\tilde{\omega}}\left[\left(1-\frac{3}{2}\frac{\Delta^2}{\tilde{\omega}^2+\Delta^2}\right)(\hat{\mathbf{h}}\mathbf{\Delta})\frac{q}{\tilde{\omega}}+\frac{i}{2}\left(\hat{\mathbf{h}}\mathbf{m}_{\tilde{\omega}}^{(1)}-3\frac{(\hat{\mathbf{h}}\mathbf{\Delta})(\mathbf{\Delta}\mathbf{m}_{\tilde{\omega}}^{(1)})}{\tilde{\omega}^2+\Delta^2}\right)\right]\right\}. \tag{36}$$

Now, according to Eq. (11) and using Eq. (33) for the renormalized frequency $\tilde{\omega}$, we easily obtain the following equations for $m_{\tilde{\omega}}^{(1)}$:

$$\left(\omega+\frac{\tilde\omega}{2\tau}\frac{\Delta^2}{(\tilde\omega^2+\Delta^2)^{3/2}}\right)m_{\tilde\omega}^{(1)}$$
$$=\frac{1}{2\tau}\frac{\tilde\omega_0\tilde\omega^2}{(\tilde\omega^2+\Delta^2)^{3/2}}\left[\frac{i}{2}\frac{\tilde\omega^2}{\tilde\omega^2+\Delta^2}\hat{\mathbf{h}}\mathbf{m}_{\tilde\omega}^{(1)}+\left(1-\frac{3}{2}\frac{\Delta^2}{\tilde\omega^2+\Delta^2}\right)\frac{Q}{\tilde\omega}\right], \tag{37}$$

where $Q=\langle(\hat{\mathbf{k}}\boldsymbol{\Delta})q\rangle=(1/3)k_F\Delta_\parallel\hat{\mathbf{l}}\mathbf{v}_s$.

A little more algebra is needed to construct an equation for $\mathbf{m}_{\tilde\omega}^{(1)}$:

$$\left(\omega\mathbf{m}_{\tilde\omega}^{(1)}+\frac{\tilde\omega}{2\tau}\left\langle\frac{(\boldsymbol{\Delta}\mathbf{m}_{\tilde\omega}^{(1)})\boldsymbol{\Delta}}{(\tilde\omega^2+\boldsymbol{\Delta}^2)^{3/2}}\right\rangle\right)$$
$$=\frac{\tilde\omega}{2\tau}\left\{-i\left\langle\frac{\tilde\omega\Delta q}{(\tilde\omega^2+\boldsymbol{\Delta}^2)^{3/2}}\right\rangle+\frac{3}{4}\frac{\tilde\omega_0^2}{(\tilde\omega^2+\Delta^2)^{3/2}}\left[\left(\frac{i\tilde\omega Q}{\tilde\omega^2+\Delta^2}-\frac{1}{3}\hat{\mathbf{h}}\mathbf{m}_{\tilde\omega}^{(1)}\right)\hat{\mathbf{h}}+\left\langle\hat{\mathbf{h}}\boldsymbol{\Delta}\frac{(\hat{\mathbf{h}}\mathbf{m}_{\tilde\omega}^{(1)})\boldsymbol{\Delta}+(\boldsymbol{\Delta}\mathbf{m}_{\tilde\omega}^{(1)})\hat{\mathbf{h}}}{\tilde\omega^2+\Delta^2}\right\rangle-\frac{5}{2}\left\langle\frac{(\hat{\mathbf{h}}\boldsymbol{\Delta})^2(i\tilde\omega q+\boldsymbol{\Delta}\mathbf{m}_{\tilde\omega}^{(1)})}{(\tilde\omega^2+\Delta^2)^2}\Delta\right\rangle\right]+\frac{i}{2}\frac{\tilde\omega_0\tilde\omega^3}{(\tilde\omega^2+\Delta^2)^{5/2}}m_{\tilde\omega}^{(1)}\hat{\mathbf{h}}\right\}. \tag{38}$$

It is evident that $\hat{\mathbf{h}}\mathbf{m}_{\tilde\omega}^{(1)}$, which appears in Eq. (37), must be calculated at $\omega_0=0$. Addressing Eq. (38), we readily find that in this case $\hat{\mathbf{h}}\mathbf{m}_{\tilde\omega}^{(1)}$ is given by

$$\left(\omega+\frac{\tilde\omega}{2\tau}\frac{\Delta^2}{(\tilde\omega^2+\Delta^2)^{3/2}}\right)\hat{\mathbf{h}}\mathbf{m}_{\tilde\omega}^{(1)}=-\frac{i}{2\tau}\frac{\tilde\omega^2}{(\tilde\omega^2+\Delta^2)^{3/2}}Q. \tag{39}$$

Since in what follows we consider the Ginzburg–Landau region (near T_c or near P_c at $T=0$), a set of equations is to be used:

$$m_{\tilde\omega}^{(1)}\simeq\frac{\tilde\omega_0}{\tilde\omega}\frac{(i/2)\hat{\mathbf{h}}\mathbf{m}_{\tilde\omega}^{(1)}+Q/\tilde\omega}{2\tau|\omega|+\Delta^2/\tilde\omega^2},$$
$$\hat{\mathbf{h}}\mathbf{m}_{\tilde\omega}^{(1)}\simeq-\frac{3iQ/\tilde\omega}{6\tau|\omega|+\Delta^2/\tilde\omega^2}. \tag{40}$$

In the denominators of these expressions the term $\Delta^2/\tilde\omega^2$ is retained. When considering the vicinity of T_c, it is to be dropped as a higher-order correction. On the other hand, at $T=0$ this term must be preserved, as well shall see shortly, in order to avoid unphysical divergences stemming from the vicinity of $\omega=0$.

The appearance of $m_{\tilde\omega}^{(1)}\neq0$ is due to a mixing of states with $S_s=0$ of spin-singlet ($S=0$) and spin-triplet ($S=1$) configurations in the presence of three factors: quasiparticle scattering ($1/\tau\neq0$), magnetic field ($\omega_0\neq0$), and supercurrent ($Q\neq0$).

According to the definition of Q, the spin-singlet part $m_{\tilde\omega}^{(1)}$ is absent from the planar phase ($\Delta_\parallel=0$). We begin our consideration of $\rho_{ij}^{(s)}$ with just this simple case, for which in the Ginzburg–Landau regime,

$$g_{\tilde\omega}^{(1)}\simeq-\frac{i}{|\tilde\omega|^3}[\boldsymbol{\Delta}_\perp^2 q+i\tilde\omega\boldsymbol{\Delta}_\perp\mathbf{m}_{\tilde\omega}^{(1)}], \tag{41}$$

$$\Delta_\perp\mathbf{m}_{\tilde\omega}^{(1)}\simeq-\frac{i}{\tilde\omega}\frac{3\langle\Delta_\perp q\rangle\boldsymbol{\Delta}_\perp}{6\tau|\omega|+\boldsymbol{\Delta}_\perp^2/\tilde\omega^2}. \tag{42}$$

Now, using Eq. (13) for the supercurrent, it can be easily shown that for the planar phase

$$\frac{\rho_\parallel^{(s)}}{\rho}=\pi T\sum_\omega\frac{2}{5}\frac{\boldsymbol{\Delta}_\perp^2}{|\tilde\omega|^3},$$
$$\frac{\rho_\perp^{(s)}}{\rho}=\pi T\sum_\omega\frac{\boldsymbol{\Delta}_\perp^2}{|\tilde\omega|^3}\left[\frac{4}{5}+\frac{1}{6\tau|\omega|+\boldsymbol{\Delta}_\perp^2/\tilde\omega^2}\right] \tag{43}$$

with $\tilde\omega=\omega+(1/2\tau)\operatorname{sign}\omega$.

Considering the limit $T=0$, we convert the ω-summation to integration and, using the frequency renormalization equation

$$1-\frac{1}{2\tau}\left\langle\frac{1}{[\tilde\omega^2+\boldsymbol{\Delta}_\perp^2(\hat{\mathbf{k}})]^{1/2}}\right\rangle=\frac{\omega}{\tilde\omega}, \tag{44}$$

pass to a new variable $z(\omega)=2\tau\tilde\omega(\omega)$. For $\epsilon_\perp=2\tau\Delta_\perp\ll1$,

$$2\tau\omega\simeq z-1+\frac{1}{3}\frac{\epsilon_\perp^2}{z^2}+\ldots, \tag{45}$$

so that at $T=0$

$$\pi T\sum_\omega(\ldots)\simeq\frac{1}{2\tau}\int_{1-\epsilon_\perp^2/3}^\infty dz\left(1-\frac{2}{3}\frac{\epsilon_\perp^2}{z^3}\right)(\ldots). \tag{46}$$

Now, from Eqs. (43) it is readily obtained that at $T=0$ and $P_c<P<P_\parallel$

$$\frac{\rho_\parallel^{(s)}}{\rho}\simeq\frac{1}{5}\epsilon_\perp^2,\quad\frac{\rho_\perp^{(s)}}{\rho}\simeq\frac{2}{5}\epsilon_\perp^2\left(1+\frac{5}{6}\ln\frac{3}{\epsilon_\perp^2}\right). \tag{47}$$

The case with $\Delta_\perp\neq0$ and $\Delta_\parallel\neq0$ needs much more effort. Here both $\mathbf{m}_{\tilde\omega}^{(1)}$ and $m_{\tilde\omega}^{(1)}$ contribute to the supercurrent. Starting from the general expression (13), linearizing with respect to q and using Eq. (32), after quite lengthy calculations the following answers for $\rho_\perp^{(s)}$ and $\rho_\parallel^{(s)}$ are obtained at $T=0$ near the critical pressure $P_\parallel(\epsilon_\parallel=2\tau\Delta_\parallel)$:

$$\frac{\rho_\perp^{(s)}}{\rho}=\frac{1}{2}\left(\frac{4}{5}\epsilon_\perp^2+\frac{1}{5}\epsilon_\parallel^2\right)+\frac{1}{3}\epsilon_\perp^2\ln\frac{3}{\epsilon_\perp^2}, \tag{48}$$

$$\frac{\rho_\parallel^{(s)}}{\rho}=\frac{1}{2}\left(\frac{2}{5}\epsilon_\perp^2+\frac{3}{5}\epsilon_\parallel^2\right)+\frac{1}{3}\epsilon_\parallel^2\ln\frac{3}{\epsilon_\perp^2}+\frac{1}{6}\epsilon_\parallel^2\frac{\omega_0^2}{\Delta^2}-\frac{3}{16}(2-\ln3)\frac{\Delta_\parallel^2}{\Delta^2}\frac{\omega_0^2}{\Delta^2}. \tag{49}$$

The last term in $\rho_\parallel^{(s)}$ is a contribution of the spin-single correlations described by $m_{\tilde\omega}^{(1)}$.

In zero magnetic field ($\omega_0=0$), from Eqs. (48) and (49) we immediately obtain that

$$\frac{\rho_{\parallel}^{(s)}}{\rho}=\frac{\rho_{\perp}^{(s)}}{\rho}=\frac{1}{2}\epsilon^2\left(1+\frac{2}{3}\ln\frac{3}{\epsilon^2}\right). \tag{50}$$

As a final remark we point out that the results found in Ref. 5 can be readily transcribed to the case of $T=0$ properties of nonmagnetic impurity-containing HTSC. In this situation, the transition temperature to the superconducting state T_{c0} for a pure sample depends on the level of the hole doping x so that the critical concentration of holes x_c at which a quantum phase transition should occur at $T=0$ is given by Eq. (1) with the pressure P being substituted by the hole concentration x. The superconducting order parameter $\Delta(x)$ near x_c is given by

$$\Delta^2(x)\simeq\frac{a}{\tau^2}\ln\frac{T_{c0}(x)}{T_{c0}(x_c)}\begin{cases}\Theta(x-x_c), & (\partial T_{c0}/\partial x)_{x_c}>0,\\ \Theta(x_c-x), & (\partial T_{c0}/\partial x)_{x_c}<0.\end{cases} \tag{51}$$

Here we have used a simple d-wave pairing model, where in the weak coupling approximation the coefficient $a=6/5$. The two possibilities in Eq. (51) take into account that in general $T_{c0}(x)$ is a nonmonotonic function of x. For instance, in $La_{2-x}Sr_xCuO_4$, $T_{c0}(x)$ is bell-shaped with a maximum at an optimal doping $x_{\rm opt}$.

We are indepted to V. P. Mineev for useful comments. This work was partly supported by Grant No. 2.16 of the Georgian Academy of Sciences.

*)E-mail: gogi@iph.hepi.edu.ge

[1] J. V. Porto III and J. M. Parpia, Phys. Rev. Lett. **74**, 4667 (1995).
[2] D. T. Sprague, T. M. Haard, J. B. Kycia, M. R. Rand, Y. Lee, P. J. Hamot, and W. P. Halperin, Phys. Rev. Lett. **75**, 661 (1995).
[3] D. T. Sprague, T. M. Haard, J. B. Kycia, M. R. Rand, Y. Lee, P. J. Hamot, and W. P. Halperin, Phys. Rev. Lett. **77**, 4568 (1996).
[4] K. Matsumoto, J. V. Porto, L. Pollack, E. N. Smith, T. L. Ho, and J. M. Parpia, Phys. Rev. Lett. **79**, 253 (1997).
[5] V. P. Mineev, JETP Lett. **66**, 693 (1997).
[6] G. Baramidze, G. Kharadze, and G. Vachnadze, JETP Lett. **63**, 107 (1996).

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