Journal of Experimental and Theoretical Physics, Vol. 91, No. 1, 2000, pp. 1–15.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 1, 2000, pp. 5–19.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Current Tensor with a Heavy Photon for Pair Production by a Longitudinally Polarized Electron¶

M. I. Konchatnij, N. P. Merenkov*, and O. N. Shekhovzova

Kharkov Institute of Physics and Technology, Ukrainian Scientific Center, Kharkov, 310108 Ukraine
**e-mail: merenkov@kipt.kharkov.ua*

Received January 18, 2000

Abstract—The electron current tensor for the scattering of a heavy photon on a longitudinally polarized electron accompanied with an additional electron–positron pair is considered. The contribution of soft and hard collinear and quasi-collinear pairs is taken into account. The full analysis of the spin-independent and spin-dependent parts of the electron current tensor is performed. The results allow us to calculate the corresponding contribution into the second-order radiative correction to cross sections for different processes with the next-to-leading accuracy. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The recent polarized experiments on deep inelastic scattering [1, 2] cover the kinematic region of the Bjorken variable $y \simeq 0.9$, where the electromagnetic corrections to the cross section are extremely large. The corresponding first-order quantum-electrodynamics (QED) correction due to one real and one virtual photon emission was computed in [3, 4]. At large values of y, this correction is of the order of the Born cross section in this region. Therefore, the calculation of the second-order QED correction becomes very important for interpreting these experiments in terms of the hadron structure functions.

In general, the cross section for the deep inelastic scattering (DIS) can be represented as a contraction of the electron current tensor $L_{\mu\nu}$ (ECT) and the hadron tensor $H_{\mu\nu}$ [5]:

$$d\sigma \propto \frac{1}{q^4} L_{\mu\nu} H_{\mu\nu},$$

where q is the momentum transfer (by definition, the ECT involves the phase space of the leptons accompanying the scattered electron). The ECT is model-independent and universal, while the hadron tensor depends on the model used for the description of hadrons and has its own specifications for different processes, event selection, and so on. Therefore, it is very important to calculate the universal quantity (the ECT) with the maximum possible accuracy, since it can be applied to many processes in scattering and annihilation channels.

The first steps in calculating the second-order QED correction to the ECT for a longitudinally polarized electron were done in [6, 7]. In [6], the one-loop corrected Compton tensor with a heavy photon was calculated with singular and finite terms in the limit $m \longrightarrow 0$, where m is the electron mass. This approximation does not take all contributions into account with the next-to-leading accuracy in the case where the radiated photon is not observed. The extension of the results of [6] to this case was performed in [8]. In [7], the ECT due to hard double-photon emission was derived keeping the electron mass finite and taking the contributions of collinear and quasi-collinear kinematics into account. An approximation of this type allows one to keep all the next-to-leading terms in the case of unobserved photons.

Besides the additional contribution to the ECT due to the virtual and real soft double-photon emission, which can be obtained in the unpolarized case [9], there is a contribution from the pair production. In this work, we calculate it with the same accuracy as in [7] for the hard double-photon emission. It is convenient to divide the corresponding ECT into parts corresponding to the contribution of virtual and real soft pairs as well as the contribution from the hard pair. The first part can be derived using the results for the unpolarized case [10], and the main problem is the calculation of the second one.

In the Born approximation, the ECT with a longitudinally polarized electron has the form

$$\begin{aligned} L^B_{\mu\nu} &= Q_{\mu\nu} + i\lambda E_{\mu\nu}, \\ Q_{\mu\nu} &= -4(p_1p_2)g_{\mu\nu} + 4(p_1p_2)_{\mu\nu}, \\ E_{\mu\nu} &= 4\epsilon_{\mu\nu\rho\sigma}p_1p_{2\sigma}, \\ (p_1p_2)_{\mu\nu} &= p_{1\mu}p_{2\nu} + p_{1\nu}p_{2\mu}, \end{aligned} \tag{1}$$

where p_1 (p_2) is the 4-momentum of the initial (final) electron and λ is the doubled initial electron helicity [it equals 1 (or –1) if the initial electron is polarized along (against) its 3-momentum direction].

¶ This article was submitted by the authors in English.

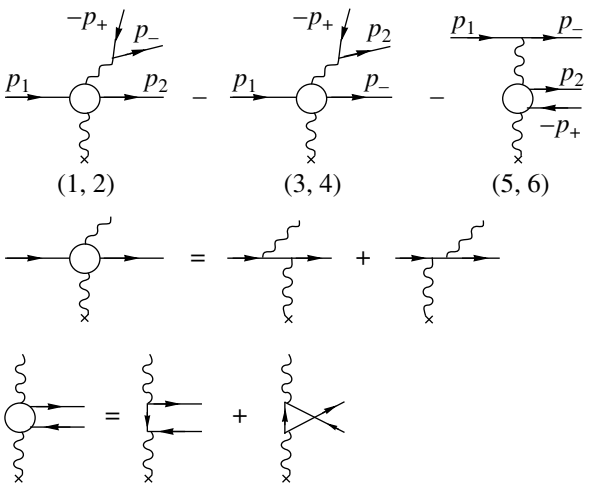


Fig. 1. The essential Feynman diagrams that contribute at $|q^2|, |u| \gg m^2$ in the collinear kinematics. Every set of diagrams is gauge-invariant relative to the heavy photon. The signs before sets are defined by the Fermi statistics for permutation of the final fermion states.

We consider corrections to $L^B_{\mu\nu}$ due to the pair production in the scattering

$$e^-(p_1) + \gamma^*(q) \longrightarrow e^-(p_2) + e^+(p_+) + e^-(p_-) \qquad (2)$$

or annihilation

$$e^-(p_1) + e^+(p_2) \longrightarrow \gamma^*(q) + e^-(p_-) + e^+(p_+)$$

processes under the condition $|q^2|, (p_1p_2) \gg m^2$ in collinear and quasi-collinear kinematics. In the unpolarized case, the corresponding calculations were performed in part for DIS [11, 12] as well as for small- and large-angle Bhabha scattering [13, 14] at the level of cross sections. Some other aspects of QED corrections due to pair production are also discussed in [15–18]. For definiteness, we investigate the scattering channel, and to obtain the corresponding result for the annihilation channel one must substitute $-p_2$ instead of p_2.

The paper is organized as follows. In Section 2, we consider the contribution due to collinear kinematics. There are two collinear kinematical regions: (a) where the created pair is emitted along the initial electron momentum direction ($\mathbf{p}_+, \mathbf{p}_- \| \mathbf{p}_1$) and (b) where the created pair moves along the final one ($\mathbf{p}_+, \mathbf{p}_- \| \mathbf{p}_2$). In collinear regions, both the photon and the fermion propagator denominators (PD) of the underlying Feynman diagrams can be small. The corresponding contribution to the ECT can be expressed in terms of the symmetric tensor $Q_{\mu\nu}$ and the antisymmetric tensor $E_{\mu\nu}$ in the same manner as in [7] for the double-photon emission process. In Section 2.1, we calculate the ECT in region *a*; in Section 2.2, in region *b*. Our final result for the collinear regions is valid with the next-to-leading accuracy, although the intermediate formulas are written with next-next-to-leading accuracy.

In Section 3, we investigate the ECT in quasi-collinear kinematics. Only the photon PDs of Feynman diagrams can be small in this case, and under the conditions considered here there are three different quasi-collinear regions [11]: $\mathbf{p}_- \| \mathbf{p}_1$, $\mathbf{p}_- \| \mathbf{p}_+$, and $\mathbf{p}_+ \| \mathbf{p}_2$. The corresponding contributions to the ECT are calculated in Sections 3.1, 3.2, and 3.3, respectively. The structure of the ECT in quasi-collinear regions is more complicated than in the collinear ones. Nevertheless, the result derived in Section 3 is valid with the next-next-to-leading accuracy. We demonstrate the elimination of the angular auxiliary parameters used to define collinear regions *a* and *b* in the sum of contributions of the collinear and quasi-collinear kinematics at the next-to-leading accuracy level. The main results, together with the virtual and soft pair contribution, are briefly formulated in Conclusions. The starting points of our calculations are given in the Appendix.

Our calculations are based, in fact, on the quasi-real electron method [19] and can be applied to processes with an electron beam energy on the order of 1 GeV and more.

2. INVESTIGATION OF COLLINEAR KINEMATICS

As noted in the Introduction, both the photon and the fermion PD of the underlying Feynman diagrams can be small in the collinear region. Because of the imposed restriction on the heavy photon mass ($-q^2 \gg m^2$) and the registration condition for the scattered electron ($-u = 2p_1p_2 \gg m^2$), only six from the total of eight diagram sets are essential. These diagrams are shown in Fig. 1.

The amplitudes corresponding to each of the diagram sets (1, 2), (3, 4), and (5, 6) in Fig. 1 are gauge invariant, and it is therefore convenient to keep these pairs together during the calculation. We refer to them as the (1, 2) set and so on.

2.1. The Contribution of Region a

In region *a*, the (1, 2) and (5, 6) sets contribute. We define the limiting angle θ_0 in this region such that

$$\theta_+, \theta_- \leq \theta_0, \quad \theta_0 \ll 1, \quad (\varepsilon_1\theta_0/m)^2 = z_0 \gg 1, \quad (3)$$

where $\theta_\pm = \widehat{\mathbf{p}_1\mathbf{p}_\pm}$ and ε_1 is the initial electron energy.

In calculating the ECT, we always neglect the terms of the order θ_0 and $1/z_0$ compared to unity. Thus, our results include, in general, terms of the orders $\ln^2 z_0$ (best), $\ln z_0$ (next-to-leading), and 1 (next-next-to-leading). In the collinear regions, we keep all the terms that contribute with the next-next-to-leading accuracy at the differential level, but we cannot perform the

angular integration analytically with the same accuracy. This is why the final result derived in Section 2 is written in only the next-to-leading approximation.

For the kinematical invariants corresponding to small PD in the relevant Feynman diagrams, we introduce the parameterization [11]

$$a = \frac{q_\pm^2}{m^2} = \frac{1}{x_+x_-}[(x_+ + x_-)^2 + x_+^2x_-^2(\mathbf{n}_+ - \mathbf{n}_-)^2],$$

$$a_1 = \frac{2(p_1p_+)}{m^2} = \frac{1}{x_+}(1 + x_+^2 + x_+^2\mathbf{n}_+^2),$$

$$a_2 = \frac{2(p_1p_-)}{m^2} = \frac{1}{x_-}(1 + x_-^2 + x_-^2\mathbf{n}_-^2),$$

$$\Delta = \frac{(p_1 - p_- - p_+)^2 - m^2}{m^2} = a - a_1 - a_2, \qquad (4)$$

$$x_+ = \frac{\varepsilon_+}{\varepsilon_1}, \quad x_- = \frac{\varepsilon_-}{\varepsilon_1},$$

$$|n_{+,-}| = \frac{\varepsilon_1\theta_{+,-}}{m}, \quad q_a = zp_1 - p_2,$$

$$z = 1 - x_+ - x_-, \quad q_\pm = p_+ + p_-,$$

where ε_+ (ε_-) is the created positron (electron) energy and $\mathbf{n}_+$, $\mathbf{n}_-$ are two-dimensional vectors that are perpendicular to the direction of the 3-momentum $\mathbf{p}_1$. To separate hard and soft pairs, we introduce an auxiliary parameter D ($D \ll 1$) and assume that $x_+ + x_- \geq \varepsilon_1 D$ for the hard pair. We define the ECT as the product of the created electron–positron pair phase space and the trace tensor of the corresponding tensor diagram (TD) given in Fig. 2,

$$L_{\mu\nu} = \left(\frac{\alpha}{4\pi^2}\right)^2 \frac{d^3p_+d^3p_-}{\varepsilon_+\varepsilon_-} T_{\mu\nu}, \qquad (5)$$

where the starting points for the calculation of $T_{\mu\nu}$ are defined in the Appendix in terms of the corresponding squared matrix element. In region a, the created pair phase space can be written in terms of the variables used in relations (4) as

$$\frac{d^3p_+d^3p_-}{\varepsilon_+\varepsilon_-} = m^4\pi^2 x_+x_-dx_+dx_-dz_+dz_-\frac{d\phi}{2\pi},$$
$$z_{+,-} = \frac{\varepsilon_1^2\theta_{+,-}^2}{m^2}, \qquad (6)$$

where ϕ is the angle between $\mathbf{n}_+$ and $\mathbf{n}_-$.

When calculating tensor $T_{\mu\nu}$, we have to keep terms with m^4 in the denominator and neglect terms of the order m^{-2}. Such an approach allows us, in principle, to compute the quantity $L_{\mu\nu}$ with the power accuracy relative to the parameter z_0, neglecting only terms of the order $O(z_0^{-1})$.

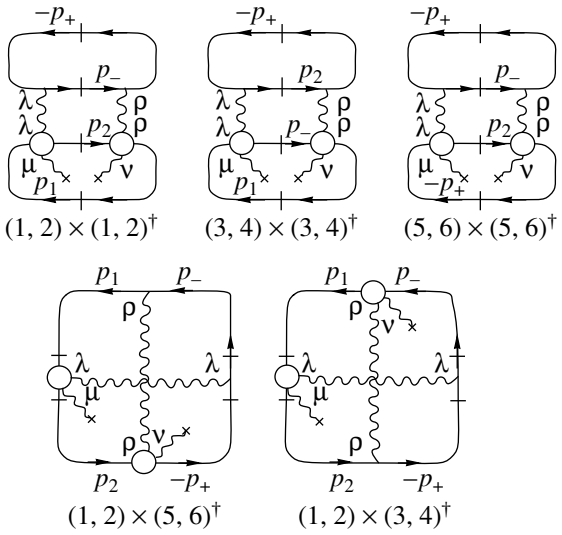


Fig. 2. Diagrams for the electron current tensor. Topological class *a* is responsible for contributions of the (1, 2), (3, 4) and (5, 6) sets that appear in both collinear and quasi-collinear kinematical regions. Topological class *b* describes the interference between (1, 2) set and (3, 4) as well as (5, 6) ones. The latter class contributes to the collinear kinematics only within the chosen accuracy.

In region a, it is convenient to separate the contributions of the (1, 2) and (5, 6) sets and their interference:

$$T_{\mu\nu}^{(a)} = \frac{16}{m^4}[T_{\mu\nu}^{(12)} + T_{\mu\nu}^{(56)} + T_{\mu\nu}^{(int)}].$$

The tensor $T_{\mu\nu}^{(12)}$ can be written as

$$T_{\mu\nu}^{(12)} = \left\{-\frac{1}{a^2\Delta^2}\left[\frac{2z}{(1-z)^2}(x_-a_1 - x_+a_2)^2 + 4az\right.\right.$$
$$\left. + 2(1-z)\Delta\right] + \frac{2x_+x_-}{\Delta^2(1-z)^2} - \frac{1+z}{a\Delta(1-z)} - \frac{1}{a\Delta^2(1-z)^2}$$
$$\times [z(1-z)(x_+ - x_-)(a_1 - a_2) \qquad (7)$$
$$\left. + 2x_+x_-(1+z)(a_1 + a_2)]\right\}(Q_{\mu\nu} + i\lambda E_{\mu\nu})$$
$$- \frac{i\lambda}{a^2\Delta^2}[2a((1-z)^2 - 2x_+x_-) + 4(1-z)^2]E_{\mu\nu}.$$

We emphasize that the last term on the right-hand side of Eq. (7) does not give a large logarithm when integrated over the angular phase space of the created pair. Therefore, within the best and next-to-leading accuracy, both tensor structures—$Q_{\mu\nu}$ and $i\lambda E_{\mu\nu}$—are

multiplied by the same function. With this accuracy, therefore, we can use the result of the corresponding calculations for the unpolarized case [11, 12], where only the $Q_{\mu\nu}$ structure is involved.

As concerns the tensor $T^{(56)}_{\mu\nu}$, its symmetric spin-independent part can be obtained from the corresponding part of $T_{\mu\nu}$ by the rule

$$\mathrm{sym}\{T^{(56)}_{\mu\nu}\} = -x_+\hat{P}\mathrm{sym}\{T^{(12)}_{\mu\nu}\}, \tag{8}$$

where the substitution operator $\hat{P}$ acts as

$$\hat{P}(x_+, x_-, \mathbf{n}_+, \mathbf{n}_-) \longrightarrow \left(x_+, -\frac{x_-}{x_+}, x_+\mathbf{n}_+, x_+(\mathbf{n}_+ - \mathbf{n}_-)\right). \tag{9}$$

It is easy to verify that

$$\hat{P}(a_1, a, a_2) \longrightarrow (a_1, 2-a_2, 2-a).$$

The rule (9) reflects the obvious topological equivalence of TD for the (1, 2) and (5, 6) sets in the unpolarized case, as follows from Fig. 2 (see also [12]).

Unfortunately, we cannot use this rule to obtain the antisymmetric spin-dependent part of $T^{(56)}_{\mu\nu}$. At the level of the TD of Fig. 2, we can explain this fact because, for the (1, 2) set, the polarized particle (the electron with the 4-momentum p_1) enters the lower block, while for the (5, 6) set it enters the upper one. It is obvious that conditions for the polarized particle are different in the lower and upper blocks.

Straightforward calculations lead to the expression

$$T^{(56)}_{\mu\nu} = S^{(56)}Q_{\mu\nu} + A^{(56)}i\lambda E_{\mu\nu}, \tag{10}$$

where

$$S^{(56)} = -\frac{2z}{(2-a_2)^2}\left(\frac{a_1}{\Delta} + \frac{1}{1-x_-}\right)^2$$
$$-\frac{1}{\Delta(2-a_2)}\left(x_+ - z - \frac{2(1-x_+ + x_+^2)}{1-x_-}\right) \tag{11}$$
$$-\frac{z}{\Delta^2} + \frac{2za_1}{\Delta^2(2-a_2)} + \frac{2(1-x_-)}{\Delta(2-a_2)^2} - \frac{4z}{\Delta^2(2-a_2)},$$

$$A^{(56)} = -\frac{1-x_- + x_-(x_+ - z)}{\Delta(2-a_2)(1-x_-)} + \frac{z}{\Delta^2}$$
$$-\frac{2za_1}{\Delta^2(2-a_2)} - \frac{2(1-x_-)(x_+ - z)}{\Delta(2-a_2)^2} \tag{12}$$
$$-\frac{2[2(1-x_+) - (1-z)(1-x_-)]}{\Delta^2(2-a_2)} - \frac{4(1-x)}{\Delta^2(2-a_2)^2}.$$

We see that in contrast with the (1, 2) set contribution, where the structures $Q_{\mu\nu}$ and $i\lambda E_{\mu\nu}$ are accompanied by the same function up to the next-to-leading accuracy, the quantities $S^{(56)}$ and $A^{(56)}$ on the right-hand side of Eq. (10) are quite different. As we see in what follows, this difference also affects the best approximation.

The interference of the (1, 2) and (5, 6) sets (see the TD in Fig. 2) can be written as

$$T^{(\mathrm{int})}_{\mu\nu} = S^{(\mathrm{int})}(Q_{\mu\nu} + i\lambda E_{\mu\nu}) + A^{(\mathrm{int})}i\lambda E_{\mu\nu}, \tag{13}$$

$$S^{(\mathrm{int})} = -\frac{1}{a\Delta}\left[2x_- + (zx_+ - x_-)\left(\frac{1}{1-x_-} + \frac{1}{1-z}\right)\right]$$
$$-\frac{1}{\Delta(2-a_2)}\left[2x_- + (z - x_+x_-)\left(\frac{1}{1-x_-} + \frac{1}{1-z}\right)\right]$$
$$-\frac{1}{a(2-a_2)}\left[-\frac{2x_+x_-}{(1-x_-)(1-z)}\right.$$
$$\left. + x_-(1+x_+)\left(\frac{1}{1-x_-} + \frac{1}{1-z}\right) - 2x_-\right] \tag{14}$$
$$-\frac{2za_1}{\Delta^2}\left(\frac{1}{a} + \frac{1}{2-a_2}\right) - \frac{2(zx_- - x_+)}{\Delta a(2-a_2)}\left(\frac{1}{1-x_-} - \frac{1}{1-z}\right)$$
$$-\frac{4z}{\Delta^2}\left(\frac{1}{a} + \frac{1}{2-a_2}\right) + \frac{8z}{a\Delta^2(2-a_2)},$$

$$A^{(\mathrm{int})} = -\frac{2(1-z)}{a\Delta(2-a_2)}(2x_+ - x_- + 3x_-^2)$$
$$-\frac{2}{\Delta^2(2-a_2)}(x_+ - x_- + 2x_+^2 + 3x_-(1-z)) \tag{15}$$
$$+\frac{2(1-z)^2}{a\Delta^2}\left(3 - \frac{4}{2-a_2}\right).$$

The quantity $S^{(\mathrm{int})}$ is invariant under the same operation that transforms the tensor $T^{(12)}_{\mu\nu}$ into $T^{(56)}_{\mu\nu}$,

$$S^{(\mathrm{int})} = -x_+\hat{P}S^{(\mathrm{int})}, \tag{16}$$

while $A^{(\mathrm{int})}$ is not invariant.

The last two lines on the right-hand side of Eq. (14), as well as $A^{(\mathrm{int})}$, do not contribute within the logarithmic accuracy framework.

The next step in our calculations is the integration of the ECT over the angular phase space of the created pair. We use parameterization (6) and integrate over z_+ and z_- from zero to z_0 and over ϕ from zero to 2π. In principle, the angular integration can be performed with a power accuracy, but we restrict ourselves with the next-to-leading accuracy in this paper. The method of integration suitable for this approximation is described in [20], and the table of the corresponding integrals is given in [11, 14].

As noted above [see Eqs. (7) and (13)], with this accuracy, the contribution to the ECT due to the (1, 2) set and the interference of the (1, 2) and (5, 6) sets contains the same function at the symmetric and antisym-

metric structures at the differential level. Therefore, we can write down the results of the angular integration in region *a* as

$$\overline{L^{(a)}_{\mu\nu}} = \frac{\alpha^2}{\pi^2} dx_+ dx_- \ln z_0 \times [(Q_{\mu\nu} + i\lambda E_{\mu\nu})\bar{S} + \overline{S^{(56)}} Q_{\mu\nu} + i\lambda \overline{A^{(56)}} E_{\mu\nu}], \quad (17)$$

where $\bar{S}$ absorbs the angular integration results on the right-hand sides of Eqs. (7) and (14):

$$\bar{S} = \frac{1+z^2}{(1-z)^4}\left[\frac{1}{2}(x_+^2 + x_-^2)\ln\frac{z_0 x_+^2 x_-^2}{z^2} - (x_+ - x_-)^2 + \frac{8z x_+ x_-}{1+z^2}\right] 2\left[-\frac{x_-^2 + z^2}{(1-x_-)(1-z)}\ln\frac{(1-z)(1-x_-)}{z x_-} + \frac{x_+ x_- - z}{(1-x_-)^2} + \frac{z x_+ - x_-}{(1-z)^2}\right]. \quad (18)$$

The corresponding symmetric pare of the contribution of the (5, 6) set is defined by

$$\overline{S^{(56)}} = \frac{x_+^2 + z^2}{(1-x_-)^4} \times \left[\frac{1}{2}(1 + x_-^2)\ln\frac{z_0 x_+^2 x_-^2}{z^2} - (1+x_-)^2 + \frac{8z x_+ x_-}{x_+^2 + z^2}\right], \quad (19)$$

and its antisymmetric part is defined by

$$\overline{A^{(56)}} = \frac{(1+x_-)(z-x_+)}{2(1-x_-)^2}\ln\frac{z_0 x_+^2 x_-^2}{z^2} - 3 + \frac{4+6z}{1-x_-} - \frac{8z}{(1-x_-)^2}. \quad (20)$$

It follows from (8) and (16) that the coefficient at the $Q_{\mu\nu}$ structure on the right-hand side of Eq. (17) is invariant under the substitution

$$(x_+, x_-, z_0) \longrightarrow \left(\frac{1}{x_+}, -\frac{x_-}{x_+}, x_+^2 z_0\right).$$

Because the heavy photon 4-momentum q_a in region *a* depends on the sum of the created electron and positron energy fractions [see (4)], we can carry out the model-independent integration of the right side of Eq. (17) over the electron (or positron) energy fraction x_- (or x_+) at fixed $x_+ + x_- = 1 - z$. The result can be written as

$$\tilde{L}^{(a)}_{\mu\nu} = \frac{\alpha^2}{\pi^2} dz \ln z_0 \times [(Q_{\mu\nu} + i\lambda E_{\mu\nu})\tilde{S} + \tilde{S}^{(56)} Q_{\mu\nu} + i\lambda \tilde{A}^{(56)} E_{\mu\nu}], \quad (21)$$

where

$$\tilde{S} = \frac{1+z^2}{3(1-z)}\ln z_0 + 4(1+z)\ln z \ln(1-z) - \frac{2(3z^2-1)}{1-z} L_{i2}(1-z) - \frac{2}{1-z}\ln^2 z + \frac{8}{3(1-z)}\ln(1-z) - \frac{20}{9(1-z)}, \quad (22)$$

$$\tilde{S}^{(56)} = \left[\frac{1-z}{6z}(4 + 7z + 4z^2) + (1+z)\ln z\right]\ln z_0 + \frac{2}{3}\left(-4z^2 - 5z + 1 + \frac{4}{z}\right)\ln(1-z) + \frac{1}{3}\left(8z^2 + 5z - 7 - \frac{13}{1-z}\right)\ln z - \frac{2}{3}z^2 + \frac{136}{9}z - \frac{107}{9} - \frac{4}{3z}, \quad (23)$$

$$\tilde{A}^{(56)} = \frac{1}{2}[5(1-z) + 2(1+z)\ln z]\ln z_0 + \frac{2}{3}(13 - 17z)\ln(1-z) + \frac{1}{3}\left(5z - 19 - \frac{13}{1-z}\right)\ln z + \frac{196}{9}z - \frac{185}{9}. \quad (24)$$

We draw the reader's attention to the leading (double logarithmic) contribution to the tensor $\tilde{L}^{(a)}_{\mu\nu}$ [the first terms on the right-hand sides of Eqs. (22), (23), and (24)]. The terms that enter Eqs. (22) and (23) are related with the electron structure function due to the pair production [17].

The first term is responsible for a non-singlet channel contribution. It has an infrared singularity as $z \longrightarrow 1$ and can be obtained by inserting the effective electromagnetic coupling $\alpha_{\text{eff}}(\ln z_0)$ (which is the integral of the running constant) into the so-called θ-term of the first-order electron structure function [17]

$$P_{1\theta}(z, \alpha_{\text{eff}}) = \frac{\alpha_{\text{eff}}(\ln z_0)}{2\pi}\frac{1+z^2}{1-z}\theta(1-z-D),$$

$$D \longrightarrow 0, \quad \alpha_{\text{eff}}(\ln z_0) = -3\pi\ln\left(1 - \frac{\alpha\ln z_0}{3\pi}\right).$$

Thus, we see that in the non-singlet channel, the spin-independent and the spin-dependent parts of the ECT have the same behavior in the leading approximation. This is true for the pair production as well as for the photon emission [7] and is a consequence of the helicity conservation in the non-singlet channel [21].

The second term describes the spin-independent part of the ECT in the singlet channel. It has a specific z^{-1} behavior at small values of *z* and tends to zero as $z \longrightarrow 1$. However, the corresponding spin-dependent

part of the ECT in Eq. (24) is described by a different structure function

$$\tilde{A}_L^{(56)}(z) = \frac{1}{2}[5(1-z) + 2(1+z)\ln z]\ln^2 z_0, \qquad \int_0^1 \tilde{A}_L^{(56)}(z)dz = 0 \tag{25}$$

that also tends to zero as $z \longrightarrow 1$ but behaves differently at small z. Therefore, we conclude that in the singlet channel, the spin-independent part of the ECT and the spin-dependent one are different in the best approximation, especially at small z. This effect occurs if the created pair is concentrated along the polarized electron momentum direction; the effect is absent (as we see in what follows) if the pair moves along the scattered electron momentum direction.

2.2. The Contribution of Region b

In considering the double-photon emission by a longitudinally polarized electron [7], we saw that in the case where both photons are pressed to the final (unpolarized) electron momentum direction, the spin-independent and spin-dependent parts of the ECT have the same behavior with the power accuracy. The analysis performed in this work clarified that a similar situation occurs for pair production.

We can explain our results using the TD in Fig. 2. In the collinear region *b*, the (1, 2) and (3, 4) sets contribute. As before (in region *a*), the (1, 2) set describes the non-singlet channel contribution into the ECT, but the (3, 4) set now describes the corresponding singlet channel contribution. As we can see from Fig. 2, TDs for the (1, 2) and (3, 4) sets are mapped into each other by the substitution

$$p_2 \longleftrightarrow -p_+.$$

This substitution does not affect the polarized electron with the 4-momentum p_1. That is why the contribution of both the spin-independent and spin-dependent parts of the ECT for the (3, 4) set can be obtained (at the differential level) from the corresponding contributions of the (1, 2) set with the help of a certain rule similar to Eqs. (8) and (9). Because both parts of the ECT have the same behavior in the non-singlet channel, they are the same in the singlet channel too. The interference of the non-singlet and singlet channels, which contributes with the next-to-leading and power accuracy, does not change this conclusion. At this point, the principal difference appears between the kinematics *a* and *b* for the pair production by polarized electron.

In region *b*, it is convenient to introduce the created electron and positron energy fractions as well as the angles relative to the final electron energy and momentum direction, respectively:

$$y_{+,-} = \frac{x_{+,-}}{y_0}, \quad y_0 = \frac{\varepsilon_2}{\varepsilon_1}, \qquad \bar{\theta}_+ (\bar{\theta}_-) = \widehat{\mathbf{p}_+\mathbf{p}_2} (\widehat{\mathbf{p}_-\mathbf{p}_2}), \tag{26}$$

where ε_2 is the scattered electron energy. The corresponding phase space of the created pair in terms of these variables reads

$$\frac{d^3p_+d^3p_-}{\varepsilon_+\varepsilon_-} = \pi^2 m^4 y_+ y_- dy_+ dy_- d\bar{z}_+ d\bar{z}_- \frac{d\phi}{2\pi}, \qquad \bar{z}_{+,-} = y_0^2 \frac{\varepsilon_1^2 \bar{\theta}_{+,-}^2}{m^2}. \tag{27}$$

We define the collinear region *b* as a cone with the opening angle $2\bar{\theta}_0$ ($\bar{\theta}_0 \ll 1$) along the scattered electron momentum direction; therefore, the maximum value of $\bar{z}_{+,-}$ is

$$\bar{z}_0 = \frac{y_0^2 \varepsilon_1^2 \bar{\theta}_0^2}{m^2} \gg 1.$$

According to the above, we can write the tensor $T_{\mu\nu}^{(b)}$ [see Eq. (5)] as

$$T_{\mu\nu}^{(b)} = \frac{16}{m^4} I^{(b)} (Q_{\mu\nu} + i\lambda E_{\mu\nu}). \tag{28}$$

To derive the quantity $I^{(b)}$, we first have to find the quantity $I^{(a)}$, which is the sum of all contributions accompanied by the structure $Q_{\mu\nu}$ in the right-hand sides of Eqs. (7), (10), and (14), and then use the rule

$$I^{(b)} = I^{(a)}(x_+ \longrightarrow -y_-, x_- \longrightarrow -y_+, a_1 \longrightarrow -b_2, a_2 \longrightarrow -b_1, a \longrightarrow a, \Delta \longrightarrow d), \tag{29}$$

where

$$b_2 = \frac{1}{y_+}(1 + y_+^2 + y_+^2 \bar{z}_+), \qquad b_2 = \frac{1}{y_-}(1 + y_-^2 + y_+^2 \bar{z}_-), \quad d = a + b_1 + b_2.$$

The result of the angular integration of the ECT in region *b* can be written as [cf. Eqs. (17), (18), and (19)]

$$\overline{L_{\mu\nu}^{(b)}} = \frac{a^2}{\pi^2} dy_+ dy_- \ln \bar{z}_0 (Q_{\mu\nu} + i\lambda E_{\mu\nu}) \overline{I^{(b)}}, \tag{30}$$

$$\overline{I^{(b)}} = \frac{1+\eta^2}{(\eta-1)^4}\left[\frac{1}{2}(y_+^2 + y_-^2)\ln\frac{\bar{z}_0 y_+^2 y_-^2}{\eta^2} - (y_+ - y_-)^2\right.$$

$$+\frac{8\eta y_+ y_-}{1+\eta^2}\Big] + \frac{y_-^2+\eta^2}{(1+y_+)^4}\Big[\frac{1}{2}(y_+^2+1)\ln\frac{\bar z_0 y_+^2 y_-^2}{\eta^2}$$

$$-(y_+-1)^2+\frac{8\eta y_+ y_-}{y_-^2+\eta^2}\Big] + 2\Big[-\frac{y_+^2+\eta^2}{(1+y_+)(1-\eta)}$$

$$\times\ln\frac{(\eta-1)(1+y_+)}{\eta y_+}+\frac{y_+y_--\eta}{(1+y_+)^2}+\frac{y_+-\eta y_-}{(1-\eta)^2}\Big],$$

where $\eta = 1 + y_+ + y_-$.

We can also integrate the right-hand side of Eq. (30) with respect to y_- or y_+ at fixed values of η—$y_+ + y_- = \eta - 1$—because in region b the 4-momentum of the heavy photon depends on η: $q_b = p_2\eta - p_1$. The corresponding expression can be obtained using the symmetric part of $\tilde L_{\mu\nu}^{(a)}$ in Eq. (21) by the rule

$$\tilde L_{\mu\nu}^{(b)}(\eta,\bar z_0) = -\mathrm{sym}\{\tilde L_{\mu\nu}^{(a)}(z\longrightarrow\eta, z_0\longrightarrow\bar z_0), \quad (31)$$

and the result is given by

$$\tilde L_{\mu\nu}^{(b)} = \frac{\alpha^2}{\pi^2}d\eta\ln\bar z_0\Big\{\ln\bar z_0\Big[\frac{1+\eta^2}{3(\eta-1)}+\frac{\eta-1}{6\eta}(4+7\eta+4\eta^2)$$

$$-(1+\eta)\ln\eta\Big]-4(1+\eta)\ln\eta\ln(\eta-1)-\frac{2}{\eta-1}\ln^2\eta$$

$$-\frac{2(3\eta^2-1)}{\eta-1}L_{i2}(1-\eta)$$

$$-\frac{2}{3}\Big[-4\eta^2-5\eta+1-\frac{4}{\eta(\eta-1)}\Big]\ln(\eta-1) \quad (32)$$

$$-\frac{1}{3}\Big(8\eta^2+5\eta-7+\frac{13}{\eta-1}\Big)\ln\eta+\frac{2}{3}\eta^2$$

$$-\frac{136}{9}\eta+\frac{107}{9}+\frac{4}{3\eta}-\frac{20}{9(\eta-1)}\Big\}(Q_{\mu\nu}+i\lambda E_{\mu\nu}).$$

At this point, we note that, for processes in which the entire energy of the initial electron transforms into the energy of the electromagnetic jet along its momentum direction (the energy is not transferred by a heavy photon), the variable η is equal to $1/z$ because in this case $\varepsilon_2 = \varepsilon_1 - \varepsilon_+ - \varepsilon_-$. This allows us to formulate the substitution law (31) in terms of the same variables z and z_0. In fact, this approximation is valid for the description of t-channel photon exchange processes within the framework of the impact factor representation [22]. This law was used in calculations of QED corrections to the small-angle Bhabha cross section at LEP1 [14].

3. QUASI-COLLINEAR KINEMATICS

Inside the collinear regions, a and b (investigated above both the photon and fermion PDs) can be small compared with the heavy photon mass q^2. In general, the smallness of each PD gives a large logarithm in the ECT after the corresponding angular integration. Therefore, we have a double-logarithmic behavior of the tensor $\tilde L_{\mu\nu}$. Besides double-logarithmic terms, the contribution of collinear regions also contains a single logarithm and constant relative to the variable z_0.

Contributions of the last kind can also arise in kinematical regions when all the fermion PDs in the underlying Feynman diagrams have the same order as q^2 and only photon PDs remain small. Traditionally, we call this the quasi-collinear kinematics.

It is easy to see that there are three quasi-collinear regions in the process under consideration: $\mathbf{p}_- \parallel \mathbf{p}_1$, $\mathbf{p}_- \parallel \mathbf{p}_+$, and $\mathbf{p}_+ \parallel \mathbf{p}_2$. The corresponding contributions to the cross sections for the unpolarized initial electron were studied in part for the DIS [11] and small-angle Bhabha [14] processes.

In what follows, we completely analyze these regions for a longitudinally polarized initial electron at the level of a universal quantity, the ECT. To simplify the calculations, we extensively use the substitution laws based on the topological equivalence of the relevant TDs in every region. The final result has a compact form, and we keep single logarithmic as well as power contributions. In the next-to-leading accuracy, we demonstrate the elimination of the angular auxiliary parameters θ_0 and $\bar\theta_0$ in the case where the separation of collinear and quasi-collinear regions has no physical sense.

3.1. Contribution of the $\mathbf{p}_- \parallel \mathbf{p}_1$ *Region*

In the quasi-collinear region $\mathbf{p}_- \parallel \mathbf{p}_1$, only the (5, 6) set of the TD contributes. In this case, the small photon PD reads

$$q_-^2 = (p_- - p_1)^2 = m^2(2-a_2)$$
$$= -x_- m^2\Big[\frac{(1-x_-)^2}{x_-^2}+z_-\Big], \quad (33)$$

and the phase space of the created electron with the 4-momentum p_- is

$$\frac{d^3p_-}{\varepsilon_-} = \frac{1}{2}m^2 x_- dx_- d\varphi dz_-, \quad (34)$$

where φ is the azimuth angle of the vector $\mathbf{p}_-$ in the coordinate system with the z-axis along the vector $\mathbf{p}_1$.

In investigating the ECT in this region, it is convenient to introduce the small 4-vector

$$p = \frac{1}{x_-}p_- - p_1 \tag{35}$$

that has the following components in the chosen coordinate system:

$$\left(0, \varepsilon_1\theta_-\cos\varphi, \varepsilon_1\theta_-\sin\varphi, \frac{m^2(1-x_-^2)}{2\varepsilon_1 x_-^2} - \frac{\varepsilon_1\theta_-^2}{2}\right).$$

We see that the longitudinal component of the 4-vector p (the last term in this expression) is parametrically smaller than the perpendicular ones and can be neglected within the chosen accuracy.

When calculating the TD for the (5, 6) set, we have to keep the terms of the types

$$\frac{1}{q_-^4}(m^2, q_-^2, (p, p)), \tag{36}$$

in order to reach the adequate accuracy (including constants relative z_0). The 4-vector p enters the ECT via scalar products and tensor structures [see in what follows Eqs. (40) and (41)].

In accordance with (36), we can write the ECT in the considered region as

$$L_{\mu\nu}^{(3.1)} = \frac{\alpha^2}{2\pi^4}\frac{d^2p_+}{\varepsilon_+}x_-dx_-d\varphi dz_-\frac{m^2}{q_-^4} \times\left[m^2L_{\mu\nu}^m + \frac{q_-^2}{(1-x_-)^2}L_{\mu\nu}^{(q_-)} + \frac{2x_-^2}{(1-x_-)^2}L_{\mu\nu}^{(p)}\right], \tag{37}$$

where the first term on the right side gives only a constant (next-next-to-leading) contribution relative to z_0 (when integrated over the created electron angular phase space, see below) and reads

$$L_{\mu\nu}^m = \frac{1}{ut_1}[(u+t_1)^2\tilde g_{\mu\nu} + 4\bar q^2\tilde p_{1\mu}\tilde p_{1\nu} - 2i\lambda E_{\mu\nu}^m]. \tag{38}$$

The second term inside the parentheses in Eq. (37) leads to only a logarithmic (next-to-leading) contribution and is defined by the formula

$$L_{\mu\nu}^{(q_-)} = \left[\frac{(1+x_-^2)(u^2+t_1^2)}{2ut_1} + 2x_- + \frac{\bar q^2}{s_1}\right]\tilde g_{\mu\nu} - \frac{2\bar q^2}{ut_1}[(\tilde p_+\tilde p_2)_{\mu\nu} - (1+x_-^2)\tilde p_{1\mu}\tilde p_{1\nu}] - \frac{i\lambda(1+x_-)}{ut_1}E_{\mu\nu}^m. \tag{39}$$

Finally, the third term on the right side of Eq. (37) leads to both the logarithmic and constant contributions:

$$L_{\mu\nu}^{(p)} = -\frac{\bar q^2}{2(1-x_-)^2}N^2g_{\mu\nu} - (KZ)_{\mu\nu},$$

$$\bar q = p_2 + p_+ - (1-x_-)p_1, \quad N = \chi_+ - \chi_2,$$
$$K_\mu = \frac{N}{1-x_-}p_{2\mu} + \chi_2 p_{1\mu} + p_\mu,$$
$$Z_\mu = -\frac{N}{1-x_-}p_{+\mu} + \chi_+ p_{1\mu} + p_\mu. \tag{40}$$

To describe the tensor $L_{\mu\nu}^{(3.1)}$, we introduced the notation

$$E_{\mu\nu}^m = (u-t_1)E_{\mu\nu}(q, p_2) + [s_1 + (1-x_-)t_1]E_{\mu\nu}(q, p_1),$$
$$E_{\mu\nu}(a, b) = \epsilon_{\mu\nu\lambda\rho}a_\lambda b_\rho,$$
$$s_1 = 2p_2p_+, \quad t_1 = -2p_1p_+, \tag{41}$$
$$\chi_+ = \frac{2p_+p}{t_1}, \quad \chi_2 = \frac{2p_2p}{u},$$
$$(ab)_{\mu\nu} = a_\mu b_\nu + a_\nu b_\mu, \quad \tilde a_\mu = a_\mu - \frac{a\bar q}{\bar q^2}\bar q_\mu,$$
$$\tilde g_{\mu\nu} = g_{\mu\nu} - \frac{\bar q_\mu\bar q_\nu}{\bar q^2}, \quad \tilde a\bar q = 0.$$

The tensor $L_{\mu\nu}^{(p)}$ satisfies the condition: $L_{\mu\nu}^{(p)}q_\nu = 0$. Therefore, we could, in principle, write it in terms of "tilde" quantities as defined in the last line of relations (41). However, our strategy (as concerns this tensor) is first to integrate it over the angular variables and then to express it through quantities of this type.

In the region $\mathbf{p}_- \parallel \mathbf{p}_1$, we can perform the model-independent integration of $L_{\mu\nu}^{(3.1)}$ over the angular variables z_- and φ. The integration of the first two terms on the right side of Eq. (37) is trivial and can be done using the formulas

$$\int\frac{m^4}{q_-^4}(dz_-)d\varphi = \frac{2\pi}{(1-x_-)^2},$$
$$\int\frac{m^2}{q_-^2}(dz_-)d\varphi = -\frac{2\pi}{x_-}\ln\frac{z_0x^2}{(1-x_-)^2}. \tag{42}$$

In the third term in Eq. (37), we use the relation

$$\int\frac{m^2}{q_-^4}dz_-d\varphi p_\mu p_\nu = -\frac{\pi}{x_-^2}\left(\ln\frac{z_0x_-^2}{(1-x_-)^2} - 1\right)g_{\mu\nu}^\perp, \tag{43}$$

where the perpendicular metric tensor $g_{\mu\nu}^\perp$ has only *xx* and *yy* components in the chosen coordinate system. It acts as

$$a_\mu g_{\mu\nu}^\perp = a_\nu^\perp, \quad a_\mu b_\nu g_{\mu\nu}^\perp = (ab)^\perp, \quad g_{\mu\nu}^\perp p_{1\nu} = 0. \tag{44}$$

It is convenient to write Eq. (43) in the symbolic form as

$$p_\mu p_\nu \longrightarrow -g^\perp_{\mu\nu}, \tag{45}$$

and the meaning of symbol "$\longrightarrow$" in Eq. (45) can be understood by comparing Eqs. (43) and (45). Using this symbol, we can write

$$\chi_+ N \longrightarrow \frac{2}{ut_1}(s_1 + y_0 t_1 - x_+ u),$$

$$\chi_2 N \longrightarrow \frac{2}{ut_1}(-s_1 + y_0 t_1 - x_+ u),$$

$$N p_\mu \longrightarrow \frac{2}{ut_1}[t_1 p_{2\mu} - u p_{+\mu} + (x_+ u - y_0 t_1) p_{1\mu}], \tag{46}$$

$$p_\mu p_\nu + \chi_2\chi_+ p_{1\mu} p_{1\nu} + (\chi_2 + \chi_+)(p_1 p)_{\mu\nu}$$

$$\longrightarrow -g_{\mu\nu} - \frac{2s_1}{ut_1} p_{1\mu} p_{1\nu} - \frac{(u p_+ + t_1 p_2, p_1)_{\mu\nu}}{ut_1}.$$

In writing the last part of relation (46), we used the representation of the metrical tensor

$$g_{\mu\nu} = g^\perp_{\mu\nu} + \frac{1}{\varepsilon_1^2} p_{1\mu} p_{1\nu} + \frac{1}{\varepsilon_1}(g_{\mu z} p_{1\nu} + g_{\nu z} p_{1\mu}) \tag{47}$$

that is valid up to terms of the order m/ε_1.

Looking at relations (46), one can see that the result of the angular integration of the tensor $L^{(p)}_{\mu\nu}$ in the considered region can be written in a covariant form. Due to gauge invariance, we can introduce the "tilde" quantities and then use the equation

$$\tilde p_+ + \tilde p_2 - (1 - x_-)\tilde p_1 = 0.$$

The result has the very simple form:

$$L^{(p)}_{\mu\nu} \longrightarrow -2\left(\frac{\bar q^2 s_1}{(1-x_-)^2 u t_1} - 1\right)\tilde g_{\mu\nu} + \frac{4\bar q^2}{(1-x_-)^2 u t_1}(\tilde p_2 \tilde p_+)_{\mu\nu}, \tag{48}$$

$$\bar q^2 = s_1 + (1 - x_-)(u + t_1).$$

Using Eqs. (42) and (43) and Eqs. (45) and (46), we can perform the angular integration on the right-hand side of Eq. (37) and write the contribution of the quasi-collinear region $\mathbf{p}_- \| \mathbf{p}_1$ to the ECT as

$$\tilde L^{(3.1)}_{\mu\nu} = \frac{\alpha^2}{\pi^3}\frac{d^3 p_+}{\varepsilon_+}\frac{dx_-}{(1-x_-)^2 u t_1} \times\left[x_- a_{\mu\nu} + b_{\mu\nu}\ln\frac{z_0 x_-^2}{(1-x_-)^2}\right], \tag{49}$$

where

$$a_{\mu\nu} = c\tilde g_{\mu\nu} + \frac{4\bar q^2}{(1-x_-)^2}(\tilde p_{2\mu}\tilde p_{2\nu} + \tilde p_{+\mu}\tilde p_{+\nu}) - 2i\lambda E^m_{\mu\nu}, \tag{50}$$

$$c = u^2 + t_1^2 + \frac{2\bar q^2 s_1}{(1-x_-)^2},$$

$$b_{\mu\nu} = (1 + x_-^2)\left[-\frac{c}{2}\tilde g_{\mu\nu} - \frac{2\bar q^2}{(1-x_-)^2}(\tilde p_{2\mu}\tilde p_{2\nu} + \tilde p_{+\mu}\tilde p_{+\nu})\right] + i\lambda(1 + x_-)E^m_{\mu\nu}. \tag{51}$$

In the situations where the created electron–positron pair is not observed, we have to sum the contributions of the collinear and quasi-collinear regions. In such cases, the parameter θ_0 that separates these regions has no intrinsic physical meaning and must disappear in the final expression for any observed physical quantity. In the next-to-leading accuracy, this fact leads to the cancellation of all terms proportional to $\ln\theta_0^2 \ln(\varepsilon_1^2/m^2)$. Let us show that this cancellation occurs in the contribution of the (5, 6) set of the TD. To extract the corresponding term in integrating the right side of Eq. (49) over the created positron angular phase space

$$\frac{d^3 p_+}{\varepsilon_+} = \varepsilon_1^2 x_+ dx_+ d\varphi_+ dc_+ \tag{52}$$

(where $c_+ = \cos\theta_+$ and φ_+ is the positron azimuth angle), it is convenient to write the tensor $b_{\mu\nu}$ on the right side of Eq. (49) as

$$b_{\mu\nu}(c_+) = [b_{\mu\nu}(c_+) - b_{\mu\nu}(1)] + b_{\mu\nu}(1). \tag{53}$$

The upper integration limit over c_+ is equal to $\cos\theta_0$. On the right-hand side of Eq. (49), only the pole-like term proportional to $b_{\mu\nu}(1)/t_1$ gives the contribution that we want to extract. Using

$$\bar q^2 = zu, \quad s_1 = -x_+ u, \quad z\tilde p_1 = \tilde p_2, \tag{54}$$

for $b_{\mu\nu}(1)$, we obtain the desired contribution in the form

$$-\frac{\alpha^2}{\pi^2} dx_+ dx_- \ln\theta_0^2 \ln\frac{\varepsilon_1^2}{m^2} \times\left[\frac{(1 + x_-^2)(z^2 + x_+^2)}{(1-x_-)^4} Q_{\mu\nu} + i\lambda\frac{(z - x_+)(1 + x_-)}{(1-x_-)^2} E_{\mu\nu}\right]. \tag{55}$$

It suffices to look at Eqs. (17), (19), and (20) to see that all the terms containing $\ln\theta_0^2 \ln(\varepsilon_1^2/m^2)$ vanish when we add the collinear and quasi-collinear contributions of the (5, 6) set of the TD. Usually, the contribution of the quasi-collinear regions added to the collinear

one restores the argument of the leading logarithm such that $\ln^2 z_0$ transforms into $\ln^2(-u/m^2)$ (see, for example, [14]). This restoration is also connected with the contribution of the lower integration limit of the pole-like term that depends on specific physical applications. In what follows, we do not concentrate on this point any more.

3.2. The Contribution of the $\mathbf{p}_- \parallel \mathbf{p}_+$ *Region*

In the quasi-collinear region $\mathbf{p}_- \parallel \mathbf{p}_+$, only the (1, 2) set of the TD contributes. The small photon PD in this region has the virtuality

$$q_\pm^2 = (p_+ + p_-)^2 = m^2 x_+ x_- \left[\frac{(1-z)^2}{x_+^2 x_-^2} + z_\pm\right], \qquad z_\pm = \frac{\varepsilon_1^2 \theta_\pm^2}{m^2}, \tag{56}$$

where $\theta_\pm = \widehat{\mathbf{p}_+ \mathbf{p}_-}$ is the angle between the created electron and positron momentum directions. Here, we use a somewhat different parameterization for $q_\pm^2$ than for the collinear region *a*. The phase space of the created electron can be parameterized as

$$\frac{d^3 p_-}{\varepsilon_-} = \frac{m^2}{2} x_- dx_- d\varphi_- dz_\pm, \tag{57}$$

where φ_- is now the azimuth angle of the vector $\mathbf{p}_-$ in the system with *z*-axis along the direction of $\mathbf{p}_+$.

If we introduce the small 4-vector

$$h = \frac{x_+}{x_-} p_- - p_+, \tag{58}$$

we can write the ECT in the considered region in complete similarity with Eq. (37) as

$$L_{\mu\nu}^{(3.2)} = \frac{\alpha^2}{2\pi^4} \frac{d^3 p_+}{\varepsilon_+} x_- dx_- d\varphi_- dz_\pm \frac{m^2}{q_\pm^4} \times \left[m^2 \bar{L}_{\mu\nu}^m + \frac{q_\pm^2 x_+^2}{(1-z)^2} L_{\mu\nu}^{(q_\pm)} + \frac{2x_-^2}{(1-z)^2} L_{\mu\nu}^{(h)}\right]. \tag{59}$$

It follows from Fig. 2 that the (1, 2) set of the TD can be obtained from the (5, 6) one by interchanging the 4-momenta p_1 and $-p_+$. This operation changes the conditions of the polarized electron. This is why the corresponding substitution law [see Eq. (9)]

$$\hat{P} = \left[-p_+ \longleftrightarrow p_1, (x_+, x_-, y_0, z_-, p) \longrightarrow \left(\frac{1}{x_+}, \frac{-x_-}{x_+}, \frac{-y_0}{x_+}, x_+^2 z_\pm, -h\right)\right], \tag{60}$$

$$\hat{P}(u, s_1, t_1) \longrightarrow (s_1, u, t_1)$$

allows us to derive from Eq. (37) only the symmetric spin-independent part of the ECT for the (1, 2) set,

$$\mathrm{sym}\left[m^2 \bar{L}_{\mu\nu}^m + \frac{q_\pm^2 x_+^2}{(1-z)^2} L_{\mu\nu}^{(q_\pm)} + \frac{2x_-^2}{(1-z)^2} L_{\mu\nu}^{(h)}\right] = \hat{P}\,\mathrm{sym}\left[m^2 L_{\mu\nu}^m + \frac{q_-^2}{(1-x_-)^2} L_{\mu\nu}^{(q_-)} + \frac{2x^2}{(1-x_-)^2} L_{\mu\nu}^{(p)}\right]. \tag{61}$$

The antisymmetric spin-dependent part of the ECT requires independent calculations. The result can be written as

$$\bar{L}_{\mu\nu}^m = \frac{1}{s_1 t_1}[(s_1 + t_1)^2 \tilde{g}_{\mu\nu} + 4Q^2 \tilde{p}_{+\mu} \tilde{p}_{+\nu} - 2i\lambda(t_1 - s_1) E_{\mu\nu}(Q, p_+)], \tag{62}$$

$$Q = p_2 - p_1 + \frac{1-z}{x_+} p_+, \quad \tilde{a}_\mu = a_\mu - \frac{aQ}{Q^2} Q_\mu,$$

$$L_{\mu\nu}^{(q_\pm)} = \frac{1}{x_+^2 s_1 t_1} \times \left\{\left[\frac{(x_+^2 + x_-^2)(s_1^2 + t_1^2)}{2} - 2x_+ x_- s_1 t_1 + x_+^2 Q^2 u\right]\tilde{g}_{\mu\nu} + 2Q^2[x_+^2 (\tilde{p}_1 \tilde{p}_2)_{\mu\nu} + (x_+^2 + x_-^2)\tilde{p}_{+\mu}\tilde{p}_{+\nu}] - i\lambda x_+[(x_+ u + (1-z)t_1) E_{\mu\nu}(Q, p_1) + (x_+ u + s_1(1-z)) E_{\mu\nu}(Q, p_2) + 2x_-(s_1 - t_1) E_{\mu\nu}(Q, p_+)]\right\}, \tag{63}$$

$$L_{\mu\nu}^{(h)} = -\frac{x_+^2 Q^2}{2(1-z)^2} N_h^2 g_{\mu\nu} - (K^h Z^h)_{\mu\nu} + i\lambda \frac{N_h x_+^2}{(1-z)^2}\left[\chi_1 E_{\mu\nu}(Q, p_2) + \chi_2^h E_{\mu\nu} - \frac{1-z}{x_+} E_{\mu\nu}(Q, h)\right],$$

$$N^h = \chi_1 + \chi_2^h, \quad \chi_1 = \frac{2p_1 h}{t_1}, \quad \chi_2^h = \frac{2p_2 h}{s_1}, \tag{64}$$

$$K_\mu^h = \frac{x_+ N^h}{1-z} p_{2\mu} + \chi_2^h p_{+\mu} - h_\mu,$$

$$Z_\mu^h = \frac{x_+ N^h}{1-z} p_{1\mu} - \chi_1 p_{+\mu} - h_\mu.$$

The tensor $L_{\mu\nu}^{(h)}$ satisfies the condition $L_{\mu\nu}^{(h)} Q_\nu = 0$ because of the gauge invariance. Obviously, we can

eliminate the structure $E_{\mu\nu}(Q, p_+)$ using the definition of the 4-vector Q given in (62). However, the expressions on the right sides of Eqs. (62) and (63) are more compact in our opinion.

The angular integration of the tensor $L_{\mu\nu}^{(3.2)}$ can be carried out with the help of the relations

$$\int \frac{m^4}{q_\pm^4} dz_\pm d\varphi_- = \frac{2\pi}{(1-z)^2},$$

$$\int \frac{m^2}{q_\pm^2} dz_\pm d\varphi_- = \frac{2\pi}{x_+x_-} \ln \frac{z_a x_+^2 x_-^2}{(1-z)^2}, \tag{65}$$

$$\int \frac{m^2}{q_\pm^4} dz_\pm d\varphi_- h_\mu h_\nu = -\frac{\pi}{x_-^2}\left(\ln \frac{z_a x_+^2 x_-^2}{(1-z)^2} - 1 \right) g_{\mu\nu}^\perp,$$

where $z_a = \varepsilon_1^2 \theta_a^2 / m^2 \gg 1$ and the perpendicular metric tensor $g_{\mu\nu}^\perp$ has only the xx and yy components in the chosen coordinate system (the z-axis along $\mathbf{p}_+$). In writing the last part of relation (65), similarly to Eq. (43), only the perpendicular components of the 4-vector h in this coordinate system was taken into account. We introduced the parameter $\theta_a \ll 1$ that defines the quasi-collinear region $\mathbf{p}_- \parallel \mathbf{p}_+$ such that $\theta_\pm \le \theta_a$. Note that Eqs. (65) can be derived from the corresponding Eqs. (42) and (43) by the application of the operation $(1/x_+^2)\hat{P}$.

Using the symbolic form of the last relation in (65), we can write

$$\chi_2^h N_h \longrightarrow \frac{2}{x_+ s_1 t_1}(x_+ u + s_1 + y_0 t_1),$$

$$\chi_1 N_h \longrightarrow \frac{2}{x_+ s_1 t_1}(x_+ u - s_1 - y_0 t_1),$$

$$N_h h_\mu \longrightarrow -\frac{2}{s_1 t_1}\left(t_1 p_{2\mu} + s_1 p_{1\mu} - \frac{s_1 + y_0 t_1}{x_+} p_{+\mu} \right), \tag{66}$$

$$h_\mu h_\nu - \chi_2^h \chi_1 p_{+\mu} p_{+\nu} + (\chi_1 - \chi_2^h)(p_+ h)_{\mu\nu} \longrightarrow -g_{\mu\nu} - \frac{2u}{s_1 t_1} p_{+\mu} p_{+\nu} - \frac{(s_1 p_1 - t_1 p_2,\, p_+)_{\mu\nu}}{s_1 t_1},$$

and the meaning of the symbol "$\longrightarrow$" in relations (66) can be understood by comparing the last line in (65) with its symbolic form

$$h_\mu h_\nu \longrightarrow -g_{\mu\nu}^\perp.$$

It is easy to see that Eqs. (66) follow from Eqs. (46) after the operator $\hat{P}$ is applied to the latter. The metric tensor in (66) is defined by analogy with (47),

$$g_{\mu\nu} = g_{\mu\nu}^\perp + \frac{1}{\varepsilon_+^2} p_{+\mu} p_{+\nu} + \frac{1}{\varepsilon_+}(g_{\mu z} p_{+\nu} + g_{\nu z} p_{+\mu}), \quad \varepsilon_+ \gg m. \tag{67}$$

From Eqs. (66), we derive

$$L_{\mu\nu}^{(h)} \longrightarrow -2\left(\frac{x_+^2 Q^2 u}{(1-z)^2 s_1 t_1} - 1 \right) \tilde{g}_{\mu\nu} - \frac{4x_+^2 Q^2}{(1-z)^2 s_1 t_1} (\tilde{p}_1 \tilde{p}_2)_{\mu\nu} + 2i\lambda \frac{x_+}{(1-z)^2 s_1 t_1} \times [(x_+ u + (1-z)t_1) E_{\mu\nu}(Q, p_2) + (x_+ u + (1-z)s_1) E_{\mu\nu}(Q, p_1)]. \tag{68}$$

The general expression for the contribution of the quasi-collinear region $\mathbf{p}_- \parallel \mathbf{p}_+$ into the ECT can be written as

$$\tilde{L}_{\mu\nu}^{(3.2)} = \frac{\alpha^2}{\pi^3} \frac{d^3 p_+}{\varepsilon_+} \frac{dx_-}{(1-z)^2 s_1 t_1} \times \left[x_- a_{\mu\nu}^{(1)} + b_{\mu\nu}^{(1)} \ln \frac{z_a x_+^2 x_-^2}{(1-z)^2} \right], \tag{69}$$

where $\tilde{L}_{\mu\nu}^{(3.2)}$ is the right-hand side of Eq. (59) integrated over the angular variables φ_- and $z_\pm$, and

$$a_{\mu\nu}^{(1)} = c^{(1)} \tilde{g}_{\mu\nu} + \frac{4x_+^2 Q^2}{(1-z)^2}(\tilde{p}_{2\mu}\tilde{p}_{2\nu} + \tilde{p}_{1\mu}\tilde{p}_{1\nu}) - 2i\lambda \frac{x_+}{(1-z)^2}[(ux_+ + (1-z)s_1) E_{\mu\nu}(Q, p_2) + (ux_+ + (1-z)t_1) E_{\mu\nu}(Q, p_1)], \tag{70}$$

$$b_{\mu\nu}^{(1)} = \frac{(x_+^2 + x_-^2)}{x_+} \left\{ \frac{c^{(1)}}{2} \tilde{g}_{\mu\nu} + \frac{2x_+^2 Q^2}{(1-z)^2}(\tilde{p}_{2\mu}\tilde{p}_{2\nu} + \tilde{p}_{1\mu}\tilde{p}_{1\nu}) - i\lambda \frac{x_+}{(1-z)^2}[(ux_+ + (1-z)t_1) E_{\mu\nu}(Q, p_1) + (ux_+ + (1-z)s_1) E_{\mu\nu}(Q, p_2)] \right\}, \tag{71}$$

$$c^{(1)} = s_1^2 + t_1^2 + \frac{2x_+^2}{(1-z)^2} u Q^2.$$

The ECT in the $\mathbf{p}_- \parallel \mathbf{p}_+$ region has a pole-like behavior both for small t_1 and small s_1. One can verify that the

terms proportional to $\ln\theta_0^2\ln(\varepsilon_1^2/m^2)$ are cancelled in the sum of the contributions of the quasi-collinear region $\mathbf{p}_- \parallel \mathbf{p}_+$ at small t_1 and collinear region *a*, while the terms proportional to $\ln\bar{\theta}_0^2\ln(\varepsilon_1^2/m^2)$ are cancelled with the contributions of region $\mathbf{p}_- \parallel \mathbf{p}_+$ at small s_1 and collinear region *b*.

In the limiting case $|t_1| \ll |u|, s_1, |Q^2|$, we can extract the corresponding terms using Eqs. (69) and (71) and relations (54) in the same way as this was done in Section 3.1, and the results give

$$-\frac{\alpha^2}{\pi^2}dx_+dx_-\ln\theta_0^2\ln\frac{\varepsilon_1^2}{m^2} \times\frac{(x_+^2+x_-^2)(z^2+1)}{(1-z)^4}[Q_{\mu\nu}+i\lambda E_{\mu\nu}]. \tag{72}$$

Looking at Eqs. (17) and (18), we see that the corresponding contribution of the (1, 2) set of the TD in the collinear region *a* has the opposite sign as compared with expression (72).

To compute the quantity $b_{\mu\nu}^{(1)}$ at $\bar{c}_+ = 1$ in another limiting case $s_1 \ll |u|, |t_1|, |q^2|$, we have to use the relations

$$t_1 = y_+u, \quad Q^2 = \eta u, \quad \tilde{p}_+ = y_+\tilde{p}_2, \quad s_1 = 2\varepsilon_1^2 y_+(1-\bar{c}_+), \quad \bar{c}_+ = \cos\bar{\theta}_+. \tag{73}$$

The result is

$$-\frac{\alpha^2}{\pi^2}dy_+dy_-\ln\bar{\theta}_0^2\ln\frac{\varepsilon_1^2}{m^2} \times\frac{(y_+^2+y_-^2)(\eta^2+1)}{(1-\eta)^4}[Q_{\mu\nu}+i\lambda E_{\mu\nu}]. \tag{74}$$

Expression (74) cancels the contribution containing $\ln\bar{\theta}_0^2\ln(\varepsilon_1^2/m^2)$ from the (1, 2) set of the TD in collinear region *b* following from Eqs. (30) and (31).

3.3. Contribution of the $\mathbf{p}_+ \parallel \mathbf{p}_2$ *Region*

In the quasi-collinear $\mathbf{p}_+ \parallel \mathbf{p}_2$ region, only the (3, 4) set of the TD contributes. As one can see from Fig. 2, the TD of that set can be obtained from the (1, 2) one by interchanging $p_2 \longleftrightarrow p_-$. This substitution does not affect any condition on the polarized electron with the 4-momentum p_1. Therefore, both the spin-independent and spin-dependent parts of the corresponding tensor $L_{\mu\nu}^{(3.3)}$ can be derived by a substitution law using the tensor $L_{\mu\nu}^{(3.2)}$. If we write

$$L_{\mu\nu}^{(3.3)} = \frac{\alpha^2}{2\pi^4}\frac{d^3p_-}{\varepsilon_-}y_+dy_+\frac{m^2}{q_+^4}d\bar{z}_+d\varphi_+L_{\mu\nu}, \qquad q_+^2 = (p_2+p_+)^2, \tag{75}$$

the tensor $L_{\mu\nu}$ on the right side of Eq. (75) can be written using the right-hand side of Eq. (59) as

$$L_{\mu\nu} = \hat{Q}\left[m^2\bar{L}_{\mu\nu}^{(m)} + \frac{q_\pm^2x_+^2}{(1-z)^2}L_{\mu\nu}^{(q_\pm)} + \frac{2x_-^2}{(1-z)^2}L_{\mu\nu}^{(h)}\right], \tag{76}$$

where the operator $\hat{O}$ is defined by

$$\hat{O} = m[p_2 \longleftrightarrow p_-, (p_+, x_-, x_+, h) \longrightarrow (y_+p_2, y_0, y_0y_+, -y_+f)], \qquad f = \frac{1}{y_+}p_+ - p_2. \tag{77}$$

(For the notation used here, see Section 2.2.) The action of $\hat{O}$ on the relevant invariants gives

$$\hat{Q}(t_1, s_1, u, q_\pm, Q) = (y_+u, y_+s_2, t_2, q_+, \bar{Q}), \qquad s_2 = 2p_2p_-, \quad t_2 = -2p_1p_-, \qquad \bar{Q} = p_2(1+y_+) + p_- - p_1. \tag{78}$$

We omit all intermediate calculations and only give the final results for the ECT in the $\mathbf{p}_+ \parallel \mathbf{p}_2$ region:

$$\tilde{L}_{\mu\nu}^{(3.3)} = \frac{\alpha^2}{\pi^3}\frac{d^3p_-}{\varepsilon_-}\frac{dy_+}{(1+y_+)^2us_2} \times\left[y_+a_{\mu\nu}^{(2)} + b_{\mu\nu}^{(2)}\ln\frac{\bar{z}_0y_+^2}{(1+y_+)^2}\right], \tag{79}$$

$$a_{\mu\nu}^{(2)} = c^{(2)}\tilde{g}_{\mu\nu} + \frac{4\bar{Q}^2}{(1+y_+)^2}(\tilde{p}_{-\mu}\tilde{p}_{-\nu} + \tilde{p}_{1\mu}\tilde{p}_{1\nu}) - 2i\lambda\frac{1}{(1+y_+)^2}[(t_2+(1+y_+)s_2)E_{\mu\nu}(\bar{Q}, p_-) + (t_2+(1+y_+)u)E_{\mu\nu}(\bar{Q}, p_1)], \tag{80}$$

$$b_{\mu\nu}^{(2)} = (1+y_+^2)\left\{\frac{c^{(2)}}{2}\tilde{g}_{\mu\nu} + \frac{2\bar{Q}^2}{(1+y_+)^2}(\tilde{p}_{-\mu}\tilde{p}_{-\nu} + \tilde{p}_{1\mu}\tilde{p}_{1\nu}) - i\lambda\frac{1}{(1+y_+)^2}[(t_2+(1+y_+)u)E_{\mu\nu}(\bar{Q}, p_1) + (t_2+(1+y_+)s_2)E_{\mu\nu}(\bar{Q}, p_-)]\right\}, \tag{81}$$

$$c^{(2)} = s_2^2 + u^2 + \frac{2t_2\bar{Q}^2}{(1+y_+)^2}.$$

To compute the quantity $b_{\mu\nu}^{(2)}$ at $\bar{c}_- = 1$ and extract [similarly to (55)] the term proportional to $\ln\bar{\theta}_0^2 \ln(\varepsilon_1^2/m^2)$ in the limiting case $s_2 \ll |t_2|, |u|, |\bar{Q}^2|$, we have to use the relations

$$\begin{aligned} t_2 &= y_- u, \quad \bar{Q}^2 = \eta u, \\ s_2 &= 2\varepsilon_2 y_-(1-\bar{c}_-), \quad \bar{c}_- = \cos\bar{\theta}_-. \end{aligned} \tag{82}$$

The result is

$$\begin{aligned} &-\frac{\alpha^2}{\pi^2} dy_+ dy_- \ln\bar{\theta}_0^2 \ln\frac{\varepsilon_1^2}{m^2} \\ &\times \frac{(1+y_+^2)(\eta^2+y_-^2)}{(1+y_-)^4}[Q_{\mu\nu} + i\lambda E_{\mu\nu}]. \end{aligned} \tag{83}$$

Expression (83) cancels the corresponding contribution of the collinear region b from the (3, 4) set of the TD [see Eq. (30)].

4. CONCLUSIONS

For completeness, we must also add the contribution of the virtual and soft (with the energy less than $\varepsilon_1 D$, $D \ll 1$) pair production to the ECT. It can be written in the Born-like form [10]:

$$\begin{aligned} L_{\mu\nu}^{S+V} &= \frac{2\alpha^2}{\pi^2}\left[L^2\left(\frac{2}{3}\ln D + \frac{1}{2}\right)\right. \\ &\left. + L\left(-\frac{17}{6} + \frac{4}{3}\ln^2 D - \frac{20}{9}\ln D - \frac{2\pi^2}{9}\right)\right](Q_{\mu\nu} + i\lambda E_{\mu\nu}), \end{aligned} \tag{84}$$

where $L = \ln(-u/m^2)$. It is easy to show that the parameter D disappears in sum of the contributions of hard (with the energy more than $\varepsilon_1 D$) collinear and quasi-collinear pairs and virtual and soft ones. For this purpose, it is enough to analyze the results of Section 2 in the limits as $z \longrightarrow 1 - D$ and $\eta \longrightarrow 1 + D/y_0$, and to take the observation made at the end of Section 3.1 into account.

In this paper, we calculated the ECT for the electron–positron pair production process at the scattering of a longitudinally polarized electron on a heavy photon. The work was stimulated by recent polarized experiments on deep inelastic scattering [1, 2], but our results can be used to compute the second-order radiative correction due to pair production for a wide class of the scattering and annihilation processes. The contribution of collinear and quasi-collinear kinematical regions of hard pair are studied, and this allows us to find the corresponding correction with the next-to-leading accuracy. The cancellation of the angular auxiliary parameters $\bar{\theta}_0$ and $\bar{\theta}_0$ for an unobserved created pair indicates that our results for collinear regions are in accordance with the quasi-collinear ones. In addition, we conclude that the auxiliary infrared parameter also disappears in the final result for the sum of the hard and soft pair contributions. We give both the spin-independent and spin-dependent parts of the ECT to make polarization effects more transparent at the level of theoretical formulas.

The spin-independent part of our ECT at the differential level coincides with the corresponding results of [10, 11] for both collinear and quasi-collinear kinematics. However, the integration of the ECT in [11] was performed with a few mistakes. The improvements due to the collinear region contribution were found in [14]. Here, we conclude that formula (17) of [10] [which is the analog of our formula (69) for the spin-independent part of the ECT] is also incorrect.

We want to draw the reader's attention to the fact that in contrast to the purely photonic corrections, the leading correction connected with hard pair production has different forms for spin-dependent and spin-independent parts of the ECT because of the singlet channel contribution. This fact indicates that, for asymmetry-like quantities, the full second-order correction will dominate just because of the pair production via the singlet channel because the whole leading non-singlet channel contribution is cancelled in this case [6]. The latter includes all photonic corrections and a non-singlet part of the corrections related with the pair production.

In the present work, we considered the effects connected with the electron–positron pair production. If the beam energy is sufficiently large, the effects of $\mu^+\mu^-$ pair production must also be taken into account. For an event selection with $|u|, |q^2| \gg m^2, \mu^2$, only the diagrams like the (1, 2) set will contribute, but the angular integration is more complicated in this case because within the next-to-leading accuracy one must keep terms of the order $\ln(\varepsilon_1^2/m^2)$, $\ln(\varepsilon_1^2/\mu^2)$ and $\ln(\mu^2/m^2)$ (here μ is the muon mass). This problem with two scales of fermion masses is open at present. It is very important, for example, for the high-precision description of small-angle events with the e^+e^- pair production at the future $\mu^+\mu^-$ collider [23].

Thus, at present all the necessary ingredients exist to compute the full model-independent second-order radiative corrections (RC) to cross sections and polarization observables of different processes with the next-to-leading accuracy. This RC can be incorporated via the ECT as mentioned in the Introduction. The corresponding two-loop photonic corrections have been found in [6–9]; pair production corrections were found in [10] and in this paper.

ACKNOWLEDGMENTS

We thank V.S. Fadin, G.I. Gakh, A.B. Arbuzov, I.V. Akushevich, and L. Trentadue for discussion. This work was supported in part by INTAS (grant no. 93-1867 ext) and Ukrainian DFFD (grant no. 24/379).

APPENDIX

We outline the starting points for the calculation of the tensor $T_{\mu\nu}$ on the right side of Eq. (5). In the case under consideration, the heavy photon virtuality modulus $|q^2|$ and the invariant $u = -2p_1p_2$ that defines the registration conditions of the scattered electron are much greater than m^2. Therefore, only the diagrams in Fig. 1 contribute in collinear and quasi-collinear kinematics, and the tensor $T_{\mu\nu}$ can be written in general form

$$T_{\mu\nu} = (M^{(12)} - M^{(34)} - M^{56})_\mu \times (M^{(12)} - M^{(34)} - M^{(56)})^+_\nu, \tag{A.1}$$

$$M_\mu^{(12)} = \bar{u}(p_2)Q_{\mu\lambda}^{(12)}u(p_1)\frac{1}{q_\pm^2}\bar{u}(p_-)\gamma_\lambda v(p_+),$$

$$M_\mu^{(34)} = \hat{O}M_\mu^{(12)}, \quad M_\mu^{(56)} = \hat{P}M_\mu^{(12)},$$

$$Q_{\mu\nu}^{(12)} = \gamma_\mu\frac{\hat{p}_1 - \hat{q}_\pm + m}{q_\pm^2 - 2p_1q_\pm}\gamma_\lambda + \gamma_\lambda\frac{\hat{p}_2 + \hat{q}_\pm + m}{q_\pm^2 + 2p_2q_\pm}\gamma_\mu, \tag{A.2}$$

where we used the notation of Section 3.

In the collinear region *a*, the (3, 4) set does not contribute and we have

$$T_{\mu\nu}^{(a)} = M_\mu^{(12)}M_\nu^{(12)+} + M_\mu^{(56)}M_\nu^{(56)+} - (M_\mu^{(12)}M_\nu^{(56)+} + M_\mu^{(56)}M_\nu^{(12)+}). \tag{A.3}$$

The first term on the right-hand side of Eq. (A.3) corresponds to the contribution of the (1, 2) set of the TD in Fig. 2 that describes only the non-singlet channel. It gives

$$M_\mu^{(12)}M_\nu^{(12)+} = \frac{1}{q_\pm^4}[-2q_\pm^2 g_{\lambda\rho} + 4(p_+p_-)_{\lambda\rho}] \times \mathrm{Tr}(\hat{p}_2 + m)Q_{\mu\lambda}^{(12)}(\hat{p}_1 + m) \times\left(1 + \lambda\gamma_5\left(1 + \frac{m\hat{k}}{p_1k}\right)\right)Q_{\nu\rho}^{(12)+}, \tag{A.4}$$

where we defined the initial electron polarization 4-vector a_μ as [7, 21]

$$a_\mu = \frac{\lambda}{m}\left(p_1 - \frac{m^2k}{p_1k}\right), \quad k = (\varepsilon_1, -\mathbf{p}_1)$$

and used the relation

$$(\hat{p}_1 + m)\left[1 - \frac{\lambda}{m}\gamma_5\left(\hat{p}_1 - \frac{m^2\hat{k}}{p_1k}\right)\right] = (\hat{p}_1 + m)\left[1 + \lambda\gamma_5\left(1 + \frac{m\hat{k}}{p_1k}\right)\right].$$

Note that the vector k vanishes in the final results because in the chosen accuracy it contributes via the scalar product (kp_1) in the same way as for the double-photon emission [7]. The multiplier inside the right brackets on the right-hand side of Eq. (A.4) describes the upper block of the corresponding TD in Fig. 2a, while the trace describes lower one. The part of trace that contains the doubled initial electron helicity λ is symmetric in the indices (λ, ρ) and asymmetric in the (μ, ν) indices.

The second term on the right side of Eq. (A.3) is responsible for the singlet channel contribution. It is related with the (5, 6) set of the TD and can be written as

$$M_\mu^{(56)}M_\nu^{(56)+} = \frac{1}{q_-^4}\left[2q_-^2 g_{\lambda\rho} + 4(p_1p_-)_{\lambda\rho} + 4i\lambda\left(E_{\lambda\rho}(p_1, p_-) + \frac{m^2}{p_1k}E_{\lambda\rho}(p_- - p_1, k)\right)\right] \times \mathrm{Tr}(\hat{p}_2 + m)Q_{\mu\lambda}^{(56)}(\hat{p}_+ - m)Q_{\nu\rho}^{(56)+}. \tag{A.5}$$

The polarized electron belongs to the upper TD block, and we see that the corresponding expression includes both symmetric and antisymmetric parts. To derive the spin-dependent part of the (5, 6) set, we must compute the part of the trace on the right side of Eq. (A.5) that is asymmetric with respect to each pair of indices (λ, ρ) and (μ, ν). Only the spin-independent parts on the right side of Eqs. (A.4) and (A.5) are transformed into one another by the operator $\hat{P}$ that implements the replacement $p_1 \longleftrightarrow -p_+$. The spin-dependent parts must be calculated independently.

The third term in the parentheses on the right side of Eq. (A.3) describes the interference of singlet and non-singlet channels. It corresponds to another class of TD (Fig. 2). If we represent it as a sum of symmetric and antisymmetric parts, we have

$$\mathrm{sym}\{M_\mu^{(12)}M_\nu^{(56)+} + M_\mu^{(56)}M_\nu^{(12)+}\} = M_\mu^{(12)}M_\nu^{(56)+} + (\mu \longleftrightarrow \nu),$$

$$\mathrm{asym}\{M_\mu^{(12)}M_\nu^{(56)+} + M_\mu^{(56)}M_\nu^{(12)+}\} = M_\mu^{(12)}M_\nu^{(56)+} - (\mu \longleftrightarrow \nu),$$

where

$$M_\mu^{(12)}M_\nu^{(56)+} = \frac{1}{q_-^2q_\pm^2}(\hat{p}_2 + m)Q_{\mu\lambda}^{(12)}(p_1 + m) \times\left[1 + \lambda\gamma_5\left(1 + \frac{m\hat{k}}{p_1k}\right)\right]\gamma_\rho(\hat{p}_- + m)\gamma_\lambda(\hat{p}_+ - m)Q_{\nu\rho}^{(56)+}. \tag{A.6}$$

Equations (A.4), (A.5), and (A.6) are the starting points of the calculation in both the collinear and quasi-collinear regions because the full contribution of the (3, 4) set of the TD can be obtained from the (1, 2) one by the operator $\hat{O}$ that changes $p_2 \longleftrightarrow p_-$.

REFERENCES

1. D. Adams *et al.* (SMC Collab.), Phys. Rev. D **56**, 5330 (1997).
2. K. Acherstaff *et al.* (HERMES Collab.), Phys. Lett. B **404**, 383 (1997).
3. T. V. Kukhto and N. M. Shumeiko, Nucl. Phys. B **219**, 412 (1983).
4. I. V. Akushevich and N. M. Shumeiko, J. Phys. G **20**, 513 (1994).
5. R. P. Feynman, *Photon–Hadron Interaction* (W. A. Benjamin, Reading, Mass., 1972).
6. I. V. Akushevich, A. B. Arbuzov, and E. A. Kuraev, Phys. Lett. B **432**, 222 (1998).
7. M. Konchatnij and N. P. Merenkov, JETP Lett. **69**, 893 (1999).
8. G. I. Gakh, M. I. Konchatnij, and N. P. Merenkov, submitted to Pis'ma Zh. Éksp. Teor. Fiz. [submitted to JETP Lett.].
9. R. Barbieri, J. A. Mignaco, and E. Remiddi, Nuovo Cimento A **11**, 824 (1972); A. B. Arbuzov, V. S. Fadin, E. A. Kuraev, *et al.*, Nucl. Phys. B **485**, 457 (1997).
10. O. Adriani, M. Aquilar-Benitez, *et al.*, Phys. Rep. **236**, 1 (1993).
11. N. P. Merenkov, Yad. Fiz. **50**, 750 (1989) [Sov. J. Nucl. Phys. **50**, 469 (1989)].
12. N. P. Merenkov, Ukr. Fiz. Zh. **34**, 1329 (1989).
13. A. B. Arbuzov, E. A. Kuraev, N. P. Merenkov, and L. Trentadue, Nucl. Phys. B **474**, 271 (1996).
14. A. B. Arbuzov, E. A. Kuraev, N. P. Merenkov, and L. Trentadue, Zh. Éksp. Teor. Fiz. **108**, 1164 (1995) [JETP **81**, 638 (1995)]; Preprint CERN-TH/95-241.
15. F. A. Berends, W. L. van Neervan, and G. J. M. Burgers, Nucl. Phys. B **297**, 429 (1988); F. A. Berends, P. De Causmaecker, R. Gastmans, *et al.*, Nucl. Phys. B **264**, 243 (1984).
16. E. A. Kuraev, A. N. Pyorishkin, and F. S. Fadin, Preprint No. 89-91, JNR (Novosibirsk, 1991).
17. E. A. Kuraev and V. S. Fadin, Yad. Fiz. **41**, 733 (1985) [Sov. J. Nucl. Phys. **41**, 466 (1985)]; S. Jadach, M. Skrzypek, and B. F. L. Ward, Phys. Rev. D **47**, 3773 (1993); S. Catani and L. Trentadue, JETP Lett. **51**, 83 (1990); A. B. Arbuzov, E. A. Kuraev, N. P. Merenkov, and L. Trentadue, JHEP (12) (1998) 009.
18. A. B. Arbuzov, E. A. Kuraev, N. P. Merenkov, and L. Trentadue, Yad. Fiz. **60**, 673 (1997) [Phys. At. Nucl. **60**, 591 (1997)].
19. V. N. Baier, V. S. Fadin, and V. A. Khoze, Nucl. Phys. B **65**, 381 (1973).
20. N. P. Merenkov, Yad. Fiz. **48**, 1782 (1988) [Sov. J. Nucl. Phys. **48**, 1073 (1988)].
21. A. P. Bukhvostov, E. A. Kuraev, and L. N. Lipatov, Preprint 83-147, INP (Budker Institute of Nuclear Physics, Siberian Division, Russian Academy of Sciences, Novosibirsk, 1983); Zh. Éksp. Teor. Fiz. **87**, 37 (1984) [Sov. Phys. JETP **60**, 22 (1984)].
22. H. Cheng and T. T. Wu, Phys. Rev. Lett. **23**, 630 (1969); V. G. Zima and N. P. Merenkov, Yad. Fiz. **24**, 998 (1976) [Sov. J. Nucl. Phys. **24**, 522 (1976)]; V. N. Baier, V. S. Fadin, V. Khoze, and E. Kuraev, Phys. Rep. **78**, 294 (1981).
23. C. M. Ankenbrandt, M. Atac, B. Autin, *et al.*, *Status of Muon Collider Research and Development and Future Plans*, BNL-65623, Fermilab-PUB-98/179, LBNL-41935 (1998).

Journal of Experimental and Theoretical Physics, Vol. 91, No. 1, 2000, pp. 106–110.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 1, 2000, pp. 119–124.

PLASMA, GASES

Levitation of Charged Macroparticles in the Anode Region of a Glow Discharge

A. A. Samaryan*, O. S. Vaulina, A. P. Nefedov, O. F. Petrov, and V. E. Fortov

Institute of Heat Physics of Extreme States, Russian Academy of Sciences, Moscow, 127412 Russia
**e-mail: samarian@hedric.msk.su*

Received January 17, 2000

Abstract—Levitation of dust particles in the anode region of a dc glow discharge was observed for the first time. A dust cloud of several tens of particles formed at a distance of several millimeters above the central part of the anode. When the discharge parameters were varied, the shape of the cloud and its position above the anode varied. An analysis of the experimental conditions revealed that these particles are positively charged in contrast to other experiments on the levitation of dust particles in a gas-discharge plasma. An estimate of the particle charge taking into account processes of electron emission from their surface is consistent with results of measurements of the electric field strength. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

A characteristic feature of plasmas containing dispersed-phase macroparticles is that particles of micron dimensions may acquire appreciable charges (of the order of 10^3–10^5 electron charges) when introduced into or formed in the plasma (for example, as a result of condensation). Under certain conditions, this can compensate for the influence of gravity and lead to levitation of the particles. So far, levitating particles have been observed in the plasma of a capacitive rf discharge [1, 2], an inductive rf discharge [3], and a dc glow discharge [4, 5]. We know that particle levitation and the formation of dusty structures take place in a discharge zone where the parameters exhibit significant inhomogeneity and there is a constant electric field component. In a capacitive rf discharge, such a zone is formed near the electrode, in an induction rf discharge it is formed in the boundary region between the homogeneous quasineutral plasma and the wall, and in a dc glow discharge it is formed in the head of the striation or in an artificially created electric double layer. In [6] Nitter studied the theoretical possibility of the levitation of negatively charged dust particles in the cathode layer of a dc glow discharge. However, an analysis of these results indicates that as a result of the substantial positive space charge in the electrode region at distances of less than $10\lambda_D$ (λ_D is the Debye length) the charge of a dust particle may be reduced substantially. For this reason the theoretically feasible levitation of macroparticles is difficult to achieve in practice. This conclusion is also confirmed by the absence of any experimental studies in which levitation of negatively charged particles was observed in the cathode layer of a dc glow discharge.

One of the main mechanisms for the charging of particles in a gas-discharge plasma is charging by fluxes of electrons and ions. As a result of the higher temperature and the mobility of the electrons the charge of the nonemitting particles in this plasma is negative. The emission of electrons from the surface of a macroparticle can provide conditions for its positive charging. In this case, processes such as thermionic emission, photoemission, and secondary electron emission may play an appreciable role. Among the various well-known studies of macroparticles charged by thermionic emission and photoemission, mention should be made of experiments to observe structures of positively charged particles in a laminar jet of thermal plasma [7] and experiments to observe the behavior of macroparticles charged by photoemission under microgravity [8]. A common feature of these studies is the formation of weakly correlated structures of macroparticles as these move in a dynamic stream. Steady-state conditions similar to the conditions existing in these experiments in a gas-discharge plasma are required to form strictly ordered structures of positively charged particles and to study these. The absence of effective electrical traps capable of achieving levitation of positively charged particles has so far prevented any experimental data from being obtained on the formation of stable structures by these particles. In the present study levitation of positively charged macroparticles in the anode region of a dc glow discharge was observed for the first time.

2. DESCRIPTION OF EXPERIMENT

The experiments were carried out using a cold-electrode glow discharge in air. The discharge was ignited in a vertical steel cylindrical chamber (see Fig. 1). The chamber had an inner diameter of 300 mm, a height of 400 mm, and incorporated circular windows (d = 100 mm) for illumination and observation. The distance between

the electrodes was 46 mm. The cathode consisted of two copper rods of length $l = 55$ mm (spacing between rods $\Delta l = 58$ mm), each interconnected by 150 μm diameter copper wire. The anode was a disk of radius $R_a = 70$ mm with edges made of wire wound onto rods (edge height $h_a = 24$ mm). Single and double probes were also inserted in the chamber. The gas-discharge chamber is shown schematically in Fig. 2. The discharge current was varied between 1 and 15 mA and the pressure between 0.2 and 2 Torr.

Thin-walled hollow spheres of borosilicate glass 10–30 μm in diameter were introduced into the plasma. The wall thickness was approximately 2–3 μm and the bulk particle density was close to 0.8 g/cm^3. The micron-size particles introduced into the plasma were located in a cylindrical container positioned in the upper part of the gas-discharge chamber. The bottom of the container was made of a metal grid with a mesh size of 40 μm. When the container was shaken, particles entered the discharge. The particles were visualized by illumination using a laser probe beam which could be moved in the vertical and radial directions of the chamber. The light scattered by the particles was observed using a CCD camera whose output signal was recorded using a video recorder.

In our experiments we observed the formation of a cloud of several tens of particles at a distance of several millimeters above the central part of the anode (see Fig. 3). The cloud formation process was as follows: after shaking the container, the particles went past the equilibrium position and were then reflected from the anode and suspended at a certain point in space. When the discharge parameters were varied, the shape of the cloud and its position above the anode changed. For example, when the discharge current was reduced, the particle cloud rose and became elongated in the vertical direction. Figure 4a gives the dependence of the effective cloud size

$$L = \sqrt{h^2 + l^2}$$

(h is the vertical dimension of the cloud and l is its horizontal dimension) as a function of the discharge current. Figure 4b shows how the distance H between the lower edge of the cloud and the anode varies as a function of the discharge current. It should be noted that as the current approaches a critical value, which depends on pressure (for example, at 0.4 Torr $I_{cr} = 2.4$ mA), the value of H increases sharply and at $I = I_{cr}$ the cloud "flies away." At a discharge current of 10 mA we usually observed the formation of several clouds suspended low above the electrode and these merged into one as the discharge current decreased.

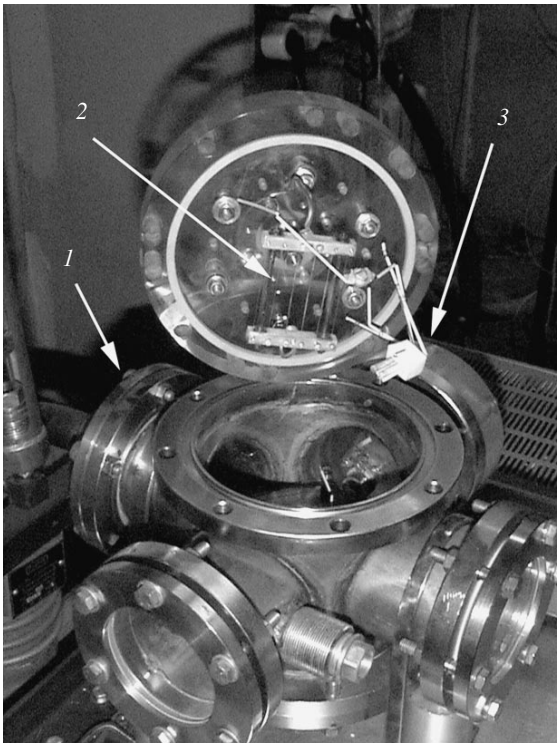


Fig. 1. Experimental apparatus: (*1*) discharge chamber, (*2*) cathode, (*3*) probe.

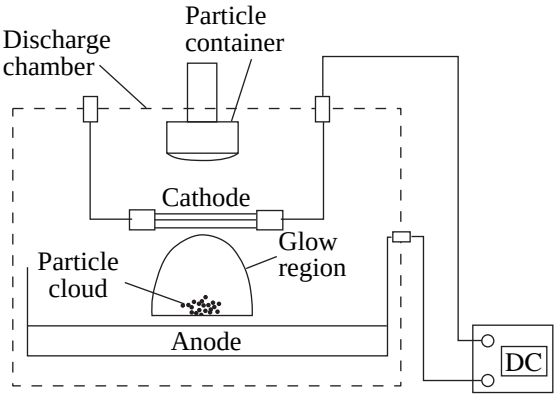


Fig. 2. Scheme of discharge chamber.

3. ANALYSIS OF EXPERIMENTAL RESULTS

One of the key problems in any experiment using a dusty plasma is determining the magnitude and polarity of the macroparticle charge. In order to analyze the experimental results we shall consider the possibility of levitation of negatively charged particles in the anode region of a glow discharge. In this case, either a negative potential drop in the anode region or any other force directed away from the anode can compensate for the gravitational force on the particle. Under our experimental conditions it is impossible for a negative anode potential drop to exist for various reasons considered in [9]. The main reason is that a narrow film of anode glow was observed experimentally near the anode. This suggests that the electrons are accelerated in the anode layer, acquiring an energy of ~15–30 eV which is suffi-

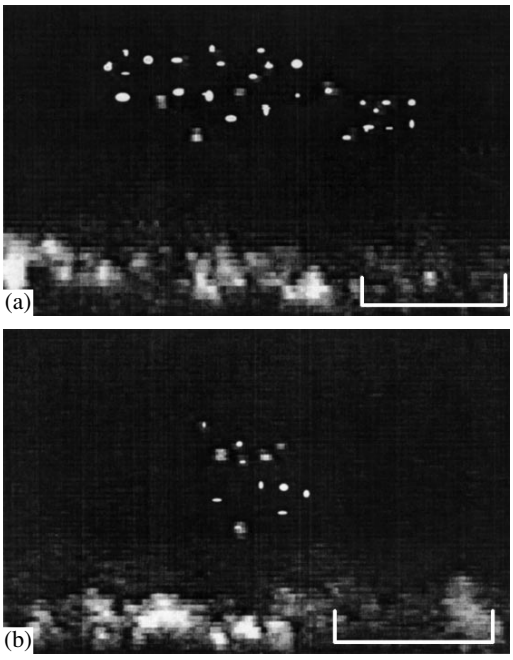


Fig. 3. Video images of particle cloud at $p = 0.3$ Torr: I_p = (a) 6.2 and (b) 3.8 mA. The scale in the figure corresponds to 5 mm.

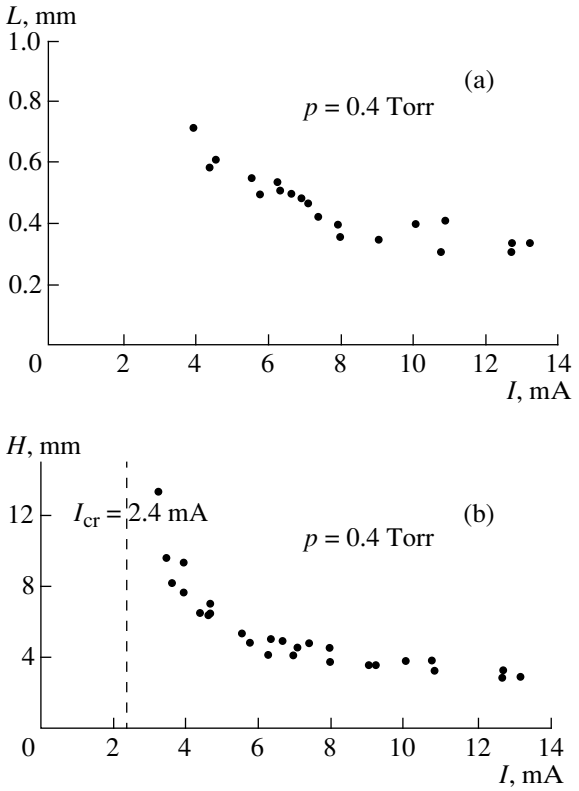


Fig. 4. Dependences of the cloud dimensions $L = \sqrt{h^2 + l^2}$ (h is the vertical dimension of the cloud and l is the horizontal dimension) (a) and the particle height above the anode H (b) on the discharge current.

cient to excite air molecules. In addition, existing experimental observations and numerical estimates indicate that it is almost impossible for a negative anode drop to occur in electronegative gases where the discharge is controlled by attachment [9]. The accuracy of the conclusions that a positive anode potential drop occurs in our case was further confirmed by measurements of the electric field using a double probe. The measured field strength E varied between 35 V/cm near the electrode and 3 V/cm in the plasma bulk.

Thus, the assumption that the particles are negatively charged necessitates the introduction of a force to compensate for the total action of the electrostatic and gravitational fields in the anode layer. The magnitude of this force F should satisfy the condition

$$F > F_g = m_d g \approx 5 \times 10^{-12}\ \text{N}$$

(m_d is the particle mass) for levitation of the smallest particles used experimentally having the radius r_p = 5 μm. As this compensating force we consider the thermophoretic force and the ion drag force [6]. An estimate of the thermophoretic force

$$F_{\text{th}} = \frac{32}{15}\sqrt{\frac{\pi m}{8T}} r_p^2 \kappa \frac{\partial T}{\partial r},$$

where T and m are the temperature and mass of a gas molecule, κ is the thermal conductivity, gives $F_{\text{th}} < 10^{-14}$ N since the temperature gradient $\partial T/\partial r$ in the layer does not exceed 5 K/cm. The upper limit of the ion drag force can be estimated using typical conditions for the electrode layer of a glow discharge, from

$$F_i \le \pi r_p^2 n_i m_i v_i^2 \left(1 - \frac{e\varphi_s}{m_i v_i^2}\right)^2,$$

where m_i is the ion mass, $n_i \approx 10^8$ cm^{-3} is the concentration of positive ions, v_i is the velocity of ions having the energy ≈3 eV, and $\varphi_s \approx 2$ eV is the surface potential of the particles. This estimate gives $F_i \sim 10^{-18}$ N for the ion drag force and this neglects the fact that in an attachment-controlled discharge, a counterpropagating flux of negative ions having a concentration close to n_i will act on the dust particle in the opposite direction [10].

On the basis of this reasoning we can conclude that under these conditions gravity will only be compensated for a positively charged particle by the force of the electric field $F \simeq ZeE$ in the region of positive anode potential drop. Taking into account the measurements of the electric field E = 3–35 V/cm and the gravitational force $F_g \approx 5 \times 10^{-12}$ N, a positive charge Z_d on a levitating particle of radius r_p = 5 μm should satisfy the condition $Z_d > 3 \times 10^3$ elementary charges.

4. ESTIMATE OF MACROPARTICLE CHARGE

In a gas-discharge plasma the particles can become charged as a result of the absorption of plasma elec-

trons and ions, and also by photoelectron and secondary electron emission processes. In a normal glow discharge secondary emission and photoemission processes are insignificant and the main charging mechanism is absorption of plasma electrons and ions by the particle, as a result of which the dust particles become negatively charged [11]. The discharge under study has parameters similar to those of an anomalous hollow-cathode glow discharge whose main characteristics are a strong electric field in the cathode space and cumulation of electrons at the discharge axis [12]. This leads to an increase in the radiation intensity in the ultraviolet part of the cathode dark space and a region of negative glow mainly attributable to nitrogen resonance radiation at 149.4 nm. Another characteristic of this type of discharge is the essentially non-Maxwellian electron energy distribution function and the presence of a fairly high number of high-energy ($\varepsilon > 50$ eV) electrons. These characteristics suggest the existence of an effective mechanism of particle charging by photoemission and secondary electron emission.

In the approximation of bounded orbital motion, which is valid for a rarefied plasma where the electron mean free path l_e before collisions with neutrals is much greater than the particle radius ($l_e \gg r_p$), the dust particle charging process is described by:

$$\frac{dZ_d}{dt} = I_e^+ + I_i^+ - I_e^- - I_i^-, \quad (1)$$

where I_e^- is the electron flux to the particle surface, I_e^+ is the flux of emitted electrons, and I_i^+ and I_i^- are the fluxes of positive and negative ions. The equilibrium potential φ_s of the macroparticle is established as a result of a balance between the electron flux to the particle surface, the flux of emitted electrons, and also the ion fluxes $I_i^\pm$ under the condition

$$\frac{dZ_d}{dt} = 0.$$

The relationship between the charge and the potential is given by

$$Z_d = r_p\varphi_s/e$$

for particles of radius $r_p \ll \lambda_D$, where λ_D is the screening radius. The sign of the charge Z_d corresponds to the sign of the derivative dZ_d/dt (1) at zero time $t = 0$.

In electronegative gases in an attachment-controlled discharge the concentrations of the positive and negative ions have similar values but the electron concentration may be one or two orders of magnitude lower than the ion concentration [10]. Despite this factor, a large difference between the ion and electron mass taking into account the higher temperature of the latter cannot lead to positive charging of the particle without taking into account processes of electron emission from their surface. We shall analyze the conditions for positive charging of macroparticles by secondary electron emission. The flux of secondary (emitted) electrons I_{se} is related to the flux of primary electrons I_e^h by the coefficient of secondary electron emission δ:

$$I_{se} = \delta I_e^h.$$

In general, the coefficient δ depends on the primary electron energy and on the dust particle material. For a monoenergetic electron beam the coefficient δ to a first approximation, following [13], may be expressed in the form

$$\delta(\varepsilon) \approx (\varepsilon_0 + e\varphi_s)/\varepsilon_m. \quad (2)$$

It is then assumed that the collision energy ε is related to the electron beam energyε_0 by

$$\varepsilon = \varepsilon_0 + e\varphi_s,$$

and ε_m is the energy at which $\delta(\varepsilon) = 1$. For glasses we have $\varepsilon_m \sim 40$ V [14].

Subject to the condition $\delta < 1$ the particle charge will be negative. In the presence of a flux I_e of low-energy electrons with $\varepsilon < 40$ eV the condition for positive particle charging will be more stringent. Neglecting the ion fluxes this condition may be written in the form

$$I_e/I_e^h < \delta - 1. \quad (3)$$

In this case, in order to ensure that the flux of emitted electrons completely compensates for the electrons absorbed by the particle ($\varphi_s = 0$, $Z_d = 0$) the following relationship must be satisfied

$$n_eT_e/n_e^h\varepsilon = \delta - 1, \quad (4)$$

where n_e is the concentration of low-energy electrons having the kinetic temperature T_e and n_e^h is the concentration of electrons having energies $\varepsilon > 40$ eV. Assuming that $T_e = 2$ eV, $\varepsilon = 50$ eV, and $\delta - 1 = 0.25$, we find that in order to completely compensate for the negative charge at the particle surface it is sufficient for the electron flux to contain high-energy electrons having a concentration n_e^h an order of magnitude lower than the concentration n_e of thermal electrons. This assumption is quite feasible under conditions of the anode layer of a glow discharge in an electronegative gas.

We shall estimate the photoemission charging of particles by a monochromatic flux at $\lambda_0 \simeq 149.4$ nm, which corresponds to the resonance radiation of nitrogen. Assuming compensated electron fluxes ($I_e/I_e^h = \delta - 1$), the value Z_p of the particle photoemission charge can be estimated from the condition that the surface potential φ_s is equal to $h\nu - W$:

$$Z_p = (h\nu - W)r_p/e, \quad (5)$$

where $h\nu = 7.9$ eV is the quantum energy which in our case corresponds to the wavelength $\lambda_0 \simeq 149.4$ nm and $W = 5$ eV is the photoemission work function of the electrons (for glass of various types $W = 4.5$–6 eV). The maximum charge of particles with $r_p = 5$ μm is equal to $Z_p \approx 10^4$ elementary charges. In order to compensate for gravity, the electric field for a particle with this charge should be ~10 V/cm, which is consistent with the values of the field in the anode region obtained from probe measurements. A more accurate estimate of the equilibrium charge of a dust particle can be obtained from formula (1), allowing for all the characteristics of the gas discharge. However, with this approach the experimental error in the measurements of the concentrations and velocities of the plasma electrons and ions, and also the intensity of the negative luminescence may not lead to the desired result and cannot give a more accurate estimate of the macroparticle charge.

5. CONCLUSIONS

In the present study, levitation of charged particles in the anode region of a dc glow discharge has been observed experimentally for the first time. A dust cloud consisting of several tens of particles formed above the central part of the anode. By varying the discharge parameters it was possible to change the shape of the cloud and control its position above the anode.

An analysis of the experimental conditions shows that the charge of the levitating particles is positive, which significantly distinguishes this experiment from normal glow discharge experiments where the dust particles are negatively charged. These estimates suggest that under the conditions of a steady-state discharge processes of electron emission from the particle surface play a significant role, resulting in the dust particles acquiring positive charge. An estimate of the charge agrees with the results of measurements of the electric field and ensures macroparticle levitation in the Earth's gravitational field.

ACKNOWLEDGMENTS

In conclusion, the authors thank A.V. Chernyshev and A.M. Lipaev for assistance with the experiments. The work was supported in part by the Russian Foundation for Basic Research (project no. 98-02-16828).

REFERENCES

1. J. H. Chu and Lin I, Phys. Rev. Lett. **72**, 4009 (1994).
2. H. Thomas, G. E. Morfill, V. Demmel, *et al.*, Phys. Rev. Lett. **73**, 652 (1994).
3. Yu. V. Gerasimov, A. P. Nefedov, V. A. Sinel'shchikov, and V. E. Fortov, Pis'ma Zh. Tekh. Fiz. **24** (19), 62 (1998) [Tech. Phys. Lett. **24**, 774 (1998)].
4. V. E. Fortov, A. P. Nefedov, V. M. Torchinsky, *et al.*, Phys. Lett. A **229** (1997).
5. A. M. Lipaev, V. I. Molotkov, A. P. Nefedov, *et al.*, Zh. Éksp. Teor. Fiz. **112**, 2030 (1997) [JETP **85**, 1110 (1997)].
6. T. Nitter, Plasma Sources Sci. Technol. **5**, 93 (1996).
7. V. E. Fortov, A. P. Nefedov, O. F. Petrov, *et al.*, Phys. Lett. A **219**, 89 (1996).
8. V. E. Fortov, A. P. Nefedov, and O. S. Vaulina, Zh. Éksp. Teor. Fiz. **114**, 2004 (1998) [JETP **87**, 1087 (1998)].
9. V. L. Granovskiĭ, *Electric Current in a Gas* (Nauka, Moscow, 1971).
10. Yu. P. Raĭzer, M. N. Shneĭder, and N. A. Yatsenko, *High-Frequency Capacitive Discharge* (Nauka, Moscow, 1995).
11. V. N. Tsytovich, Usp. Fiz. Nauk **167**, 57 (1997) [Phys. Usp. **40**, 53 (1997)].
12. B. I. Moskalev, *Discharge with Full Cathode* (Énergiya, Moscow, 1969).
13. I. M. Bronshteĭn and B. S. Fraĭman, *Secondary Electron Emission* (Nauka, Moscow, 1969).
14. B. Walch, M. Horanyi, and S. Robertson, Phys. Rev. Lett. **75**, 838 (1995).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 1, 2000, pp. 111–125.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 1, 2000, pp. 125–141.

PLASMA, GASES

Decay of Solitons in an Isotropic Collisionless Quasineutral Plasma with Isothermal Pressure

I. B. Bakholdin*, A. A. Zharkov, and A. T. Il'ichev*****

**Keldysh Institute of Applied Mathematics, Russian Academy of Sciences, Moscow, 125047 Russia*
***Moscow State Food Production University, Moscow, 125080 Russia*
****Steklov Mathematics Institute, Russian Academy of Sciences, Moscow, 117966 Russia*
e-mail: ilichev@mi.ras.ru

Received January 20, 2000

Abstract—Soliton-type solutions of the complete unreduced system of transport equations describing the plane-parallel motions of an isotropic collisionless quasineutral plasma in a magnetic field with constant ion and electron temperatures are studied. The regions of the physical parameters for fast and slow magnetosonic branches, where solitons and generalized solitary waves—nonlocal soliton structures in the form of a soliton "core" with asymptotic behavior at infinity in the form of a periodic low-amplitude wave—exist, are determined. In the range of parameters where solitons are replaced by generalized solitary waves, soliton-like disturbances are subjected to decay whose mechanisms are qualitatively different for slow and fast magnetosonic waves. A specific feature of the decay of such disturbances for fast magnetosonic waves is that the energy of the disturbance decreases primarily as a result of the quasistationary emission of a resonant periodic wave of the same nature. Similar disturbances in the form of a soliton core of a slow magnetosonic generalized solitary wave essentially do not emit resonant modes on the Alfvén branch but they lose energy quite rapidly because of continuous emission of a slow magnetosonic wave. Possible types of shocks which are formed by two types of existing soliton solutions (solitons and generalized solitary waves) are examined in the context of such solutions. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

A small number of works where the nonlinear plane-parallel wave motions of an isotropic collisionless quasineutral magnetized plasma are studied on the basis of a hydrodynamic description using transport equations has now been published. The first works in this field are probably [1–6], where particular solutions of the transport equations in the cold-plasma limit are studied. Attempts to describe not individual solutions but rather entire classes of solutions of the equations for a cold plasma have been made along the conventional path of further simplifying these equations (which are still too complicated for general investigation) by the many-scale method. Using a unified form of this method [7], the authors of [8] derive the Korteweig–de Vries (KdV) equation for long wavelength magnetosonic waves branching from the state of rest. As a result, it was concluded that solitary waves—solitons—in a cold plasma exist for two values of the angle $0 < \theta \le \pi/2$ of inclination of the unperturbed magnetic field with respect to the direction of propagation of the wave (x-axis). For $\theta < \theta_c$, where θ_c is a critical value of the angle θ, solitons correspond to a rarefaction wave, while for $\theta_c < \theta \le \pi/2$ solitons correspond to a compression wave. However, investigations of the complete system of equations [9, 10] have revealed that solitons do not exist for $\theta < \theta_c$, but in this range of angles of inclination they are replaced by generalized solitary waves—unlocalized solutions, which are products of a nonlinear resonance of a solitary wave and a periodic wave. In [11], where the complete system of equations for a cold plasma is likewise analyzed, families of solitons of envelopes, which are a bifurcation from a wave number for which the phase velocity is equal to the group velocity, are found.

Of the few works concerned with the transport equations for an isotropic collisionless quasineutral plasma with isothermal pressure, i.e., with constant ion and electron temperatures, we call attention first to [12, 13], where particular solutions are sought for the system of transport equations with hot electrons and cold ions. In [14], a nonlinear Schrödinger-type equation was obtained, in an investigation of a two-temperature isotropic collisionless quasineutral plasma, for the slowly varying amplitudes of Alfvén waves; this equation supports the modulation instability, i.e., it possesses soliton solutions. Questions concerning the correspondence between such a theoretical description and observations of Alfvén soliton structures in the plasma of the Earth's magnetosphere are discussed.

The present paper is organized as follows. The transport equations for plane-parallel motions of an isotropic collisionless quasineutral plasma with constant ion and electron temperatures are presented in Section 2. Linear resonances of long- and short-wavelength waves are studied in Section 3, and in Section 4 the dynamical

system for traveling waves is obtained and the types of bifurcations from the state of rest are described. Solutions of the solitary wave type—soliton solutions, which exist in an infinite region in the plane of physical parameters—are studied in Section 5. In this region the evolution of localized compressive disturbances results in the formation of solitons, specifically, a disturbance in the form of a soliton does not undergo distortions in the course of evolution. Solitons occur only for the fast magnetosonic branch of the dispersion relation. The existence of generalized solitary waves, which replace solitons in a bounded region of the physical parameters for fast magnetosonic waves and occur for all parameters for slow magnetosonic waves, is established Section 6.

For a more complete analysis of all possible stationary solutions, in Section 7 shocks are also examined. Solutions in which a section of oscillatory type—a wave zone—separates two uniform states are encountered in nondissipative systems. In time the extent of the wave zone increases and its envelope becomes self-similar. For a plasma such a wave zone is called a nonstationary structure of a collisionless shock wave [15]. For the simplest dispersion model—the KdV equation—a transition which does not expand with time (a local transition) occurs at the boundary of the wave zone between the uniform and periodic states (a sequence of solitary waves) in the limit $t \longrightarrow \infty$. In the description of the phenomenon using the averaged equations for the envelope, this transition can be treated as a shock. We shall treat shocks in nondissipative models as any local transitions between uniform, periodic, and quasiperiodic states. A general approach, making it possible to predict on the basis of the form of the dispersion curve the possible type of jump with a stationary structure, is developed in [16, 17]. This approach to classification of possible shocks is also used in the present paper.

A brief analysis and a discussion of the results are presented in Section 8.

2. FORMULATION OF THE PROBLEM

An isotropic collisionless plasma with constant ion and electron temperatures can be described on the basis of a hydrodynamic model of two fluids—electronic and ionic. In this model the basic equations have the form

$$\begin{gathered}
\mathrm{rot}\mathbf{B} - \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t^*} = \frac{4\pi e}{c}(n_i\mathbf{v}_i - n_e\mathbf{v}_e),\\
\mathrm{rot}\mathbf{E} + \frac{1}{c}\frac{\partial \mathbf{B}}{\partial t^*} = 0,\\
\mathrm{div}\mathbf{B} = 0,\\
\mathrm{div}\mathbf{E} = 4\pi e(n_i - n_e),\\
\frac{\partial n_i}{\partial t^*} + \mathrm{div}(n_i\mathbf{v}_i) = 0,\\
m_i\frac{d^{(i)}\mathbf{v}_i}{dt^*} = e\left\{\mathbf{E} + \frac{1}{c}(\mathbf{v}_i\times\mathbf{B})\right\} - \nabla p_i,\\
\frac{\partial n_e}{\partial t^*} + \mathrm{div}(n_i\mathbf{v}_e) = 0,\\
m_e\frac{d^{(e)}\mathbf{v}_e}{dt^*} = -e\left\{\mathbf{E} + \frac{1}{c}(\mathbf{v}_e\times\mathbf{B})\right\} - \nabla p_e,
\end{gathered} \tag{2.1}$$

where

$$\frac{d^{(i)}}{dt^*} = \frac{\partial}{\partial t^*} + \mathbf{v}_i\cdot\nabla, \quad \frac{d^{(e)}}{dt^*} = \frac{\partial}{\partial t^*} + \mathbf{v}_e\cdot\nabla,$$

E and **B** are the electric and magnetic field intensities, t^* is the physical time, $\mathbf{v}_{e,i}$, $n_{e,i}$, and $m_{e,i}$ are the velocity, number density of the particles and the mass of the electrons (e) and ions (i), $p_{e,i}$ is the pressure in the electronic and ionic liquids, c is the speed of light, and $-e$ is the electron charge.

Eliminating the electron velocity $\mathbf{v}_e$ and the electric field intensity **E** from Eqs. (2.1), using the quasineutrality condition $n_e \approx n_i$ and neglecting the displacement current,[1] for plane waves propagating along the x-axis [12, 14], we obtain

$$\begin{gathered}
\frac{dn}{dt} = -n\frac{\partial u}{\partial x},\\
\frac{du}{dt} = -\frac{n^{-1}}{2}\frac{\partial(B_y^2 + B_z^2)}{\partial x} - b^2 n^{-1}\frac{\partial n}{\partial x},\\
\frac{dv}{dt} = n^{-1}B_x\frac{\partial B_y}{\partial x} - R_e^{-1}\frac{d}{dt}\left(n^{-1}\frac{\partial B_z}{\partial x}\right),\\
\frac{dw}{dt} = n^{-1}B_x\frac{\partial B_z}{\partial x} + R_e^{-1}\frac{d}{dt}\left(n^{-1}\frac{\partial B_y}{\partial x}\right),\\
\frac{dB_y}{dt} = B_x\frac{\partial v}{\partial x} - B_y\frac{\partial u}{\partial x} + R_i^{-1}\frac{\partial}{\partial x}\frac{\partial w}{\partial t},\\
\frac{dB_z}{dt} = B_x\frac{\partial w}{\partial x} - B_z\frac{\partial u}{\partial x} - R_i^{-1}\frac{\partial}{\partial x}\frac{\partial v}{\partial t},
\end{gathered} \tag{2.2}$$

where

$$\frac{d}{dt} = \frac{\partial}{\partial t} + u\frac{\partial}{\partial x},$$

$$b = \sqrt{\frac{m_i V_i^2/V_A^2 + m_e V_e^2/V_A^2}{m_i + m_e}}$$

is the dimensionless effective speed of sound, $V_{e,i}$ are the thermal velocities of the electrons and ions,

[1] The displacement current can be neglected if $R_i^{-1}V_A^2/c^2 \ll 1$ [12].

$V_A = |\mathbf{B}_0|/\sqrt{4\pi n_0(m_e + m_i)}$ is the Alfén velocity, $R_{e,i} = \omega_{e,i}/\omega_0$ are the dimensionless dispersion parameters, $\omega_{e,i} = e|\mathbf{B}_0|/m_{e,i}c$ are the electron and ion cyclotron (Larmor) frequencies. The spatial and temporal variables (x, t), the ion or electron density n, and the components of the magnetic field intensity (B_x, B_y, B_z) and the ion velocity (u, v, w) are scaled to, respectively, the characteristic length L, the frequency ω_0, the density n_0 of the unperturbed plasma, the modulus of the unperturbed magnetic field vector $\mathbf{B}_0$, and the Alfén velocity V_A. For one-dimensional motions the component B_x of the magnetic field always remains a constant of the motion. The variables n, u, v, w, B_x, B_y, and B_z in a state of rest are, respectively, 1, 0, 0, 0, $\cos\theta$, $\sin\theta$, and 0.

3. LINEAR WAVE RESONANCES

For the analysis below we shall require certain properties of the dispersion relation, which will be enumerated in the present section, for linear waves. A detailed analysis of the propagation of linear waves in a magnetized plasma in the hydrodynamic approximation can be found, for example, in [18].

The dispersion relation for the system of equations (2.2) has the form

$$\begin{aligned}(R_eR_i + k^2)^2V^6 - \{R_e^2R_i^2 + R_e^2R_i^2\cos^2\theta \\ + R_eR_i[1 + \cos^2\theta(\rho + 1)]k^2 + (R_eR_i + k^2)^2b^2\}V^4 \\ + \cos^2\theta[R_e^2R_i^2 + 2R_e^2R_i^2b^2 + R_eR_i(\rho + 2)b^2k^2]V^2 \\ - R_e^2R_i^2\cos^4\theta b^2 = 0,\end{aligned} \tag{3.1}$$

where

$$V = \frac{\omega}{k}, \quad \rho = \left(\sqrt{\frac{R_e}{R_i}} - \sqrt{\frac{R_i}{R_e}}\right)^2.$$

The system (2.2) is invariant under the transformation $x \longrightarrow -x$, $t \longrightarrow -t$. Consequently, the branches of the dispersion relation $\omega = \omega(k)$ are odd functions of the wave number and symmetric with respect to the origin of coordinates in the $k\omega$ plane, so that it is sufficient to study the case $\omega \geq 0$, $k \geq 0$. For $b \neq 0$ the dispersion relation possesses three branches: slow magnetosonic, $\omega_s = kV_s(k)$; fast magnetosonic, $\omega_f = kV_f(k)$; and, Alfén $\omega_A = kV_A(k)$. The relative arrangement of these branches is shown in Fig. 1. For $b = 0$ we have a cold plasma and the dispersion curve possesses two branches—Alfén and magnetosonic. The determination of the explicit functions $\omega = \omega(k)$ from Eq. (3.1) involves complicated calculations and is immaterial for the exposition below. Only the following should be noted.

(1) It is evident from Eqs. (3.1) that for a fixed value of the phase velocity V the straight line $\omega = kV$ on the positive k semiaxis possesses no more than two points in common with all branches of (3.1), since for fixed V the relation (3.1) is a biquadratic equation for k.

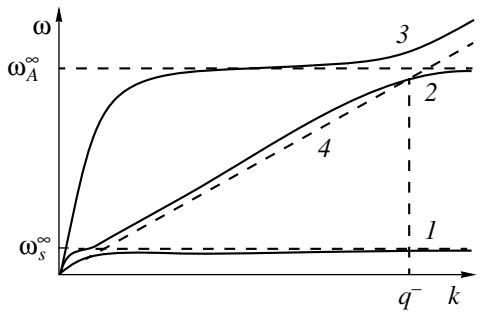


Fig. 1. Relative arrangement of the curves of the branches of the dispersion relation $\omega = \omega(k)$ of Eqs. (2.1) (the case $b < b_c$). Curve *1* corresponds to a slow magnetosonic branch, curve *2* corresponds to the Alfvén branch, and curve *3* corresponds to a fast magnetosonic branch. A resonance between the long-wavelength slow magnetosonic wave and an Alfvén wave occurs: the straight line *4*, tangent to *1*, always intersects *2* at $k = q^-$.

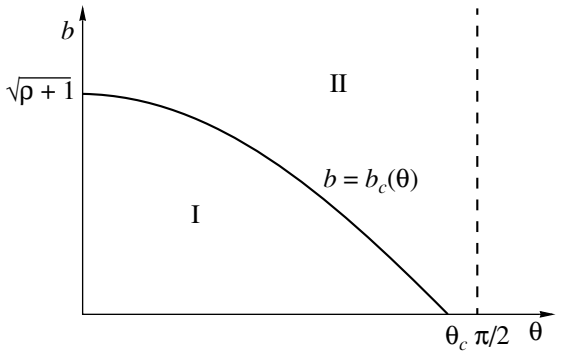


Fig. 2. Position of the curve $b = b_c(\theta)$ separating the region of existence of generalized solitary waves (I) and solitons (II).

(2) The values of ω on the Alfén and slow branches in the short-wavelength limit ($k \longrightarrow \infty$) are bounded in modulus by the constants $\omega_A^\infty = R_e\cos\theta$ and $\omega_s^\infty = R_i\cos\theta$, which do not depend on the parameter b, and the straight lines $\omega = \omega_{A,s}^\infty$ are horizontal asymptotes of the plots of these branches (Fig. 1).

(3) For $b \neq 0$ the plot of the fast magnetosonic branch lies everywhere above the inclined asymptote $\omega = bk$.

(4) The phase velocities of infinitely long-wavelength waves for Alfén, fast and slow branches are given, respectively, by the expressions

$$V_A(0) = \cos\theta, \quad V_f(0) = V^+, \quad V_s(0) = V^-,$$

$$V^\pm = \sqrt{[1 + b^2 \pm \sqrt{(1 + b^2)^2 - 4\cos^2\theta b^2}]/2}, \tag{3.2}$$

$$0 < V^- < V_A < V^+.$$

(5) The curve in the parameter space $b = b_c(\theta)$ (Fig. 2),

$$b_c(\theta) = \sqrt{(\rho^{-1} + 1)[(\rho + 1)\cos^2\theta - 1]},$$

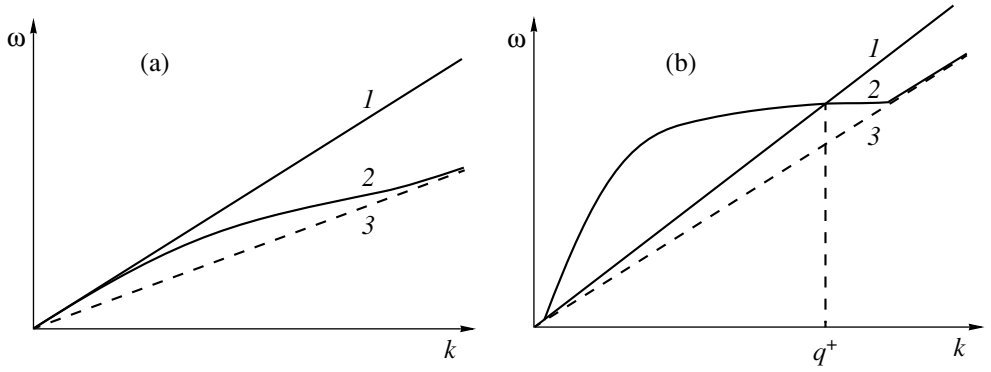


Fig. 3. Curve *2* is fast magnetosonic branch for (a) $b < b_c$ and (b) $b > b_c$, approaching the asymptote (dashed line *3*) as $k \longrightarrow \infty$. In the case (b) we have a resonance of long- and short-wavelength magnetosonic waves: the straight line *1*, tangent to *2* at 0, intersects *2* at $k = q^+$. There is no resonance in the case (a).

is critical in the following sense. For $b > b_c(\theta)$ (region II in Fig. 2) the straight line $\omega = V^+k$ tangent to the fast magnetosonic branch 0 no longer intersects the dispersion curve (simple resonance, Fig. 3a). For $b < b_c$ (region I in Fig. 2) there exist additional points of intersection $k = \pm q^+$, $q^+ > 0$, of this straight line with the magnetosonic branch of the dispersion curve (Fig. 3b); this corresponds to a linear resonance of the long- and short-wavelength waves. We note that for $b = b_c$ the coefficient of the cubed wave number in the expansion of ω for small k vanishes. This makes it necessary, specifically, to take into account the next terms in the long-wavelength expansion of the frequency.

(6) The straight line $\omega = V^-k$ tangent to the slow branch always intersects the Alfén branch at a single point (for $k > 0$). Thus there exists a linear resonance of the long-wavelength slow magnetosonic wave and the short-wavelength mode of the Alfén branch.

(7) For $b < b_c$ the fast magnetosonic branch possesses a point (k_r, V_r) where the phase velocity is equal to the group velocity. Geometrically, this means that the straight line $\omega = V_rk$ and the curve of the fast magnetosonic branch are tangent at the point (k_r, V_r) (so-called 1 : 1 resonance).

Thus, for the fast and slow magnetosonic branches we have four types of resonances: a simple resonance, a resonance of long- and short-wavelength fast magnetosonic waves, a resonance of a long-wavelength slow magnetosonic and short-wavelength mode of the Alfén branch, and a 1 : 1 resonance.

4. DYNAMICAL SYSTEM

The traveling wave solutions of the system of equations (2.2) are functions of $\xi = x - Vt$. In the present paper we shall consider waves moving with velocity close to one of the magnetosonic velocities. We shall write $V = V^\pm + \mu$, where μ is a small parameter and $V^\pm$ is one of the velocities (3.2). Traveling waves are described by a system of equations obtained by a single integration from (2.2). The density n and x component of the velocity can be expressed in terms of other unknown functions according to the formulas

$$n^{-1} = \frac{1}{2}\left[1 + \frac{b^2}{V^2} - \frac{B}{V^2} - \operatorname{sgn}\left(1 - \frac{b^2}{V^2}\right) \times \sqrt{\left(1 - \frac{B}{V^2} + \frac{b^2}{V^2}\right)^2 - 4\frac{b^2}{V^2}}\right],$$

$$u = V(1 - n^{-1}),$$

where

$$B = \frac{1}{2}(b_y^2 + 2b_y\sin\theta + B_z^2), \quad b_y = B_y - \sin\theta.$$

The constants of integration are chosen so that the state of rest satisfies equations for the traveling waves, and the equations themselves have the form

$$\begin{aligned} \frac{dv}{d\xi} &= -R_iB_z - \frac{R_i\cos\theta}{V}nw, \\ \frac{dw}{d\xi} &= \frac{R_i\cos\theta}{V}nv + R_ib_y + R_in(n^{-1} - 1)\sin\theta, \\ \frac{db_y}{d\xi} &= R_enw + nB_z\frac{R_e\cos\theta}{V}, \\ \frac{dB_z}{d\xi} &= -R_env - \frac{R_e\cos\theta}{V}nb_y. \end{aligned} \tag{4.1}$$

Near $V = V^\pm$ the system (4.1) can be written in the form

$$\frac{d\mathbf{w}}{d\xi} = A\mathbf{w} + \mathbf{F}(\mu, \mathbf{w}), \quad \mathbf{w} = (v, w, b_y, B_z)^{\mathrm{T}}, \tag{4.2}$$

where

$$A = \begin{pmatrix} 0 & -\frac{R_i\cos\theta}{V^\pm} & 0 & -R_i \\ \frac{R_i\cos\theta}{V^\pm} & 0 & R_i\left(1-\frac{\sin^2\theta}{V^{\pm 2}-b^2}\right) & 0 \\ 0 & R_e & 0 & \frac{R_e\cos\theta}{V^\pm} \\ -R_e & 0 & -\frac{R_e\cos\theta}{V^\pm} & 0 \end{pmatrix},$$

and the nonlinear vector function $\mathbf{F} = (F_1, F_2, F_3, F_4)^{\mathrm{T}}$, $F_i = O(\mu\mathbf{w}, |\mathbf{w}|^2)$, $i = 1, 4$, has the form

$$F_1 = -\frac{R_i\cos\theta}{V}nw + \frac{R_i\cos\theta}{V^\pm}w,$$

$$F_2 = \frac{R_i\cos\theta}{V}nv - \frac{R_i\cos\theta}{V^\pm}v + R_i(1-n)\sin\theta,$$

$$F_3 = R_enw - R_ew + \frac{R_e\cos\theta}{V}nB_z - \frac{R_e\cos\theta}{V^\pm}B_z,$$

$$F_4 = -R_env + R_ev - \frac{R_e\cos\theta}{V}nb_y + \frac{R_e\cos\theta}{V^\pm}b_y.$$

Invertibility of the system of equations (4.2) implies the following equalities:

$$AR = -RA, \quad \mathbf{F}(\mu, R\mathbf{w}) = -R\mathbf{F}(\mu, \mathbf{w}),$$

where R is a diagonal matrix and $R = \mathrm{diag}(1, -1, 1, -1)$. These equalities mean that among the solutions of the system (4.2) there are solutions with even v and b_y and odd w and B_z.

The characteristic equation

$$|A_1(V) - \lambda I| = 0,$$

where I is the unit matrix and $A_1(V)$ is obtained from A by substituting $V^\pm$ for V, can be written in the form

$$\lambda^4 - R_iR_e\left[2 - \frac{\sin^2\theta}{V^2-b^2} - (\rho+2)\left(\frac{\cos\theta}{V}\right)^2\right]\lambda^2 + (R_iR_e)^2\left[1-\left(\frac{\cos\theta}{V}\right)^2\right] \times \left[1-\left(\frac{\cos\theta}{V}\right)^2 - \frac{\sin^2\theta}{V^2-b^2}\right] = 0. \quad (4.3)$$

The resonances at $V = V^\pm$ and $V = V_r$ described in the preceding section give rise to bifurcations from the zero solution, which in the first case give solitary and generalized solitary waves and in the second case magnetosonic envelope solitons, for which the velocities of the envelope and the high-frequency fill are the same [11, 19].

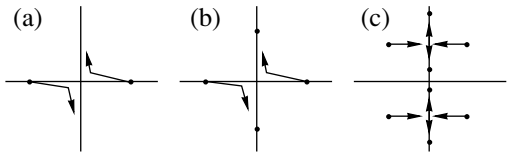


Fig. 4. Dynamics of resonance eigenvalues for a simple resonance (a), resonance of long- and short-wavelength waves (b), and 1 : 1 resonance (c).

The dynamics of the characteristic values in the complex plane of the spectral parameter λ with μ passing through zero is depicted in Fig. 4: Fig. 4a illustrates the situation $V = V^+ + \mu$ and $b > b_c$; Fig. 4b illustrates $V = V^+ + \mu$, $b < b_c$ or $V = V^- + \mu$; and, Fig. 4c illustrates $V = V_r + \mu$. Only the first two cases will be considered in the present paper.

For $V = V^\pm$ the characteristic equation (4.3) can be represented in the form

$$\lambda^2\left\{\lambda^2 - \frac{R_eR_i}{V^{\pm 2}}[(V^{\pm 2} - \cos^2\theta) - \rho\cos^2\theta]\right\} = 0. \quad (4.4)$$

The following roots of Eq. (4.4) lie on the imaginary axis:

a) for $V = V^+$ and $b > b_c$, a second-order zero;

b) for $V = V^+$ and $b < b_c$ or for $V = V^-$, a second-order zero and two nonzero values $\lambda = \pm iq^\pm$, $q^\pm > 0$, where

$$q^{\pm 2} = -\frac{R_eR_i}{V^{\pm 2}}(V^{\pm 2} - \cos^2\theta - \rho\cos^2\theta). \quad (4.5)$$

5. SOLITARY WAVES

For $V = V^+$ and $b > b_c$ we have a simple bifurcation, for which the motion of the eigenvalues of the matrix A_1 (roots of Eq. (4.3)), which move from the real to the imaginary axis as μ passes through zero, is depicted in Fig. 4a. Let $\mathbf{w} = \mathbf{w}_0 + \mathbf{w}_1$, where $\mathbf{w}_0 \in E_0$, $\mathbf{w}_1 \in E_1$, and E_0 and E_1 are, respectively, the central and hyperbolic spaces of the zeroth eigenvalue. According to the central manifold theorem [20], for sufficiently small μ and $\mathbf{w}_0$ we have $\mathbf{w}_1 = \Phi(\mu, \mathbf{w}_0)$, where the function $\Phi(\mu, \mathbf{w}_0) = O(\mu|\mathbf{w}_0|, |\mathbf{w}_0|^2)$. In what follows, we shall represent $\mathbf{w}_0$ in the form

$$\mathbf{w}_0 = a_0\boldsymbol{\phi}_0 + a_1\boldsymbol{\phi}_1, \quad (5.1)$$

where $\boldsymbol{\phi}_0$ and $\boldsymbol{\phi}_1$ are the eigenvector and the adjoint vector of the matrix A ($A\boldsymbol{\phi}_0 = 0$, $A\boldsymbol{\phi}_1 = \boldsymbol{\phi}_0$), and are

given by the expressions

$$\boldsymbol{\phi}_0 = \begin{pmatrix} -\dfrac{\cos\theta}{V^+} \\ 0 \\ 1 \\ 0 \end{pmatrix},$$

$$\boldsymbol{\phi}_1 = \frac{V^{+2}}{V^{+2}-\cos^2\theta}\begin{pmatrix} 0 \\ R_e^{-1} - R_i^{-1}\left(\dfrac{\cos\theta}{V^+}\right)^2 \\ 0 \\ (R_i^{-1} - R_e^{-1})\dfrac{\cos\theta}{V^+} \end{pmatrix}.$$

For sufficiently small μ and $\mathbf{a} = (a_0, a_1)^{\mathrm{T}}$, in accordance with the central manifold theorem the dynamical fourth-order system (4.2) reduces to a second-order system, which has the following form in the variables a_0 and a_1:

$$\begin{aligned} \dot a_0 &= a_1 + O(\mu|\mathbf{a}|, |\mathbf{a}|^2), \\ \dot a_1 &= \mu c_1^+ a_0 - c_2^+ a_0^2 + O(a_1^2) + o(\mu|\mathbf{a}|, |\mathbf{a}|^2), \end{aligned} \tag{5.2}$$

where

$$c_1^+ = 2\frac{R_e R_i}{\Delta^+}\left(\frac{\cos^2\theta}{V^{+3}} + \frac{V^+\sin^2\theta}{(V^{+2}-b^2)^2}\right),$$

$$c_2^+ = \frac{R_e R_i \sin\theta}{\Delta^+(V^{+2}-b^2)}\left(\frac{\cos^2\theta}{V^{+2}} + \frac{1}{2} + \frac{V^{+2}\sin^2\theta}{(V^{+2}-b^2)^2}\right),$$

$$\Delta^+ = \frac{V^{+2} - \cos^2\theta - \rho\cos^2\theta}{V^{+2}-\cos^2\theta} > 0.$$

The system (5.2) was obtained by scalar multiplication of the expression (4.2) by the adjoint vector $\boldsymbol{\psi}_0$ and the eigenvector $\boldsymbol{\psi}_1$ of the adjoint matrix A^{T}: $A^{\mathrm{T}}\boldsymbol{\psi}_1 = 0$, $A^{\mathrm{T}}\boldsymbol{\psi}_0 = \boldsymbol{\psi}_1$. We note that the vectors $\boldsymbol{\psi}_i$, $i = 1, 2$, are normalized so that $\langle\boldsymbol{\phi}_i, \boldsymbol{\psi}_j\rangle = \delta_{ij}$, $i, j = 1, 2$, δ_{ij} is the Kronecker delta function, and $\langle\cdot,\cdot\rangle$ denotes the standard scalar product in $\mathbb{C}^4$. In the leading approximation in μ, for $\mu > 0$ the system of equations (5.2) possesses the solution

$$\begin{aligned} a_0 &= \mu\frac{3}{2}\frac{c_1^+}{c_2^+}\cosh^{-2}\frac{\sqrt{c_1^+\mu}}{2}\xi + O(\mu^2), \\ a_1 &= \dot a_0 + O(\mu^2). \end{aligned} \tag{5.3}$$

Moreover, it turns out that for small μ the solution (5.3) is exact for the complete system (5.2) (and hence for (4.2)), which is even and decreases exponentially at both infinities.

We denote by $\mathbf{a}^* = (a_0^*, a_1^*)^{\mathrm{t}}$ the leading part with respect to μ in (5.3). Then, for $\mu_0 > 0$ and sufficiently small $\mu \in (0, \mu_0]$ there exists a family of soliton solutions $\mathbf{a} = (a_0, a_1)^{\mathrm{T}}$ of the complete system (5.2). Moreover, the following estimate is valid:

$$|\mathbf{a} - \mathbf{a}^*| \le \alpha_0\mu^2\exp(-\sigma_0\sqrt{c_1^+\mu}|\xi|), \tag{5.4}$$

where α_0 depends only on μ_0 and $\sigma_0 < 1$. We shall prove this assertion.

We make in (5.2) the following scale transformation:

$$b_0 = \frac{2}{3}\frac{c_2^+}{c_1^+}\mu^{-1}a_0, \quad b_1 = \frac{2}{3}\frac{c_2^+}{c_1^+\sqrt{c_1^+}}\mu^{-3/2}a_1,$$

$$\zeta = \sqrt{c_1^+\mu}\,\xi.$$

Then Eqs. (5.2) become

$$b_0'' = b_0 - \frac{3}{2}b_0^2 + O(\mu), \tag{5.5}$$

where a prime denotes differentiation with respect to ζ. The equation (5.5) possesses the solution

$$b_0 = \tilde b_0 + O(\mu), \quad \tilde b_0 = \operatorname{sech}^2\zeta. \tag{5.6}$$

Next, let us assume that $b_0 = \tilde b_0 + \hat b_0$, where $\tilde b_0$ is given by Eq. (5.6) and $\hat b_0$ is a small nonlinear perturbation. The equation (5.5) can be written in the form

$$\mathcal{M}\hat b_0 = \mu N(b_0, \mu), \tag{5.7}$$

where

$$\mathcal{M} = \frac{d^2}{d\zeta^2} - 1 + 3\tilde b_0^2,$$

and $N(b_0, \mu)$ is an even function for even b_0 (in accordance with the property that the initial equations are invertible) and $N(b_0, \mu) \le c$, where the constant c does not depend on μ for $\mu \in (0, \mu_0)$.

In accordance with the implicit function theorem, Eq. (5.7) possesses a unique solution with a quite low amplitude, if the operator $\mathcal{M}$ is invertible in suitable functional spaces.

We shall determine the space of exponentially decreasing functions C_ω^j, $j = 0, 1, 2$, and $\sigma < 1$, as follows:

$$C_\sigma^j = \left\{ f \in C^j(\mathbb{R}), \sup_\zeta \exp(\sigma\zeta|)\left|\partial_\zeta^m f\right| < \infty, m \le j \right\}.$$

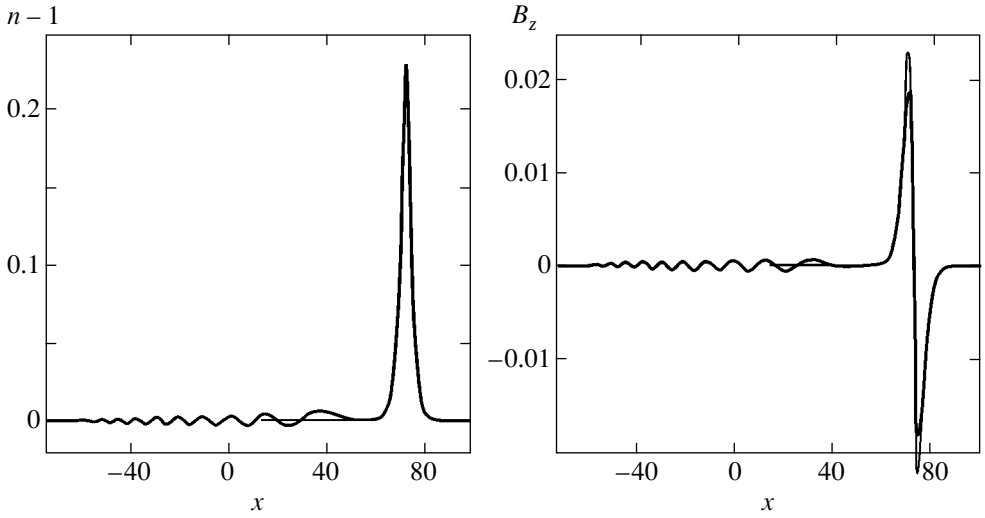


Fig. 5. Evolution of the initial data of the fast magnetosonic solitary wave type for $\theta > \theta_c$, $\mu = 0.05$, $\theta = 1.555$, $b^2 = 0.5$, $t = 0$ (fine line) and $t = 250$ (heavy line). The coordinate system moves with velocity $V^+ + \mu$.

We introduce the notations $X \subset C^2_\sigma$ and $Y \subset C^0_\sigma$ for Banach spaces of even functions with the norms $\|\cdot\|_X$ and $\|\cdot\|_Y$, respectively. We shall prove that the equation

$$\mathcal{M}\hat{b}_0 = f, \tag{5.8}$$

where $\hat{b}_0 \in X$ and $f \in Y$, possesses a unique bounded solution. Equation (5.8) has the form

$$\hat{b}_0'' - \hat{b}_0 + 3\tilde{b}_0^2\hat{b}_0 = f. \tag{5.9}$$

The homogeneous equation (5.9) does not have a fundamental system of solutions in the function space X: $v_1 = c_1\tilde{b}_0'$ is an odd solution (c_1 is determined from the condition $v_1'(0) = 1$), and the linearly independent solution $v_2 = c_2 v_1 + c_3 v_1 \int v_1^{-2}$ (with constants c_2 and c_3, determined from the conditions $v_2(0) = 1$ and $v_2'(0) = 0$) is even but increasing as $\exp\zeta$ at infinity. The absence of a fundamental system of solutions in X signifies that the solution (5.9) is unique if it exists. The existence of a solution of the inhomogeneous equation (5.9) can be verified directly: the solution is given by

$$v = v_2 \int_\zeta^\infty v_1 f\, d\zeta + v_1 \int_0^\zeta v_2 f\, d\zeta. \tag{5.10}$$

Estimates for $\|v\|_X$ in terms of $\|f\|_Y$, from which (5.4) follows directly, can be easily obtained from the solution (5.10).

An expression for the principal part of the soliton for physical variables can be restored using the formula (5.1). Figure 5 shows the evolution of a symmetric (density) and antisymmetric (B_z component of the magnetic field) solitary waves corresponding to (5.3). A solitary wave propagating without a change in shape is formed; its profile is essentially identical to the initial profile. The physical constants in this investigation correspond to a hydrogen plasma, $R_e^{-1} = 0.02341352$ (the characteristic frequency was taken as the geometric-mean of the Larmor frequencies).

6. FAST AND SLOW GENERALIZED SOLITARY WAVES

In the present section we shall examine nonlocal analogs of solitary waves—generalized solitary waves which are formed as a result of bifurcations corresponding to resonances of long- and short-wavelength fast magnetosonic waves as well as a long-wavelength slow magnetosonic and a short-wavelength mode of the Alfén branch. In other words we shall study waves with velocities near V^+ for $b < b_c$ (a case in addition to the one considered in the preceding section) and also near V^-. The diagram presented in Fig. 4b illustrates the bifurcations leading to the appearance of the waves considered here.

As already noted, in the cases considered all roots of Eq. (4.4) are purely imaginary: a second-order zero and $\pm iq^\pm$, $q^\pm > 0$, where $q^\pm$ is given by Eq. (4.5). We now make the following substitution of variables in Eqs. (4.2):

$$\mathbf{w} = a_0(\xi)\boldsymbol{\phi}_0 + a_1(\xi)\boldsymbol{\phi}_1 + a_+(\xi)\boldsymbol{\phi}_+ + a_-(\xi)\boldsymbol{\phi}_-, \tag{6.1}$$

where $a_- = \bar{a}_+$, $A\boldsymbol{\phi}_0 = 0$, $A\boldsymbol{\phi}_1 = \boldsymbol{\phi}_0$, $A\boldsymbol{\phi}_+ = iq\boldsymbol{\phi}_+$, and $A\boldsymbol{\phi}_- = -iq\boldsymbol{\phi}_-$ (bar indicates complex conjugation). The eigen-

vectors and the adjoint vectors of the matrix A are given by the formulas

$$\boldsymbol{\phi}_0 = \begin{pmatrix} -\dfrac{\cos\theta}{V^{\pm}} \\ 0 \\ 1 \\ 0 \end{pmatrix},$$

$$\boldsymbol{\phi}_1 = \frac{V^{\pm 2}}{V^{\pm 2} - \cos^2\theta} \begin{pmatrix} 0 \\ R_e^{-1} - R_i^{-1}\left(\dfrac{\cos\theta}{V^{\pm}}\right)^2 \\ 0 \\ (R_i^{-1} - R_e^{-1})\dfrac{\cos\theta}{V^{\pm}} \end{pmatrix},$$

$$\boldsymbol{\phi}_+ = \begin{pmatrix} R_i q^{\pm -1}\left[1 - R_i R_e^{-1}\left(\dfrac{\cos\theta}{V^{\pm}}\right)^2\right] \\ iR_i R_e^{-1}\dfrac{\cos\theta}{V^{\pm}} \\ q^{\pm -1}(R_i - R_e)\dfrac{\cos\theta}{V^{\pm}} \\ -i \end{pmatrix}, \quad \boldsymbol{\phi}_- = \bar{\boldsymbol{\phi}}_+.$$

The vector function $\mathbf{a_1} = (a_0, a_1, a_+, a_-)^t$ satisfies the equations

$$\begin{aligned} \dot{a}_0 &= a_1 + O(\mu|\mathbf{a_1}|, |\mathbf{a_1}|^2), \\ \dot{a}_1 &= \mu c_1^{\pm} a_0 - c_2^{\pm} a_0^2 + f_1^{\pm}(a_0, a_1, a_+, a_-), \\ \dot{a}_+ &= iq^{\pm} a_+ + i(\mu d_1^{\pm} a_0 + d_2^{\pm} a_0^2) \\ &\quad + g_1^{\pm}(a_0, a_1, a_+, a_-), \\ \dot{a}_- &= -iq^{\pm} a_- + i(\mu d_1^{\pm} a_0 + d_2^{\pm} a_0^2) \\ &\quad + \bar{g}_1^{\pm}(a_0, a_1, a_+, a_-), \end{aligned} \tag{6.2}$$

where

$$c_1^{\pm} = 2\frac{R_e R_i}{\Delta^{\pm}}\left[\frac{\cos^2\theta}{V^{\pm 3}} + \frac{V^{\pm}\sin^2\theta}{(V^{\pm 2} - b^2)^2}\right],$$

$$c_2^{\pm} = \frac{R_e R_i \sin\theta}{\Delta^{\pm}(V^{\pm 2} - b^2)}\left[\frac{\cos^2\theta}{V^{\pm 2}} + \frac{1}{2} + \frac{V^{\pm 2}\sin^2\theta}{(V^{\pm 2} - b^2)^2}\right],$$

$$\Delta^{\pm} = \frac{V^{\pm 2} - \cos^2\theta - \rho\cos^2\theta}{V^{\pm 2} - \cos^2\theta},$$

$$\begin{aligned} d_1^{\pm} &= -\frac{R_e^2 \cos\theta(R_i V^{\pm 2} - R_e\cos^2\theta)}{2q^{\pm 2}V^{\pm 4}} \\ &\quad + \frac{R_e(R_e - R_i)\cos\theta}{2q^{\pm 2}V^{\pm}} \\ &\quad \times\left[\frac{R_i\cos^2\theta}{V^{\pm 3}} + \frac{2R_i V^{\pm}\sin^2\theta}{(V^{\pm 2} - b^2)^2}\right], \\ d_2^{\pm} &= \frac{R_e R_i(R_i - R_e)\cos\theta\sin\theta}{2q^{\pm 2}V^{\pm}(V^{\pm 2} - b^2)} \\ &\quad \times\left[\frac{\cos^2\theta}{V^{\pm 2}} + \frac{1}{2} + \frac{V^{\pm 2}\sin^2\theta}{(V^{\pm 2} - b^2)^2}\right], \end{aligned} \tag{6.3}$$

and $f_1^{\pm}(a_0, a_1, a_+, a_-)$ and $g_1^{\pm}(a_0, a_1, a_+, a_-)$ are complex nonlinear functions of the arguments, where $f_1(a_0, 0, 0, 0) = O(a_0^3)$ and $g_1(a_0, 0, 0, 0) = O(a_0^3)$.

Once again, Eqs. (6.2) were obtained by scalar multiplication of (4.2) by the eigenvectors and the adjoint vector $\boldsymbol{\psi}_0$, $\boldsymbol{\psi}_1$, $\boldsymbol{\psi}_+$, and $\boldsymbol{\psi}_-$ of the transposed matrix A^{T}: $A^{\mathrm{T}}\boldsymbol{\psi}_1 = 0$, $A^{\mathrm{T}}\boldsymbol{\psi}_0 = \boldsymbol{\psi}_1$, $A^{\mathrm{T}}\boldsymbol{\psi}_+ = -iq^{\pm}\boldsymbol{\psi}_+$, and $A^{\mathrm{T}}\boldsymbol{\psi}_- = iq^{\pm}\boldsymbol{\psi}_-$. The normalization of these vectors is chosen from the conditions

$$\langle\boldsymbol{\phi}_i, \boldsymbol{\psi}_j\rangle = \delta_{ij}, \quad j = 1, 2,$$

$$\langle\boldsymbol{\phi}_+, \boldsymbol{\psi}_j\rangle = \langle\boldsymbol{\psi}_+, \boldsymbol{\phi}_j\rangle = 0, \quad j = 1, 2,$$

$$\langle\boldsymbol{\phi}_+, \boldsymbol{\psi}_+\rangle = 1, \quad \langle\boldsymbol{\phi}_+, \boldsymbol{\psi}_-\rangle = 0.$$

The adjoint basis vectors have the form

$$\boldsymbol{\psi}_0 = \left[\Delta^{\pm} R_i\left(1 - \frac{\cos^2\theta}{V^{\pm 2}}\right)\right]^{-1} \times \begin{pmatrix} (R_e - R_i)\dfrac{\cos\theta}{V^{\pm}} \\ 0 \\ R_i\left[1 - R_i R_e^{-1}\left(\dfrac{\cos\theta}{V^{\pm}}\right)^2\right] \\ 0 \end{pmatrix},$$

$$\boldsymbol{\psi}_1 = \Delta^{\pm -1} \begin{pmatrix} 0 \\ R_e \\ 0 \\ R_i\dfrac{\cos\theta}{V^{\pm}} \end{pmatrix},$$

$$\boldsymbol{\psi}_+ = -\frac{R_e}{2q^\pm}\begin{pmatrix} 1 \\ -iq^{\pm-1}(R_i - R_e)\dfrac{\cos\theta}{V^\pm} \\ \dfrac{\cos\theta}{V^\pm} \\ -iR_i q^{\pm-1}\left[1 - R_e R_i^{-1}\left(\dfrac{\cos\theta}{V^\pm}\right)^2\right] \end{pmatrix}$$

$$\boldsymbol{\psi}_- = \overline{\boldsymbol{\psi}}_+.$$

The advantage of the substitution (6.1) lies in the fact that the unknown vector function **w** decomposes into the long-wavelength (a_0, a_1) and short-wavelength a_+ parts. In this connection, it is natural to make the following scale transformation in (6.2)

$$a_0 = \mu b_0, \quad a_1 = \mu|\mu|^{1/2} b_1,$$
$$a_+ = \mu^2 z, \quad \zeta = |\mu|^{1/2}\xi.$$

In the new variables Eqs. (6.2) become

$$b_0' = b_1 + O(\mu),$$

$$\begin{aligned} b_1' &= \operatorname{sgn}(\mu)(c_1^\pm b_0 - c_2^\pm b_0^2) + O(\mu), \\ \dot{z} &= iq^\pm z + i(d_1^\pm b_0 + d_2^\pm b_0^2) + O(|\mu|^{1/2}), \end{aligned} \tag{6.4}$$

$$\dot{\bar{z}} = -iq^\pm \bar{z} - i(d_1^\pm b_0 + d_2^\pm b_0^2) + O(|\mu|^{1/2}),$$

where a prime denotes differentiation with respect to ζ. It follows from (6.2) that to lowest order in μ the system of equations (6.2) has the solution

$$a_0^\pm = \mu\frac{3}{2}\frac{c_1^\pm}{c_2^\pm}\cosh^{-2}\frac{\sqrt{c_1^\pm\mu}}{2}\xi, \quad a_1^\pm = \dot{a}_0^\pm. \tag{6.5}$$

We note that in the present case $\Delta_+ < 0$, $c_2^+ < 0$ and $c_1^+ < 0$ in (6.3), and $\Delta^- > 0$, $c_2^- < 0$, and $c_1^- > 0$. Hence we have that in (6.5) $\mu < 0$ for fast magnetosonic and $\mu > 0$ for slow magnetosonic waves.

To determine the asymptotic behavior of (6.5) at infinity, we shall examine the local structure of these solutions in the spectral region. Dropping higher order terms, we rewrite the last pair of equations in Eqs. (6.2) in the form

$$\begin{aligned} &\mathcal{L}a_+ = if^\pm(a_0^{\pm *}), \quad \overline{\mathcal{L}}a_- = -if^\pm(a_0^{\pm *}), \\ &\mathcal{L} = \frac{\partial}{\partial\xi} - iq^\pm, \end{aligned} \tag{6.6}$$

where $f^\pm(y) = (\mu d_1^\pm y + d_2^\pm y^2)$, and $a_0^{\pm *}$ is identical to (6.5). Applying the Fourier transform

$$\hat{a}_\pm = \frac{1}{2}\pi\int_{-\infty}^{\infty} a_\pm \exp(-ik\xi)d\xi,$$

we obtain from (6.6)

$$\hat{a}_+ = -\frac{\hat{f}_{2s}}{k + q^\pm}, \quad \hat{a}_- = \frac{\hat{f}_{2s}}{k - q^\pm}, \tag{6.7}$$

where

$$\hat{f}_{2s} = (c_1 k + c_2 k^3)\operatorname{cosech}\left(-\frac{\pi k}{\sqrt{\mu c_1^\pm}}\right)$$

and c_1 and c_2 are constants. We shall seek the solutions of the system (6.2), expressed by even functions, which is compatible with the property of invertibility. Specifically, the leading part of the solution under study is (6.5) and its asymptotic behavior is the same at both infinities. We have

$$a_\pm = \frac{1}{2}\int_{\Gamma = \Gamma_1 \cup \Gamma_2} \hat{a}_\pm e^{ik\xi} dk, \tag{6.8}$$

where the integration contour Γ_1 passes above the poles $\hat{a}_\pm$ on the real axis, and the contour Γ_2 passes below these poles. The contours Γ_1 and Γ_2 contribute to the asymptotic behavior at the plus and minus infinities, respectively. It follows from (6.7) and (6.8) that to lowest order in μ, as $\xi \longrightarrow \pm\infty$

$$a_+ \longrightarrow \pm 2\pi C \exp\left(-\frac{\pi q^\pm}{\sqrt{\mu c_1^\pm}}\right)\sin(q^\pm\xi), \tag{6.9}$$

where C is a constant. The exact value of the constant C can be determined only by analyzing the complete system (6.2). It turns out that the higher order infinitesimals in the parameter μ make the same contribution, in order of magnitude, to C. In other words, instead of a polynomial, infinite series in powers of k appear in the expression for $\hat{f}_\pm$. However, we are interested only in the qualitative behavior of the solution (6.5). Consequently, we shall not determine here the exact expressions for the constants. The asymptotic behavior at minus infinity is determined from the condition that the solutions (6.5) are even.

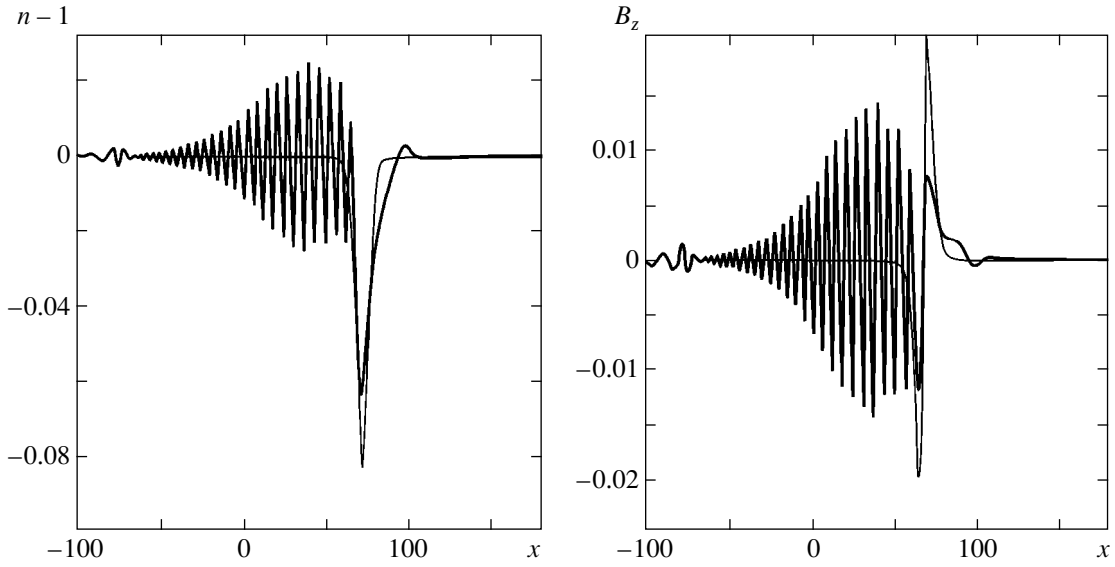


Fig. 6. Evolution of the initial data of the fast magnetosonic solitary wave type for $\theta < \theta_c$, $\mu = -0.05$, $\theta = 1.535$, $b^2 = 0.5$, $t = 0$ (fine line) and $t = 250$ (heavy line). The coordinate system moves with velocity $V^+ + \mu$.

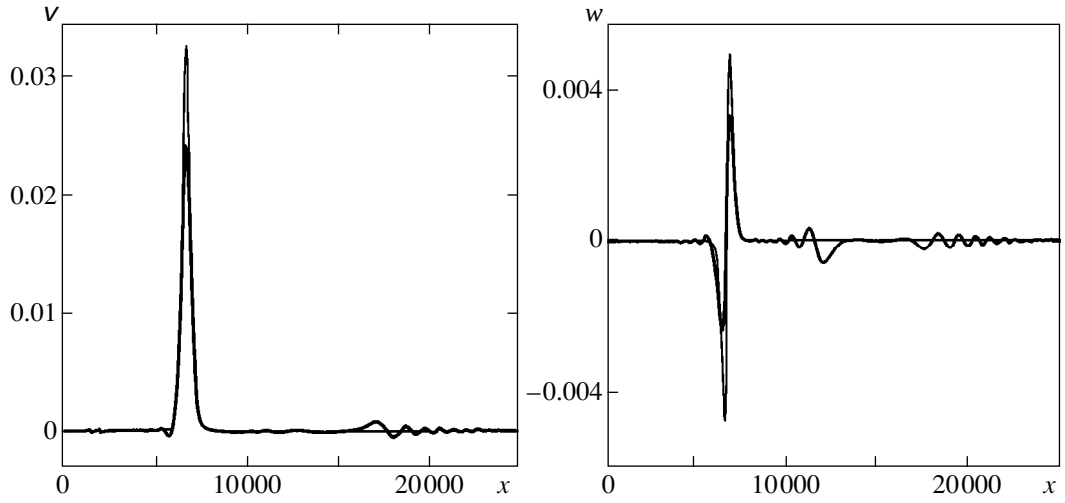


Fig. 7. Evolution of the initial data of the slow magnetosonic solitary wave type for $\mu = 0.0125$, $\theta = 0.5$, $b^2 = 0.5$, $t = 0$ (fine line) and $t = 20000$ (heavy line). The coordinate system moves with velocity $V^- + \mu$.

The expression (6.9) determines an exponentially small oscillating "tail" of the solution corresponding to (6.5). The transformation formula (6.1) gives

$$\mathbf{w}^{\pm}(+\infty) = \mathbf{D}^{\pm}\exp\left(-\pi\frac{1}{\sqrt{c_1^{\pm}\mu}}\right)\sin(q^{\pm}x) + O\left[\mu\exp\left(-\pi\frac{1}{\sqrt{c_1^{\pm}\mu}}\right)\right]. \quad (6.10)$$

The exact value of the components of the constant vectors $\mathbf{D}^{\pm}$ can be calculated only taking into account the terms of all orders in μ in (6.2). The solution, in the lowest order approximation in μ described by the expression (6.5), with the asymptotic behavior (6.10) is a generalized solitary wave.

Figure 6 shows the numerical solutions illustrating the evolution of the initial data of the symmetric (density) and antisymmetric (B_z component of the magnetic field) solitary wave type, corresponding to (6.5) for a fast magnetosonic branch. The soliton decays slowly as a result of the short-wavelength emission. The emission process is quasistationary.

Figure 7 shows how the initial data of the slow magnetosonic wave soliton type evolve for a sufficiently small value $\mu = 0.0125$. The modes of the Alfén wave are not seen; the wave number at the point of intersec-

tion of the straight line $\omega = (V^- + \mu)k$ and the dispersion curve is approximately 63.9. The calculation does not permit studying such short-wavelength waves. It is obvious that the amplitude of the resonance mode of the Alfén branch is extremely small here.

The soliton part of the solution decays exponentially as $x \longrightarrow \pm\infty$. This corresponds to purely imaginary values of k with the simultaneous solution of the equations (3.1) and $V = V^- + \mu$. The maximum admissible value of μ for which the purely imaginary value of k still exists for given values of θ and b is approximately 0.1337; for larger values of μ the straight line $\omega = (V^- + \mu)k$ simultaneously intersects the Alfén and fast magnetosonic branches. Consequently, all values of k are real. As the critical value of μ is approached, the value $k = q^-$ at the point of intersection with the Alfén branch decreases and approaches a value approximately equal to 0.82, and the absolute value of the purely imaginary k approaches infinity. It is possible that in this case the wavelength of the resonance mode of the Alfén branch is comparable to the wavelength of the soliton "core" of a generalized solitary wave with finite amplitude, if it exists, but the formulas (6.5) used for the initial data do not describe the soliton core near the critical value of μ, as one can already see for $\mu = 0.1$ ($q^- \approx 55.6$).

Qualitatively, the evolution of the initial data for large and small values of μ is of the same type. The difference lies only in the character of the decay time and the relative amplitude of the waves emitted. The decay time increases as μ decreases, and the relative amplitude decreases. At first, a fast magnetosonic wave moving leftward is emitted, after which a slow wave moving leftward is emitted, and then a fast wave moving rightward is emitted. These waves move away from the soliton. Essentially, this is a correction to the error in the initial data, since the asymptotic formulas (6.5) of first-order accuracy in μ are used to describe the solitary wave. As a result, only a slow magnetosonic wave remains to the left of the soliton. This is the wave responsible for the subsequent decay of the soliton. We note that the emission of a wave of this type is observed in a similar calculation for a fast magnetosonic branch as well as for a generalized KdV equation [21]. In the calculations indicated, the wavelength at the point of intersection of the straight line $\omega = Vk$ is comparable to the length of the soliton, and consequently quasistationary resonance radiation, which makes the dominant contribution to soliton decay, is also observed.

7. SHOCKS FOR THE FAST MAGNETOSONIC BRANCH AND THEIR RELATION TO SOLITARY WAVES

In nondissipative systems, shocks of a more general form than in dissipative systems can arise. The concept of shock can include, aside from shocks between uniform states, shocks between periodic, quasiperiodic, and stochastic states. Such shocks are realized as elements of self-similar solutions in the problem of the decay of an arbitrary discontinuity. The stationary solutions studied above are treated as shock structures. The general theory of such shocks is described in [16, 17]. In order to apply this theory the model must be described by a symmetric, invertible, and conservative system of equations. After this system is integrated once, the system of ordinary differential equations for stationary solutions (or solutions of the traveling-wave type) can be represented as a dynamic system of the form (4.2), $\mathbf{w} = \{w_q\}$, invariant under the transformations $x \longrightarrow -x$ and $w_q \longrightarrow w_q$ with $q = 1, \ldots, 2n - 1$ (symmetric unknowns) and $w_q \longrightarrow -w_q$ with $q = 2, \ldots, 2n$ (antisymmetric unknowns).

According to this theory, a preliminary assessment is made of the possibility of the existence of a stationary structure of a shock based on the properties of invariant manifolds. Then, the satisfaction of the condition of evolution is analyzed for this structure: the number of boundary conditions at the shock should be one greater than the number of outgoing characteristics. The evolutionary nature signifies stability and therefore observability of the shock in numerical and physical experiments. A numerical experiment confirming the existence of the type of shock under study completes the investigation. The method is applicable for dynamical systems of arbitrary order. The distortions introduced by the numerical scheme are equivalent to including higher order derivatives with respect to x in the initial equations. Consequently, if a numerical scheme preserving the properties of conservativeness and symmetry is used, then the numerical method for checking the existence of stationary solutions is not critical with respect to approximation errors. A brief description of the numerical method employed is presented in the Appendix.

The numerical experiment consists of the following. Initial data are taken for v and B_y, proportional to the principal eigenvector for magnetosonic waves, multiplied by a function of the type $\tanh(x - x_0)/\delta)$ and the initial data for w and B_z are taken to be zero. Aside from a magnetosonic shock, different weaker shocks are also found. A fragment of a solution with one magnetoshock is singled out, and the computational region is extended up to the required dimensions. Figure 8 shows the evolution of different types of solutions as a function of θ with fixed b. Solutions with all possible types of shocks (Figs. 8a–8c) are found. The dashed rectangle shows the section of the solution that is investigated for stationariness. Figure 9 shows schematically how the type of solution depends on the amplitude of the global discontinuity $\Delta n = n_{x \to -\infty} - n_{x \to +\infty}$ and the parameter θ in the region near the critical value $\theta_c(b)$.

The simplest type of shock is a shock between a sequence of solitary waves (as $t \longrightarrow \infty$) and the uniform state, Fig. 8c (soliton-type shock). Shocks of this type have been studied for the KdV equation in [15]. The conditions for the existence of such a shock are

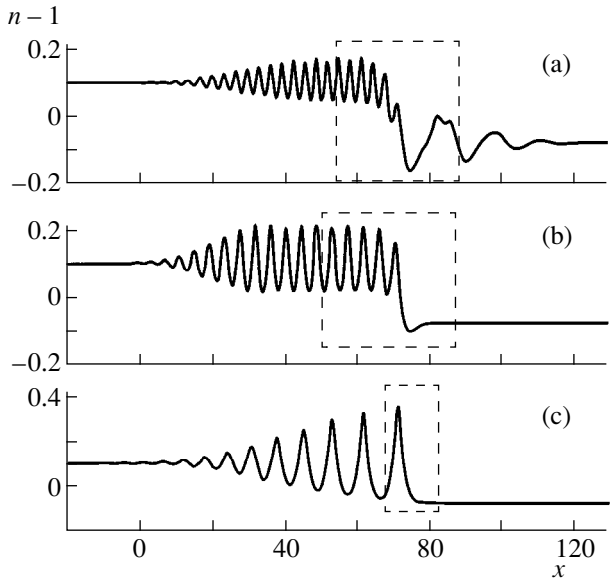


Fig. 8. Various types of solutions with a fast magnetosonic shock for θ = 1.52, 1.53, 1.55 from top to bottom, respectively; $b^2 = 0.5$ and $t = 250$.

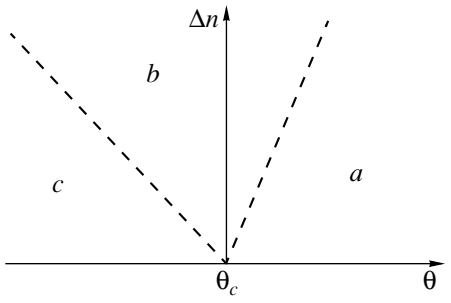


Fig. 9. Diagram showing the arrangement of regions of various types of solutions with a fast magnetosonic shock near the critical value $\theta = \theta_c$ (*b*): (*a*) shocks with soliton structure; (*b*) shocks with radiation emission; (*c*) shocks with a nonstationary structure.

obvious from the above discussion. The necessary condition for the existence of a low-amplitude soliton is the absence of intersection of the straight line $U = \omega/k$ and the dispersion curve in the region to the right of the shock. For existence of single-hump solitons of finite amplitude, and these are the solitary waves that form a sequence in soliton-type shocks, it is also necessary to require that the value of k in the simultaneous solution of the equation $U = \omega/k$ and the equation for the dispersion curve be purely imaginary and not complex.

In the present model it is also possible to have a shock between uniform and periodic states (shock with radiation); Fig. 8b. We shall explain why and under what conditions the structure of such a shock will exist. Let us assume that the system of ordinary differential equations for traveling waves has at least two stationary points. We linearize the initial system with respect to the uniform states near each of these points. We examine the corresponding dispersion curves $\omega = \omega(k)$ and the equations for the wavelengths of solutions of the traveling-wave type $R(U, k) = 0$, where U is the velocity of the shock. The equation $R(U, k) = 0$ possesses $2n$ roots, and $k_{2r} = -k_{2r-1}$, $r = 1, \ldots, n$. The solution of the linearized dynamical system depends on $2n$ parameters c_j and can be written in the form

$$\mathrm{Re}\left[\sum_{j=1}^{2n} c_j \exp(ik_j x)\right].$$

Waves which grow as $x \longrightarrow +\infty$ are associated with the values $\mathrm{Im}(k_j) < 0$, and waves which grow as $x \longrightarrow -\infty$ are associated with values $\mathrm{Im}(k_j) > 0$; waves associated with real values of k_j are purely periodic waves. In what follows, it is assumed that these qualitative properties are also preserved for nonlinear variants of systems with the difference that the growing waves can be growing only near the point of equilibrium and are bounded away from it. Let us consider a subset of an invariant subspace formed by the phase trajectories $\{\mathbf{w}(x), -\infty < x < +\infty\}$, passing through an ϵ neighborhood of the point C at $x = x_0$ and bounded or periodic at $+\infty$. We shall refer to such subspaces as $S(C, \epsilon; |\mathbf{w}| < M, x \longrightarrow +\infty)$ or $S(C, \epsilon; \mathbf{w}(x + T) \longrightarrow \mathbf{w}(x), x \longrightarrow +\infty)$. The quantity x_0 plays no role, since the solution is determined to within a phase shift.

Let there be two points of equilibrium. The straight line $\omega = kU$ for one of them (C_1) does not intersect the dispersion curve, and for the other (C_2) it intersects once. At the point of equilibrium C_1 there are n roots with $\mathrm{Im}\,k > 0$ and n roots with $\mathrm{Im}\,k < 0$; at the point C_2 there are $n - 1$ roots with $\mathrm{Im}\,k > 0$ and $n - 1$ roots with $\mathrm{Im}\,k < 0$ and two roots with $\mathrm{Im}\,k = 0$. Therefore the invariant manifolds $S_1 = S(C_1, \epsilon_1; \mathbf{w} \longrightarrow C_1, x \longrightarrow -\infty)$ and $S_2 = S(C_2, \epsilon_2; \mathbf{w}(x + T) \longrightarrow \mathbf{w}(x), x \longrightarrow +\infty)$ possess, respectively, the dimensions n (number of decreasing waves as $t \longrightarrow -\infty$) and $n + 1$ (number of periodic waves plus the number of decreasing waves as $t \longrightarrow +\infty$). Their total dimension is $2n + 1$. In this case the presence of lines of intersection is a case of the general position. The lines of intersection correspond to a shock with radiation emission. In this model such a shock is possible for $b < b_c$ and for $b > b_c$.

Numerical solutions of the problem of the decay of an initial discontinuity for the generalized KdV and Schrödinger equations show that for fourth-order dynamical systems such a shock arises only if all values of k are complex at the point C_1. The numerical experiments showed that in the case of a shock with an emitted wave the roots are complex for the present model as well. Spatially decaying oscillations in front of a shock can be seen in Fig. 8b.

In [16] it is shown that shocks with radiation emission and shocks with a soliton structure are evolution-

ary. They are indeed observed in the numerical experiment.

One would assume that a shock in which there is a sequence of generalized solitary waves is possible when an intersection of the straight line $U = \omega/k$ and the dispersion curve is present for the region to the right of the shock. However, the numerical experiment shows that such shocks are replaced by shocks with nonstationary structure (Fig. 8a). Apparently, such shocks are nonevolutionary. For a fixed value $\theta < \theta_c$ there exists a critical value $\Delta n_c(\theta)$ for which a transition occurs from shocks with stationary structure to shocks with radiation emission. If $\Delta n \ll \Delta n_c$, then the solution in the framed region is similar to a generalized solitary wave, but the amplitudes of the soliton and periodic components fluctuate slowly with time. As $\Delta n \longrightarrow 0$ the amplitude of the periodic component approaches zero and the solution becomes similar to a solution with a soliton-type shock. Conversely, if Δn is close to Δn_c, then the solution becomes chaotic.

Comparing the condition for the existence of solitary waves with the conditions for the existence of shock structures, it can be concluded that a soliton-type shock corresponds to the existence of a solitary wave, a shock with nonstationary structure corresponds to a generalized solitary wave, and a shock with radiation corresponds to the situation where there is no solitary wave, i.e., the case of relatively large amplitudes.

8. DISCUSSION

As a result of resonances between long- and short-wavelength waves, the equations of a isotropic collisionless quasineutral plasma which are studied in the present paper admit as solutions the following soliton-like structures.

8.1. Fast Waves

(a) Simple resonance. For $b > b_c$ there exists a family of solitons corresponding to compression waves (Fig. 5). A perturbation in the form of a soliton propagates without a change in shape. For example, for a hydrogen isotropic collisionless plasma, which is one of the models for the plasma in the Earth's magnetosphere (see, for example, [14]), in the absence of magnetic storms $b \simeq 0.1$, and solitons occur for $\cos\theta < 1.01 m_e/m_i$, i.e., for angles θ in the range (88.68°, 90°]. It had been assumed (see, for example, [8]) that soliton families branching from the state of rest in a magnetized plasma exist for all angles of inclination $0 < \theta \le \pi/2$ of the unperturbed magnetic field with respect to the direction of propagation of the wave, since the motion of the plasma was assumed to be described by the model KdV equation. At the same time, the results presented in the present paper show that solitons exist only for a very narrow range of angles θ and moderate values of the parameter b in a physically real plasma.

(b) Resonance of short- and long-wavelength waves. For $b < b_c$ the case of the general position is the absence of solitons. They are replaced by soliton-like structures—generalized solitary waves, which are the result of a superposition of a soliton "core" (6.5) and a periodic resonant wave. For small amplitudes of the core the amplitude of the periodic wave is exponentially small. The generalized solitary waves are rarefaction waves (Fig. 6).

The substitution of solitons by nonlocalized objects—generalized solitary waves—has a large effect on the character of the decay of localized disturbances. In this case, just as in a cold plasma [22], the decay of disturbances in the form of a soliton core, for example, leads to quasistationary emission of a periodic wave whose period is close to the period of the resonant periodic wave. The radiation is all the more intense, the larger the amplitude of the initial wave and the closer the effective velocity of sound b to the value b_c. In this case the radiation carries off the main, and significant, energy of the initial wave (see Fig. 6). Quasistationary decay of solitary waves as a result of emission, associated with the substitution of soliton solutions by generalized solitary waves, is typical for wave motions in a number of other dispersive media (see, for example, [21, 23, 24]).

8.2. Slow Waves

For all values of the physical parameters b and θ there are no solitons, and generalized solitary waves, formed as a result of a nonlinear resonance of a slow soliton core and a periodic wave with wave numbers and frequencies on the Alfén branch, occur. Just as in the case of fast waves, for algebraically small core amplitudes the amplitude of the resonant periodic component is exponentially small. The resonant wave number q^- in this case lies far from zero. Numerical calculations of the decay of a slow soliton core did not reveal any emitted waves (their amplitude is very small; the reason is that q^- lies far from zero). As the core decays, at first a fast magnetosonic wave moving leftward is emitted; this is followed by a slow leftward moving wave and then a fast rightward moving wave. In time these waves travel away from the soliton core. As a result, only a slow magnetosonic wave is constantly emitted; this wave is responsible for the subsequent decay of the soliton (Fig. 7). The indicated mechanism of decay of soliton disturbances with the replacement of soliton solutions by generalized solitary waves is new and requires further investigation.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project no. 99-01-00277.

APPENDIX

Numerical Scheme

Efficient numerical simulation of symmetric solutions—solitons, generalized solitary waves, shocks with radiation emission, and soliton-type shocks—requires that the numerical solutions preserve the basic properties of the initial model: conservativeness and symmetry. In this connection it is desirable to put the initial system into a conservative form and to approximate it by central differences:

$$\frac{\partial n}{\partial t} + \frac{\partial nu}{\partial x} = 0, \quad \frac{\partial nu}{\partial t} + \frac{\partial}{\partial x}\left(\frac{B_y^2 + B_z^2}{2} + b^2 n\right) = 0,$$

$$\frac{\partial nv}{\partial t} + \frac{\partial}{\partial x}\left(nuv - B_x B_y + R_e^{-1}\frac{dB_z}{dt}\right) = 0,$$

$$\frac{\partial nw}{\partial t} + \frac{\partial}{\partial x}\left(nuw - B_x B_z + R_e^{-1}\frac{dB_y}{dt}\right) = 0,$$

$$\frac{\partial B_y}{\partial t} + \frac{\partial}{\partial x}\left(uB_y - B_x v - R_i^{-1}\frac{\partial}{\partial x}\frac{dw}{dt}\right) = 0,$$

$$\frac{\partial B_z}{\partial t} + \frac{\partial}{\partial x}\left(uB_z - B_x w - R_i^{-1}\frac{\partial}{\partial x}\frac{dv}{dt}\right) = 0.$$

However, it was found that problems associated with the instability at the boundary arise in the approximation of this system. Consequently, it was transformed into a form where each equation contained time derivatives only of one unknown. This ensures stability and simplifies the implementation of an algorithm for solving implicit discrete equations:

$$\frac{\partial n}{\partial t} + \frac{\partial nu}{\partial x} = 0,$$

$$\frac{\partial nu}{\partial t} + \frac{\partial}{\partial x}\left(nu^2 + \frac{B_y^2 + B_z^2}{2} + b^2 n\right) = 0,$$

$$\frac{\partial nv}{\partial t} + \frac{\partial}{\partial x}\left\{-\frac{1}{R_e R_i}\frac{\partial^2 v}{\partial x \partial t} + nuv - B_x B_y + \frac{1}{R_e}\left[B_x\frac{\partial w}{\partial x} - B_z\frac{\partial u}{\partial x} - \frac{1}{R_i}\left(\frac{\partial u}{\partial x}\frac{\partial}{\partial x}(v) + u\frac{\partial^2 v}{\partial x^2}\right)\right]\right\} = 0.$$

$$\frac{\partial nw}{\partial t} + \frac{\partial}{\partial x}\left\{-\frac{1}{R_e R_i}\frac{\partial^2 w}{\partial x \partial t} + nuw - B_x B_z - \frac{1}{R_e}\left[B_x\frac{\partial v}{\partial x} - B_y\frac{\partial u}{\partial x} + \frac{1}{R_i}\left(\frac{\partial u}{\partial x}\frac{\partial}{\partial x}(w) + u\frac{\partial^2 w}{\partial x^2}\right)\right]\right\} = 0,$$

$$\frac{\partial B_y}{\partial t} + \frac{\partial}{\partial x}\left\{-\frac{1}{R_e R_i n}\frac{\partial^2 B_y}{\partial x \partial t} + B_y u - B_x v - \frac{1}{R_i n}\left[B_x\frac{\partial B_z}{\partial x} + \frac{1}{R_e}\left(u\frac{\partial^2 B_y}{\partial x^2} + \frac{\partial u}{\partial x}\frac{\partial}{\partial x}(B_y)\right)\right]\right\} = 0,$$

$$\frac{\partial B_z}{\partial t} + \frac{\partial}{\partial x}\left\{-\frac{1}{R_e R_i n}\frac{\partial^2 B_z}{\partial x \partial t} + B_z u - B_x w + \frac{1}{R_i n}\left[B_x\frac{\partial B_y}{\partial x} - \frac{1}{R_e}\left(u\frac{\partial^2 B_z}{\partial x^2} + \frac{\partial u}{\partial x}\frac{\partial}{\partial x}(B_z)\right)\right]\right\} = 0.$$

A cross-type three-layer scheme with a second-order approximation in time or a two-layer scheme with a first-order approximation (control scheme) can be used. The two-layer scheme makes it possible to include, if so desired, dissipative terms in the initial equations. Comparing the results obtained with these two schemes did not show any substantial differences because of the small time step, and the condition $\Delta t = c_0 \Delta x^3$, where c_0 is a constant, was maintained in the calculations.

We shall use a superscript to denote the temporal layers and a subscript to denote the number of nodes in space. These equations are regularized: each equation contains time derivatives only of one variable with the exception of the variable n, which is determined explicitly. This makes it possible to solve four implicit difference equations for nv, nw, B_y, and B_z separately. The method for approximating the first-order derivatives is obvious:

$$f_x \longrightarrow \frac{(f_{i+1} - f_{i-1})/2}{\Delta x}, \quad f_t \longrightarrow \frac{(f^{m+1} - f^{m-1})/2}{\Delta t}.$$

To preserve in the numerical solution the symmetry properties used in the theory presented above, central differences must be used to approximate the x derivatives. Templates for approximating the even and odd conservative high-order derivatives are presented below:

$$(gf_x)_x \longrightarrow [(g_{i+1} + g_i)(f_{i+1} - f_i) - (g_i + g_{i-1})(f_i - f_{i-1})]/2/\Delta x^2,$$

$$(gf_x h_x)_x \longrightarrow [(g_{i+1} + g_i)(f_{i+1} - f_i)(h_{i+1} - h_i) - (g_i + g_{i-1})(f_i - f_{i-1})(h_i - h_{i-1})]/2/\Delta x^3,$$

$$(gf_{xt})_x \longrightarrow \{[(g_{i+1}^m + g_i^m)(f_{i+1}^{m+1} - f_i^{m+1}) - (g_i^m + g_{i-1}^m)(f_i^{m+1} - f_{i-1}^{m+1})]/2/\Delta x^2 - [(g_{i+1}^m + g_i^m)(f_{i+1}^{m-1} - f_i^{m-1}) - (g_i^m + g_{i-1}^m)(f_i^{m-1} - f_{i-1}^{m-1})]/2/\Delta x^2\}/2/\Delta t,$$

$$(gf_{xx})_x \longrightarrow [g_{i+1}(f_{i+2} + f_i - 2f_{i+1})$$
$$- g_{i-1}(f_{i-2} + f_i - 2f_{i-1})/2/\Delta x^3.$$

Here f, g, and h are arbitrary unknowns. The purpose of such complicated templates is to decrease the effect of the dispersion due to the numerical scheme and to ensure stability. The quantity Δt was chosen as $\sim\Delta x^3$, since four of the six equations of the simulated system are similar to KdV equations.

An iteration method was used to solve the implicit discrete equations:

$$f^{m+1,j+1} = f^{m+1,j} + c_f$$
$$\times F_f(f^{m+1,j}, f^{m-1}, n^m, u^m, v^m, w^m, B_y^m, B_z^m),$$
$$f^{m+1,0} = f^{m-1}, \quad f \longrightarrow nv, nw, B_y, B_z.$$

Here F_f is a finite-difference approximation of the corresponding equation and j is the number of the iteration step. The iteration process was stopped when $|f^{m+1,j+1} - f^{m+1,j}| \leq \epsilon$, where ϵ is a parameter determining the desired accuracy. The values of c_f were obtained in the course of the numerical experiment.

REFERENCES

1. D. Montgomery, Phys. Fluids **2**, 585 (1959).
2. P. G. Saffman, J. Fluid Mech. **11**, 16 (1961).
3. P. G. Saffman, J. Fluid Mech. **11**, 552 (1961).
4. R. Z. Sagdeev, in *Reviews of Plasma Physics*, Ed. by M. A. Leontovich (Atomizdat, Moscow, 1964; Consultants Bureau, New York, 1968), Vol. 4.
5. P. J. Kellogg, Phys. Fluids **7**, 1555 (1964).
6. T. Kakutani, J. Phys. Soc. Jpn. **21**, 385 (1966).
7. T. Kakutani, H. Ono, T. Taniuti, and C. Wei, J. Phys. Soc. Jpn. **24**, 1159 (1968).
8. T. Kakutani and H. Ono, J. Phys. Soc. Jpn. **26**, 1305 (1969).
9. A. Il'ichev, J. Plasma Phys. **55**, 181 (1996).
10. A. T Il'ichev, Mat. Zametki **59**, 719 (1996).
11. A. T. Il'ichev, Izv. Akad. Nauk, Mekh. Zhidk. Gaza **3**, 154 (1996).
12. T. Kakutani, T. Kawahara, and T. Taniuti, J. Phys. Soc. Jpn. **23**, 1138 (1967).
13. T. Kawahara, J. Phys. Soc. Jpn. **27**, 1331 (1969).
14. V. L. Patel and B. Dasgupta, Physica D (Amsterdam) **27**, 387 (1987).
15. A. V. Gurevich and L. P. Pitaevskiĭ, Zh. Éksp. Teor. Fiz. **65**, 590 (1973) [Sov. Phys. JETP **38**, 291 (1973)].
16. I. B. Bakholdin, Prikl. Mat. Mekh. **63**, 52 (1999).
17. I. B. Bakholdin, Izv. Akad. Nauk, Mekh. Zhidk. Gaza **4**, 95 (1999).
18. V. D. Shafranov, in *Reviews of Plasma Physics*, Ed. by M. A. Leontovich (Gosatomizdat, Moscow, 1963; Consultants Bureau, New York, 1967), Vol. 3.
19. A. T. Il'ichev, Izv. Akad. Nauk, Mekh. Zhidk. Gaza **2**, 3 (2000).
20. G. Iooss and M. Adelmeyer, *Topics on Bifurcation Theory and Applications* (World Scientific, Singapore, 1992).
21. I. Bakholdin and A. Il'ichev, in *Contemporary Mathematics*, Ed. by F. Dias and J.-M. Ghidaglia (Providence, Rhode Island, 1996), Vol. 200, p. 1.
22. I. Bakholdin and A. Il'ichev, J. Plasma Phys. **60**, 569 (1998).
23. T. R. Akylas and R. H. J. Grimshaw, J. Fluid Mech. **242**, 279 (1992).
24. E. S. Benilov, R. Grimshaw, and E. P. Kuznetsova, Physica D (Amsterdam) **69**, 270 (1993).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 1, 2000, pp. 126–131.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 1, 2000, pp. 142–148.

FLUIDS

Pyroelectric Investigation of the Surface Polarization in the Isotropic Phase of a Nematic Liquid Crystal

N. M. Shtykov*, V. P. Panov, and M. I. Barnik

Shubnikov Institute of Crystallography, Russian Academy of Sciences, Moscow, 117333 Russia
**e-mail: lev@glasnet.ru*

Received March 15, 2000

Abstract—Measurements of the temperature dependence of the pyroelectric coefficients in cells with homeotropic and planar boundary conditions were performed in order to study the surface polarization and determine its nature in the isotropic phase of a nematic liquid crystal. Analysis of these data established that the ordering polarization arising as a result of the nonuniformity of the order parameter in the near-boundary region of the liquid crystal and the polar monolayer due to the bifilar properties of the molecules of the liquid crystal at the boundary with the solid wall of the cell are responsible for the surface polarization in the isotropic phase of a nematic liquid crystal. The values of the pyroelectric coefficients of monolayers and the coefficients of the ordering polarization for planar and homeotropic orientations are estimated. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

It is well known that the ordering of the molecules of a nematic liquid crystal (NLC) at a boundary with a solid substrate can differ substantially from the ordering in the interior volume not only quantitatively with respect to the magnitude of the order parameter but also qualitatively as a formation of a polar orientational order at the interface. The macroscopic surface polarization which arises influences the characteristics of the electro-optic effects [1, 2], determines the preinclination angles as a result of the dipole contribution to the bonding energy between the liquid-crystal molecules and the surface [3], and creates the conditions for the realization of surface orientational transitions [4, 5]. Thus, investigations of surface phenomena in NLC are of interest from the applied standpoint as well as from the standpoint of studying the fundamental properties of the liquid-crystal state.

Molecular ordering at the boundary of a liquid crystal has been studied by ellipsometry [6], X-ray spectroscopy [7], and nuclear magnetic resonance [8]. However, these methods cannot distinguish the polar and nonpolar types of molecular ordering. One method of investigating the surface properties of liquid crystals which are sensitive to the polarity of the molecular order is optical second-harmonic generation. In the dipole approximation second harmonic generation is forbidden in centrosymmetric media, and it can be observed only in noncentrosymmetric media, i.e., polar media. In [9] it was determined, using second-harmonic generation, that polar ordering of cyanobiphenyl molecules is absent at a free boundary of a NLC (at the boundary with air), but it exists at a boundary with a glass substrate. The surface polarization is localized in a monomolecular layer at the interface, and the degree of polar ordering of the monolayer is independent of temperature and remains constant in a wide range, including both the nematic and isotropic phases of the liquid crystal. Second-harmonic generation has been observed at the boundary with glass substrates with both homeotropic and planar orientations of the NLC [10]. In [11, 12] surface polarization was investigated using the method of modulation spectroscopy of electroreflection from a NLC–semiconductor interface. The polarization observed depended on the temperature and was explained by the presence of boundary-induced polar ordering of the liquid-crystal molecules.

In the present work the pyroelectric method was used to study the surface polarization in the isotropic phase of a NLC in thin layers with planar and homeotropic boundary conditions. This method, just as the second-harmonic generation method, makes it possible to distinguish polar and nonpolar types of molecular ordering. The most likely contributions to surface polarization, which are due to the presence of a nonuniform orientational ordering of the molecules and the bifilar nature of the liquid-crystal molecules, i.e., the difference in the interaction of the different ends of polar molecules with a solid surface [9, 10, 13], were analyzed.

2. THEORETICAL FOUNDATIONS OF THE METHOD

There are at least two possible mechanisms for the appearance of surface polarization in the isotropic phase of a NLC. One is the difference in the interaction of different ends of the polar molecules of a NLC with a solid surface; this difference could lead to the formation of a polar monolayer at the interface [11]. In [9, 10]

1063-7761/00/9101-0126$20.00

it was established that such a monolayer exists on a clean glass surface, on a surface treated with a surfactant in order to obtain a homeotropic orientation, and on a surface coated with a polyimide film, which usually used to obtain planar orientation. In all cases the molecules of the experimental cyanobiphenyl liquid crystal were bonded with the surface by a polar CN group, i.e., the polarization vector of the monolayer was directed along the normal to the surface inside the liquid crystal. The angles of inclination θ_m of the liquid-crystal molecules in the monolayers with respect to the normal are different and equal to $\theta_m \sim 67°$ on a surfactant film and $\theta_m \sim 80°$ on the polyimide film.

The second reason for the appearance of surface polarization in the isotropic phase of a NLC, especially near a nematic–isotropic liquid ($N \rightleftarrows I$) phase transition, could be the difference between the surface-induced scalar order parameter S_0 and the volume order parameter S. This leads to the appearance of the so-called "ordering polarization" (order-electric polarization), predicted in [14]:

$$\mathbf{P}_0 = r\,\mathrm{grad}S, \tag{1}$$

where r is the ordering polarization constant, which is equal in order of magnitude to the flexoelectric coefficients e_1 and e_3 [15].

It is obvious from symmetry considerations that in the isotropic phase of a NLC the surface-induced order parameter depends only on the coordinate z. The Z-axis is directed along the normal to the orienting surface and therefore the gradient of the order parameter is parallel to the normal. Then the expression for the normal component of the ordering polarization can be written in the form

$$P_{0z}(z) = rdS/dz. \tag{2}$$

The phenomenological Landau–de Gennes theory [8, 16, 17] is ordinarily used to describe the properties of orientational ordering of a liquid crystal at a boundary with a solid surface. According to this theory, the free-energy density of a liquid crystal with uniform orientation of the director can be written as an expansion in powers of the order parameter and its gradient [16]:

$$f = f_0 + \frac{a}{2}(T - T^*)S^2 - \frac{B}{3}S^3 + \frac{C}{4}S^4 + \frac{L}{2}\left(\frac{dS}{dz}\right)^2 + f_s, \tag{3}$$

where f_0 is the part of the free energy that is independent of the order parameter, a, B, C, and L are the material parameters, T^* is the maximum temperature of supercooling of the isotropic phase, f_s is the surface density of the free energy, the simplest expression for which, according to [16], is $f_s = -GS_0$ (here G is the interaction energy between the liquid crystal and the surface and S_0 is the order parameter at the surface of the liquid crystal). Minimizing the integral of the free-energy density (3) gives an expression for the order parameter of a liquid crystal near $S(z)$ and at the interface with the substrate S_0. In the isotropic phase of the liquid crystal, neglecting the terms $\sim S^3$ and $\sim S^4$ in the expansion, the following expressions are obtained for $S(z)$ and S_0 [8]:

$$S(z) = S_0\exp\left(-\frac{z}{\xi}\right), \quad S_0 = S_{00}\left(\frac{T^*}{T - T^*}\right)^{1/2}, \tag{4}$$

where $\xi = (L/aT^*)^{1/2}(T/T^* - 1)^{-1/2}$ is the correlation length and $S_{00} = G/(aLT^*)^{1/2}$.

Averaging over the thickness of the cell the ordering polarization $\langle P_{op}\rangle$, induced by one bounding surface, $z = 0$ (we assume that the order parameter S at the second surface, $z = d$, corresponds to the bulk value (in the isotropic phase $S = 0$)) can be expressed as

$$\langle P_{op}\rangle = \frac{1}{d}\int_0^d r\frac{dS}{dz}dz = \frac{1}{d}\int_{S_0}^{S} r\,dS = -\frac{r}{d}S_0. \tag{5}$$

The total polarization on a single boundary surface in the isotropic phase of the NLC averaged over the thickness of the cell is equal to the sum of the polarization $\langle P_m\rangle$ of the surface layer and the ordering polarization $\langle P_{op}\rangle$. Using the relations (4) the pyroelectric coefficients for cells with planar and homeotropic boundary conditions can be expressed as

$$\begin{aligned}\gamma_j &= \frac{d}{dT}(\langle P_{mj}\rangle + \langle P_{opj}\rangle)\\ &= \gamma_{mj} + \frac{rG_j}{4d\sqrt{aL}}(T - T^*)^{-3/2}\\ &= \gamma_{mj} + \gamma_{opj}(T - T^*)^{-3/2},\end{aligned} \tag{6}$$

where γ_j is the total pyroelectric coefficient, γ_{mj} is the pyroelectric coefficient of the monolayer, γ_{opj} is the pyroelectric coefficient due to the ordering polarization, and the index j corresponds to different types of boundary conditions at the cell walls.

The total polarization of the cell $\langle P\rangle = \langle P_m\rangle + \langle P_{op}\rangle$ is determined by contributions from both boundary surfaces of the liquid crystal in the cell $\langle P\rangle^{(1)}$ and $\langle P\rangle^{(2)}$, which under isothermal conditions are equal in modulus and opposite in direction. Thus, under stationary isothermal conditions the total polarization is zero.

The pyroelectric current arising with modulation of the temperature of a liquid crystal in the cell is determined by the contributions of the polarization currents from both cell boundaries, and if the temperature modulation amplitudes $\Delta T_0^{(1)}$ and $\Delta T_0^{(2)}$ at the boundaries are different, then the pyroelectric current through the cell is also different from zero. According to the conditions of the experiment described below, where the

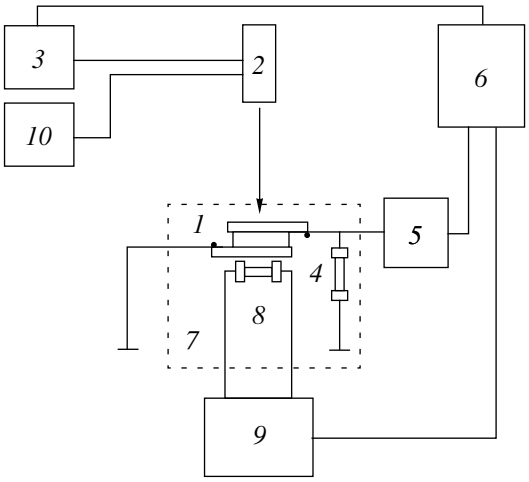


Fig. 1. Block diagram of the experimental setup for measuring the polarization by the pyroelectric effect method: (*1*) cell, (*2*) laser, (*3*) pulse generator, (*4*) load resistance, (*5*) preamplifier and selective voltmeter, (*6*) computer, (*7*) thermally insulated chamber, (*8*) platinum resistance thermometer, (*9*) multimeter, (*10*) laser power supply.

temperature modulation in the liquid crystal layer is achieved by modulating the power of the laser radiation, its amplitude at the first boundary surface of the cell, on which the laser radiation is incident, is much greater than on the second surface because of the damping of the light wave in the absorbing layer of the liquid crystal. Consequently, it can be assumed that the experimentally measured pyroelectric current i_p is determined primarily by the polarization only at the first boundary surface:

$$i_p = \frac{d\langle P\rangle}{dT}\left(\frac{dT^{(1)}}{dt} - \frac{dT^{(2)}}{dt}\right) \approx \frac{d\langle P\rangle}{dT}\omega\Delta T_0^{(1)} = \gamma\omega\Delta T_0^{(1)}, \quad (7)$$

where γ is the pyroelectric coefficient, and $\omega = 2\pi F$ is the circular frequency of the temperature modulation.

3. EXPERIMENT

A block diagram of the experimental setup for investigating the pyroelectric properties of a liquid crystal is shown in Fig. 1. Radiation from an FP-82/30 semiconductor laser *2* with wavelength $\lambda = 820$ nm and power $W_0 = 30$ mW in the continuous-wave regime is incident on the sample *1*. The power of the laser radiation is modulated by a signal from the square-pulse generator *3* with frequency $F = 125$ Hz and duty factor *2*. The radiation is absorbed in a thin layer of the liquid crystal, giving rise to temperature modulation in the layer. The pyroelectric voltage from the resistance of the load *4* is amplified by a preamplifier and a selective nanovoltmeter *5* and fed into the input of the acoustic chart in the computer *6*.

The measurement and processing of the pyroelectric signal were performed using the PhysLab4.0 program system and the multimedia resources of a personal computer. The amplitude of the first harmonic of the pyroelectric signal at the modulation frequency and the phase of the pyroelectric signal relative to the reference signal, fed into the second input of the acoustic chart of the computer from the pulsed generator *3*, are measured.

To measure the temperature dependences of the pyroelectric signal the sample is placed in a thermally insulated chamber *7*, where the temperature is measured using a platinum resistance thermometer *8* and a multimeter *9*, which transmits the measurement results to the serial port (COM2) of the computer.

The liquid-crystal cells were fabricated from glass plates with a transparent current-conducting tin dioxide (SnO_2) layer on one side. Planar orientation of the liquid crystal was obtained by rubbing a polyimide film deposited on an electrode [17]. Homeotropic orientation of the NLC was obtained by cleaning the surface of the glass with a polishing electrode. Two glass plates prepared in this manner were used to assemble a flat capacitor type cell in which the distance between the electrodes was set by teflon interlayers with thickness $d = 10$ μm.

The surface polarization was investigated in the liquid crystal 4-pentyl-4'-cyanobiphenyl (5CB). To increase the absorption of the laser radiation and thereby the pyroelectric signal and to produce a temperature gradient in the direction of the normal to the cell, vanadyl phthalocyanin dye, which has an absorption maximum at wavelength $\lambda = 805$ nm, was dissolved in the liquid crystal. The dye concentration in the liquid crystal was 1 wt %, which gives optical density $D = 1.05$ at the wavelength of the radiation from the semiconductor laser.

The procedure for calibrating a measurement system using a ferroelectric liquid crystal with known spontaneous polarization was used to calculate the absolute value of the pyroelectric coefficient $\gamma(T)$ from the measured pyroelectric voltage $U_p = i_p R_L$.

4. RESULTS AND DISCUSSION

Figure 2 shows the temperature dependences of the pyroelectric voltage which were measured by cooling samples with planar and homeotropic boundary conditions. The vertical line in the figure marks the isotropic–nematic phase transition temperature. One can see in the figure that the pyroelectric signal is different from zero in the nematic and in the isotropic phases of the liquid crystal. Figure 3 shows the temperature dependences, computed from the experimental results, of the pyroelectric coefficients in the isotropic phase of 5CB for cells with planar and homeotropic boundary conditions. The figure also shows the results of an analysis of the experimental data using the dependence (6). The corresponding parameters are presented in the table.

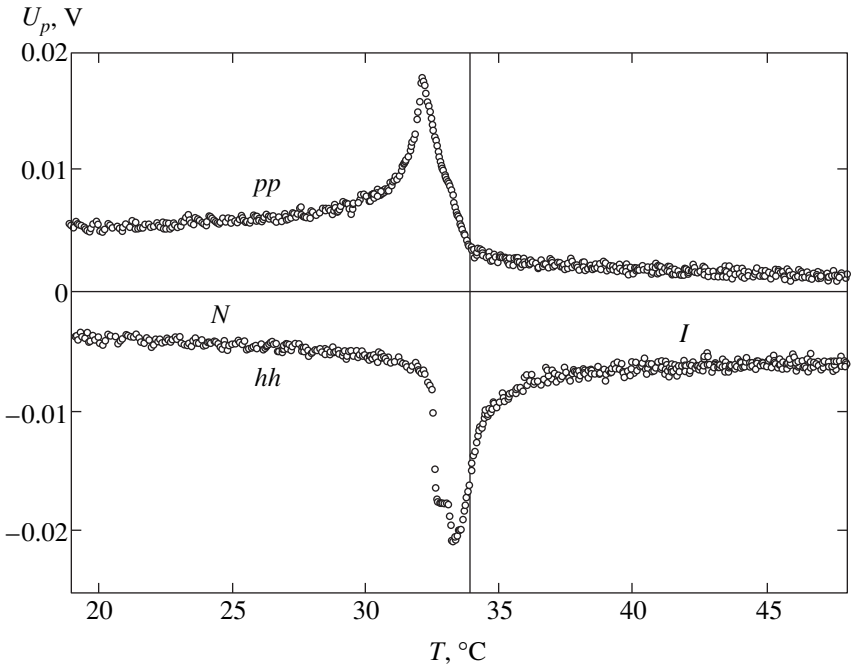


Fig. 2. Pyroelectric voltage U_p versus temperature for cells with planar (*pp*) and homeotropic (*hh*) boundary conditions: (*I*) isotropic phase, (*N*) nematic phase.

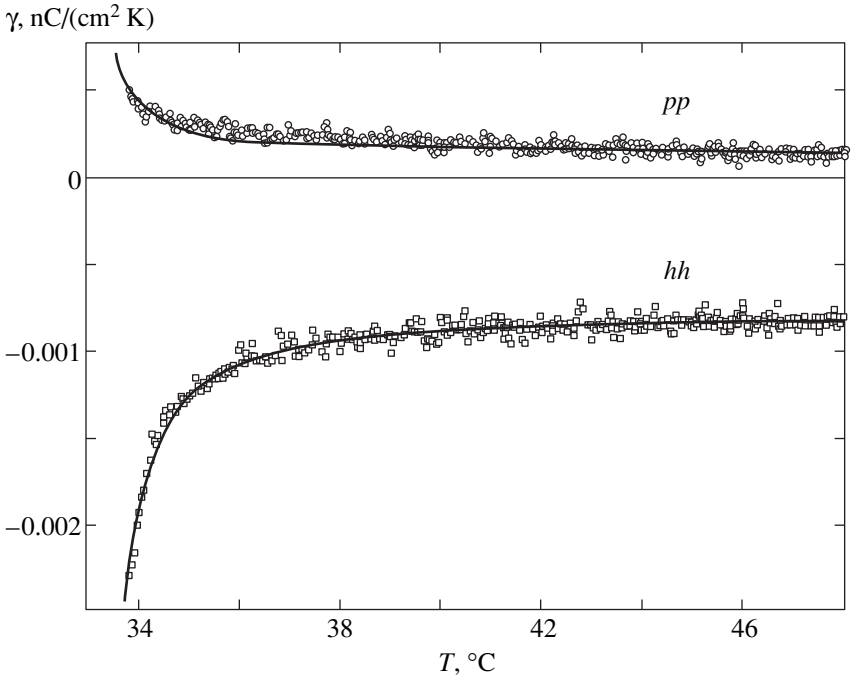


Fig. 3. Pyroelectric coefficient γ versus temperature in the isotropic phase of 5CB for cells with planar (*pp*) and homeotropic (*hh*) boundary conditions. The solid lines show the results of the analysis of the experimental data.

In the isotropic phase, far from the nematic–isotropic liquid transition temperature, when transient processes leading to orientational ordering of the liquid-crystal molecules had still not appeared, the pyroelectric coefficients are essentially constant in magnitude. This means that the corresponding surface polarization varies linearly with temperature, as, for example, the density of the liquid crystal.

Table

Type of cell	γ_{mj}, nC/(cm^2 K)	γ_{opj}, nC/(cm^2 K)	T^*, °C
Planar	1.3×10^{-4}	4.6×10^{-4}	32.7
Homeotropic	-8.2×10^{-4}	-1.5×10^{-3}	32.8

We shall determine the polarization of the surface monolayer of the liquid crystal of thickness d_m as the ratio of its dipole moment to the volume of the monolayer Ad_m, where A is the area of the monolayer. If it is assumed that all longitudinal dipoles in the monolayer are oriented in the same direction at the angle θ_m, then the normal component of the dipole moment of the monolayer will be $D_m = \Delta N m \cos\theta_m$, where m is the dipole moment of the molecule and ΔN is the number of molecules in a monolayer. In this case the polarization P_m of the monolayer will be $D_m/Ad_m = (\Delta N/Ad_m)m\cos\theta_m$. The number of liquid-crystal molecules in the monolayer is $\Delta N = A/\sigma$, where σ is the area per molecule. Then the surface density of molecules in the monolayer is $N_m = 1/d_m\sigma$, and the polarization of the monolayer, averaged over the cell thickness, is

$$\langle P_m \rangle = (N_m m \cos\theta_m) d_m / d.$$

Assuming $m \sim 5$ D $= 1.6 \times 10^{-27}$ C cm, $d_m \sim 2$ nm, and $\sigma = 20$ Å^2 and $\theta_m = 10°$ under homeotropic boundary conditions and $\sigma = 60$ Å^2 and $\theta_m = 80°$ under planar boundary conditions, we estimate the polarization of the monolayer $\langle P_{mp} \rangle \sim 0.05$ nC/cm^2 for planar and $\langle P_{mh} \rangle \sim 0.8$ nC/cm^2 for homeotropic boundary conditions. If the temperature dependence of the surface polarization is determined primarily by the temperature variation of the density of liquid-crystal molecules in the monolayer and the temperature coefficient $(1/N_m)(dN_m/dT)$ is of the same order of magnitude as for the density of molecules in the volume of the liquid crystal, i.e., $\sim 4 \times 10^{-4}$, then the pyroelectric coefficient of the monolayer is $\gamma_{mp} \sim 0.2 \times 10^{-4}$ nC/(cm^2 K) and $\gamma_{mh} \sim 3.2 \times 10^{-4}$ nC/(cm^2 K), respectively, for the planar and homeotropic cases. Thus the computed values of the pyroelectric coefficients for the surface monolayer agree in order of magnitude with the experimental results (see table). The best agreement is obtained for the homeotropic cell.

Near a nematic–isotropic liquid phase transition the pyroelectric coefficient increases sharply with decreasing temperature, especially in a homeotropic cell. As follows from Eq. (6), this is due to the increase in the ordering polarization as a result of an increase in the surface order parameter S_0 (4), induced by the interaction of the liquid-crystal molecules with the orienting surface. The fact that the pyroelectric coefficient has the opposite sign shows that the ordering polarization vectors are oppositely directed at the interface for planar and homeotropic orientations. This means, taking account of Eq. (5), that the surface order parameter S_0 relative to the direction of the normal to the surface for planar and homeotropic orientations have opposite signs. This result agrees with the results of ellipsometric measurements [18] of the surface order parameter in the isotropic phase of a NLC for different types of orienting substituents. As follows from Eqs. (4) and (6), the signs and magnitudes of the surface order parameter S_0 and the pyroelectric coefficient of the ordering polarization are determined by G, the interaction energy of the liquid crystal with the surface. From the results of the analysis of the experimental data presented in the table, we obtain $|G_h/G_p| \approx 3$ for the ratio of the moduli of the interaction energy of a liquid crystal with the surface for homeotropic and planar orientations.

The ordering polarization coefficient r in the isotropic phase of 5CB can be estimated directly from the results obtained. Assuming the interaction energy G of the liquid crystal with the surface to be 4×10^{-5} J/m^2 [19, 20], $a = 2.2 \times 10^5$ J/(m^3 K), and $L = 4.5 \times 10^{-12}$ J/m [21] we obtain for the ordering polarization coefficient $r = 1.5 \times 10^{-11}$ C/m or 4.5×10^{-4} CGSE for homeotropic orientation and $r = 0.5 \times 10^{-11}$ C/m or 1.5×10^{-4} CGSE for planar orientation. Thus, the ordering polarization coefficient is estimated to be $r = (1.0 \pm 0.5) \times 10^{-11}$ C/m or $(3.0 \pm 1.5) \times 10^{-4}$ CGSE, which is indeed comparable to the flexoelectric coefficients $(e_1 + e_3) \sim 5 \times 10^{-11}$ C/m [22].

ACKNOWLEDGMENTS

We thank L.M. Blinov for assisting in setting up the experimental apparatus and for helpful discussions, S.P. Palto for providing the PhysLab4.0 software and consultations concerning its use, and E.A. Luk'yants for providing the dye. This work was supported by the Russian Foundation for Basic Research (project no. 98-02-17071).

REFERENCES

1. L. M. Blinov, E. I. Kats, and A. A. Sonin, Usp. Fiz. Nauk **152**, 449 (1987) [Sov. Phys. Usp. **30**, 604 (1987)].
2. O. D. Lavrentovich, V. G. Nazarenko, V. M. Pergamenshchik, *et al.*, Zh. Éksp. Teor. Fiz. **99**, 777 (1991) [Sov. Phys. JETP **72**, 431 (1991)].
3. J. D. Parson, Phys. Rev. Lett. **41**, 877 (1978).
4. G. Barbero, Z. Gabbasova, and M. A. Osipov, J. Phys. II **1**, 691 (1991).
5. G. Barbero and A. G. Petrov, J. Phys.: Condens. Matter **6**, 2291 (1994).
6. K. Miyano, Phys. Rev. Lett. **43**, 51 (1979).
7. P. S. Pershan and J. Als-Nielsen, Phys. Rev. Lett. **52**, 759 (1984).
8. G. P. Crawford, R. Stannarius, and J. W. Doane, Phys. Rev. A **44**, 2558 (1991).
9. P. Guyot-Sionnest, H. Hsiung, and Y. R. Shen, Phys. Rev. Lett. **57**, 2963 (1986).

10. W. Chen, M. Feller, and Y. R. Shen, Phys. Rev. Lett. **63**, 2665 (1989).
11. A. N. Chuvyrov and A. N. Lachinov, Kristallografiya **25**, 795 (1980) [Sov. Phys. Crystallogr. **25**, 455 (1980)].
12. A. P. Krekhov, O. A. Scaldin, and A. N. Chuvyrov, Mol. Cryst. Liq. Cryst. **212**, 245 (1992).
13. A. Derzhanski, A. G. Petrov, and M. D. Mitov, J. Phys. (Paris) **39**, 273 (1978).
14. J. Prost and J. P. Marcerou, J. Phys. (Paris) **38**, 315 (1977).
15. G. Barbero, I. Dozov, I. Palierne, and G. Durand, Phys. Rev. Lett. **56**, 2056 (1986).
16. P. Sheng, Phys. Rev. A **26**, 1610 (1982).
17. M. I. Barnik, V. V. Lazarev, E. E. Maltzev, and N. M. Shtykov, Mol. Mater. **6**, 129 (1996).
18. J. Tarczon and K. Miyano, J. Chem. Phys. **73**, 1994 (1980).
19. G. P. Crawford, D. K. Yang, S. Zumer, *et al.*, Phys. Rev. Lett. **66**, 723 (1991).
20. T. Moses and Y. R. Shen, Phys. Rev. Lett. **67**, 2033 (1991).
21. T. Moses, Mol. Cryst. Liq. Cryst. **319**, 121 (1998).
22. L. M. Blinov, M. Ozaki, and K. Yoshino, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 220 (1999) [JETP Lett. **69**, 236 (1999)].

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 1, 2000, pp. 132–143.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 1, 2000, pp. 149–162.

SOLIDS
Structure

Application of the Principle of Maximum Entropy Production to the Analysis of the Morphological Stability of a Growing Crystal

L. M. Martiouchev*, **, V. D. Seleznev*, and I. E. Kuznetsova**

**Ural State Technical University, Yekaterinburg, 620002 Russia*
***Institute of Industrial Ecology, Ural Division, Russian Academy of Sciences, Yekaterinburg, 620219 Russia*
**, **e-mail: martiouchev@mailcity.com; mlm@ecko.uran.ru*

Received December 20, 1999

Abstract—The morphological stability of spherical and cylindrical crystals and an infinite plane growing from a supersaturated solution is studied using the principle of maximum entropy production in the Mullins and Sekerka approximation. In contrast to the first two geometries, the computational results for a plane agree completely with the results obtained on the basis of the classical linear perturbation theory. The concept of the binodal of a morphological transition is introduced in order to interpret the results for the sphere and cylinder. The boundaries of the metastable region are investigated. Morphological phase diagrams of stable–unstable growth are presented in terms of the variables surface energy and supersaturation as well as the variables crystal size and supersaturation. The physical nature of the appearance of metastability in this system is discussed. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Questions concerning structure formation during continuous growth of a crystal are attracting a great deal of attention in connection with their theoretical and practical importance. Even though great progress has been made in this field after the publication of the classic works [1] and [2], many problems remain unresolved. We note here only two facts directly concerning the subject of the present paper.

1. A complete understanding of the relation between the conventional stability analysis (see, for example, [2, 3]) and the general principles of nonequilibrium thermodynamics is absent in the literature on crystallization. In most cases these approaches are either contrasted to one another or they are developed independently. Indeed, according to the theoretical works [4, 5] a fundamental principle in the development of a nonequilibrium system is the principle of maximum production of entropy. This principle can be formulated as follows: an arbitrary nonequilibrium system evolves toward a locally equilibrium state at an extremal rate (with maximum entropy production). Since this principle is important for the exposition, we shall discuss it in somewhat greater detail. The principle of the maximum rate of entropy production is proposed in [4] as a basis for the deductive construction of the thermodynamics of nonequilibrium processes. In this differential principle the fluxes which for fixed thermodynamic forces maximize entropy production are sought under quite general assumptions concerning the relations between the thermodynamic fluxes and forces. It is shown that this principle makes it possible to obtain all of the basic equations and laws of nonequilibrium thermodynamics and, specifically, other well-known variational formulations—the principles of Onsager, Biot, and Prigogine [4]. In [5] the apparatus of the phenomenological thermodynamics of irreversible processes is analyzed critically and another variational principle generalizing the preceding principles is advanced on the basis of the principle of local equilibrium. One consequence of the approach proposed in [5] is the confirmation of the extremal character of the rate of a transition of a dynamical system to a locally equilibrium state.

The principle studied in the present paper reduces, as will be shown below, in a particular case to the principle of the maximum possible local rate of growth of a crystal. However, the attitude of scientists concerned with crystallization toward the principle of maximum rate is extremely contradictory and changes with time. This is due primarily to the fact that the works [4, 5] have remained unnoticed, and this principle arose purely intuitively in the theory of crystal growth. The history of the application of the maximum-rate principle, specifically, with respect to the growth of a dendrite, is briefly as follows. Temkin seems to have been one of the first to use this principle in analytical calculations [6]. This principle was a criterion for selecting a definite solution from an entire family of possible solutions obtained in a phenomenological model. However, the experimental work [7] showed a substantial difference from the result obtained in [6] and the theoretical works [8, 9], based on a linear analysis of the stability of a growing paraboloid assuming isotropic surface tension, conversely, led to good agreement with exper-

iment. As a result the latest theory [8, 9], which has been termed the theory of marginal stability, is contrasted with the principle of maximum rate of growth and, in consequence, the attitude developed that this principle is incorrect. However, in the opinion of the authors the discrepancy between the theory based on a maximality principle for selecting a definite solution and experiment could indicate, first and foremost, the coarseness of the model used in [6] as a basis for this theory. Approximately eight years later there appeared theoretical works indicating contradictions in the theory of marginal stability itself, specifically, the absence of a stationary solution corresponding to a needle-shaped dendrite. This led to the development of an improved theory in which weak anisotropy of surface tension was introduced (see the reviews [10, 11]). The new approach, known as the theory of microscopic solvability, also employs stability analysis. One result of this analysis was the assertion that a solution with the maximum rate of growth is the only stable solution from a discrete spectrum of stationary needle-like solutions. As the theory of microscopic solvability continued to advance, the problems of dendritic growth appeared to be largely solved. However, an experiment performed in an anisotropic Hele-Shaw cell and theoretical calculations in a boundary-layer model revealed a new problem: dendrites are not always observed in the case of weak anisotropy with decreasing supercooling/supersaturation [12–14]. As a result the tip of a dendrite splits. In [12–14], in order to get out of this contradiction, it is proposed that the criterion of solvability must be replaced by a more general criterion: the dynamically selected morphology is the most rapidly growing morphology. In other words, if the existence of more than one morphology is possible, only the most rapidly growing morphology is nonlinearly stable and therefore observable.

In summary, during the 30 years of study of the nonequilibrium crystal growth two approaches to the problem of morphology selection competed with one another: stability analysis and application of a maximality principle. Both principles are intuitively quite plausible. Even though the stability analysis in many cases leads to conclusions that the rate of growth is maximum, it seems obvious that in most cases each approach will lead, if not to qualitative then to quantitative differences. In our opinion, the search for the more "correct" approach is not promising; they should not be contrasted to one another, but rather an attempt should be made to find a logical relationship between them.

2. A large number of experimental and computer simulation works show that there exist ranges of the parameters where different morphologies can coexist [12–26]. Thus, in [22] two different types of disturbances started to develop simultaneously during crystallization of a melt of succinonitrile with additions of acetone, in the presence of definite supercooling, on a crystalline nucleus and coexistence of two forms of growth was observed: dendrites growing in the $\langle 100 \rangle$ and $\langle 111 \rangle$ directions. It is also remarked in the literature that the transition from one morphology to another accompanying a change in the parameters (for example, the supersaturation) can occur both with a jump in the rate or with a jump in its derivative [13–18]. On this basis an analogy is often drawn between phase and morphological diagrams and the concepts of first- and second-order morphological transitions are introduced [13–17, 27]. The most fundamental question arising in this formulation of the problem is finding the principle for selecting a possible morphology which would enable a complete calculate of a morphological diagram (with boundaries of metastable and labile regions). However, this problem has not been solved. In the literature it is hypothesized that far from equilibrium entropy production determines morphological selection but no specific calculations have been performed [13–15].

Thus, the goal of the present work can be formulated on the basis of the forgoing analysis (paragraphs 1 and 2): to study, using the principle of maximum entropy production, the problem of morphological selection during nonequilibrium growth of a crystal and to show, using the concept of morphological diagrams, the relationship between this approach and the stability analysis. For clarity this analysis is performed for the simplest problems: growth of a sphere, an infinite cylinder, and an infinite plane from a solution in the classic Mullins–Sekerka (MS) approximation [2].

2. PRINCIPLE OF MAXIMUM ENTROPY PRODUCTION IN CRYSTALLIZATION PROBLEMS

We shall study isothermal–isobaric crystallization as the most typical case and we shall assume that the growing crystal completely displaces the solvent. It is well known [28] that local entropy production σ for the system under study equals

$$\sigma = I\nabla\mu, \tag{1}$$

where $\nabla\mu$ and I are, respectively, the gradient of the chemical potential and the flux of the crystallizing component. The expression (1) can be used for any elements of the volume under study and, specifically, for the region near the surface of a growing crystal. In this case

$$I = D(\nabla C)_{\text{int}} = (C - C_{\text{int}})V, \tag{2}$$

where C is the density of the crystal, C_{int} is the concentration of the dissolved substance near a surface of arbitrary type, V is the local growth rate, and D is the diffusion coefficient.

We note that the entropy production (1) and (2) is proportional to the rate of increase of the mass of the growing crystal or, specifically, the velocity of the boundary. At the same time it contains an additional factor equal to the gradient of the chemical potential. Based on the investigations performed in [4, 5], the fol-

lowing local principle (maximality principle) can be formulated for the system studied in this work: when disturbances with adequate amplitude are present in the system, a state characterized by maximum local entropy production is realized. We note that this assertion generalizes the assertion, made in the crystallization literature, that the crystal growth rate [12–14] and the rate of mass increase [15] are maximum. We shall make several important remarks concerning this principle.

1. The principle formulated seems, at first glance, to contradict the well-known Prigogine principle of minimum entropy production. However, the regions of application of these principles are absolutely different. Thus, the Prigogine principle is valid for weakly nonequilibrium stationary processes occurring under fixed boundary conditions [28], whereas the principle of maximum entropy production concerns a nonequilibrium system moving toward equilibrium or a stationary state; a self-developing nonequilibrium system is considered and not all boundary conditions can be constant. Correspondingly, these principles concern different types of variational problems (in Prigogine's formulation the fluxes at a boundary are not varied), but, as shown in [4] for nonlinear processes, the Prigogine principle is a consequence of the principle of maximum entropy production.

2. The principle formulated above is local. It is obvious that when this principle is satisfied at each point of the system the principle of maximum entropy production will be valid even for the entire system studied. As shown in [29], entropy production for an isothermal–isobaric process is directly proportional to the change in the Gibbs potential, taken with the opposite sign. It is well known that the system under study, developing spontaneously, moves toward a minimum of the Gibbs potential, so that on the basis of the principle of maximum entropy production it can be concluded that this occurs, when possible, at the maximum rate. Thus, the maximum principle formulated generalizes the principles proposed in [30–32].

3. In the special literature devoted to crystal growth it is asserted that a crystal growing in a stationary manner is edged by the most slowly growing faces [33]. This is a consequence of the so-called wedging out of the faces, where on account of this growth the most rapidly growing face becomes degenerate, and as a result two neighboring faces remain slowly growing. This is not a rejection of the general maximality principle, since the principle talks about the manner in which a system moves toward an equilibrium or a stationary state. The opposite could signify rejection of this principle—the development at each stage of growth only of the most slowly growing faces in the presence of faces capable of more rapid growth.

The consequences of the maximality principle will be analyzed below for various crystal growth geometries.

3. ENTROPY PRODUCTION AND MORPHOLOGICAL SELECTION DURING THE GROWTH OF A SPHERE

We shall study diffusion-controlled growth of a spherical particle from a supersaturated solution with initial concentration C_∞. In the MS approximation it is assumed that local equilibrium exists near each element of a phase boundary. The crystallographic factors are neglected. It is assumed that the diffusion field is described by the Laplace equation, i.e., the condition of quasistationary diffusion is satisfied:

$$\left|\frac{C_\infty - C_{\rm int}}{C - C_{\rm int}}\right| \le \left|\frac{C_\infty - C_0}{C - C_0}\right| \ll 1, \tag{3}$$

where C_0 is the concentration of dissolved matter near a flat boundary. The behavior of an infinitesimal distortion of a sphere by a single spherical harmonic $Y_{lm}(\theta, \varphi)$ is investigated in this approximation. The equation of the distorted surface of the sphere has the form

$$r(\theta, \varphi, t) = R(t) + \delta(t) Y_{lm}(\theta, \varphi), \quad \delta(t) \ll R(t), \tag{4}$$

where R is the radius of the unperturbed particle, δ is the amplitude of the disturbance, and t is the time. The solution of the Laplace equation is constructed under the assumption that the equilibrium concentration near the surface of the crystal satisfies the equation

$$C_{\rm int} = C_0 + C_0 \Gamma K, \tag{5}$$

where Γ is the surface tension (proportional to the free surface energy), and K is the average curvature. Using a more general solution of the Laplace equation for a weakly distorted particle and neglecting all powers of δ above the first power gives [2]

$$V = \dot R + \dot\delta Y_{lm}, \tag{6}$$

$$\dot\delta(t) = \frac{C_0 D(l-1)}{(C - C_R)R^2} \times \left\{\frac{C_\infty - C_0}{C_0} - \frac{\Gamma}{R}[(l+1)(l+2)+2]\right\}\delta, \tag{7}$$

where

$$C_R = C_0(1 + 2\Gamma/R), \quad \dot\delta \equiv d\delta/dt,$$

$$\dot R \equiv dR/dt \quad l = 1, 2, 3, \ldots$$

It follows from (7) that disturbances will grow if the radius of the crystal is greater than the critical radius R_s^s [2]:

$$R_s^s = [(l+1)(l+2)/2 + 1]R_s^*, \tag{8}$$

where $R_s^* = 2\Gamma C_0/(C_\infty - C_0)$ is the radius of the critical nucleus (here and below the subscript "*s*" denotes the sphere).

The formulas (7) and (8) obtained above completely determine the stability of the growing spherical particle with respect to an infinitesimal disturbance [2].

We shall now apply the thermodynamic approach to the analysis of this problem. Using (1) and (2), we shall find the difference between the entropy productions between the growth of perturbed (Σ_{sp}) and unperturbed (Σ_{sn}) spherical crystals; in so doing, we shall calculate the local energy production per unit time for a volume element with unit thickness and area cut out by the solid angle $d\Omega$ near the crystal surface. For definiteness, we shall use the ideal-solution approximation, i.e., $\nabla\mu \sim (\nabla C)_{\text{int}}/C_{\text{int}}$. As a result we obtain

$$\Delta\Sigma_s \equiv \Sigma_{sp} - \Sigma_{sn} = (\sigma_p r^2 - \sigma_n R^2)d\Omega$$
$$\sim \left[\frac{(C - C_{\text{int}})^2}{C_{\text{int}}} V^2 r^2 - \frac{(C - C_R)^2}{C_R}\dot{R}^2 R^2\right] d\Omega.$$

Using (3)–(7) and the fact that the rate of growth of the unperturbed sphere is determined by the formula [2]

$$\dot{R} = \frac{D(C_\infty - C_R)}{R(C - C_R)}$$

we obtain

$$\Delta\Sigma_s \sim \left\{ l - \left[\frac{(l+1)(l+2)+2}{2}(l-1) + 1\right]\frac{R_s^*}{R} - \left[\frac{(l-1)(l+2)}{4}\frac{C_\infty - C_R}{C_R}\right]\frac{R_s^*}{R}\right\}\delta Y_{lm} d\Omega. \quad (9)$$

Assuming the relative supersaturation to be small, the expression (9) can be rewritten as

$$\Delta\Sigma_s \sim (R\dot{\delta} + \delta\dot{R})d\Omega$$
$$\sim \left\{ l - \left[\frac{l(l+1)^2}{2} - 1\right]\frac{R_s^*}{R}\right\}\delta Y_{lm} d\Omega. \quad (10)$$

We now choose a direction (θ, φ) corresponding to the maximum value of $Y_{lm}(\theta, \varphi)$. The most dangerous (from the standpoint of disruption of spherical growth) state obtains for this direction. Using the expression (10) it can be shown that for the direction under study in the range of possible values of the radius of the sphere $[R_s^*, R_s^s]$ the function $\Delta\Sigma_s$ is positive for $R > R_s^b$:

$$R_s^b = R_s^* \frac{l^3 + 2l^2 + l - 2}{2l}. \quad (11)$$

A plot of the function $\Delta\Sigma_s(R)$ for various perturbing harmonics is presented in Fig. 1.

Thus, the difference between the entropy production between the growth of perturbed and unperturbed spherical crystals changes sign at the point R_s^b, which is different from the point $R = R_s^s$ obtained using the Mullins–Sekerka theory. The difference ($R_s^b < R_s^s$) arose because the entropy production for the volume element under study depends not only on the linear rate of growth of the crystal but also on the change in the area of the crystal (Fig. 2)—this reflects the second term in (10), which is proportional to the amplitude δ of the disturbance. In accordance with the Mullins–Sekerka theory, for $R = R_s^s$ the crystal becomes unstable with respect to an infinitesimal disturbance, and in the terminology of equilibrium thermodynamics this

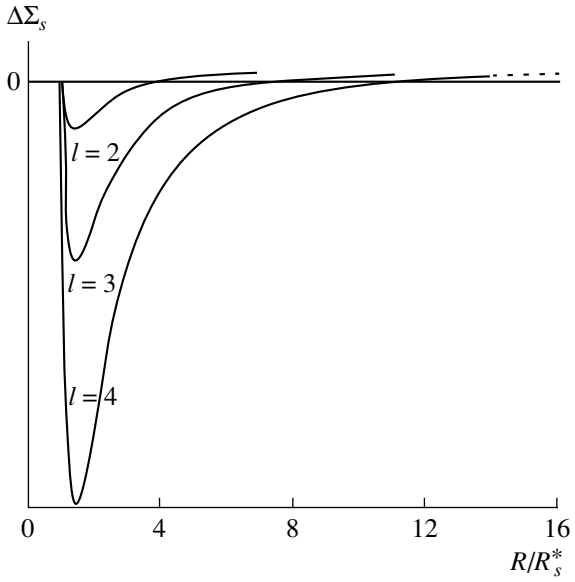


Fig. 1. Plot of $\Delta\Sigma_s$ as a function of the relative size of the sphere R/R_s^* for different perturbing harmonics, $l = 2, 3, 4$.

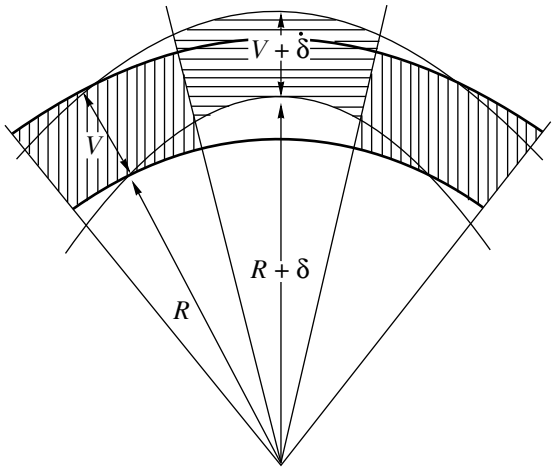


Fig. 2. Schematic diagram of the perturbed surface of a sphere (the central solid angle contains a convexity, and the peripheral solid angles contain an unperturbed surface).

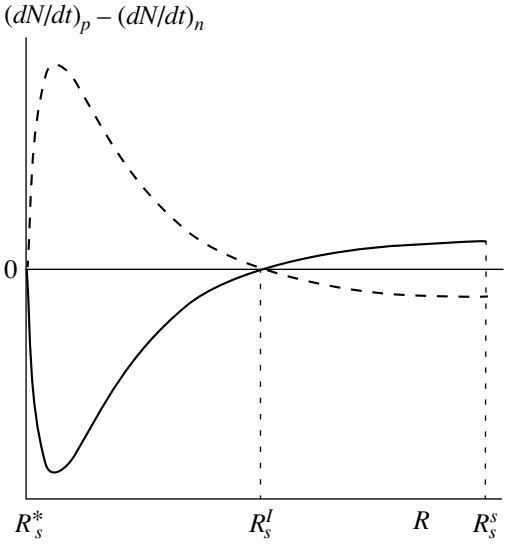


Fig. 3. Difference between the fluxes during growth of perturbed (containing a convexity, solid line, or dip, dashed line) and unperturbed sections of a sphere versus the size of the crystal.

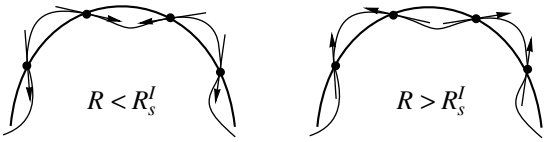


Fig. 4. Directions of tangential fluxes before and after the critical radius R_s^I.

point can be called the spinodal of a nonequilibrium transition. According to the Mullins–Sekerka theory, in the interval $[R_s^b, R_s^s)$ the growing sphere is stable with respect to infinitesimal perturbations. However, in this interval entropy production in the volume element under study is greater in the presence of the disturbance, and therefore growth of a distorted sphere is preferable on the basis of the principle of maximum entropy production. This contradiction vanishes if it is assumed that the particle is metastable, unstable with respect to certain small but finite disturbances. On this basis we shall call the region $[R_s^b, R_s^s)$ metastable and the point R_s^b the binodal of a morphological transition.

As shown above the existence of metastable behavior for a growing particle is a logical consequence of the results of the simultaneous application of the principle of maximum entropy production and linear stability analysis. However, aside from this deductive approach, the origin of metastability in this system can be explained differently on the basis of an analysis of the tangential fluxes of crystallizing component which arise near the surface of the sphere. For this, we shall consider the difference between the increment to the mass of the crystal (or, in other words, the flux of matter from the solution to the crystal surface) in the perturbed $(dN/dt)_p$ and unperturbed $(dN/dt)_n$ cases. We shall calculate this change per unit time for a volume element of unit thickness with area cut out by the solid angle $d\Omega$ near the crystal surface. A calculation similar to the one performed in the preceding case gives

$$\begin{aligned}\left(\frac{dN}{dt}\right)_p - \left(\frac{dN}{dt}\right)_n &= [(C - C_{\text{int}})Vr^2 - (C - C_R)\dot{R}R^2]d\Omega \\ &\sim (2\dot{R}\delta + R\dot{\delta})Y_{lm}d\Omega \\ &\sim \left[l + 1 - \frac{R_s^* l}{2R}(l^2 + 2l + l)\right]\delta Y_{lm}d\Omega.\end{aligned} \tag{12}$$

This difference is positive for $R > R_s^I$:

$$R_s^I = R_s^* l(l-1)/2. \tag{13}$$

We shall assume that the perturbed and unperturbed surface elements belong not to different particles, as assumed previously, but rather to the same particle and a convexity (where the value of $Y_{lm}(\theta, \varphi)$ is maximum) is present on one of the solid angles $d\Omega$, while a surface close to the unperturbed surface is present on the other (Fig. 2). Then, starting with the crystal size R_s^I, the flux of matter per solid angle containing the perturbed section of the surface is greater than the analogous flux toward the unperturbed section. It is evident that for dips (where $-|Y_{lm}(\theta, \varphi)|$ is minimum) the situation is exactly opposite: the flux of matter toward dips will exceed the flux to the unperturbed parts of the surface for $R < R_s^I$ and will be less than the flux for $R > R_s^I$. All this is illustrated in Fig. 3. In consequence, the tangential fluxes of the crystallizing component in the system up to the point R_s^I are directed from convexities toward dips and thereby decrease the disturbance, while at a radius greater than R_s^I the tangential fluxes are, conversely, directed toward convexities and thereby cause them to grow (see Fig. 4). This result is also confirmed by direct differentiation with respect to the angle $i = (\varphi, \theta)$ of the concentration field near the perturbed growing spherical crystal, found in [2]:

$$\left.\frac{\partial c}{\partial i}\right|_{r=R} = -\delta\frac{(C_\infty - C_0)R - C_0\Gamma l(l+1)}{R^2}\frac{\partial Y_{lm}}{\partial i}.$$

It follows directly from this expression that the direction of the tangential fluxes first changes for a sphere of size R_s^I away from dips toward peaks in the perturbed surface.

Thus, when the radius of the sphere reaches R_s^I a new destabilizing force—a tangential flux of matter—

capable of developing the disturbance appears near the surface of the sphere. As follows from (12), this flux is proportional to the amplitude of the perturbing harmonic. It is also interesting to note that for R close to R_s^I the expression (12) is formally different from zero only for the harmonic amplitude approaching infinity. These factors motivated the introduction of the concept of a metastable region and a binodal of a morphological transition. The behavior of fluxes of matter near the surface of a perturbed crystal, which make the decisive contribution in the calculation of the difference of the entropy productions $\Delta\Sigma_s$, was studied above. The point R_s^b lies somewhat to the right of the point R_s^I, since the expression for entropy production contains an additional factor which somewhat damps the contribution arising when only the matter fluxes are taken into account. A closer position of R_s^b to the point of the spinodal R_s^s ($R_s^I < R_s^b < R_s^s$) indicates that metastability is due primarily to the appearance of tangential fluxes. If the point R_s^b was located to the left of R_s^I, then this would indicate the existence of another thermodynamic force, which was neglected and which is capable of increasing the perturbation.

Figure 5 shows the dependence of entropy production in the volume element under study on the size of the crystal. As follows from (8) and (11), the metastable region

$$R_s^s - R_s^b = \frac{l^2 + 3l + 2}{2l} R_s^*$$

exists for any physically possible parameters. The width of this interval increases with the surface energy and decreases with increasing relative supersaturation $(C_\infty - C_0)/C_0$.

Figure 6a shows the morphological phase diagram of stable (spherical)–unstable (dendritic) growth in the variables supersaturation and surface tension. The diagram was constructed using (8) and (11) with a fixed value of the radius. The supersaturation and surface tension variables are independent of one another. The advantage of this construction is that in real experiments experimenters always work with a specific material and on a specific spatial scale ("with a fixed magnification of a microscope"). By varying the parameters (for example, the supersaturation) they observe a particular morphology (nonequilibrium phase) [20]. The diagram shows the regions of existence and coexistence of various morphologies corresponding to stable growth and growth with developing perturbations $l = 2$ and $l = 3$. We note that metastable regions referring to different harmonics do not intersect. Figure 6b shows the same morphological diagram but in the variables crystal radius and supersaturation. Diagrams of this type are convenient for analyzing the temporal evolution of a surface (with increasing R) under prescribed external conditions. The region of coexistence of different morphological phases (stable and unstable) can also be seen in the diagram.

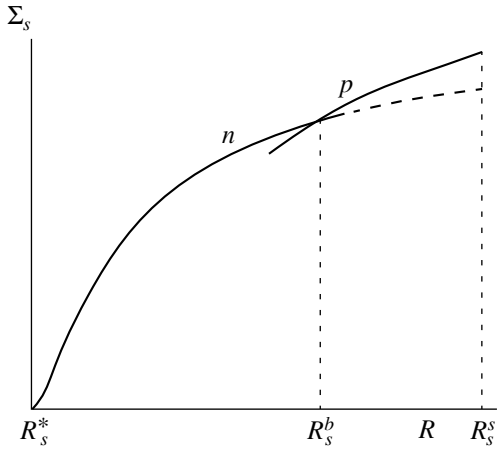


Fig. 5. Local entropy production in a distinguished volume element as a function of the radius R of the sphere. The metastable region is denoted by the dashed line (n—unperturbed growth, p—perturbed growth on the section of the surface of the sphere with a convexity, $\delta/R_s^* = 0.05$).

4. ENTROPY PRODUCTION AND MORPHOLOGICAL SELECTION DURING GROWTH OF AN INFINITE CYLINDER

We shall now study the growth of an infinite cylinder in the Mullins–Sekerka approximation. A classic linear stability analysis of this problem is given in [34]. In the present paper we investigate the behavior of an infinitesimal distortion of the shape of a particle

$$F(\varphi, z) = \cos(k\varphi)\cos(k_z z/R),$$

where k is a positive integer and k_z can assume any real positive value. Just as in the preceding case, we shall give a brief presentation of the results obtained in the present work which are required for further calculations:

$$r(\varphi, z, t) = R(t) + \delta(t)F(\varphi, z), \quad \delta(t) \ll R(t), \tag{14}$$

$$V = \dot{R} + \dot{\delta}F(\varphi, z), \tag{15}$$

$$\dot{\delta}(t) = \frac{C_0 D}{(C - C_R)R^2}\frac{1}{A_\lambda}\left\{\frac{C_\infty - C_0}{C_0}(H-1) - \frac{\Gamma A_\lambda}{R}\left[H(k^2 + k_z^2 - 1) + \frac{H-1}{A_\lambda}\right]\right\}\delta, \tag{16}$$

$$R_c^s = \left[1 + \frac{HA_\lambda}{H-1}[k^2 + k_z^2 - 1]\right]R_c^*, \tag{17}$$

where $C_R = C_0(1 + \Gamma/R)$, $R_c^* = \Gamma C_0/(C_\infty - C_0)$ is the radius of the critical nucleus of the cylinder (here and

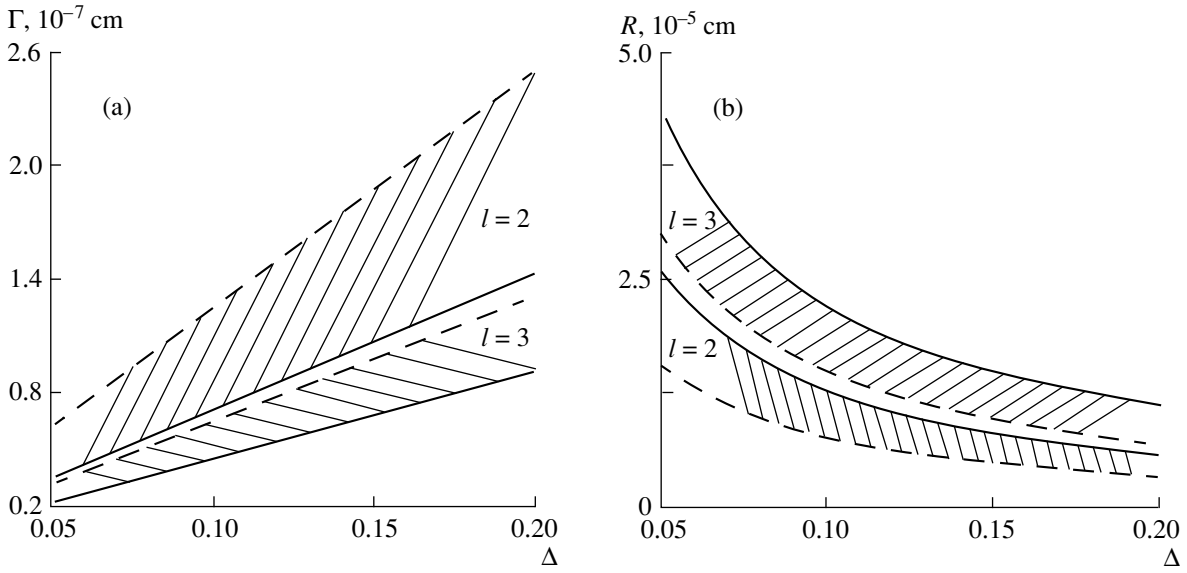


Fig. 6. (a) Morphological phase diagram of stable–unstable growth of a sphere in the variables relative supersaturation and surface tension for $l = 2, 3$. Solid lines—spinodals, dashed lines—binodals. The metastable region is hatched. Stable growth occurs above the binodal and unstable growth occurs below the spinodal. The curves were constructed for $R = 10^{-5}$ cm. (b) Radii of the binodal R_s^b (dashed lines) and spinodal R_s^s (solid lines) of a morphological transition as a function of the relative supersaturation $\Delta = (C_\infty - C_0)/C_0$ for $\Gamma = 10^{-7}$ cm, $l = 2, 3$.

below the index "c" denotes the cylinder), R_c^s is the radius at which the cylindrical crystal becomes unstable with respect to an infinitesimal disturbance of its shape,

$$H = H(k, k_z) = \frac{k_z}{2K_k(k_z)}[K_{k-1}(k_z) + K_{k+1}(k_z)],$$

$K_k(k_z)$ is a modified Hankel function, $A_\lambda = -0.5\ln(\eta^2\lambda^2)$, λ is found from the equation

$$\lambda^2\ln(\eta^2\lambda^2) + \frac{C_\infty - C_R}{C - C_R} = 0,$$

and $\ln(\eta^2)$ is the Euler constant.

Just as in the case of a sphere, we shall use the ideal-solution approximation to calculate the difference between the entropy production in cases of growth of the perturbed (Σ_{cp}) and unperturbed (Σ_{cn}) cylinders (we shall find the entropy production of an element of volume of unit thickness with area cut out by the angle $d\varphi$ and element of length dz, near the crystal surface):

$$\Delta\Sigma_c \equiv \Sigma_{cp} - \Sigma_{cn} = (\sigma_{cp}r - \sigma_{cn}R)d\varphi dz$$
$$\sim\left[\frac{(C - C_{\text{int}})^2}{C_{\text{int}}}(\dot R + \dot\delta)^2 r - \frac{(C - C_R)^2}{C_R}\dot R^2 R\right]d\varphi dz. \quad (18)$$

Using (14)–(17) and assuming the relative supersaturation to be small, the expression (18) can be written as

$$\Delta\Sigma_c \sim (2R\dot\delta + \delta\dot R)d\varphi dz \sim \left\{2H - 1 - \left[2A_\lambda\left(H(k^2 + k_z^2 - 1) + \frac{H-1}{A_\lambda}\right) + 1\right]\frac{R_c^*}{R}\right\} \times \delta F(\varphi, z)d\varphi dz. \quad (19)$$

Just as in the preceding case, we shall choose the direction (φ, z), corresponding to the state that is most dangerous from the standpoint of breakdown of cylindrical growth. For this direction, in the range $[R_c^*, R_c^s]$ of the possible values of the radius of the cylinder, the function $\Delta\Sigma_c$ is greater than zero for $R > R_c^b$:

$$R_c^b = R_c^*\left[1 + \frac{2A_\lambda H(k^2 + k_z^2 - 1)}{2H - 1}\right]. \quad (20)$$

Arguments completely similar to those presented in Section 3 show that all conclusions concerning the binodal and the metastable region for growth of a cylinder are qualitatively similar to the results obtained for a

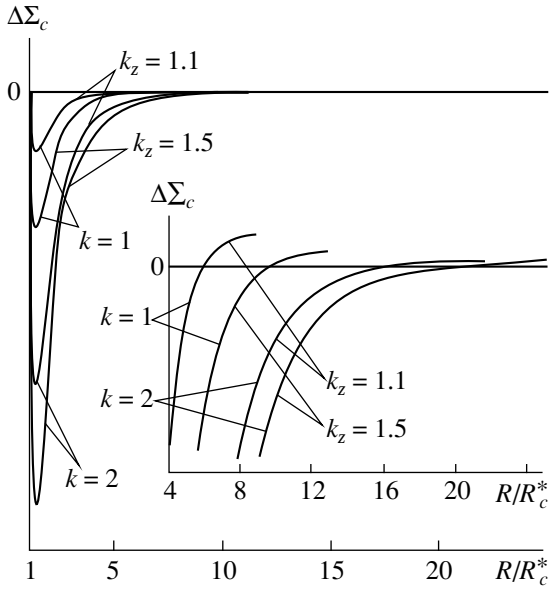


Fig. 7. Plot of $\Delta\Sigma_c$ as a function of the relative size of the cylinder R/R_c^* for $k = 1$, $k_z = 1.1$, 1.5 and $k = 2$, $k_z = 1.1$, 1.5.

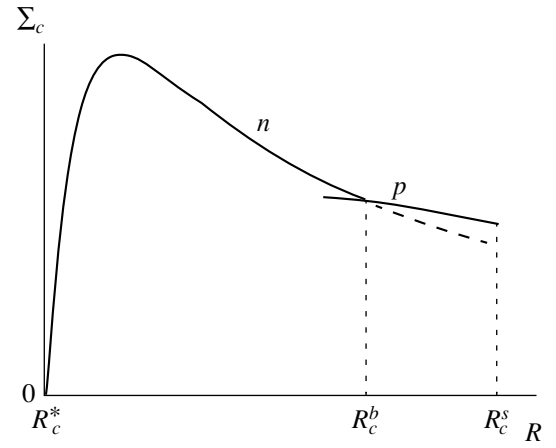


Fig. 8. Local entropy production for a distinguished volume element as a function of the radius R of the cylinder. The metastable interval is shown with dotted lines (n—unperturbed growth, p—perturbed growth on a section of the surface of the cylinder with a convexity, $\delta/R_c^* = 0.1$).

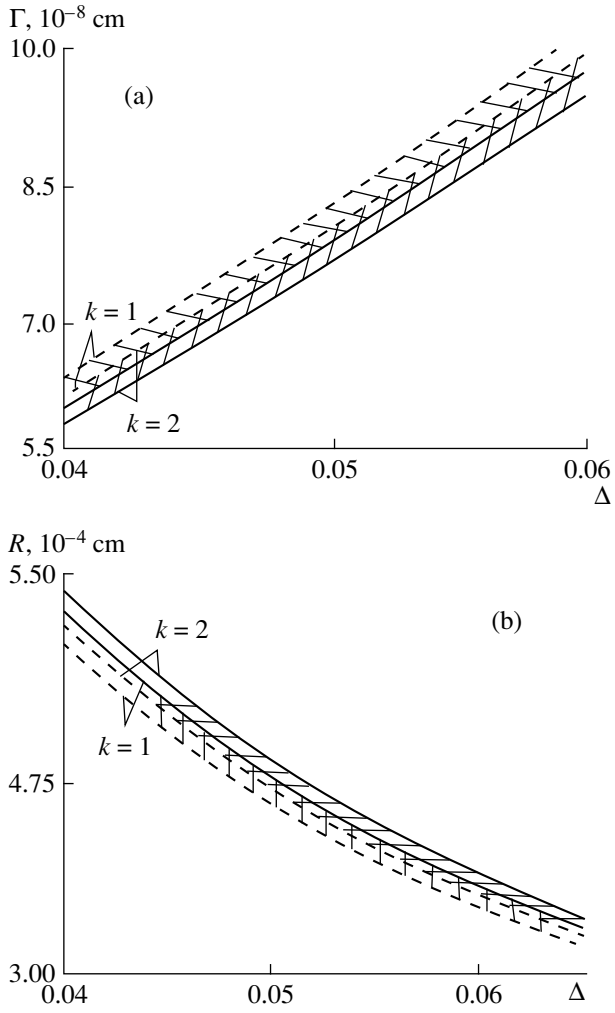


Fig. 9. (a) Morphological phase diagram of stable–unstable growth of a cylinder in the variables relative supersaturation and surface tension for $k_z = 10$, $k = 1$, 2. Solid lines—spinodals, dashed lines—binodals. The metastable region is hatched. Stable growth occurs above the binodal and unstable growth occurs below the spinodal. The curves were constructed for $R = 5 \times 10^{-4}$ cm. (b) Radii of the binodal R_c^b (dashed lines) and spinodal R_c^s (solid lines) of a morphological transition as a function of the relative supersaturation $\Delta = (C_\infty - C_0)/C_0$ for $\Gamma = 7.5 \times 10^{-8}$ cm, $k_z = 10$, $k = 1$, 2.

growing sphere. Thus, for example, the analog of the point R_s^I, the point

$$R_s^I = R_c^*[1 + A_\lambda(k^2 + k_z^2 - 1)],$$

also lies to the left of R_c^b. The analogs of Figs. 1, 5, and 6 for a cylinder are, respectively, Figs. 7, 8, and 9. As one can see by comparing Figs. 5 and 8, the curve of Σ_c in the case of cylindrical growth passes through a maximum. This difference arose because the surface of a cylinder increases in proportion to the radius, while the surface of the sphere increases in proportion to the squared radius of the crystal. We also note that for most cases (except $k = 0$, $1 < k_z < 1.2$, and $k = 1$, $k_z < 0.6$) the point R_c^b lies to the right of the extremum.

Another important difference of the cylindrical from the spherical case is the possibility of intersection of metastable regions corresponding to different perturbing harmonics. This leads to the coexistence of a large number of morphological phases. This feature can be analyzed more clearly using Fig. 10. Figure 10 shows

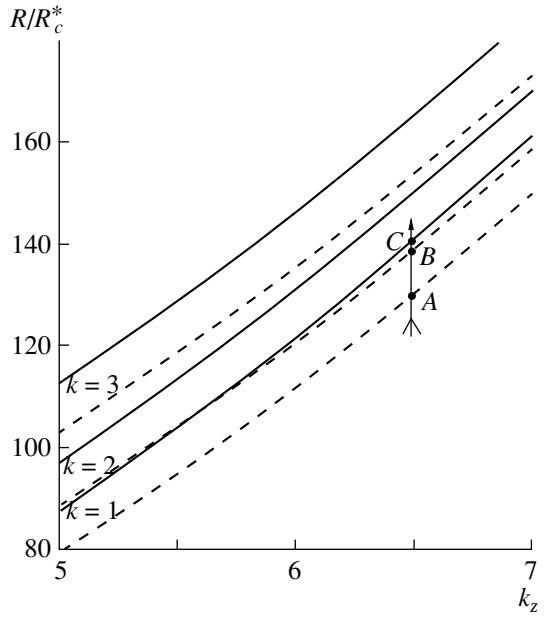


Fig. 10. Radii of the binodal and spinodal in units of R_c^* versus k_z for $k = 1$, 2, 3, and $A_\lambda = 2.8$ (corresponds to $\Delta = 0.05$). Dashed lines—binodals, solid lines—spinodals. Details are presented in the text.

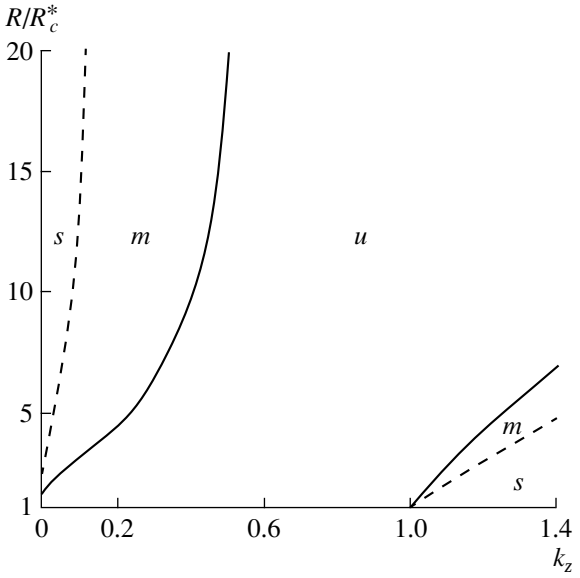


Fig. 11. Radii of the binodal and spinodal (in units of R_c^*) versus k_z in cases of long-wavelength disturbances only in the direction z ($k = 0$, $k_z \le 1$). The symbols *s*, *m*, and *u* designate, respectively, stable, metastable, and unstable cylindrical growth. Dashed lines—binodals, solid lines—spinodals.

the radii R_c^b and R_c^s in units of R_c^* as functions of k_z for $k = 1$, 2, and 3. It is evident from the figure that the binodals and spinodals for various harmonics can intersect. We shall use this plot to clarify the evolution of the system. Let us consider the growth of a crystal in the presence of a definite perturbation along the *z*-axis (for example, $k_z = 6.5$) and the possibility of the appearance of arbitrary disturbances along φ, i.e., the cylindrical crystal develops along the line *AC*. This growth is stable up to the point *A*. Crystal growth becomes metastable with respect to disturbances with $k = 1$ if the radius of the cylinder lies in the interval *AB*. In this interval instability is possible only in the presence of a perturbation with a large enough amplitude. If growth of several cylindrical crystals occurs under these conditions, some of them will continue cylindrical growth while others will become unstable. It is obvious that the fraction of stable cylindrical crystals decreases as *R* increases. At the point *B* perturbations can develop with $k = 1$ and $k = 2$. In this case, when several crystals grow simultaneously, some of them (most) will become unstable with the development of the $k = 1$ perturbation while others will also become unstable but with respect to the perturbation $k = 2$, and finally a minority will remain cylindrical. We note that inside this interval simultaneous development of perturbations with $k = 1$ and 2 is also possible on the same crystal. At the point *C* all crystals with a cylindrical shape become unstable with respect to an infinitesimal disturbance with $k = 1$. If high-frequency perturbations are present along the *z*-axis, then, as follows from (17) and (20), coexistence of an unbounded number of morphological phases is possible. Thus, for example, for $k_z = 700$ the spinodals with $k = 1$, 10, and 20 intersect 17, 11, and 7, respectively, neighboring binodals; this can result in the coexistence of 18, 12, and 8 morphological phases. We underscore that in the case of the growth of a sphere no more than two different morphologies can coexist (stable spherical and with a developing perturbation, see Fig. 6).

Following [34] and using (17) and (20) we shall consider the case (Fig. 11) of long-wavelength perturbations only in the direction *z* ($k = 0$, $k_z \le 1$).

(1) If $k_z = 1$, then $R_c^b = R_c^s = R_c^*$ and the crystal is unstable from the outset of growth.

(2) If $0.6 \le k_z < 1$, then $R_c^s < R_c^b < R_c^*$ and the cylindrical surface is always unstable. This result is similar to the result obtained using the theory of morphological stability [34]. It is explained by the fact that the surface energy and the concentration gradient favor the growth of the perturbation.

(3) According to [34], the most specific case is $k_z < 0.6$, when the concentration gradient in (16) changes sign and therefore leads to a decrease of the perturbations, while the surface energy favors growth. As a result, for $k_z < 0.6$ a cylindrical surface is unstable for $R < R_c^s$ and stable for $R > R_c^s$ [34]. Our calculation in this case also showed the presence of a metastable region separating stable and unstable regions ($R_c^s < R < R_c^b$).

For some perturbations of a special form ($k = 1$ and $k_z < 1$) the metastable interval $[R_c^b, R_c^s]$ can contain the maximum of the function $\Delta\Sigma_c$. Therefore, as the crystal grows from R_c^b to R_c^s instability at first increases and then decreases, i.e., reentrant behavior occurs.

Another special perturbation is $k = 0$, $k_z \gg 1$. In this case $H(k, k_z) \simeq k_z + 0.5$ [34] and $R_c^s \longrightarrow R_c^b$, i.e., the metastable region vanishes. The specific nature of this perturbation makes the problem of the growth of an infinite cylinder similar to the problem of the growth of an infinite plane, which is considered below.

5. ENTROPY PRODUCTION AND MORPHOLOGICAL SELECTION DURING GROWTH OF AN INFINITE PLANE

We shall now consider the growth of a plane in the Mullins–Sekerka approximation. This case is fundamentally different from the two preceding cases in that the stationary approximation for a plane leads to a constant velocity v of the plane (previously the velocity was inversely proportional to the size of the crystal) and conservation of the area of any distinguished surface element in time (previously the surface area within a certain solid angle increased).

For simplicity, here we shall discuss only the one-dimensional problem (the solution for the two-dimensional problem is similar but more complicated). An example is crystallization from solution in a cell between two plates separated by a quite small distance. In [2, 9] a linear stability analysis is performed for a growing plane with a superposed disturbance

$$\xi = \delta[\sin(kx) + \cos(kx)], \tag{21}$$

where $\delta = \hat{\xi} e^{\omega t}$ is the amplitude of the disturbance, $\hat{\xi} \ll 1$, and k is the wave vector. A linearized solution was found for the rate of growth $\omega(k)$ of the disturbance determining the stability of the surface

$$\omega(k) = vk(1 - d_0 L k^2) \tag{22}$$

in the approximation $kL \gg 1$, where $L = 2D/v$ is the diffusion length and d_0 is the capillary length, proportional to the surface tension.

It follows from (22) that the flat crystallization front is unstable with respect to infinitesimal disturbances with wave vectors less than k_s [2, 9]:

$$k_s = 1/\sqrt{d_0 L}. \tag{23}$$

Just as in the two preceding cases, we shall find the difference between the entropy productions in the cases

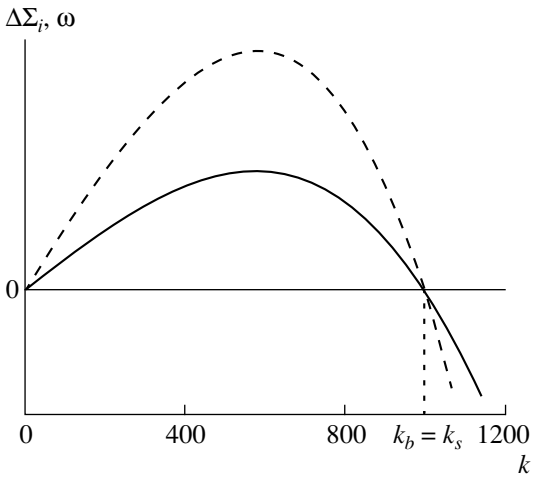


Fig. 12. $\Delta\Sigma_i$ (solid line) and ω (dashed line) as a function of the wave vector k.

of growth of the perturbed Σ_{ip} and unperturbed Σ_{in} surfaces for an element of length dx:

$$\Delta\Sigma_i \equiv \Sigma_{ip} - \Sigma_{in} = (\sigma_{ip}\sqrt{1 + (\xi_x')^2} - \sigma_{in})dx$$
$$\sim \left[(v + \dot{\xi})^2 \sqrt{1 + (\xi_x')^2}\frac{(C - C_{\text{int}})^2}{C_{\text{int}}} - v^2\frac{(C - C_0)^2}{C_0}\right]dx, \tag{24}$$

where

$$\dot{\xi} \equiv \frac{\partial\xi}{\partial t}, \quad \xi_x' \equiv \frac{\partial\xi}{\partial x}, \quad C_{\text{int}} = C_0 + C_0\Gamma k^2\xi.$$

Using (22), the expression (24) can be written in the form

$$\Delta\Sigma_i \sim v^2 k\xi\left(2 + \frac{C - C_0}{C_0}d_0 k + 2d_0 k - 2d_0 L k^2\right)dx.$$

By analogy to the preceding cases we shall choose a point on the surface that corresponds to the most dangerous (from the standpoint of the disruption of stable growth of the plane) state. The behavior of the function $\Delta\Sigma_i(k)$ is qualitatively similar to the behavior of $\omega(k)$ (see Fig. 12). It passes through zero at the points $k = 0$ and k_b:

$$k_b = \left[d_0\left(1 + \frac{C - C_0}{2C_0}\right) + \sqrt{d_0^2\left(1 + \frac{C - C_0}{2C_0}\right)^2 + 4d_0 L}\right]$$
$$\times \frac{1}{2d_0 L} \simeq \frac{1}{\sqrt{d_0 L}} + \left(1 + \frac{C - C_0}{2C_0}\right)\frac{1}{2L} \tag{25}$$
$$\times\left[1 + \sqrt{d_0 L}\left(1 + \frac{C - C_0}{2C_0}\right)\frac{1}{4L}\right].$$

Using (23), we find the ratio k_b/k_s:

$$\frac{k_b}{k_s} = 1 + \left(1 + \frac{C - C_0}{2C_0}\right)\frac{1}{2k_sL}$$
$$\times\left[1 + \left(1 + \frac{C - C_0}{2C_0}\right)\frac{1}{4k_sL}\right] \simeq 1.$$

Hence one can see that in the approximation employed ($kL \gg 1$) k_b is equal to k_s. Thus, the stability boundaries found by the thermodynamic method and the perturbation theory method are the same.

This result can also be obtained by a different method employing the results of the calculations presented above for the sphere and the approach proposed in [2]. As shown in [2], in the limit of high-frequency perturbations ($l \gg 1$) under certain restrictions the results obtained for the sphere can be used to describe the stability of an infinite plane, and the expression (7) becomes

$$\dot{\delta} = \frac{D\omega_1}{C - C_R}(G - C_0\Gamma\omega_1^2)\delta, \qquad (26)$$

where $\omega_1 = l/R$ and $G = \nu(C - C_{\text{int}})/D$ is the concentration gradient at the interface.

From the expressions (26) and (10) we obtain

$$\Delta\Sigma_i \sim (R\dot{\delta} + \delta\dot{R}) \sim \frac{\delta lD}{C - C_R}\left(G - C_0\Gamma\omega_1^2 + \frac{G}{l}\right)$$
$$= \frac{\delta lD}{C - C_R}\left[G\left(1 + \frac{1}{l}\right) - C_0\Gamma\omega_1^2\right] \simeq R\dot{\delta}.$$

Thus, $\Delta\Sigma_i$ changes sign at the same point as $\dot{\delta}$. Therefore the metastable region is absent in the case of the growth of an infinite plane. We note that this result can also be obtained by analyzing a cylindrical surface. As shown, for $k = 0$ and $k_z \gg 1$ the binodal and spinodal points coincide.

6. CONCLUSIONS

In summary, in this paper a relation has been proposed, on the basis of a study of the morphological stability of a crystal growing from solution, between the thermodynamic approach employing the principle of maximum entropy production and a linear analysis for morphological stability. The maximality principle appears here not as an alternative but rather as a supplement to the conventional perturbation theory and makes its own contribution to the solution of the problem of a morphological transition.

It was shown in this paper that instability in the growth of a sphere and an infinite cylinder from a supersaturated solution is a first-order morphological transition (the existence of a metastable region and an abrupt increase in the production of entropy or, specifically, rate of growth of the crystal mass for $R > R^b$). The calculation was performed in the Mullins–Sekerka approximation for an ideal solution and small supersaturations. It is difficult to check experimentally the boundary R^b found for stable growth because these values and also R^s, R^b, and $R^s - R^b$ are of the order of R^* (i.e., about 1 μm). It follows from the explicit expressions obtained for these boundaries that in such experiments small supersaturations crystal–solution system with large surface tension should be used and, if possible, low-frequency disturbances should be eliminated. A full nonlinear analysis of the problems studied, which is quite difficult and laborious, could serve as a theoretical check. We note that a weakly nonlinear stability analysis performed in [35, 36] for a sphere and cylinder made it possible to calculate the nonlinear critical radius, which was found to lie in the metastable zone found in the present paper.

The possibility of the coexistence of a large number of morphological phases is a distinguishing feature of the problem with cylindrical symmetry as compared with the spherical problem. However, it seems that this result is not fundamental and follows from the approximations employed; under different conditions (imperfection of the solution, arbitrary supersaturations) several morphological phases can coexist during growth of a sphere.

Instability during crystallization of an infinite plane was also studied in this paper in the Mullins–Sekerka approximation for $kL \gg 1$, and it was shown that in this case the thermodynamic approach and the perturbation theory approach lead to the same result. In consequence, there is no metastable region, and entropy production accompanying a transition undergoes only a kink with a change in wave vector. The difference arising, as compared with the growth of a sphere and cylinder, can be explained by the fact that in the present calculations a quasistationary approximation was used in which the plane, in contrast to the two other problems, moves at a constant velocity and the surface area remains unchanged.

In conclusion, we underscore the fact that the coexistence of different morphologies is described in many works, but they all use either experiment or computer simulation methods [13–26]. The advantage of the approach proposed here is that it permits performing an analytic calculation of morphological diagrams (with stable, unstable, and metastable regions).

ACKNOWLEDGMENTS

We thank L.D. Son, E.B. Treivos, and V.P. Koverda for a fruitful discussion of this work, constructive remarks, and criticism. We also thank Yu.O. Punin for support of this work and for helpful bibliographical citations.

REFERENCES

1. G. P. Ivantsov, Dokl. Akad. Nauk SSSR **58**, 567 (1947).
2. W. W. Mullins and R. F. Sekerka, J. Appl. Phys. **34**, 323 (1963).
3. S. R. Coriell and G. B. McFadden, in *Handbook of Crystal Growth*, Ed. by D. T. J. Hurle (North-Holland, Amsterdam, 1993), Vol. 1, Part B, p. 785.
4. H. Ziegler, in *Progress in Solid Mechanics*, Ed. by I. N. Sneddon and R. Hill (North-Holland, Amsterdam, 1963; Mir, Moscow, 1966), Vol. 4, Chap. 2.
5. I. P. Vyrodov, Zh. Fiz. Khim. **6**, 1329 (1982).
6. D. E. Temkin, Dokl. Akad. Nauk SSSR **132**, 1307 (1960).
7. S.-C. Huang and M. E. Glicksman, Acta Metall. **29**, 701 (1981).
8. J. S. Langer and H. Müller-Krumbhaar, Acta Metall. **26**, 1681 (1978).
9. J. S. Langer, Rev. Mod. Phys. **52**, 1 (1980).
10. D. Kessler, J. Koplik, and H. Levine, Adv. Phys. **37**, 255 (1988).
11. E. A. Brener and V. I. Melnikov, Adv. Phys. **40**, 53 (1991).
12. E. Ben-Jacob, P. Garik, T. Mueller, *et al.*, Phys. Rev. A **38**, 1370 (1989).
13. E. Ben-Jacob and P. Garik, Nature **343**, 523 (1990).
14. E. Ben-Jacob, Contemp. Phys. **34**, 247 (1993).
15. Mu Wang and Nai-ben Ming, Phys. Rev. Lett. **71**, 113 (1993).
16. J. L. Hutter and J. Bechhoefer, Phys. Rev. Lett. **79**, 4022 (1997).
17. J. L. Hutter and J. Bechhoefer, Physica A (Amsterdam) **239**, 103 (1997).
18. S. K. Chan, H. H. Reimer, and M. J. Kahlweit, J. Cryst. Growth **32**, 303 (1976).
19. O. Shochet and E. Ben-Jacob, Phys. Rev. E **48**, R4168 (1993).
20. Y. Sawada, A. Dougherty, and J. P. Gollub, Phys. Rev. Lett. **56**, 1260 (1986).
21. Y. Sawada, Physica A (Amsterdam) **140**, 134 (1986).
22. Y. Sawada, B. Perrin, P. Tabeling, *et al.*, Phys. Rev. A **43**, 5537 (1991).
23. O. Shochet, K. Kassner, E. Ben-Jacob, *et al.*, Physica A (Amsterdam) **187**, 87 (1992).
24. T. Ihle and H. Müller-Krumbhaar, Phys. Rev. E **49**, 2972 (1994).
25. D. Grier, E. Ben-Jacob, R. Clarke, *et al.*, Phys. Rev. Lett. **56**, 1264 (1986).
26. H. Honjo, S. Ohta, and M. Matsushita, Phys. Rev. A **36**, 4555 (1987).
27. E. A. Brener, H. Müller-Krumbhaar, and D. E. Temkin, Phys. Rev. E **54**, 2714 (1996).
28. P. Glansdorff and I. Prigogine, *Thermodynamic Theory of Structure, Stability and Fluctuations* (Wiley, New York, 1971; Mir, Moscow, 1973).
29. B. Ya. Lyubov, *The Theory of Crystallization in Large Volumes* (Nauka, Moscow, 1975).
30. M. I. Shakhparovov, Zh. Fiz. Khim. **12**, 3043 (1979).
31. R. W. Bene, J. Appl. Phys. **61**, 1826 (1987).
32. K. N. Tu, S. R. Herd, and U. Gosele, Phys. Rev. B **43**, 1198 (1991).
33. R. A. Laudise and R. Parker, *The Growth of Single Crystals. Crystal Growth Mechanisms: Energetics, Kinetics, and Transport* (Prentice-Hall, New York, 1970; Mir, Moscow, 1974).
34. S. R. Coriell and R. L. Parker, J. Appl. Phys. **36**, 632 (1965).
35. P. P. Debroy and R. F. Sekerka, Phys. Rev. E **51**, 4608 (1995).
36. P. P. Debroy and R. F. Sekerka, Phys. Rev. E **53**, 6244 (1996).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 1, 2000, pp. 144–149.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 1, 2000, pp. 163–169.

SOLIDS
Structure

Surface Precrystallization of Normal C24 Alkane in Porous Glass

V. P. Voronov

Petroleum and Gas Institute, Russian Academy of Sciences, Moscow, 117917 Russia
e-mail: vika@fluid.msk.ru

Received January 31, 2000

Abstract—The behavior of the enthalpy and specific heat of normal C24 alkane in the bulk and in porous glasses is investigated using an adiabatic scanning microcalorimeter. Enthalpy jumps, which precede a phase transition in the entire volume of the pores, are found in porous glass with characteristic pore size 1000 Å at a transition from the isotropic liquid into the rotator phase *R*II. The enthalpy jumps are interpreted as a layerwise growth of a crystal phase on the surface of porous glass. It is also found that porous glass substantially changes the phase behavior of alkanes. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The formation of layers of a crystal phase on an interphase boundary at temperatures above the temperature of a volume phase transition into an isotropic liquid (precrystallization phenomenon) has been observed in liquid crystals at a smectic *A*–isotropic liquid (*SmA–I*) transition on a free surface [1] and on a solid substrate [2, 3] treated with surfactants as well as on a free surface in normal alkanes [4, 5]. However, the observed processes leading to the formation of surface layers in liquid crystals are substantially different from those in normal alkanes. In liquid crystals the penetration depth of the crystal phase from the surface into the interior volume increased as the temperature of the volume phase transition was approached (jumps were observed in the intensity of the reflected X-rays), whereas in normal alkanes only one monolayer crystalline phase, whose thickness and properties remained unchanged as the temperature of the volume phase transition into the crystal phase was approached, arose. These phenomena are of a different physical nature.

It has recently been shown on the basis of the Brazovskiĭ phenomenological theory of weak crystallization [6] that in an isotropic phase of liquid crystals close to the volume phase transition *SmA–I* conditions arise for the appearance of a finite contact layer of the smectic phase on the interphase boundary. For this, the value of the surface-induced order parameter of the low-temperature phase on an interphase boundary should be greater than the order parameter arising with a volume phase transition. The properties of the layer of the crystal phase and the mechanism of the formation of the layer do not depend on which interphase boundary, liquid–gas or liquid–solid, it forms. As the temperature of the volume phase transition is approached, the thickness of the layer increases as the logarithm of the dimensionless temperature.

As shown in [7], the precrystallization on a free surface of normal alkanes is associated with fluctuations of the displacements δ_s of molecules in the surface layer along their axes; these fluctuations are much greater than the fluctuations of the displacements δ_b of molecules in the bulk crystal phase. The proposed mechanism of surface precrystallization assumes that the ordered bulk phase possesses a high three-dimensional positional ordering, while in liquid crystals the ordered bulk phase (*SmA*) possesses weak crystalline order. Precrystallization in alkanes differs from precrystallization in liquid crystals by the fact that only one monolayer forms on the free surface of alkanes, and this occurs for a limited range of lengths of molecules with the number of carbon atoms in the molecule ranging from 14 to 50.

The basic condition $\delta_s/\delta_b \gg 1$ in the model of [7], determining the appearance of a monolayer of a crystalline phase on the free surface of normal alkanes, may not be satisfied on a liquid–solid interphase boundary, since the solid wall limits fluctuations of the displacements in the surface layer. In this case precrystallization, if it is possible, should differ from precrystallization on a free surface.

To check this proposition the behavior of the enthalpy and specific heat of normal C24 alkane was investigated in two cases: in the bulk and on porous glasses. In bulk C24 the following sequence of phases is observed on heating: completely ordered crystalline phase (*Cr*)–the rotor phase *R*II–isotropic liquid phase (*IL*). On cooling from the isotropic liquid the sequence of phases is: *IL–R*II–*R*I–*R*V–*Cr*. Two new metastable rotor phases—*R*I and *R*V—arise in the region of the supercooled state of the sample. All intermediate phases *R*I, *R*II, and *R*V possess a highly ordered layered structure [8]. In the *R*II phase the axes of the molecules in a layer are perpendicular to the plane of the layer and

form a hexagonal structure. The layers are packed so that the arrangement of the molecules in a layer repeats every three layers, forming the sequence ABCABC … . In the *R*I phase the hexagonal structure of the molecules in the layer is distorted and the arrangement of molecules in the layer repeats every two layers forming a sequence of layers ABAB … . The structure of the *R*V phase is similar to that of the *R*I phase, but the axes of the molecules are tilted with respect to the plane of the layer.

Precrystallization of C24 on the interface between a liquid and glass near an *IL–R*II phase transition should be accompanied by a jump in the temperature dependence of the enthalpy, just as at a first-order transition. Porous glass was chosen in order to increase the total heat effect, since the jump in enthalpy is proportional to the surface of the interphase boundary.

A series of jumps at temperatures above the phase transition *IL–R*II in the entire volume of the pores was observed in the temperature dependence of the enthalpy of C24 placed in porous glass with characteristic pore size 1000 Å. The metastable phases *R*I and *R*V vanished, while in the porous glass with 100 Å pores the equilibrium phase *R*II also vanished and a single transition from a crystal into an isotropic liquid (*Cr–IL*) was observed.

2. EXPERIMENT

The measurements of the enthalpy and specific heat of normal alkane C24 in the bulk and in porous glass were performed using an adiabatic scanning microcalorimeter with maximum cell volume 0.29 cm^3. The cell was surrounded with two isothermal screens whose temperature strictly "followed" the temperature of the cell. A platinum resistance thermometer was placed on the outer screen, whose temperature was maintained equal to the temperature of the cell. The construction of the calorimeter is described in greater detail in [9].

The measurement process was performed and monitored with a computer system. The scanning measurement regime was conducted with heating and cooling. For heating a constant electric power p_0 which was measured during scanning, was supplied to the calorimeter. In addition, the cell temperature T and the amount of heat (enthalpy) $H(T)$ supplied to the cell from the onset of scanning were measured. The cooling regime required preliminary adjustment and was conducted as follows. The cell was heated with constant power p_0, and the temperature of the inner screen was set $\Delta T(p_0)$ below the temperature of the cell so that the heat transfer between the cell and the inner screen would compensate the heating power supplied to the cell. During scanning the heating was switched off and the temperature difference between the cell and the inner screen was maintained constant. As a result the calorimeter cooled down at a rate determined by the heat-transfer power $q_0 = -p_0$.

In a real experiment the measurement regime is quasiadiabatic, i.e., weak heat transfer, ordinarily not exceeding 10^{-5} W, which results in temperature drift of the cell $(dT/dt)_0$, occurs between the cell and the inner screen. During scanning only the electric power p_0 supplied to the cell (or removed during cooling) is measured and integrated. Consequently, the heat transfer, which is not known beforehand, must be taken into account when calculating the total amount of heat acquired by the cell. Before scanning starts, at temperature T_s, the heat-transfer power of the cell q_s can be calculated from the expression $q_s = C_{sum}(dT/dt)_0$. Since the total specific heat of the cell C_{sum} is unknown, two independent measurements must be performed in order to determine the heat transfer: the temperature drift of the cell $(dT/dt)_0$ and the rate of change of the cell temperature dT/dt with weak heating with power q. The power q and the measurement time were chosen so that the cell temperature would change very little and the specific heat C_{sum} could be taken as constant. The total power $q + q_s$ acquired by the cell determines the rate of change of the temperature:

$$q + q_s = C_{sum}\frac{dT}{dt}.$$

The power of heat transfer between the cell and the surrounding medium was calculated, using the following expression, from the results of these measurements:

$$q_s = \frac{q}{(dT/dt)/(dT/dt)_0 - 1}.$$

To take into account the temperature dependence of heat transfer the heat-transfer power of the cell q_f was measured once again at temperature T_f when scanning stopped and before scanning started. Finally, since the heat transfer of the cell in the measurement temperature interval $[T_s, T_f]$ varies linearly, the following expression for the power was used to calculate the temperature dependence of the enthalpy $H(T)$ (total amount of heat flowing into the cell):

$$p(T) = p_0 + \left[(q_f - q_s)\frac{T - T_s}{T_f - T_s} + q_s\right].$$

The total heat capacity of the cell C_{sum} was calculated by numerically differentiating the enthaply $H(T)$. To determine the specific heat of the sample, the specific heat of an empty cell, which was measured beforehand, was subtracted from the total specific heat.

Porous glasses which contained 1000 Å and 100 Å pores and consisted of networks of channels connected randomly with one another were used in the experiment. Cylindrical matrices, 5 mm in diameter and 14 mm long, which were saturated with normal alkane at a temperature above the temperature of a phase transition into an isotropic liquid were prepared from these glasses and inserted into the cell. The surface of the pores was not specially treated. The matrices were first

Temperature and heat of phase transitions in normal alkane C24 in the bulk and in porous glasses

$T_{Cr–RII}$, °C	$\Delta H_{Cr–RII}$, J g^{-1}	$T_{RII–IL}$, °C	$\Delta H_{RII–IL}$, J g^{-1}	$T_{Cr–IL}$, °C	$\Delta H_{Cr–IL}$, J g^{-1}
In the bulk					
48.20	91.5	50.72	155.1	–	–
±0.05	±1	±0.01	±1	–	–
In porous glass, 1000 Å					
47.60	69.5	50.41*	153.6	–	–
±0.05	±0.8	±0.01	±2.5	–	–
In porous glass, 100 Å					
–	–	–	–	46.90	106.5
–	–	–	–	±0.05	±2

* The temperature of the transition is determined according to the enthalpy jump ΔH_3 in Fig. 4 (curve *2*).

washed with a solvent and dried at a temperature near 200°C. At total saturation approximately 0.14 g of sample was placed in the matrix with 1000 Å pores, and 0.06 g was placed into a matrix with 100 Å pores. Normal alkanes, acquired from the Aldrich Chemical Company, without any additional purification were used in the experiment.

3. EXPERIMENTAL RESULTS

Figure 1 (curve *1*) shows the temperature dependence of the enthalpy, obtained on heating a bulk C24 sample from the crystal state into an isotropic liquid. The enthalpy jumps correspond to the transitions *Cr*–*R*II and *R*II–*IL*. Both transitions are strong first-order transitions with heat characteristic of melting of porous liquids (see table).

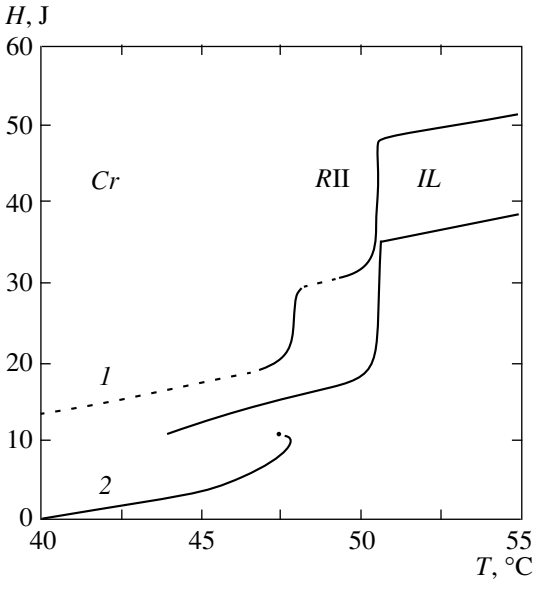


Fig. 1. Temperature dependences of the enthalpy which were obtained by heating (*1*) and cooling (*2*) C24 in the bulk. For convenience in making comparisons, the curves are displaced relative to one another by a constant amount ΔH.

The temperature dependence obtained for the enthalpy on cooling is also presented in Fig. 1 (curve 2). The section of $H(T)$ with a negative derivative corresponds to a transition from the supercooled state of the sample into a crystal. The scale of the figure does not permit distinguishing in the temperature dependence $H(T)$ weak variations of the enthalpy corresponding to transitions into the metastable phases *R*I and *R*V[1] against the background of large jumps in enthalpy, which are accompanied by an *IL*–*R*II transition and a transition into the crystalline phase. Figure 2 shows separately the behavior of the specific heat of C24 in the temperature range where the sample is in a supercooled state. It is evident that in this interval there are two specific-heat peaks, at T = 44.55 and 47.19°C, which correspond to transitions between the phases *R*V–*R*I and *R*I–*R*II.

The phase behavior of C24 placed in a porous glass changes substantially. The temperature dependence obtained for the enthalpy on heating C24 in a porous glass matrix with characteristic pore size 1000 Å is presented in Fig. 3 (curve *1*). One can see that the phase-transition temperatures (the table gives the transition temperatures determined on heating of the sample) shift in the direction of lower temperatures compared with transition temperatures in the interior volume of the sample (curve *2*). The specific heat of the phase transition from a crystal into the *R*II phase, $\Delta H_{Cr–RII}$ (see table), also changes. In porous glass the quantity $\Delta H_{Cr–RII}$ is much smaller than the specific heat of the transition in the bulk sample. This could be due to the

[1] The *R*V–*R*I transition is a second-order transition. The *R*I–*R*II transition, by virtue of the symmetry of the phases, should be a first-order transition. However, the experimental data show that there is no heat of transition or this heat is very small and does not exceed 7×10^{-2} J g^{-1}. The smallness of the transition heat is probably due to the closeness of the transition *R*I–*R*II in C24 to the tricritical point on the line of transitions *R*II–*R*I.

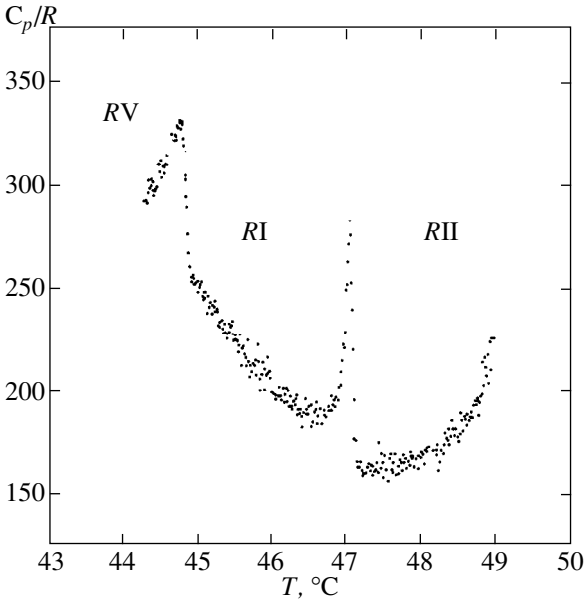


Fig. 2. Temperature dependence of the specific heat of C24 in units of the universal gas constant R near the transitions RV–RI and RI–RII.

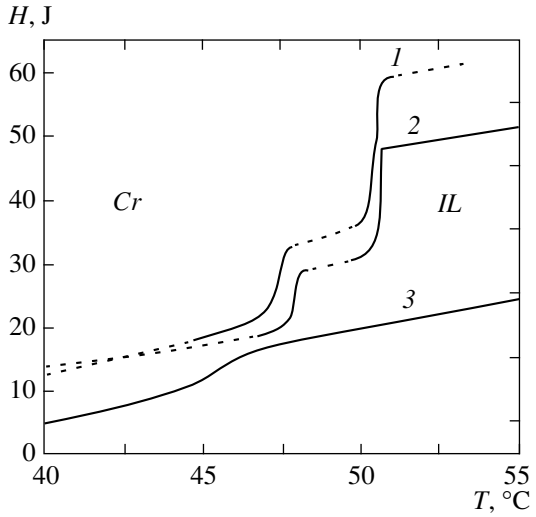


Fig. 3. Temperature dependences of the enthalpy which were obtained by heating C24 placed in porous glasses with 1000 Å (*1*) and 100 Å (*3*) pores. For comparison, the temperature dependence of the enthalpy of bulk C24 (*2*) with the same scan rate as in porous glasses is presented. For convenience in making comparisons, the curves are displaced relative to one another by a constant amount along the H-axis.

difference in the crystal structures formed by C24 in the volume and in porous glass. At the same time the specific heat $\Delta H_{RII–IL}$ of the transition from the phase RII into the isotropic liquid remained, to within the limits of the measurement error, the same as in bulk C24.

When a sample placed in porous glass with 1000 Å pores is cooled, the temperature interval of the supercooled state remains practically unchanged. However, the metastable phases RI and RV vanish and only the equilibrium phase RII remains. Moreover, it was found that in porous glass with 100 Å pores the equilibrium phase RII also vanishes and a single transition is observed out of the crystal into the isotropic phase (curve *3* in Fig. 3). This behavior is quite obvious. The porous medium destroys the layered structure of the mesophases and for pore sizes comparable to the interplanar separation the layered structure is completely destroyed. The specific heat $\Delta H_{Cr–IL}$ of the transition in a matrix with 100 Å pores is much smaller than the total heat of the transition out of the crystal into the isotropic liquid in a bulk sample and in porous glass with 1000 Å pores.

The porous glass also changes the character of the transition from the RII phase into the isotropic liquid. Figure 4 (curve *2*) shows the temperature dependence of the enthalpy near the RII–IL transition, obtained on heating C24 placed in porous glass with 1000 Å pores. In contrast to a transition in the bulk sample (curve *3*), the transition into an isotropic liquid is accompanied by three enthalpy jumps in an interval of about 0.3 K for the same value of the total enthalpy change $\Delta H_{RII–IL}$. This effect is also qualitatively reproduced on cooling (curve 1). The position of the enthalpy jump ΔH_1 does not depend on the direction of change of the temperature. The jumps ΔH_2 and ΔH_3 have hysteresis, and the temperature interval between them changes but the magnitude of the jumps remains the same as on heat-

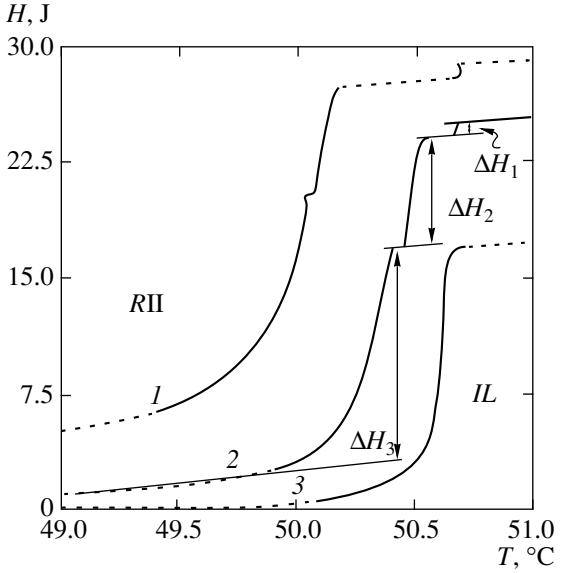


Fig. 4. Temperature dependences of the enthalpy near the transition RII–IL in C24 placed in porous glass with 1000 Å pores on heating (*2*) and cooling (*1*). The curve *3* corresponds to a change in enthalpy on heating of the bulk sample. For comparison the curves are shifted relative to one another along the H-axis. The scan rate is the same for all cases.

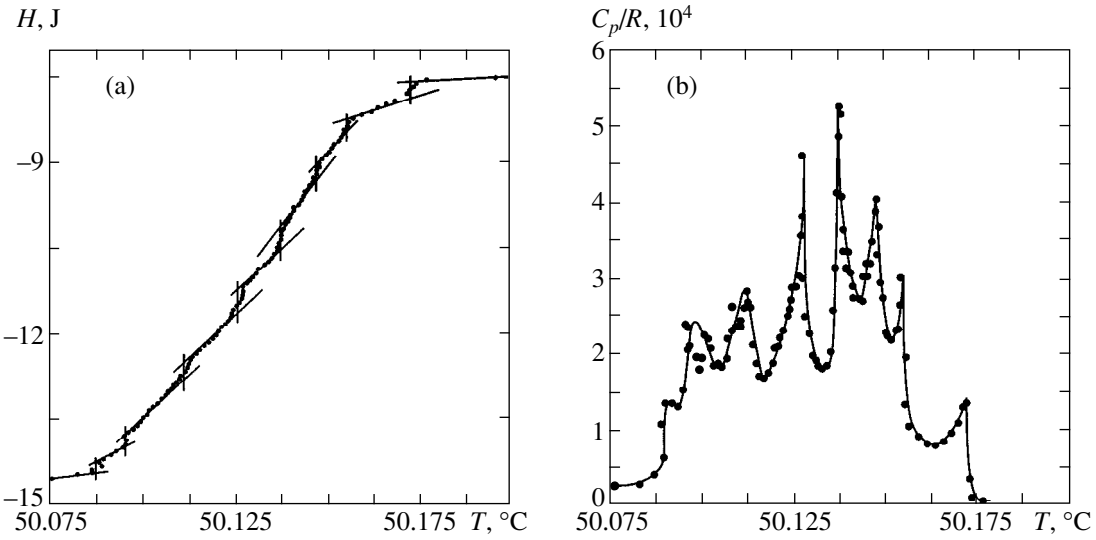


Fig. 5. Behavior of the enthalpy (a) and specific heat in units of R (b) on cooling of C24 placed in porous glass with 1000 Å pores in the temperature range corresponding to the enthalpy jump ΔH_2 in Fig. 4 (curve *1*). The negative values of H in Fig. 5a mean that the measurements were performed on cooling of the sample. The solid lines are drawn for better visual perception of the experimental points.

ing. This behavior of the enthalpy at a *R*II–*IL* transition was reproduced repeatedly with different rates of heating and cooling of the sample.

The "fine structure" of the enthalpy jump ΔH_2, consisting of a series of eight weak jumps in an interval <0.1 K, can be distinguished on an extended temperature scale. The fine structure (Fig. 5a) can be seen most clearly with slow cooling of the sample (<0.4 K/h far from a transition). On heating the jumps become diffuse and indistinguishable. Each enthalpy jump corresponds to a peak of the specific heat (Fig. 5b), similar to the peak in the specific heat accompanying a diffuse first-order phase transition.

4. DISCUSSION

It is natural to infer that the enthalpy jumps at an isotropic liquid–*R*II phase transition in C24 placed in porous glass with 1000 Å pores are associated with the formation of monolayers of the phase *R*II on the surface of the pores. The jump ΔH_1 is related with the formation of the first monolayer. The jump ΔH_2 corresponds to the formation of the next eight monolayers on the surface of the pores, and the final transition into the *R*II phase in the remaining volume of the pores is accompanied by the enthalpy jump ΔH_3.

It is obvious that the magnitude of the enthalpy jumps, corresponding to the formation of monolayers on the surface of the pores, should be proportional to the pore volumes. Treating the porous medium as a network of randomly interconnected cylinders with radius r_0, the following relation can be written down between the enthalpy jumps ΔH_1 and $\Delta H_{RII–IL}$ and the corresponding volumes V_1 and V:

$$\frac{\Delta H_1}{\Delta H_{RII–IL}} = \frac{V_1}{V} = 1 - \frac{(r_0 - \delta)^2}{r_0^2}, \qquad (1)$$

where V_1 is the volume of the first monolayer, V is the total volume of the pores in the matrix, δ is the thickness of a monolayer, and $r_0 = 500$ Å.

It follows from the experimental data that the ratio of the enthalpy jumps is $\Delta H_1/\Delta H_{RII–IL} = 0.0387 \pm 0.001$. Then we find from Eq. (1) the thickness of the monolayer $\delta = 9.8$ Å. This value is almost three times less than the thickness of the C24 monolayer formed on a free surface [5], where the molecules are oriented perpendicular to the surface and form extended chains. However, taking account of the complicated geometry of the porous glass, it is difficult to expect that the long C24 molecules in a monolayer on the surface of the pores will be oriented in the same manner. More likely, the molecules are deformed and their axes are tilted with respect to the surface of the monolayer. Generally speaking, this could cause the heat of the transition *R*II–*IL* in the bulk to be different from the value in porous glass. However, it follows from the experimental results that this difference falls within the measurement error (see table).

A relation similar to Eq. (1) can also be written down for the enthalpy jumps ΔH_2, ΔH_1 and volumes V_2, V_1:

$$\frac{\Delta H_2}{\Delta H_1} = \frac{V_2}{V_1}, \qquad (2)$$

where ΔH_2 is the enthalpy jump corresponding to the formation of eight monolayers on the surface of the

pores, and V_2 is the volume which they occupy. It follows from the experimental data that

$$\frac{\Delta H_2}{\Delta H_1} = 7.34 \pm 0.2.$$

Assuming the thicknesses of the monolayers to be the same and equal to 9.8 Å, we find that the ratio of the volumes is

$$\frac{V_2}{V_1} = \frac{1-[1-8\delta/(r_0-\delta)]^2}{[r_0/(r_0-\delta)]^2-1} = 7.28.$$

Thus, the equality (2) in the approximations considered holds with good accuracy. Under the same assumptions the equality $\Delta H_3/\Delta H_{RII-IL} = V_3/V$, where $\Delta H_3/\Delta H_{RII-IL} = 0.70 \pm 0.02$ and $V_3/V = 0.68$, also holds to within the measurement error in the heat of the transitions. All this confirms the assumption that the enthalpy jumps ΔH_1 and ΔH_2 at the transition *IL*–*R*II in C24 placed in porous glass with 1000 Å pores are related with layerwise crystallization on the surface of porous glass, preceding a transition in the entire volume of the pores.

The precrystallization process observed in C24 on the surface of porous glass is different from the precrystallization of normal alkanes on a free surface. In order for a monolayer of a crystal phase to appear on a free surface the fluctuations of the displacements of molecules in the surface layer must be much greater than the fluctuations of the displacement of the molecules in a bulk crystal phase: $\delta_s \gg \delta_b$ [7]. At the very least the condition $\ln(\delta_s/\delta_b) > 1$ must be satisfied. On a solid–liquid interphase boundary a solid wall limits the fluctuations of the displacements of the molecules along their axes in the surface layer and they are of the same order of magnitude as in the bulk crystal phase. Consequently, the necessary condition for the development of precrystallization according to the scenario of [7] is not satisfied. In this case the situation is superficially similar to precrystallization in liquid crystals.

The phenomenon of precrystallization on a solid substrate depends to a large degree on the properties of the surface. In all experiments with liquid crystals, where the formation of smectic layers in an isotropic phase preceding a volume phase transition was observed, the solid surface was treated with surfactants, which encouraged the formation of monolayers. Consequently, the observed layerwise precrystallization of C24 on the surface of porous glass whose surface was not specially treated is unexpected.

The behavior of the enthalpy and specific heat was also investigated in normal alkanes C22 and C28 in the bulk and in a porous glass matrix with 1000 Å pores. Two intermediate phases *R*I and *R*II are present in C22. Just as in C24, the phase *R*I is metastable. In C28 the mesophases *R*IV and *R*III, in which the molecules are tilted with respect to the plane of the layer and form a weakly perturbed hexagonal structure in the layer, are observed [8]. The phase *R*III is metastable. Precrystallization at a transition into an isotropic liquid, similar to the precrystallization observed in C24 placed in a porous glass matrix with 1000 Å pores, was not observed in these alkanes. This is probably explained by the change in the properties of the surface of the pores. However, coarser effects, such as displacement of the temperatures of the phase transitions and vanishing of the metastable phases in porous glass, have been observed in all alkanes investigated.

It can be asserted on the basis of the results obtained that precrystallization in normal alkanes at a liquid–solid boundary—if it arises—occurs by a mechanism that is different from the mechanism of precrystallization on a free surface.

ACKNOWLEDGMENTS

I thank E.E. Gorodetskiĭ and V.É. Podnek for helpful discussions and critical remarks and J. Thoen for helpful discussions of the results of this work during a visit to the Catholic University in Leuven (Belgium). I also thank V.M. Buleĭko for assisting in the measurements. This work was supported by the Russian Foundation for Basic Research (project no. 98-01-00380).

REFERENCES

1. B. M. Ocko, A. Breslau, P. S. Pershan, *et al.*, Phys. Rev. Lett. **57**, 94 (1986).
2. B. M. Ocko, Phys. Rev. Lett. **64**, 2160 (1990).
3. G. P. Crowford, R. Stannarius, and J. W. Doane, Phys. Rev. A **44**, 2558 (1991).
4. J. C. Earushau and C. J. Hughes, Phys. Rev. A **46**, R4494 (1992).
5. X. Z. Wu, E. B. Sirota, S. K. Sinha, *et al.*, Phys. Rev. Lett. **70**, 958 (1993).
6. E. E. Gorodetskiĭ, E. S. Pikina, and V. É. Podnek, personal communication.
7. A. Tkachenko and Y. Rabin, Phys. Rev. Lett. **76**, 2527 (1996).
8. E. B. Sirota, H. E. King, Jr., D. M. Singer, and Henry H. Shao, J. Chem. Phys. **98**, 5809 (1993).
9. L. V. Entov, V. A. Levchenko, and V. P. Voronov, Int. J. Thermophys. **14**, 221 (1993).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 1, 2000, pp. 150–166.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 1, 2000, pp. 170–188.

SOLIDS
Structure

The Twice-Renormalized Rouse Formalism of Polymer Dynamics: Segment Diffusion, Terminal Relaxation, and Nuclear Spin-Lattice Relaxation¶

N. F. Fatkullin*, **, R. Kimmich*, and M. Kroutieva***

**Kazan State University, Kazan, 420008 Russia*
***e-mail: Nail.Fatkullin@ksu.ru*
****Sektion Kernresonanzspektroskopie, Universitat Ulm, 89069, Ulm, Germany*

Received August 18, 1999

Abstract—The twice-renormalized Rouse formalism, a refined version of Schweizer's renormalized Rouse treatment of chain dynamics in entangled polymers, is presented. The time scale of validity is extended to include the terminal chain relaxation and center-of-mass diffusion. In clear contrast to the laws concluded from other polymer dynamics concepts (such as the reptation (tube) model or the polymer mode–mode coupling formalism), the predictions perfectly coincide with all the results of recent spin-lattice relaxation dispersion and diffusion experiments as well as with computer simulations. On the other hand, the twice-renormalized Rouse formalism fails to explain the rubber–elastic plateau of stress relaxation. It is inferred that this is a consequence of the single-chain nature of the present approach not accounting for the fact that viscoelasticity is largely a manifestation of collective multichain modes. In the rigorous sense, no such multichain treatment has yet been established to our knowledge. The necessity to consider interchain cooperativity in any real comprehensive polymer dynamics theory is concluded from low-frequency spin-lattice relaxation data, which are shown to reflect fluctuations of long-distance intermolecular dipole–dipole interactions. © *2000 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION

The term "chain entanglement" globally represents the topological constraints and the excluded-volume effect of a matrix consisting of polymer chains longer than the critical length (L_c). The general challenge of entangled polymer dynamics then is, how to reduce the multichain problem to a tractable formalism. Basically, two approaches are under discussion in the literature. In the reptation (tube) model [1, 2], the entanglement effect is intuitively thought to result in a tube in which the tagged chain is confined on a time scale shorter than the "disengagement time" (τ_d). This tube must be considered as a fictitious object, the diameter of which is a parameter to be fitted to the experimental data. There are no means to directly visualize or measure the tube geometry. That is, irrespective of any successful description of the experimental facts, the merit of this model (or any of the numerous modifications of it) is that an illustrative notion is provided.

Another approach is to generalize the equations of motion so that the entanglement effect is represented by a certain memory function term. Examples of such theories are the renormalized Rouse (RR) formalism [3], the polymer mode–mode coupling concept [4], and others [5, 6].

A shortcoming to any approach is that it cannot be derived rigorously from the fundamental principles. This point is directly connected with the closure problem for an infinite set of Bogolubov–Born–Green–Kirkwood–Yvon equations: the so-called BBGKY hierarchy closure problem. Generally speaking, one cannot rigorously derive a set of equations (which, as a rule, are of the integro-differential type) for the quantities of interest. At any step of derivation, one will get terms containing new, unknown quantities.

The closure problem can be considered to be solvable only for a narrow class of problems that have explicitly distinguishable small parameters reflecting some small deviation from ideal, or weak interactions. Well-known examples of systems, which belong to the aforementioned class, are gases of real particles (atoms or molecules) with a low concentration and quasiparticles (phonons in solids at temperatures much lower than melting point). With small parameters, a perturbation theory can be established, which in principle can provide the possibility for an approximate treatment controlled by internal means of theory.

In the case of strong interactions, the closure procedure in fact means the introduction of additional postulates that are based on intuitive physical ideas or certain analogies. These additional postulates or the respective equations of motion should be considered as uncon-

¶ This article was submitted by the authors in English.

trolled approximations. What is really required here is that

(1) the additional postulates must not contradict physical principles and facts;

(2) the mathematical treatment must be consistent in the chosen framework.

The final criterion for such an approach can only be the comparison of the theoretical predictions with the experimental data. This part of theoretical physics essentially is the search for and investigation of all possible more or less successful approximations. Since we cannot resolve the closure problem rigorously, it is clear that no approximate theory or model exists that can describe all experimental data. Consequently, one should expect that different approaches are required to describe different aspects of nature. Correspondence between these approaches should therefore be considered mainly as a supplement rather than competition.

In the reptation model, dynamical confinement inside the tube and the Rouse dynamics for curvilinear motions are postulated in an illustrative but phenomenological way. On the other hand, treatments based on different memory-function formalisms are closer to the initial principles because their starting point is the exact microscopic generalized Langevin equation that was obtained with the aid of the Zwanzig–Mori projection operator technique. The BBGKY hierarchy here appears through the memory function, which contains multiparticle correlation functions. Their behavior is determined by the so-called "projected" dynamics, which so far cannot be illustrated in a simple way. Some additional postulates must be introduced at this stage of "derivation" that are not as transparent as those in reptation model. They are based on analytical intuition and analogies coming from liquid-state physics rather than on geometrical reasons.

The appealing feature of the memory function formalism is its analytical beauty and simplicity. For example, the Rouse model, which is the simplest model of polymer dynamics, can be derived from the generalized Langevin equation by assuming a diagonal form of the memory matrix, an isotropic motion of polymer segments, and fast (relative to the real segment motion) projected dynamics. The comparison with experimental data shows that the Rouse model does not describe chain dynamics in entangled polymer melts. This means that at least one of the postulates is too far from reality in this case. Therefore, one is proned to examine more slowly decaying memory functions. Schweizer preserves the assumption of isotropy of motion on any time scale in his renormalized Rouse formalism and postulates ordinary Rouse dynamics as an approximation for the projected dynamics that governs the decay of the memory kerner. Some predictions achieved on this basis represent experimental data very well, especially those that concern the short time dynamics [7]. However, they fail to provide the proper explanation for the data referring to slow and chain-length-dependent modes.

The objective of the present work is to find theoretical descriptions of experimentally well established phenomena specifically connected with single-chain correlation functions. Other features such as viscoelasticity, which appears to be largely based on collective multichain modes, are not considered. In the class of dynamic observables differentiated in this way, the entanglement effect then reveals itself by modified modes of a single "tagged" chain. Prominent examples of such single-chain-mode-sensitive experiments are the spin-lattice relaxation (e.g., [8]) and the segment (center-of-mass) diffusion (not to speak of computer simulations (e.g., [9, 10])). With certain reservations, dielectric relaxation also falls into this category [11, 12]. No attempt will be undertaken to account for mechanical stress relaxation. Viscoelasticity probes feature another physical quality. It appears that there is no way to self-consistently and comprehensively represent collective multichain modes by modifications of single(tagged)-chain correlations. As a matter of fact, no single-chain mode formalism is known that consistently describes the pronounced phenomena of spin-lattice relaxation dispersion, anomalous segment diffusion, and stress relaxation all at one time.

The outline of the paper is as follows. In Sections 2 and 3, we will review the basic theory of polymer dynamics and the essentials of the renormalized Rouse formalism in the full-mode-number-dependent version [7]. Using the results for the mean squared segment displacement derived in Section 3, the twice-renormalized Rouse formalism will be established in Section 4. In Section 5, the theoretical predictions will be compared with typical NMR experiments from various laboratories (including ours). In the final section, the impact of collective multichain modes on polymer dynamics will be sketched out as a perspective for a comprehensive polymer-dynamics theory.

2. BASIC THEORY

2.1. The Equation of Motion

The generalized Langevin equation for a tagged chain was first derived by Zwanzig and Bixon [13, 14] in the frame of the Zwanzig–Mori projection operator formalism (see, e.g., [15]). Following Schweizer's treatment based on the assumption of isotropy [3, 16], the equation of motion of the nth Kuhn segment at a position $\mathbf{R}_n(t)$ reads

$$\zeta\left[\frac{\partial \mathbf{R}_n(t)}{\partial t} + \int_0^t d\tau \sum_m \Gamma_{nm}(t-\tau)\frac{\partial \mathbf{R}_m(\tau)}{\partial \tau}\right] = \frac{3k_BT}{b^2}\frac{\partial^2 \mathbf{R}_n(t)}{\partial n^2} + \mathbf{F}_n^Q(t), \quad (1)$$

where ζ is the "bare" friction coefficient of a Kuhn segment, k_B is Boltzmann constant, T is the absolute temperature, b is the Kuhn segment length, and $\mathbf{F}_n^Q(t)$ is the stochastic force acting on the Kuhn segment number n at time t. The upper index Q indicates that the time evolution is governed by the so-called projected dynamics, i.e., by the propagator e^{iQLt}, where L is a Liouvillian taking into account all interactions and determining the real dynamics through the propagator e^{iLt}. The operator Q is given by $Q = 1 - P$, where P is a projection operator on the most general N-body polymer field of the tagged chain [3, 16].

The projection operator P acts on an arbitrary function which is defined in the phase space of the whole system. It extracts the part lying in the phase space of the tagged macromolecule by equilibrium averaging over all coordinates and momenta of matrix chains, preserving the variables of the tagged chain. Similarly, the operator Q extracts the fluctuating part of a function, maintaining the variables of the tagged chain constant.

The term

$$\frac{3k_BT}{b^2}\frac{\partial^2\mathbf{R}_n(t)}{\partial n^2}$$

represents the intrachain entropic elastic force, where the polymer is treated in the continuum limit [2]. The memory matrix can be written as

$$\Gamma_{nm}(t) = \frac{1}{2k_BT\zeta}\langle \mathbf{F}_n^Q(0)\mathbf{F}_m^Q(t)\rangle. \qquad (2)$$

It must be considered as the crucial constituent of (1). Its existence follows directly from the exact set of microscopic equations of motion for the whole system. $\Gamma_{nm}(t)$ describes dynamical correlations of the fluctuating part of intermolecular forces acting on segments with numbers n and m at moments 0 and t. These forces alone are the actual cause of entanglement effects in polymer systems. The first term on the right side of (1) describes the local friction effects on the dynamics of the tagged chain. It can be included in the memory matrix as well, where it would correspond to the fast decaying diagonal part $\zeta\delta(t)\delta_{km}$. The separation of the local friction term from the long-living part $\Gamma_{nm}(t)$ of the whole memory matrix follows the historical convention in context with the Rouse model.

The precise knowledge of the memory matrix would in turn yield the exact single-chain equation of motion accounting rigorously for all entanglement effects. However, because of the BBGKY problem mentioned in the introduction, we do not have a straightforward way to compute it. One can only construct models by anticipating different assumptions.

In the Rouse model, no excluded volume of the polymer chains is assumed; therefore, they can cross each other. The local density fluctuations relax on the time scale of the segmental relaxation time. The Rouse equations of motion can thus be obtained from (1) by neglecting the long-living memory term. However, experimental data indicate that the dynamic behavior of polymer melts with a large molecular mass is different from that predicted by the Rouse model. The reason is that polymers cannot cross each other in reality. Therefore, density fluctuations of matrix chains are long living and generally nonlocal. This fact is reflected in (2) by correlations between the forces with which the matrix chains act on the segments of the tagged chain at times 0 and t.

The exact expression for the memory matrix [3, 16] is not as simple as (2). It is based on a tensor product instead of a scalar product of stochastic forces and is in general a functional of the instantaneous conformation of the tagged chain. By postulating an isotropic character of the motions at all times, the memory kernel adopts the form given in (2).

The force–force correlation functions on the right side of (2) contain three-chain projected dynamical correlation functions. Using the superposition approximation, it can be written as an integral over the k-space [3, 16]:

$$\Gamma_{nm}(t) = \frac{8}{27}\frac{\rho_m d^6 g^2(d)}{k_BT\zeta}\int_0^{b^{-1}} dk k^4 \hat{\omega}_{nm}^Q(k, t)\hat{S}^Q(k, t), \qquad (3)$$

where we have corrected a mistake in the numerical factor given in Schweizer's original treatment.

The quantity ρ_m is the segment number density, $g(r)$ is the radial intermolecular distribution function averaged over all segments, and d is the hard-core segment diameter. The intrachain projected dynamical structure factor as a function of the wave number k and time t is defined by

$$\hat{\omega}_{nm}^Q(k, t) = \hat{\omega}_{nm}(k)\exp\left\{-\frac{k^2}{6}\langle r^2(t)\rangle_Q\right\}, \qquad (4)$$

where $\hat{\omega}_{nm}(k)$ is the static intrachain structure factor. Likewise, the projected collective correlation function of the matrix density surrounding the tagged chain is given by

$$\hat{S}^Q(k, t) = \hat{S}(k)\exp\left\{-\frac{k^2}{6}\langle r^2(t)\rangle_Q\right\}. \qquad (5)$$

The quantity $\hat{S}(k)$ is the static collective structure factor. Here, we have assumed identical exponential functions on the right-hand sides of (4) and (5).

Single chain $\hat{\omega}_{nm}^Q(k, t)$ and collective $\hat{S}^Q(k, t)$ projected dynamical structure factors in (3) reflect the fact that local density fluctuations around the tagged chain are relaxed by the projected motions both of the tagged chain

and the matrix chains. The projected nature of the dynamics is hidden in Eqs. (4) and (5), namely, in the projected mean squared displacement, which qualitatively describes typical distances over which elementary density fluctuations become dispersed around the tagger chain during time t.

In the frame of this approach, the projected mean squared segment displacement $\langle r^2(t)\rangle_Q$ is now a function representing the unknown features of the multiparticle dynamics. Even worse, one could even expect different projected dynamics for intrachain and collective chain dynamical structure factors. However, we will neglect any potential differences for simplicity.

In order to further evaluate Eqs. (4) and (5), we now postulate that the projected mean squared segment displacement may be replaced by the "real" mean squared segment displacement predicted by the Rouse model [2]. This is the essence of Schweizer's renormalized Rouse concept [3]. In the following section, Schweizer's treatment will be augmented by taking the full mode number dependence of the memory kernel into account [7]. Furthermore, in the approach of the present study, the projected mean squared displacements Eqs. (4) and (5) are replaced by the real mean squared displacements, as suggested by the renormalized Rouse concept. This intuitive second iteration is the crucial assumption of the twice-renormalized Rouse formalism to be described in the following.

2.2. Normal-Coordinate Analysis

The generalized Langevin equation Eq. (1) is linear in the segment coordinates. Therefore, a normal-coordinate analysis as usual in Rouse-type theories suggests itself. The pth normal coordinate of a chain consisting of N Kuhn segments reads

$$\mathbf{X}_p(t) \equiv \frac{1}{N}\int_0^N \cos\left(\frac{\pi}{N}pn\right)\mathbf{R}_n(t)\,dn, \tag{6}$$

where $p = 0, 1, 2, \ldots, N-1$. The inverse transformation is

$$\mathbf{R}_n = \mathbf{X}_0 + 2\sum_{p=1}^{N-1}\mathbf{X}_p\cos\left(\frac{\pi}{N}pn\right). \tag{7}$$

In terms of normal coordinates, the generalized Langevin equation takes the form

$$\zeta\left[\frac{\partial\mathbf{X}_p(t)}{\partial t} + \int_0^t \Gamma_p(t-\tau)\frac{\partial\mathbf{X}_p}{\partial\tau}d\tau\right] = -\frac{p^2\zeta}{\tau_1^R}\mathbf{X}_p(t) + \mathbf{F}_p^Q(t). \tag{8}$$

The memory function with respect to the pth normal coordinate is defined by

$$\Gamma_p(t) = \int_0^N \Gamma_m(t)\cos\left(\frac{\pi}{N}pm\right)dm, \tag{9}$$

where we have neglected chain end effects (see, e.g., [17]) so that

$$\Gamma_{nk}(t) = \Gamma_{|n-k|}(t) \equiv \Gamma_m(t) \tag{10}$$

with $m = |n-k|$. The so-called Rouse time, the longest Rouse relaxation time applying to the mode for $p = 1$,

$$\tau_1^R = N^2\tau_s, \tag{11}$$

is based on the relaxation time of a Kuhn segment

$$\tau_s = \frac{\zeta b^2}{3\pi^2 k_B T}, \tag{12}$$

that is, the shortest Rouse relaxation time referring to the mode for $p = N-1$.

Combining Eqs. (3), (4), (5), (9), and (10) leads to

$$\Gamma_p(t) = \frac{8}{27}\frac{\rho d^6 g^2(d)}{k_B T\zeta}\int_0^{1/b} dk\,k^4\hat{S}(k)\exp\left\{-\frac{k^2}{3}\langle r^2(t)\rangle_Q\right\} \times \int_0^N dn\,\hat{\omega}_{nm}(k)\cos\left(\frac{\pi}{N}pm\right). \tag{13}$$

The intrachain static structure factor for a Gaussian coil is given by [2]

$$\hat{\omega}_{nm}(k) = \exp\left\{-\frac{k^2b^2}{6}|n-m|\right\}. \tag{14}$$

The collective static structure factor may be approached by

$$\hat{S}(k) \simeq \hat{S}(0) = \rho_n k_B T\kappa_T, \tag{15}$$

where κ_T is the isothermal compressibility. Furthermore, in the limit $N \gg 1$, we may use the approximation

$$\int_0^N dm\cos\left(\frac{\pi}{N}pm\right)\exp\left\{-\frac{k^2b^2}{6}|n-m|\right\} \simeq \frac{k^2b^2/6}{(k^2b^2/6)^2 + (\pi p/N)^2}. \tag{16}$$

That is, the memory function of the pth normal mode obeys

$$\Gamma_p(t) \simeq \frac{8}{81\pi^2}\frac{\psi b^5}{\tau_s} \times \int_0^{1/b} k^4 \frac{k^2b^2/6}{(k^2b^2/6)^2 + (\pi p/N)^2}\exp\left\{-\frac{k^2}{3}\langle r^2(t)\rangle_Q\right\}dk, \quad (17)$$

where $\psi = \rho_m d^3(d/b)^3 g^2(d)\hat{S}(0)$ is a dimensionless parameter characterizing the strength of the entanglement effect on the chain dynamics.

With the new dimensionless variable

$$q \equiv \sqrt{\frac{1}{3}k^2\langle r^2(t)\rangle_Q}, \quad (18)$$

Eq. (17) becomes

$$\Gamma_p(t) = \frac{16}{3\sqrt{3}\pi^2}\frac{\psi b^3}{\langle r^2(t)\rangle_Q^{3/2}}\frac{1}{\tau_s} \times \int_0^{\sqrt{\langle r^2(t)\rangle_Q/(3b^2)}} \frac{q^6\exp\{-q^2\}}{q^4 + \left(\frac{2\pi p\langle r^2(t)\rangle_Q}{Nb^2}\right)^2}dq. \quad (19)$$

Using the normal-coordinate autocorrelation function

$$C_p(t) = \langle \mathbf{X}_p(t)\mathbf{X}_p(0)\rangle, \quad (20)$$

Eq. (8) can be converted into

$$\frac{\partial}{\partial t}C_p(t) + \int_0^t \Gamma_p(t-\tau)\frac{\partial C_p(\tau)}{\partial \tau}d\tau = -\frac{p^2}{\tau_1^R}C_p(t). \quad (21)$$

Converting this equation with the aid of the Fourier–Laplace transform

$$\hat{C}_p(\omega) = \int_0^\infty C_p(t)e^{i\omega t}dt, \quad (22)$$

leads to the solution

$$\hat{C}_p(\omega) = \frac{C_p(0)}{\frac{p^2}{\tau_1^R[1+\hat{\Gamma}_p(\omega)]} - i\omega}, \quad (23)$$

where $C_p(0) = Nb^2/(2\pi^2p^2)$ and $\hat{\Gamma}_p(\omega) = \int_0^\infty \Gamma_p(\omega)e^{i\omega t}dt$ [2]. The inverse transformation results in

$$C_p(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{C}'_p(\omega)e^{-i\omega t}d\omega, \quad (24)$$

where $\hat{C}'_p(\omega) = 2\mathrm{Re}\{\lim_{\epsilon\to 0}\hat{C}_p(\omega + i\epsilon)\}$.

2.3. General Expressions for the Normal-Mode Relaxation Times

The relaxation time of normal mode number p is defined as

$$\tau_p \equiv \frac{1}{C_p(0)}\int_0^\infty C_p(t)dt = \frac{\hat{C}_p(0)}{C_p(0)} = \frac{\tau_1^R[1+\hat{\Gamma}_p(0)]}{p^2} = \tau_s\left(\frac{N}{p}\right)^2[1+\hat{\Gamma}_p(0)]. \quad (25)$$

In the limit of small entanglement effects, i.e., $\hat{\Gamma}_p(0) \ll 1$, Rouse-like dynamics is expected with the mode relaxation times $\tau_p \simeq \tau_s(N/p)^2$. In the opposite case (when the memory function term $\hat{\Gamma}_p(0)$ on the right-hand side of Eq. (25) dominates), we have $\tau_p \simeq \tau_s(N/p)^2\hat{\Gamma}_p(0)$. The factor $\hat{\Gamma}_p(0)$ slows the relaxation process down due to the entanglement effect. The so-called terminal relaxation time defined for the mode $p = 1$ is

$$\tau_{\max} \equiv \tau_1 \simeq \tau_s N^2\hat{\Gamma}_1(0). \quad (26)$$

2.4. General Expression for the Segment Mean Squared Segment Displacement

The mean squared displacement of the nth Kuhn segment, $\left\langle(\mathbf{R}_n(t) - \mathbf{R}_n(0))^2\right\rangle$, can be expressed in normal coordinates with the aid of Eq. (7). The result is

$$\left\langle(\mathbf{R}_n(t) - \mathbf{R}_n(0))^2\right\rangle = \left\langle(\mathbf{X}_0(t) - \mathbf{X}_0(0))^2\right\rangle + 8\sum_{p=1}^{N-1}[C_p(0) - C_p(t)]\cos^2\left(\frac{\pi}{N}pn\right) \equiv g_{cm}(t) + g_{rel}^n(t). \quad (27)$$

Note that only autocorrelation functions of the normal coordinates occur, whereas cross-correlations between normal coordinates of different orders vanish. The first contribution to Eq. (27) represents the mean squared displacement of the center-of-mass, $g_{cm}(t)$; the second term refers to displacements relative to the center-of-mass. Averaging Eq. (27) gives

$$\langle r^2(t)\rangle \equiv \left\langle\left\langle(\mathbf{R}_n(t) - \mathbf{R}_n(0))^2\right\rangle\right\rangle = \left\langle(\mathbf{X}_0(t) - \mathbf{X}_0(0))^2\right\rangle + 4\sum_{p=1}^{N-1}[C_p(0) - C_p(t)] \equiv g_{cm}(t) + g_{rel}(t), \quad (28)$$

where $g_{rel}(t) \equiv (1/N)\sum_{n=1}^N g_{rel}^n$.

2.5. General Expression for the Center-of-Mass Mean Squared Displacement

On the time scale $t \gtrsim \tau_{\max} \equiv \tau_1$ (when the center-of-mass vector $\mathbf{X}_0(t)$ begins to vary perceptibly), the zero-order memory function $\Gamma_0(t)$ is already approaching

zero. The left-hand side of Eq. (8) for $p = 0$ may therefore be approximated in the so-called Markovian limit, i.e., $t \gg \tau_{\max} \equiv \tau_1$, as

$$\begin{aligned}\zeta\left[\frac{\partial \mathbf{X}_0(t)}{\partial t} + \int_0^t \Gamma_0(t-\tau)\frac{\partial \mathbf{X}_0(\tau)}{\partial \tau}d\tau\right] \\ \simeq \zeta\left[\frac{\partial \mathbf{X}_0(t)}{\partial t} + \left(\int_0^\infty \Gamma_0(t)d\tau\right)\frac{\partial \mathbf{X}_0(t)}{\partial \tau}\right] \\ = \zeta[1 + \hat{\Gamma}_0(0)]\frac{\partial \mathbf{X}_0(t)}{\partial \tau} = F_0^Q(t).\end{aligned} \tag{29}$$

In this long time limit, the center-of-mass motion obviously follows the Rouse equation of motion if the segment friction coefficient is replaced by

$$\zeta^* = \zeta[1 + \hat{\Gamma}_0(0)]. \tag{30}$$

The center-of-mass self-diffusion coefficient therefore obeys

$$D = \frac{k_B T}{N\zeta^*} = \frac{k_B T}{N\zeta[1 + \hat{\Gamma}_0(0)]} \simeq \frac{k_B T}{N\zeta\hat{\Gamma}_0(0)}, \tag{31}$$

where the latter equality is reached in the limit of a strong entanglement effect, that is, $\hat{\Gamma}_0(0) \gg 1$.

It may be of interest to consider the product

$$D\tau_1 = \frac{k_B T}{N\zeta}\frac{[1 + \hat{\Gamma}_1(0)]}{[1 + \hat{\Gamma}_0(0)]}N^2\tau_s = \frac{[1 + \hat{\Gamma}_1(0)]}{[1 + \hat{\Gamma}_0(0)]}\frac{Nb^2}{3\pi^2}. \tag{32}$$

For the original Rouse model, $\Gamma_p(t) \equiv 0$ applies so that Eq. (32) takes the form $D_R\tau_1^R = Nb^2/(3\pi^2)$. An equivalent expression is valid for the reptation model [1]: $D_{rep}\tau_d = Nb^2/(3\pi^2)$, where τ_d is the so-called tube disengagement time. These relations are to be compared with the result of the present treatment for the strong entanglement limit,

$$D\tau_1 \simeq \frac{\hat{\Gamma}_1(0)}{\hat{\Gamma}_0(0)}\frac{Nb^2}{3\pi^2} < \frac{Nb^2}{3\pi^2} = \frac{2}{\pi^2}R_g^2, \tag{33}$$

where R_g is the radius of gyration of the polymer chain. The above inequality is the consequence of $\hat{\Gamma}_1(0) < \hat{\Gamma}_0(0)$ [compare with Eq. (19)]. That is, the present formalism predicts a smaller mean squared displacement in the terminal relaxation time τ_1 than the Rouse and the reptation models.

On a time scale short in comparison to the decay time of the memory function $\Gamma_0(t)$, center-of-mass diffusion becomes anomalous. The reasons for this may be elucidated by considering the relation

$$g_{cm}(t) = \lim_{p\to 0}(4[C_p(0) - C_p(t)]). \tag{34}$$

Rewriting Eq. (21) in the pseudo-Markovian approximation gives

$$\begin{aligned}\frac{\partial}{\partial t}C_p(t) + \int_0^t \Gamma_p(t-\tau)\frac{\partial}{\partial \tau}C_p(\tau)d\tau \\ \simeq \frac{\partial}{\partial t}C_p(t) + \left[\int_0^t \Gamma_p(\tau)d\tau\right]\frac{\partial}{\partial t}C_p(t) = -\frac{p^2}{\tau_1^R}C_p(t).\end{aligned} \tag{35}$$

As a formal solution, we may write

$$C_p(t) = C_p(0)\exp\left\{-\frac{p^2}{\tau_1^R}\int_0^t \frac{d\tau}{\tilde{\Gamma}_0(\tau)}\right\}, \tag{36}$$

where

$$\tilde{\Gamma}_0(\tau) \equiv 1 + \int_0^t \Gamma_0(\tau)d\tau. \tag{37}$$

Substituting $C_p(t)$ in Eq. (34) correspondingly leads to the center-of-mass mean squared displacement

$$\begin{aligned}g_{cm}(t) = \lim_{p\to 0}\left[4\frac{Nb^2}{2\pi^2 p^2}\left(1 - \exp\left\{-\frac{p^2}{\tau_1^R}\int_0^t \frac{d\tau}{\tilde{\Gamma}_0(\tau)}\right\}\right)\right] \\ = \frac{2}{\pi^2}\frac{b^2}{\tau_s N}\int_0^t \frac{d\tau}{\tilde{\Gamma}_0(\tau)}.\end{aligned} \tag{38}$$

In the limit $\int_0^t \Gamma_0(\tau)d\tau \ll 1$, ordinary diffusion with the Rouse diffusion coefficient results in

$$g_{cm}(t) = 6Dt = \frac{2}{\pi^2}\frac{b^2}{\tau_s N}t. \tag{39}$$

This limit may also be characterized by defining a sort of entanglement time for the center-of-mass diffusion (τ_e^{cm}) given by the equation $\tilde{\Gamma}_0(\tau_e^{cm}) = 1$. The condition for Eq. (39) then reads $t \ll \tau_e^{cm}$.

On the basis of Eq. (37), another characteristic time,

$$\tau^{cm} \equiv (\Gamma_0(0))^{-1}\int_0^\infty \Gamma_0(\tau)d\tau,$$

may be defined. For $t \gg \tau^{cm}$, the mean squared displacement by center-of-mass diffusion again becomes normal but with a diffusion coefficient determined by Eq. (31):

$$g_{cm}(t) = 6Dt = \frac{2}{\pi^2}\frac{b^2}{\tau_s N}\frac{t}{1 + \hat{\Gamma}_0(0)}. \tag{40}$$

By contrast, in the intermediate range ($\tau_e^{cm} \ll t \ll \tau^{cm}$), center-of-mass diffusion is expected to deviate from this linear time dependence. This finding may be compared with the anomalous center-of-mass diffusion behavior predicted in the frame of the reptation model [2] in the limit $\tau_e \ll t \ll \tau_1^R$, namely, $g_{cm}^{rep}(t) \propto (b^2 N_e/N)\sqrt{t/\tau_s}$, where τ_e and N_e represent the so-called entanglement time and the number of Kuhn segments in the so-called entanglement length.

2.6. General Expression for the Autocorrelation Function of the Tangent Vector

In our previous papers [7, 18], we showed that the spin-lattice relaxation time (see, e.g., [8]) of polymers in the limit $t \gg \tau_s$ can be traced back to the autocorrelation function of the Kuhn-segment tangent vector in the continuum limit:

$$\mathbf{b}_n(t) \equiv \frac{\partial \mathbf{R}_n(t)}{\partial n} = -\frac{2\pi}{N}\sum_{p=1}^{N-1} \mathbf{X}_p p \sin\left(\frac{\pi p n}{N}\right). \quad (41)$$

The autocorrelation function of the tangent vector thus becomes

$$\langle \mathbf{b}_n(t)\mathbf{b}_n(0)\rangle = \frac{4\pi^2}{N^2}\sum_{p=1}^{N-1} p^2 C_p(t)\sin^2\left(\frac{\pi}{N}pn\right). \quad (42)$$

Averaging over all Kuhn segments in a chain leads to

$$\overline{\langle \mathbf{b}_n(t)\mathbf{b}_n(0)\rangle} = \frac{2\pi^2}{N^2}\sum_{p=1}^{N-1} p^2 C_p(t). \quad (43)$$

3. RENORMALIZED ROUSE FORMALISM

All expressions derived so far more or less depend directly on the memory function $\Gamma_p(t)$, that is, on the projected mean squared segment displacement $\langle r^2(t)\rangle_Q$ via Eq. (17). Schweizer, in his Renormalized Rouse (RR) Formalism [3], suggested for the limit $t \ll \tau_1^R$ that

$$\langle r^2(t)\rangle_Q \simeq \langle r^2(t)\rangle_R = \frac{6D_0(t)}{N} + \frac{6Nb^2}{3\pi^2}\sum_{p=1}^{N-1}\frac{1}{p^2}\left(1 - \exp\left\{-\frac{tp^2}{\tau_1^R}\right\}\right), \quad (44)$$

where $D_0 = k_B T/\zeta = b^2/(3\pi^2\tau_s)$ is the segmental diffusion coefficient. That is, the projected mean squared segment displacement $\langle r^2(t)\rangle_Q$ is replaced by that expected on the basis of the original Rouse model [2]: $\langle r^2(t)\rangle_R$.

The Renormalized Rouse Formalism [3] was introduced by Schweizer *et al.* as a basic element for the Polymer Mode–Mode Coupling concept [4, 16]. In those papers, the mode-number-dependent term $(2\pi p\langle r^2(t)\rangle_Q/(nb^2))^2$ in expressions corresponding to Eq. (19) was neglected. However, in our previous paper [7], we showed that this term is quite essential for the normal-mode relaxation of entangled polymers. The theoretically predicted mean squared segment displacement takes a considerably modified form, and the experimentally well established regions (I and II) of the frequency dispersion of the entangled-polymer melts [19] can only be explained on this basis [7, 20].

Equation (44) for the Rouse mean squared segment displacement has the limits

$$\langle r^2(t)\rangle_R \simeq \frac{2}{\pi^{3/2}}b^2\left(\frac{t}{\tau_s}\right)^{1/2} \quad \text{for } \tau_s \ll t \ll \tau_1^R, \quad (45)$$

$$\langle r^2(t)\rangle_R \simeq \frac{1}{3}Nb^2 + \frac{2}{\pi^2}\frac{b^2}{N}\left(\frac{t}{\tau_s}\right) \quad \text{for } t \gg \tau_1^R. \quad (46)$$

Here, we have made use of the asymptotic expressions

$$\frac{6Nb^2}{3\pi^2}\sum_{p=1}^{N-1}\frac{1}{p^2}\left(1 - \exp\left\{-\frac{tp^2}{\tau_1^R}\right\}\right) \simeq \frac{2}{\pi^{3/2}}b^2\left(\frac{t}{\tau_s}\right)^{1/2} \quad (47)$$

$$\text{for } \tau_s \ll t \ll \tau_1^R,$$

$$\frac{6Nb^2}{3\pi^2}\sum_{p=1}^{N-1}\frac{1}{p^2}\left(1 - \exp\left\{-\frac{tp^2}{\tau_1^R}\right\}\right) \simeq \frac{1}{3}Nb^2 = 2R_g^2 \quad (48)$$

$$\text{for } t \gg \tau_1^R.$$

In combination, Eqs. (45), (46) may be expressed as

$$\langle r^2(t)\rangle_R \simeq \frac{2}{\pi^2}\frac{b^2}{N}\left(\frac{t}{\tau_s}\right) + \frac{2}{\pi^{3/2}}b^2\left(\frac{t}{\tau_s}\right)^{1/2}\frac{1}{1 + \dfrac{6(t/\tau_s)^{1/2}}{\pi^{3/2}N}}. \quad (49)$$

Below, explicit evaluations of expressions for the observables will be carried out in the frame of the high- and low-number limits $p > N/(6\pi)$ and $p < N/(6\pi)$, respectively.

3.1. High-Mode-Number Limit of the RR Formalism

The maximum value of the integration variable in Eq. (19) is given by $q_{\max} = \sqrt{\langle r^2(t)\rangle_Q/(3b^2)}$. The condition that the q^4 term in the denominator of the integrand can safely be neglected therefore is

$$\left(\frac{2\pi\langle r^2(t)\rangle_Q}{Nb^2}\right)^2 \gg q_{\max}^4 \quad \text{or} \quad p \gg \frac{N}{6\pi}. \quad (50)$$

This is the definition of a high-mode-number limit. The maximum mode number possible is $p_{\max} = N - 1 \simeq N$, so that the range $N/(6\pi) < p < N$ is under consideration. The range in the time or frequency domains in which certain power laws show up in the experiments carried out when the high-mode-number limit is relevant is called "region I" [7, 20]. It will be defined below. Furthermore, it will be shown that the region-I power laws appear only if the parameter ψ complies to the condition given in Eq. (65).

For short times ($t \longrightarrow 0$ or $\langle r^2(t)\rangle_Q \longrightarrow 0$), the memory function in the high-mode-number limit asymptotically approaches the value

$$\Gamma_p(0) = \frac{4}{7\pi^4 3^5}\frac{\psi}{\tau_s}\left(\frac{N}{p}\right)^2 \simeq 2.4\times 10^{-5}\frac{\psi}{\tau_s}\left(\frac{N}{p}\right)^2. \quad (51)$$

In the case of longer times ($\tau_s \ll t \ll \tau_1^R$), the exponential function $\exp\{-q^2\}$ in the integrand of the memory function at Eq. (19) approaches zero before the upper integration limit is reached. The upper integration limit may therefore be replaced by ∞. Again, neglecting the q^4 term permits one to solve the integral analytically. The resulting memory function for the high-mode-number limit is

$$\Gamma_p^{RR}(t) = \frac{5}{4\pi^{7/2}3^{1/2}}\frac{\psi b^7}{\tau_s\langle r^2(t)\rangle_Q^{7/2}}\left(\frac{N}{p}\right)^2. \quad (52)$$

Replacing $\langle r^2(t)\rangle_Q$ by the Rouse mean squared segment displacement for $\tau_s \ll t \ll \tau_1^R$ [see Eq. (45)] in Eq. (52) gives

$$\begin{aligned}\Gamma_p^{RR}(t) &= \frac{5\pi^{7/4}}{32\sqrt{6}}\frac{\psi}{\tau_s}\left(\frac{N}{p}\right)^2\left(\frac{\tau_s}{t}\right)^{7/4}\\ &\simeq 0.47\frac{\psi}{\tau_s}\left(\frac{N}{p}\right)^2\left(\frac{\tau_s}{t}\right)^{7/4}.\end{aligned} \quad (53)$$

This result may be combined with the limiting expression Eq. (51) in the form

$$\begin{aligned}\Gamma_p^{RR}(t) &\simeq \frac{4}{7\pi^4 3^5}\frac{\psi}{\tau_s}\left(\frac{N}{p}\right)^2\frac{1}{1+\dfrac{128\sqrt{6}}{35\times 3^5\pi^{23/4}}\left(\dfrac{t}{\tau_s}\right)^{7/4}}\\ &\simeq 2.4\times 10^{-5}\frac{\psi}{\tau_s}\left(\frac{N}{p}\right)^2\frac{1}{1+5.1\times 10^{-1}\left(\dfrac{t}{\tau_s}\right)^{7/4}}.\end{aligned} \quad (54)$$

For calculating the normal-mode relaxation times [Eq. (25)], we need $\hat{\Gamma}_p^{RR}(0)$:

$$\begin{aligned}\hat{\Gamma}_p^{RR}(0) &= \int_0^\infty \Gamma_p^{RR}(t)dt = \frac{4}{7\pi^4 3^5}\psi\left(\frac{N}{p}\right)^2\\ &\times\left[\frac{\pi^{23/4}3^5 35}{128\sqrt{6}}\right]^{4/7}\frac{\pi}{\dfrac{7}{4}\sin\left(\dfrac{4\pi}{7}\right)} \simeq 0.05\psi\left(\frac{N}{p}\right)^2.\end{aligned} \quad (55)$$

The normal-mode relaxation times consequently are

$$\tau_p^{RR} \simeq \tau_s\left(\frac{N}{p}\right)^2\left[1+0.05\psi\left(\frac{N}{p}\right)^2\right]. \quad (56)$$

In the limit

$$\frac{N}{4\pi} < p < 0.22\sqrt{\psi}N, \quad (57)$$

the normal-mode relaxation times calculated on the basis of the renormalized Rouse formalism are much longer than those predicted by the original Rouse model. As a rough scaling law, one may write

$$\tau_p^{RR} \simeq 0.05\psi\tau_s\left(\frac{N}{p}\right)^4. \quad (58)$$

Equation (54) shows that the memory function $\Gamma_p^{RR}(t)$ decays rapidly in the high-mode-number limit. Therefore, it is permissible to approximate $\hat{\Gamma}_p^{RR}(\omega) \simeq \hat{\Gamma}_p^{RR}(0)$ (compare [7]). This is equivalent to writing for the normal-mode correlation function

$$C_p(t) \simeq C_p(0)\exp\left\{-\frac{t}{\tau_p^{RR}}\right\}. \quad (59)$$

Using the formalism described in the appendix of [7], the combination of Eqs. (28), (57), and (59) leads to the mean squared segment displacement

$$\left\langle(\mathbf{R}_n(t)-\mathbf{R}_n(0))^2\right\rangle_{RR} \simeq g_{rel} \simeq 0.5b^2\left(\frac{t}{\psi\tau_s}\right)^{1/4}, \quad (60)$$

where the brackets represent all averages considered above in this context. The autocorrelation function of the segment tangent vector of the nth Kuhn segment becomes

$$\langle\mathbf{b}_n(t)\mathbf{b}_n(0)\rangle_{RR} \simeq 0.4b^2\left(\frac{\tau_s\psi}{t}\right)^{1/4}. \quad (61)$$

This power law was derived for the normal-mode-number range Eq. (57). For experimental purposes, it is of interest to express this range in terms of frequency and

time domains by recalling that the main contribution to the integral in Eq. (24) arises from frequencies satisfying

$$\frac{p^2}{\tau_1^R \hat{\Gamma}_p^{RR}(\omega)} \simeq \left(\frac{p}{N}\right)^4 \frac{1}{0.05\psi}\frac{1}{\tau_s} \propto \omega \tag{62}$$

[see Eq. (23)]. On this basis, the inequality given in Eq. (57) can be converted into

$$\frac{1}{(6\pi)^4 0.05\psi} \ll \omega\tau_s \ll 0.05\psi \tag{63}$$

or

$$\frac{1}{0.05\psi} \ll \frac{t}{\tau_s} \ll (6\pi)^4 0.05\psi. \tag{64}$$

This defines "region I" of the frequency or time domains in the frame of the high-mode-number limit [$p > N/(6\pi)$]. Note that the corresponding mode-number range [Eq. (57)] exists only for satisfactorily strong entanglement effects, that is,

$$\frac{N}{6\pi} \ll 0.22\sqrt{\psi} \quad \text{and} \quad \psi \gg \frac{0.05}{(6\pi)^2} \simeq 1.4\times 10^{-4}. \tag{65}$$

Region I shows up only in time or frequency domain experiments if this condition is fulfilled.

3.2. Low-Mode-Number Limit of the RR Formalism

The low-mode-number limit, $p \ll N/(6\pi)$, universally leads to certain power laws in the time or frequency domains in a range called "region II." More illustratively, this limit is defined by the condition that the root mean squared segment displacements exceed the Kuhn segment length *b*. The upper integration limit in Eq. (19) can therefore be equated with ∞. Following the renormalized Rouse approach and replacing $\langle r^2(t)\rangle_Q$ with the Rouse expression given in Eq. (45) for the time limit $\tau_s \ll t \ll \tau_1^R$ gives

$$\Gamma_p^{RR}(t) = \frac{8\pi^{1/4}}{3\sqrt{6}}\frac{\psi}{\tau_s}\left(\frac{\tau_s}{t}\right)^{3/4}\int_0^\infty \frac{q^6 \exp\{-q^2\}}{q^4 + \frac{16}{\pi}\frac{p^2 t}{N^2\tau_s}}dq. \tag{66}$$

For the calculation of the normal-mode relaxation times [Eq. (56)], we need the quantity

$$\begin{aligned}\hat{\Gamma}_p^{RR}(0) &= \int_0^\infty \Gamma_p^{RR}(t)dt \\ &= 2\left(\frac{\pi}{3}\right)^{3/2}\psi\sqrt{\frac{N}{p}} \simeq 2.14\psi\sqrt{\frac{N}{p}}.\end{aligned} \tag{67}$$

According to Eq. (56), this is the term representing the entanglement effects in the normal-mode friction coefficient. In the low-mode-number limit of the renormalized Rouse formalism, the normal-mode relaxation times consequently obey

$$\tau_p^{RR} = \tau_s\left(\frac{N}{p}\right)^2\left[1 + 2.14\psi\sqrt{\frac{N}{p}}\right] \simeq 2.14\psi\tau_s\left(\frac{N}{p}\right)^{2.5}, \tag{68}$$

where the latter approximation holds for the case of very strong entanglement effects.

In order to calculate the mean squared segment displacement in the low-mode-number limit of the renormalized Rouse formalism, we need the normal-coordinate correlation functions $C_p(t)$ [see Eqs. (28) and (20)]. These are given in Eqs. (23) and (24). The decay time scale of $C_p(t)$ is determined by the normal-mode relaxation times τ_p^{RR}. According to Eq. (66), the fast decay of the memory functions $\Gamma_p^{RR}(t)$ begins for $t \gg \frac{\pi}{16}\tau_p^R$. For large *N*, the inequality $\tau_p^{RR} \gg \frac{\pi}{16}\tau_p^R$, so that $\hat{\Gamma}_p(\omega) \simeq \hat{\Gamma}_p(0)$ in Eq. (23). This is equivalent to use the exponential approximation for $C_p(t)$, i.e., Eq. (59), where the normal-mode relaxation times are now given by Eq. (68). In the limit $N \longrightarrow \infty$, this exponential approximation asymptotically approaches the exact expression. For finite *N*, the normal-coordinate correlation function decays nonexponentially, as revealed by numerical evaluations. This is even more true with a larger *N*/*p* ratio. The detailed behavior $C_p(t)$ will be investigated in a forthcoming study.

Inserting Eq. (68) into the exponential function given in Eq. (59) and this into Eq. (28) leads to

$$\begin{aligned}\langle r^2(t)\rangle_{RR} &\equiv \left\langle\left\langle (\mathbf{R}_n(t) - \mathbf{R}_n(0))^2\right\rangle\right\rangle_{RR} \simeq g_{rel}^{RR}(t) \\ &= \left(\frac{6}{\pi}\right)^{3/5}\frac{\Gamma(3/5)}{\pi^2}\left(\frac{t}{\psi\tau_s}\right)^{2/5} \simeq 0.23 b^2\left(\frac{t}{\psi\tau_s}\right)^{2/5}\end{aligned} \tag{69}$$

for $t \ll \tau_1^{RR}$. With the accuracy of the numerical coefficients, this result coincides with Eq. (69) of our previous treatment [7] but is at variance with Schweizer's original prediction [3]. This discrepancy demonstrates that the mode-number dependence, which was neglected in Schweizer's formalism, is quite crucial.

The correlation function of the segment tangent vector, Eq. (43), is analogously obtained with the aid of exponential normal-coordinate correlation functions $C_p(t)$ [Eq. (59)] and the normal-mode relaxation times Eq. (68) in the form

$$\langle \mathbf{b}_n(t)\mathbf{b}_n(0)\rangle_{RR} \simeq 1.03 b^2\left(\frac{\tau_s\psi}{t}\right)^{2/5} \tag{70}$$

for $t \ll \tau_1^{RR}$.

3.3. RR Theoretical Mean Squared Displacement for the Whole Time Scale

The mean squared segment displacement given in Eq. (69) applies to the short time limit $t \ll \tau_1^{RR}$. For the twice-renormalized Rouse formalism to be described in the subsequent section, we need an expression valid in the whole time scale of the relevant range. In the long time limit ($t \longrightarrow \infty$), the mean squared displacement relative to the center of mass generally obeys

$$\langle r^2(t \longrightarrow \infty)\rangle_{rel} = g_{rel}(t \longrightarrow \infty) = 2R_g^2 = \frac{1}{3}Nb^2. \quad (71)$$

The long time limit of the mean squared center-of-mass displacements can be derived from

$$\langle r^2(t \longrightarrow \infty)\rangle_{cm} = g_{cm}(t \longrightarrow \infty) = 6Dt, \quad (72)$$

where the center-of-mass diffusion coefficient is given according to Eq. (31) as a function of the frequency-domain memory kernel term $\hat{\Gamma}_0^{RR}(0)$. Taking Eq. (19) for $p = 0$ and approximating $\langle r^2(t)\rangle_Q$ by the mean squared displacement found in the frame of the ordinary Rouse model [Eq. (49)] gives

$$\hat{\Gamma}_0^{RR}(0) = \frac{16}{3\sqrt{3}\pi^2}\frac{\psi b^2}{\tau_s}\int_0^\infty \frac{dt}{\langle r^2(t)\rangle_R^{3/2}} \times \int_0^{\sqrt{\langle r^2(t)\rangle_R/(3b^2)}} q^2\exp\{-q^6\}dq \simeq 4.8\psi\sqrt{N}, \quad (73)$$

so that the center-of-mass diffusion coefficient in the Renormalized Rouse approach reads

$$D^{RR} = \frac{k_BT}{\zeta N[1+4.8\psi\sqrt{N}]} \simeq \frac{k_BT}{4.8\zeta\psi\sqrt{N^{3/2}}} \simeq \frac{0.21}{3\pi^2}\frac{b^2}{\psi\tau_s N^{3/2}}. \quad (74)$$

Within the accuracy of the numerical coefficient, this result coincides with that reported by Schweizer [3]. The anomalous part of center-of-mass diffusion can be derived from Eq. (38) as

$$g_{cm}^{RR}(t) = \frac{\sqrt{6}}{\pi^{13/4}}\frac{1}{\psi N}\left(\frac{t}{\tau_s}\right)^{3/4} \quad \text{for } \tau_e^{cm} \ll t \ll \tau^{cm}, \quad (75)$$

again in accordance with [3]. In total, the renormalized Rouse formalism predicts the following regimes for the mean squared center-of-mass diffusion:

$$g_{cm}^{RR} = \frac{1}{3\pi^2}\frac{b^2}{N}\frac{1}{\tau_s} \quad \text{for } t \ll \tau_e^{cm}, \quad (76)$$

$$g_{cm}^{RR} = \frac{0.06}{\psi N}\left(\frac{t}{\tau_s}\right)^{3/4} \quad \text{for } \tau_e^{cm} \ll t \ll \tau^{cm}, \quad (77)$$

$$g_{cm}^{RR} = 6D^{RR}t \quad \text{for } t \gg \tau^{cm}. \quad (78)$$

The anomalous part of the mean squared center-of-mass displacements and diffusion in the high-mode-number limit may be neglected in the following because the mean squared displacement relative to the center of mass, $g_{rel}(t)$—that is, the second term in Eq. (79)—dominates when the anomalous segment diffusion becomes perceptible. Thus, combining the limits given in Eqs. (69), (71), and (74) in analogy to the derivation of Eq. (49) then leads to the average mean squared displacement in the frame of the renormalized Rouse formalism

$$\langle r^2(t)\rangle_{RR} \equiv \left\langle\left\langle (\mathbf{R}_n(t) - \mathbf{R}_n(0))^2\right\rangle\right\rangle_{RR} = g_{cm}^{RR}(t) + g_{rel}^{RR}(t) \simeq \frac{0.21}{3\pi^2}\frac{b^2t}{\psi\tau_s N^{3/2}} + \frac{0.23b^2(t/(\psi\tau_s))^{2/5}}{1+\dfrac{0.69(t/(\psi\tau_s))^{2/5}}{N}}. \quad (79)$$

4. TWICE-RENORMALIZED ROUSE FORMALISM

The renormalized Rouse approach fails to describe the chain dynamics at long times, that is, in the realm of the longest Rouse mode relaxation times, when molecular-weight dependences come into play. We therefore try a second renormalization step. This twice-renormalized Rouse (TRR) formalism is based on the general memory function expression in Eq. (19). However, the projected mean squared displacement in that function $\langle r^2(t)\rangle_Q$ is now replaced by the result of the renormalized Rouse formalism $\langle r^2(t)\rangle_{RR}$ given in Eq. (79) instead of formula (44) for $\langle r^2(t)\rangle_R$ derived on the basis of the ordinary Rouse model. We again distinguish the high- and low-number limits as before.

4.1. High-Mode-Number Limit of the TRR Formalism

Equation (51) for the limit $t \longrightarrow 0$ is still valid because it is independent of chain dynamics, which becomes important only on the time scale $t \gg \tau_s$. That is, the second renormalization only affects the time dependence of the memory function given in Eq. (52). This expression is now modified according to the mean squared displacement of the renormalized Rouse the-

ory, i.e., Eq. (79) [implying the result Eq. (69)]. Instead of Eq. (53) we now have for $t \gg \tau_s$

$$\Gamma_p^{TRR}(t) = \sqrt{\frac{\pi}{3}}\frac{5}{4\pi^4}\frac{\psi}{\tau_s}\frac{b^7\left(\frac{N}{p}\right)^2}{[0.23b^2(t/(\psi\tau_s))^{2/5}]^{7/2}} \simeq 2.13\frac{\psi}{\tau_s}\left(\frac{N}{p}\right)^2\left(\frac{\psi\tau_s}{p}\right)^{7/5}. \quad (80)$$

Taking into account the short and long time limits, Eqs. (51), (80) lead to the combined approach

$$\Gamma_p^{TRR}(t) \simeq 2.4\times 10^{-5}\frac{\psi}{\tau_s} \times\left(\frac{N}{p}\right)^2\frac{1}{1+1.12\times 10^{-5}\left(\frac{t}{\psi\tau_s}\right)^{7/5}}. \quad (81)$$

Thus,

$$\hat{\Gamma}_p^{TRR}(0) = \int_0^\infty \Gamma_p^{TRR}(t)dt \simeq 0.148\psi^2\left(\frac{N}{p}\right)^2. \quad (82)$$

According to Eq. (95), the normal-mode relaxation times thus become in the high-mode-number limit

$$\tau_p^{TRR} \simeq \tau_s\left(\frac{N}{p}\right)^2\left[1+0.148\psi^2\left(\frac{N}{p}\right)^2\right]. \quad (83)$$

Analogous to relations (60) and (61), the mean squared segment displacement and the segment tangent vector correlation function are described in the frame of the twice-renormalized Rouse formalism by

$$\left\langle(\mathbf{R}_n(t)-\mathbf{R}_n(0))^2\right\rangle_{TRR} \simeq g_{rel}(t) \simeq 0.4b^2\left(\frac{t}{\psi\tau_s}\right)^{1/4} \quad (84)$$

and

$$\langle \mathbf{b}_n(t)\mathbf{b}_n(0)\rangle_{TRR} \simeq 0.5b^2\left(\frac{\tau_s\psi}{t}\right)^{1/4}, \quad (85)$$

respectively. Equations (83), (84), and (85) obviously deviate from their renormalized Rouse theoretical counterparts—Eqs. (56), (60) and (61), respectively—merely by numerical factors. That is, the renormalized Rouse approach already provides a good description of the phenomena to be expected in the time/frequency domain region I [7].

4.2. Low-Mode-Number Limit of the TRR Formalism

The memory function is obtained by inserting the expression for the mean squared segment displacement in the frame of the renormalized Rouse treatment in the low-mode-number limit—Eq. (69)—into Eq. (19). The result is

$$\Gamma_p^{TRR}(t) = \frac{16}{3\sqrt{3}\pi^2}\frac{\psi b^2}{\kappa^{3/2}}\frac{1}{\tau_s}\left(\frac{\psi\tau_s}{t}\right)^{3/5} \times\int_0^\infty \frac{q^6\exp\{-q^2\}}{q^4+\frac{4\pi^2p^2\kappa^2}{NB^4}\left(\frac{t}{\psi\tau_s}\right)^{4/5}}dq, \quad (86)$$

where $\kappa = (6/\pi)^{3/5}\Gamma(3/5)/\pi^2 \simeq 0.23$ is a numerical constant. Thus,

$$\hat{\Gamma}_p^{TRR}(0) = \int_0^\infty \Gamma_p^{TRR}(t)dt = \frac{16}{3\sqrt{3}\pi^2}\frac{\psi b^2}{\kappa^{3/2}} \times\int_0^\infty q^2\exp\{-q^2\}dq\int_0^\infty \frac{1}{\tilde{t}^{3/5}}\frac{1}{1+\frac{4\pi^2p^2\kappa^2}{q^4}\tilde{t}^{4/5}}d\tilde{t}, \quad (87)$$

where $\tilde{t} = t/(\psi\tau_s)$. The integration gives

$$\hat{\Gamma}_p^{TRR}(0) = \frac{5\pi^5}{72\sqrt{2}(\Gamma(3/5))^{5/2}}\psi^2\frac{N}{p} \simeq 5.52\psi^2\frac{N}{p}. \quad (88)$$

Combining this result with Eq. (25) leads to the normal-mode relaxation times

$$\tau_p^{TRR} = \tau_s\left(\frac{N}{p}\right)^2\left[1+5.52\psi^2\frac{N}{p}\right] \simeq 5.52\psi^2\left(\frac{N}{p}\right)^3. \quad (89)$$

The dependence on the normal-mode number p here is obviously much stronger than predicted by the renormalized Rouse treatment [see Eq. (68)]. It also differs from what one would expect on the basis of the reptation model [2], $\tau_p^{rep} \propto n^3/p^2$.

The terminal relaxation time predicted by the low-mode-number limit of the TRR concept thus is

$$\tau_1^{TRR} \simeq 5.52\tau_s\psi N^3. \quad (90)$$

Remarkably, the chain length dependence is the same as that expected on the basis of the reptation model [1, 2] and the polymer mode coupling treatment [4].

In the long time limit ($t \gg (N^2/(4\pi^2\kappa^2p^2))^{5/4} = 0.4\tau_s\psi(N/p)^{2.5} \simeq 0.2\tau_p^{RR}$), the normal-mode autocorrelation function starts to decay substantially. Then, it is justified to take the normal-mode relaxation times as the time constants of exponential decays, i.e.,

$$C_p(t) \simeq C_0(0)\exp\left\{-\frac{t}{\tau_p^{TRR}}\right\}, \quad (91)$$

on the time scale $\tau_s\psi^{-4} \ll t \ll \tau_1^{TRR}$. Inserting this function into Eq. (28) and replacing the sum by an integral

calculated in the Appendix of [7] leads to the mean squared segment displacement in the low-mode-number limit of the TRR concept:

$$\langle r^2(t)\rangle_{TRR} \simeq g_{rel}(t) = \frac{4}{\pi^3}\frac{\Gamma(2/3)(\Gamma(3/5))^{5/6}9^{1/3}2^{1/6}}{(5\pi^2)^{1/3}} \times b^2\left(\frac{t}{\psi^2\tau_s}\right)^{1/3} \simeq 0.16b^2\left(\frac{t}{\psi^2\tau_s}\right)^{1/3}. \quad (92)$$

Likewise, inserting Eq. (91) into Eq. (43), we find the expression

$$\overline{\langle \mathbf{b}_n(t)\mathbf{b}_n(0)\rangle}_{TRR} = \frac{\pi}{6}\frac{\Gamma(1/3)}{(\Gamma(3/5))^{5/6}}\left(\frac{5\pi^2}{9\sqrt{2}}\right)^{1/3} \times b^2\left(\frac{\psi^2\tau_s}{t}\right)^{1/3} \simeq 1.58b^2\left(\frac{\psi^2\tau_s}{t}\right)^{1/3} \quad (93)$$

for the tangent vector autocorrelation function.

According to Eq. (31), we need the quantity $\hat{\Gamma}_p^{TRR}(0)$ in order to calculate the center-of-mass self-diffusion coefficient. Inserting Eq. (79) into Eq. (19) for $p = 0$ gives

$$\hat{\Gamma}_0^{TRR}(0) = \int_0^\infty \Gamma_0^{TRR}(t)dt = \frac{16}{3\sqrt{3}\pi^2}\frac{\psi b^2}{\tau_s} \times \int_0^\infty \frac{1}{\left[0.04b^2\dfrac{t}{\tau_s\psi N^{1.5}} + \dfrac{0.23b^2(t/(\tau_s\psi)^{2/5})}{1+\dfrac{0.69}{N}\left(\dfrac{t}{\tau_s\psi}\right)^{2/5}}\right]^{3/2}}dt. \quad (94)$$

The substitution

$$y = \frac{1}{N}\left(\frac{t}{\psi\tau_s}\right)^{2/5}$$

leads to

$$\hat{\Gamma}_0^{TRR}(0) \simeq 0.345\psi^2 N \times \int_0^\infty \frac{1}{\left[0.04y^{3/2} + \dfrac{0.23}{1+0.63y}\right]^{3/2}}dy \simeq 19.1\psi^2 N. \quad (95)$$

Together with Eq. (31), we thus find for the center-of-mass diffusion coefficient

$$D^{TRR} = \frac{k_BT}{N\zeta[1+19.1\psi^2N]} = \frac{1}{3\pi^2}\frac{b^2}{\tau_sN} \times \frac{1}{[1+19.1\psi^2N]} \simeq \frac{0.052}{3\pi^2}\frac{b^2}{\psi^2N^2\tau_s}. \quad (96)$$

According to Eqs. (38) and (86) for p = 0, the center-of-mass mean squared displacement obeys

$$g_{cm}^{TRR} = \frac{1}{3\pi^2}\frac{b^2}{N}\frac{t}{\tau_s} \text{ for } t \ll \tau_e,$$
$$g_{cm}^{TRR} = 0.1\frac{b^2}{\tau_sN\psi}\left(\frac{t}{\psi\tau_s}\right)^{3/5} \text{ for } \tau_e \ll t \ll \tau_1^{RR},$$
$$g_{cm}^{TRR} = 6D^{TRR}t \text{ for } t \gg \tau_1^{RR}. \quad (97)$$

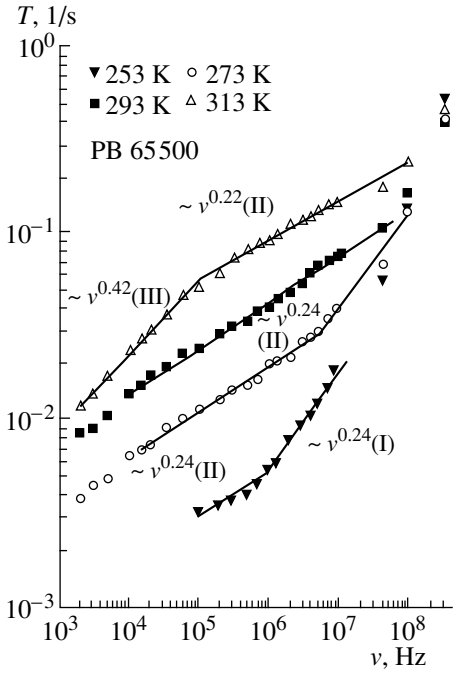
5. COMPARISON WITH EXPERIMENTAL DATA

5.1. Field-Cycling NMR Relaxometry

^{1}H and ^{2}H field-cycling NMR relaxometry [8, 21, 22] (in combination with conventional NMR methods) is a technique suitable for probing chain dynamics on the time scale 10^{-10} s $< t < 10^{-3}$ s. In the case of ^{2}H NMR, the length scale of the underlying interactions (that is, the extension of the fluctuating areas to be probed) is local throughout ("intrasegmental"). The same applies to high-frequency ^{1}H NMR, whereas low-frequency ^{1}H NMR is indicative for fluctuating intersegment and interchain nuclear dipole–dipole interactions as well. In the high-frequency/low-temperature regime (typically $\nu \gtrsim 10^8$ Hz; T < 300 K) intra-Kuhn-segment fluctuations due mainly to rotational isomerism dominate ("component A") [20]. However, these motions are intrinsically anisotropic according to the definition of a Kuhn segment. That is, residual correlations of the dipolar or quadrupolar interactions survive the intra-Kuhn-segment fluctuations. The further decay of these correlations directly reflects chain modes ("component B"). The latter relaxation phenomenon is of particular interest for comparisons with the above formalism.

To what extent the whole range of normal modes can be probed depends on the chain length of course. If the largest normal-mode relaxation time falls on the time scale of the technique, spin-lattice relaxation tends to become molecular-weight-dependent at low frequencies. However, in the context of this study, we are interested in particularly long and strongly "entangled" polymers so that the chain length does not affect the modes relevant in the experimental time/frequency window. Note that nuclear dipole–dipole interactions in the experimental time/frequency window. Note that nuclear dipole–dipole interactions which dominate ^{1}H spin-lattice relaxation are subject to fluctuations of the dipole–dipole distance and the orientation relative to the external magnetic field, whereas ^{2}H spin-lattice relaxation is governed solely by molecular reorientations.

The frequency dispersion of the spin-lattice relaxation time T_1 in melts of entangled polymers is subdivided into a number of regions in which specific power



Frequency dependence of the proton spin-lattice relaxation time of a linear-polybutadiene melt ($M_w = 65500$) at various temperatures. Regions I, II and III of the proton spin-lattice relaxation are indicated. The data above about 3×10^7 Hz are influenced by fluctuations within the Kuhn segments.

laws of a universal nature occur [19, 23]. At frequencies $\omega \ll \tau_s^{-1}$, where τ_s^{-1} is the fluctuation rate within a Kuhn segment, the following sequence of proton T_1 dispersions have been found from high to low frequencies (and low to high temperatures):

$$\begin{aligned} T_1 &\propto \omega^{0.5 \pm 0.05} \quad \text{(region I)}, \\ T_1 &\propto \omega^{0.25 \pm 0.05} \quad \text{(region II)}, \\ T_1 &\propto \omega^{0.45 \pm 0.05} \quad \text{(region III)}. \end{aligned} \tag{98}$$

For molecular weights larger than the critical value ($M_w \gg M_c$), no molecular weight dependence was observed. The crossover frequencies are functions of the polymer species and the temperature. The figure shows the T_1 frequency dispersion of a polybutadiene melt as an example where all three regions appear in our frequency window with the same sample, albeit at different temperatures.

Region III was shown to be due to fluctuations of the inter-Kuhn-segment nuclear dipole–dipole interaction by comparing the proton data with the deuteron T_1 dispersion observed in deuterated polymer melts of similar chain lengths [23]. In the deuteron case, the spin-lattice relaxation is of an entirely local, intra-Kuhn-segment nature, so that region III does not appear in contrast to the proton case.

In our previous papers [7, 18], the spin-lattice relaxation rate induced by chain modes was shown to be related to the segment tangent vector correlation function according to

$$\frac{1}{T_1} \propto \int_{-\infty}^{\infty} \langle \mathbf{b}_n(t)\mathbf{b}_n(0)\rangle^2 e^{-i\omega t} dt \quad (\omega\tau_s \ll 1). \tag{99}$$

Note that the segment tangent vector is indicative for the orientation of the residual dipolar coupling tensor (relevant for proton relaxation) as well as for the orientation of the residual electric-field gradient tensor (determining the coupling of nuclear quadrupole moments to the electric field gradients in the molecules). Inserting the expressions for $\langle \mathbf{b}_n(t)\mathbf{b}_n(0)\rangle$ in the high-mode-number limit of the renormalized Rouse treatment [Eq. (61)] or of the twice-renormalized Rouse concept [Eq. (85)] into Eq. (99) leads to

$$\begin{aligned} T_1^{RR} &\propto T_1^{TRR} \propto \omega^{1/2} \\ \text{for } \frac{20}{(6\pi)^4 \psi \tau_s} &\ll \omega \ll \frac{\psi}{20\tau_s} \end{aligned} \tag{100}$$

in both approximations. The comparison with the experimental high-frequency results [Eq. (98)] suggests that region I of the spin-lattice relaxation dispersion manifests the high-mode-number limits of the once- or twice-renormalized Rouse formalisms.

Likewise, combining Eqs. (70) and (93) based on the low-number limit of the RR and TRR theories, respectively, with Eq. (99) gives

$$\begin{aligned} T_1^{RR} &\propto \omega^{1/5} \text{ for } \frac{1}{\tau_1^{RR}}, \frac{1}{\tau_1^{TRR}} \ll \omega \ll \frac{20}{(6\pi)^4 \psi \tau_s}, \\ T_1^{TRR} &\propto \omega^{1/3} \text{ for } \frac{1}{\tau_1^{RR}}, \frac{1}{\tau_1^{TRR}} \ll \omega \ll \frac{20}{(6\pi)^4 \psi \tau_s}. \end{aligned} \tag{101}$$

The experimental proportionality indicated as region II in Eq. (98) can thus be considered as a manifestation of the low-mode-number limits of the once- or twice-renormalized Rouse theories. The low-frequency spin-lattice relaxation dispersion found for deuterons (purely intra-Kuhn-segment quadrupolar interactions) fits the prediction of the TRR formalism better, whereas that for protons (nuclear dipole–dipole interactions) favors the RR result. Anyway, all power law exponents evaluated from the region II T_1 dispersions of at least seven different polymer species turned out to be between the two theoretical exponents derived for the RR and TRR formalisms in the low-mode-number limit. Region II thus appears to represent just the crossover to the situation when the second renormalization becomes relevant.

Numerical evaluations of the absolute values of the spin-lattice relaxation time of polyisobutylene and polydimethylsiloxane have been carried out by Fenchenko for different frequencies within regions I and II [24]. The coincidence with the experimental data is good.

Region III of the proton spin-lattice relaxation dispersion appears when the fluctuations of intersegment nuclear dipole–dipole coupling become relevant, i.e., at very low frequencies. Fluctuations of this sort are a consequence of translational displacements of the segments bearing the interacting dipoles. That is, the time dependence of the mean squared segment displacements begins. The intersegment contribution to the spin-lattice relaxation rate is given by [23]

$$\frac{1}{T_1^{inter}} \propto \rho_s \int_{-\infty}^{\infty} \langle r^2(t)\rangle^{-3/2} e^{-i\omega t} dt, \qquad (102)$$

where ρ_s is the spin number density. Inserting the RR and TRR low-mode number limit results—Eqs. (69) and (92), respectively—into Eq. (102) gives

$$T_1^{RR} \propto \omega^{2/5} \text{ and } T_1^{TRR} \propto \omega^{1/2}. \qquad (103)$$

That is, the experimental finding represented by the third power law in Eq. (98) is again just between the predictions of the two approaches. We conclude that region III of the experimental proton T_1 dispersion reflects fluctuations of intersegment nuclear dipole–dipole interactions in the low-mode-number limit of the RR and TRR theories.

The power laws given in Eq. (103) have been derived under the assumption that the displacements of the segments on which the interacting nuclei are sitting are independent of each other. That is, multichain correlations of the displacement modes have been ignored. However, the evaluation of absolute values of the relaxation rates indicates that the displacements of the interacting nuclei are correlated. This leads to the important conclusion that, on a microscopic length scale, collective multichain modes contribute.

The experimental T_1 dispersions given in Eq. (98) for regions I, II, and III follow from the TRR formalism in quite a natural way. This is in contrast to the prediction of the reptation model [1]

$$T_1 \propto \omega^{3/4} M^0 \text{ for } 1/\tau_1^R \ll \omega \ll 1/\tau_e, \qquad (104)$$

$$T_1 \propto \omega^{1/2} M^{-1/2} \text{ for } 1/\tau_d \ll \omega \ll 1/\tau_1^R, \qquad (105)$$

where τ_e and τ_d are the entanglement and disengagement times, respectively. These combined molecular-weight and frequency dependences were never observed in isotropic polymer melts. However, preliminary results for polymer chains confined in porous glasses with pore diameters smaller than twice the radius of gyration indicate some features of the reptation concept [25]. More recently, the law given in Eq. (104) was verified with polyethylene oxide chains confined in 10-nm channels of a rigid polymer matrix [26].

There is no analytical formula predicting the spin-lattice relaxation behavior based on the polymer mode–mode coupling concept [4]. However, one can state that the molecular-weight dependence generally suggested by this concept definitely contradicts the experimental facts. It also appears that the absolute values of T_1 estimated with the polymer mode–mode coupling treatment are too large.

5.2. Residual Dipolar Coupling NMR Effects

Fluctuations of residual dipolar couplings, i.e., dipole–dipole interactions not subject to motional averaging by fast motions on the NMR time scale, have been subjected to polymer studies using NMR spin-echo methods for a while [27]. More recently, several new methods have been suggested on this basis with the aim to probe features specific to the diverse polymer dynamics theories. For instance, the dipolar correlation effect on the stimulated echo was shown to be sensitive to the power-law dipolar functions, which are of interest in this context [8, 28, 29]. Other labs have proposed to employ two-pulse spin-echo sequences [27, 30, 31] or multiple-quantum coherence selective NMR experiments [32]. All these methods have the potential to probe chain dynamics on a time scale extending that of T_1 dispersion experiments in a complementary manner. The physical limit is determined in principle by spin-lattice relaxation which, at a high enough magnetic energy, takes place on a 1- to 10-s time scale. The prospects are promising, but there are still some fundamental questions to be answered before evaluations can be considered accurate enough for distinctions between different models. The least important one is whether the usual Anderson–Weiss approximation is applicable [8].

Ball *et al.* [30] have described their polydimethylsiloxane (PDMS) data using the residual second NMR line moment and the tube disengagement time τ_d defined in the tube/reptation model as fitting parameters. For M_w = 540 000 and T = 300 K, a value of τ_d = 0.019 s came out. However, this value is much too short, as judged on the basis of known experimental diffusion data. For example, in PDMS 340 000 at 305 K, the diffusion coefficient was reported to be 2.7×10^{-15} m^2/s [34]. Using the reptation formalism, one calculates from this the disengagement time for PDMS 540 000 at 300 K as τ_d = 0.1 s, which is more than five times as long as the experimental data. The evaluation by Ball *et al.* appears to reflect the character of the dipolar correlation function $\langle \mathcal{H}_d(t)\mathcal{H}_d(0)\rangle$ so that any model predicting $\langle \mathcal{H}_d(t)\mathcal{H}_d(0)\rangle \propto t^{-\alpha}$ with $\alpha \simeq 0.5$ should qualitatively describe the data on the level reported.

An attempt to interpret the data by Graf *et al.* [32] with the reptation concept leads to a molecular-weight dependence unexpected according to this very concept. Another point one must consider is the temperature/frequency master curve superposition that these authors used in order to increase the dynamic window of their study. The basis for recalibrating the time axis is the entanglement time (τ_e) data taken from viscoelasticity studies. However, it is known from Ngai *et al.* [33] that the apparent activation energies evaluated from stress

relaxation and diffusion data do not coincide (as they intrinsically should, according to the reptation concept). This discrepancy may be regarded as a direct hint to the collective multichain dynamics responsible for viscoelasticity.

C. Field-Gradient NMR Diffusometry

Mean squared segment displacements can directly be evaluated from field-gradient NMR diffusometry experiments [8, 35] after compensating potential influences of immaterial spin diffusion by spin flip-flop processes [36, 37, 41]. Examples are given in [38, 39, 40, 42, 43].

In polyethylene oxide melts, for molecular weights much greater than the critical value ($M_w \gg M_c$), the influence of spin diffusion by flip-flop processes becomes negligible for diffusion times [37] t < 0.1 s. An apparent (time dependent) diffusion coefficient may be defined which obeys [38, 40, 42]

$$D_a \propto t^{-0.5\pm0.1} \tag{106}$$

in the experimental range 0.01 s < t < 0.1 s. This compares to the prediction of the twice-renormalized Rouse formalism

$$D_a^{TRR}(t, M) \simeq 0.22\left(R_g^{(0)4} D_a^{(0)}\right)^{1/3} M^0 t^{-3/2} \quad \text{for } t < \tau_1^{TRR}, \tag{107}$$

where we have expressed the parameters of Eq. (99) by the radius of gyration ($R_g^{(0)}$) and the diffusion coefficient ($D_a^{(0)}$) of some known reference polymer with the molecular weight $M^{(0)} \gg M_c$. By contrast, the reptation model suggests [2]

$$D_a^{rep}(t, M) = \frac{1}{6}\left(\frac{4}{\pi}\right)^{1/4} R_g^{(0)} a^{1/2} D_a^{(0)1/4} t^{-3/4} \quad \text{for } \tau_e \ll t \ll \tau_1^R, \tag{108}$$

$$D_a^{rep}(t, M) = \sqrt{\frac{D_a^{(0)} M^{(0)}}{3\pi M}}\, R_g^{(0)} t^{-1/2} \quad \text{for } \tau_1^R \ll t \ll \tau_d, \tag{109}$$

where the quantities a, τ_e, τ_d, and τ_1^R are the segment length of the primitive path, the entanglement time, the disengagement time, and the Rouse relaxation time, respectively.

These formulas may be evaluated for polyethylene oxide (PEO) melts taking the data [44] for PEO12 300 ($M_w^{(0)}$, $D_a^{(0)} = 3.1 \times 10^{-13}$ m^2/s at $T = 80°$C, $R_g^{(0)} = 10.5$ nm, $a = 4$ nm, $M_c = 3600$) as a reference.

The values estimated for PEO melts ($M = 5000000$ and $M = 438000$) on the basis of the reptation formulas given in Eqs. (108) and (109) are at least a factor of one-third below the experimental values for these molecular weights on the indicated time scale [38, 42]. On the other hand, the values predicted on the basis of the TRR formula given in Eq. (107) deviate only by about 30% from the experimental data [42].

If the whole time scale including diffusion times > 0.1 s is considered, when spin diffusion by flip-flop processes must be taken into account, the discrepancy between the reptation model, on the one hand, and the TRR concept and the experimental data, on the other, becomes quite dramatic: in order to describe the experimental data with the reptation model, either a $D_a^{(0)}$ two orders of magnitude larger than measured must be assumed, or the segment length of the primitive path must be assumed to take an unrealistic value $a = 40$ nm.

The polymer mode–mode coupling approach [4] also fails to describe the experimental findings. Although there is no analytical prediction with numerical coefficients so that quantitative estimations might be possible, one can state that the molecular-mass dependence already predicted the time scale of the anomalous segment diffusion regimes, while our experimental data do not display any such dependence.

6. DISCUSSION

The twice-renormalized Rouse formalism perfectly describes the empirical power laws of regions I, II, and III of the spin-lattice relaxation dispersion in melts of entangled polymers. It also reproduces the segment diffusion behavior as far as it can be probed by field-gradient NMR diffusometry. The main predictions are also in accordance with recent computer simulations [9, 10, 45], which reveal the crossover from $\langle r^2(t)\rangle_{RR} \propto t^{2/5}$ [see Eq. (69)] to $\langle r^2(t)\rangle_{TRR} \propto t^{1/3}$ [see Eq. (99)] with increasing time on a scale beyond that of Rouse behavior. In Shaffer's computer simulation [10], the normal-mode relaxation times were found to scale as $\tau_p \propto (N/p)^3$ in agreement with the TRR approach. By contrast, the reptation model predicts $\tau_p \propto N^3/p^2$.

The anomalous center-of-mass diffusion was examined in a computer simulation by Smith [9]. With increasing time, crossovers from $g_{cm}(t) \propto t^{0.71}$ to $g_{cm}(t) \propto t^{0.61}$ and then to $g_{cm}(t) \propto t$ were analyzed. These findings can be interpreted as transitions from the RR behavior given in Eq. (75) to the laws found in the frame of the TRR concept, Eq. (97). This is contrasted by the reptation prediction $g_{cm}^{rep} \propto t^{1/2}/N$ for $\tau_e \ll t \ll \tau_1^R$.

The striking coincidence of the TRR treatment with experimental findings and computer simulations in an extremely wide dynamic range is contrasted by the failure to account for the viscoelastic plateau of polymer melts, whereas the reptation model and the polymer mode–mode coupling concept describe this plateau at least qualitatively based on intrachain contributions to

the stress tensor. Comparisons to experimental data can be found in [48–50]. Under the same assumptions, the TRR treatment leads to a viscosity scaling according to $\eta \propto N^2$, and no plateau is predicted. On the other hand, as pointed out several times in the course of the text, both the reptation model and the polymer mode–mode coupling concept do not explain the T_1 dispersion laws and anomalous segment diffusion.

Despite the different nature of the reptation model and the polymer mode–mode coupling concept, both concepts have one common feature: in both cases, viscoelasticity is treated as a single-chain effect. Formally, this means that all collective interchain contributions to the stress tensor are neglected. So, the multibody behavior of entangled polymers reveals itself merely by correspondingly modified single-chain dynamics. This, however, is an anticipation which was never proved. On the other hand, proton and deuteron T_1 dispersion and NMR field-gradient diffusometry for the detection of anomalous segment displacements definitely probe the real single-chain dynamics as expressed by time correlation functions. In this respect, spin-lattice relaxometry and NMR diffusometry are particularly suitable for testing any single-chain theory.

Mechanical relaxation is of a more complicated nature. One may suspect that it specifically reflects collective multichain effects. That is, if the theory is "pressed" into a single-chain "corset" to account for viscoelasticity, the true single-chain dynamics, as probed by the experiments mentioned before, may be lost. The proper description of viscoelasticity appears to require a concept including correlated multichain dynamics. No such theory exists so far. However, a number of phenomenological approaches have been reported [33, 46, 47]. A review can be found in [16]. One of these phenomenological concepts—the lateral diffusion model by Herman [47]—interestingly predicts anomalous segment displacements scaling as $\langle r^2(t)\rangle \propto t^{1/3}$ in coincidence with Eq. (92) of the TRR treatment.

A statistical mechanical theory taking into account multichain correlations in entangled polymer systems is an extremely complex challenge waiting to be solved. From the experimental point of view, field-cycling NMR relaxometry [8] for studies of the low-frequency T_1 dispersion and spin diffusion measurements [41] can show the way to the right direction, because these methods probe relative displacements of segments sitting on different chains.

ACKNOWLEDGMENTS

We thank Professor Ken Schweizer for several stimulating discussions. This work was supported by the Deutsche Forschungsgemeinschaft, INTAS, the Volkswagen-Stiftung, and the Russian Foundation for Basic Research (project no. 98-03-33307a).

REFERENCES

1. P. G. de Gennes, J. Chem. Phys. **55**, 572 (1971).
2. M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics* (Clarendon, Oxford, 1986).
3. K. S. Schweizer, J. Chem. Phys. **91**, 5802 (1989).
4. K. S. Schweizer, J. Chem. Phys. **91**, 5822 (1989).
5. G. Ronca, J. Chem. Phys. **79**, 1031 (1983).
6. W. Hess, Macromolecules **21**, 2620 (1988).
7. N. Fatkullin and R. Kimmich, J. Chem. Phys. **101**, 822 (1994).
8. R. Kimmich, *NMR Tomography, Diffusometry, and Relaxometry* (Springer-Verlag, Berlin, 1997).
9. S. W. Smith, C. K. Hall, and B. D. Freeman, J. Chem. Phys. **104**, 5616 (1996).
10. J. S. Shaffer, J. Chem. Phys. **103**, 761 (1995).
11. U. Kirst, F. Kremer, and V. Litviniov, Macromolecules **26**, 975 (1993).
12. R. Hofmann, F. Kremer, and E. W. Fischer, Physica A (Amsterdam) **201**, 106 (1993).
13. R. Zwanzig, J. Chem. Phys. **60**, 2717 (1974).
14. M. Bixon and R. Zwanzig, J. Chem. Phys. **68**, 1890 (1978).
15. B. J. Berne and R. Pecora, *Dynamic Light Scattering* (Wiley, New York, 1976), Chap. 11.
16. K. S. Schweizer, M. Fuchs, G. Szamel, *et al.*, Macromol. Theory Simul. **6**, 1037 (1997).
17. R. Kimmich, M. Köpf, and P. Callaghan, J. Polym. Sci. B **29**, 1025 (1991).
18. N. Fatkullin, R. Kimmich, and H. W. Weber, Phys. Rev. E **47**, 4600 (1993).
19. H. W. Weber and R. Kimmich, Macromolecules **26**, 2597 (1993).
20. R. Kimmich, N. Fatkullin, H. W. Weber, and S. Stapf, J. Non-Cryst. Solids **172–174**, 689 (1994).
21. R. Kimmich, Bull. Magn. Reson. **1**, 195 (1980).
22. F. Noack, Prog. Nucl. Magn. Reson. Spectrosc. **18**, 171 (1986).
23. R. Kimmich, N. Fatkullin, R.-O. Seitter, and K. Gille, J. Chem. Phys. **108**, 2173 (1998).
24. K. V. Fenchenko, J. Polymer Sci. B **39**, 146 (1997).
25. S. Stapf and R. Kimmich, Macromolecules **29**, 1638 (1996).
26. R. Kimmich, R.-O. Seitter, U. Beginn, *et al.*, Chem. Phys. Lett. **307**, 147 (1999).
27. J. P. Cohen-Addad, Prog. Nucl. Magn. Reson. Spectrosc. **18**, 1 (1993).
28. R. Kimmich, E. Fischer, P. Callaghan, and N. Fatkullin, J. Magn. Reson. A **117**, 53 (1995).
29. E. Fischer, F. Grinberg, R. Kimmich, and S. Hafner, J. Chem. Phys. **409**, 846 (1998).
30. R. C. Ball, P. T. Callaghan, and E. T. Samulski, J. Chem. Phys. **106**, 7352 (1997).
31. P. T. Callaghan and E. T. Samulski, Macromolecules **30**, 113 (1997).
32. R. Graf, A. Heuer, and H. W. Spiess, Phys. Rev. Lett. **80**, 5738 (1998).
33. K. L. Ngai, S. L. Peny, and J. Skolnick, Macromolecules **25**, 2184 (1992).

34. S. Paul, G. Fleischer, F. Fujara, and B. Geise, Macromolecules **30**, 1414 (1997).
35. J. Kärger, H. Pfeifer, and W. Heink, Adv. Magn. Reson. **12**, 1 (1988).
36. N. F. Fatkullin, Zh. Éksp. Teor. Fiz. **99**, 1013 (1991) [Sov. Phys. JETP **72**, 563 (1991)].
37. E. Fischer, R. Kimmich, and N. Fatkullin, J. Chem. Phys. **106**, 9883 (1997).
38. E. Fischer, R. Kimmich, and N. Fatkullin, J. Chem. Phys. **104**, 9174 (1996).
39. P. T. Callaghan and A. Coy, Phys. Rev. Lett. **68**, 3176 (1992).
40. G. Fleischer and F. Fujara, Macromolecules **25**, 4210 (1992).
41. N. F. Fatkullin, G. A. Yatsenko, R. Kimmich, and E. Fischer, Zh. Éksp. Teor. Fiz. **114**, 538 (1998) [JETP **87**, 294 (1998)].
42. E. Fischer, N. Fatkullin, R. Kimmich, *et al.*, Phys. Rev. E **59**, 4079 (1999).
43. E. Rommel, R. Kimmich, M. Spülbeck, and N. Fatkullin, Prog. Colloid Polym. Sci. **93**, 155 (1993).
44. M. Appel and G. Fleischer, Macromolecules **26**, 5520 (1993).
45. K. Binder and W. Paul, J. Polym. Sci., Part B: Polym. Phys. **35**, 1 (1997).
46. J. F. Douglas and J. B. Hubbard, Macromolecules **24**, 3163 (1991).
47. M. F. Herman, Macromolecules **25**, 4925 (1992).
48. *Rheological Fundamentals of Polymer Processing*, Ed. by H. H. Winter, J. A. Covas, *et al.* (Kluwer, Dordrecht, 1995), p. 61.
49. M. Baumgartel, A. Schausberger, and H. H. Winter, Rheol. Acta **29**, 400 (1990).
50. R. M. Kannan and T. P. Lodge, Macromolecules **30**, 3694 (1997).

Journal of Experimental and Theoretical Physics, Vol. 91, No. 1, 2000, pp. 16–30.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 1, 2000, pp. 20–35.

ATOMS, SPECTRA, RADIATION

Acoustic Activity Effect for Picosecond Soliton-Like Pulses

S. V. Sazonov

Kaliningrad State Technical University, Kaliningrad, 236000 Russia
e-mail: nst@alg.kaliningrad.ru

Received November 11, 1999; in final form, March 2, 2000

Abstract—A theoretical analysis is made of the acoustic activity for interfering picosecond acoustic soliton-like pulses of down to a single oscillation period. An analysis is made of the case where these pulses propagate parallel to an external magnetic field and one of the acoustic axes in a cubic crystal containing paramagnetic impurities having effective spin $S = 1$. Allowance is made for natural, magnetic (Faraday), and cross acoustic activity. This cross activity is caused by the significant spatial nonlocality of the spin–phonon interaction for such short pulses in crystals having no center of inversion in the presence of paramagnetic impurities. A system of nonlinear equations is obtained for the transverse and longitudinal components of the strain in the form of a coupling between the "differentiated" nonlinear Schrödinger equation (with nonlinearity after the derivative sign) and the Korteweg–de Vries equation which generalizes the known systems of long–short-wavelength resonance to the case where the slowly varying envelope approximation is not valid. An approximate solution of this system is used to study the structure of an elastic soliton-like pulse whose transverse component has a rotating plane of polarization, which propagates under conditions of nonlinear coupling with the longitudinal strain.

1. INTRODUCTION

Acoustic activity, being an analog of the corresponding optical effect, is observed as the elliptic polarization of transverse elasticity waves as they propagate in a crystal [1]. Both natural and artificial acoustic activity occurs. In the first case, the effect occurs when the free crystal has no center of symmetry [1] and in the second, artificial anisotropy may be created, for example, by an external magnetic field in a crystal containing paramagnetic impurities (acoustic Faraday effect) [2].

The development of laser optics and coherent physical acoustics has involved producing increasingly shorter optical and acoustic pulses, down to a single period of the corresponding vibrations (videopulses) [3–5]. Typical durations of optical videopulses are $\tau_p \sim$ 5–10 fs [6] and of acoustic pulses $\tau_p \sim 10$ ps [4, 5].

Historically, it has been found that after a certain time transient resonant optical effects have found acoustic analogs in paramagnetic crystals. For example, the effect of optical self-induced transparency [7] stimulated the prediction and observation of a corresponding acoustic effect [8, 9]. In many phenomena the optical–acoustic analogies were so close that they have led to the creation of various fields of physical science such as magnetic quantum acoustics [10]. In this context the appearance of theoretical studies on the interaction of elastic videopulses with matter following corresponding studies for optical videopulses is quite natural. In [11] the present author reported a theoretical study of the effect of magnetooptic activity for laser videopulses of femtosecond duration in a Faraday geometry (direction of propagation of the signal parallel to the external magnetic field $\mathbf{B}_0$). Substantial differences were identified between this phenomenon and the similar effect for weak monochromatic signals. In view of this, it is interesting to study the nonlinear propagation of elastic videopulses in acoustically active media and to identify distinguishing features compared with the similar propagation of monochromatic acoustic signals and optical videopulses. The present paper reports such a study and will take into account both natural and magnetic acoustic activity.

We shall subsequently bear in mind the important factor that in studies of the interaction of acoustic videopulses with matter, as in femtosecond optics, the approximation of slowly varying amplitudes and phases [4] traditionally used for resonant nanosecond acoustics [8–10] cannot be applied in the wave and material equations.

We note some characteristics which distinguish the propagation of acoustic picosecond videopulses from femtosecond optical pulses when these propagate in a crystal.

(1) An optical pulse is transverse. However, in the nonlinear regime interaction takes place between the longitudinal and transverse components of the strain tensor of an acoustic pulse, which leads to energy exchange between these components.

(2) The characteristic spatial dimension of an optical video signal is $l_e \sim c\tau_p$, where c is the velocity of light. Assuming that $\tau_p \sim 10^{-15}$ s we obtain $l_e \sim 10^{-4}$ cm $\gg h$, where h is the crystal lattice constant in the direction of propagation of the videopulse ($h \sim 10^{-8}$–10^{-7} cm). As a result of this last inequality, as a femtosecond optical

pulse propagates in a crystal it "sees" it as a continuous medium. The spatial scale l_s of an elastic videopulse of duration $\tau_p \sim 10$ ps is of the order of 10^{-7}–10^{-6} cm [12] because of the smallness of the velocity of sound compared with c. Thus, spatial dispersion effects caused by the lattice structure become significant under these conditions. If the crystal has no center of symmetry, the natural acoustic activity for videopulses of this duration may be very appreciable. The spatial dispersion should also be manifested in the spin–phonon interaction as the videopulse propagates in a paramagnetic crystal: the discreteness of the lattice has the result that the effective spin of the paramagnetic impurity "senses" the strain not only at its location but also in a certain vicinity thereof [13] (nonlocal spin–phonon coupling).

(3) The pressure P_s at the location of an acoustic picosecond videopulse may reach 1 kbar [4] which corresponds to a relative strain $\mathscr{E} \sim P_s/\rho a^2 \sim 10^{-3}$ (ρ is the average density of the medium and a is the velocity of sound in it). For these strains the anharmonicity of the vibrations of crystal lattice sites is a significant factor (acoustic anharmonicity) [12, 13].

We shall now study the nonlinear propagation of an acoustic videopulse in a crystal, generally having no center of symmetry, containing impurity paramagnetic centers and located in an external magnetic field. We shall assume that the paramagnetic centers possess the effective spin $S = 1$. Examples may be iron group ions, Fe^{2+}, Ni^{2+}, etc. which exhibit strong dynamic coupling with lattice vibrations [2].

2. NONLOCAL HAMILTONIAN AND SEMICLASSICAL EQUATIONS OF MOTION

We set ourselves the goal of deriving equations for the picosecond physical acoustics of paramagnetic centers bearing in mind the three main observations made above.

We first write the Hamiltonian of a system comprising "paramagnetic centers + strain field" in the form

$$\hat{H} = \hat{H}_\mathrm{s} + \hat{H}_\mathrm{ph} + \hat{H}_\mathrm{int}, \tag{1}$$

where $\hat{H}_\mathrm{s}$, $\hat{H}_\mathrm{ph}$, and $\hat{H}_\mathrm{int}$ are the Hamiltonians of the paramagnetic spin system, the elastic (phonon) field, and the interaction between the spin system and the elastic field. In this case we have [1, 14]

$$\begin{aligned}\hat{H}_\mathrm{s} &= \sum_{\alpha=1}^{N} \hat{H}_\mathrm{s}^\alpha = -\sum_{\alpha=1}^{N} \mathbf{B}_0 \hat{\boldsymbol{\mu}}^\alpha \\ &= \mu_B \sum_{\alpha=1}^{N} \mathbf{B}_0 \hat{\mathbf{g}}^\alpha(0) \hat{\mathbf{S}}^\alpha,\end{aligned} \tag{2}$$

$$\begin{aligned}H_\mathrm{ph} = \frac{1}{2}\int\Bigg[&\frac{\hat{p}_i\hat{p}_i}{\rho} + \lambda_{ijkl}\frac{\partial \hat{U}_i}{\partial x_j}\frac{\partial \hat{U}_k}{\partial x_l} \\ &+ \left(\lambda_{ijkl}\delta_{pq} + \frac{1}{2}\lambda_{ijklpq}\right)\frac{\partial \hat{U}_i}{\partial x_j}\frac{\partial \hat{U}_k}{\partial x_l}\frac{\partial \hat{U}_p}{\partial x_q} \\ &+ \xi_{ijkls}\frac{\partial \hat{U}_i}{\partial x_j}\frac{\partial^2 \hat{U}_k}{\partial x_s \partial x_l} - \nu_{ijklsm}\frac{\partial^2 \hat{U}_i}{\partial x_i \partial x_l}\frac{\partial^2 \hat{U}_k}{\partial x_s \partial x_m}\Bigg] d\mathbf{r}.\end{aligned} \tag{3}$$

Here $\hat{\boldsymbol{\mu}}^\alpha$ is the magnetic moment created by the effective spin, $\hat{\mathbf{S}}^\alpha \equiv \hat{\mathbf{S}}(\mathbf{r}_\alpha) = (\hat{S}_x^\alpha, S_y^\alpha, S_z^\alpha)$ are the spin operators of the α paramagnetic ion having the coordinates of the position vector $\mathbf{r}_\alpha$ (where $\hat{\boldsymbol{\mu}}^\alpha = -\mu_B \hat{\mathbf{g}}^\alpha(0)\hat{\mathbf{S}}^\alpha$ is the Bohr magneton, $\hat{\mathbf{g}}^\alpha(0)$ is the Landé tensor in the absence of a strain field,

$$\hat{\mathscr{E}}_{ij} = \frac{1}{2}\left(\frac{\partial \hat{U}_i}{\partial x_j} + \frac{\partial \hat{U}_j}{\partial x_i}\right)$$

is the operator of the linear component of the strain tensor, $\hat{U}_i$ and $\hat{p}_i$ are the operators of the ith component of the displacement vector and the momentum density of the medium, ρ is the average density of the medium, N is the total number of paramagnetic ions, δ_{pq} is the Kronecker symbol, λ_{ijkl} and λ_{ijklpq} are the tensors of the second- and third-order elastic moduli, respectively (the latter describes the elastic cubic anharmonicity); the tensors ξ_{ijkls} and ν_{ijklsm} describe the spatial acoustic dispersion. In this case ξ_{ijkls} is antisymmetric with respect to the two first pairs of indices [1]. This is responsible for the natural acoustic activity [1] (for further details see below). In crystals having a center of symmetry we find $\xi_{ijkls} = 0$. Note that in (3) and everywhere below, summation is implied over the recurrent Latin indices and integration is performed over the entire crystal volume.

We shall clarify the form of the spin–phonon interaction Hamiltonian based on the semiphenomenological "spin Hamiltonian" approach [2, 14]. Adding to $\hat{H}_\mathrm{s}^\alpha$ an operator which commutes over all matrices $\hat{S}_i^\alpha$ ($i = z, y, z$; $\alpha = 1, \ldots, N$) does not change the state of the spin subsystem. In the absence of an elastic field this operator can be any function f of the dimensions of energy of the well-known Casimir operator $\hat{\mathbf{S}}^{\alpha 2} = S(S+1)\hat{I}$ ($\hat{I}$ is the unit matrix) in $SU(3)$ algebra. Then we have

$$\hat{H}_\mathrm{s} = \sum_{\alpha=1}^{N} [\mu_B \mathbf{B}_0 \hat{\mathbf{g}}(0)\hat{\mathbf{S}}^\alpha + f(\hat{\mathbf{S}}^\alpha \hat{\mathbf{R}}(0)\hat{\mathbf{S}}^\alpha)], \tag{4}$$

where in the absence of a strain field $\hat{\mathbf{R}}(0) = \hat{I}$. On account of the homogeneity of the crystal the functional dependence $f(\hat{\mathbf{S}}^\alpha \hat{\mathbf{R}}(0)\hat{\mathbf{S}}^\alpha)$ is the same for all spins.

The presence of an elastic field leads to modulation of the tensors $\hat{\mathbf{g}}$ and $\hat{\mathbf{R}}$, i.e., they become dependent on the components of the strain tensor. Expanding $\hat{\mathbf{R}}$ as a Taylor series in terms of the components of the linear relative strain and their spatial derivatives, we obtain

$$\hat{\mathbf{R}} = \hat{\mathbf{R}}(0) + \left(\frac{\partial\hat{\mathbf{R}}}{\partial U_{i,j}}\right)_0\frac{\partial\hat{U}_i}{\partial x_i} + \left(\frac{\partial\hat{\mathbf{R}}}{\partial U_{i,jk}}\right)_0\frac{\partial^2\hat{U}_i}{\partial x_j\partial x_k} + \left(\frac{\partial\hat{\mathbf{R}}}{\partial U_{i,jkl}}\right)_0\frac{\partial^3\hat{U}_i}{\partial x_j\partial x_k\partial x_l} + \ldots, \quad (5)$$

where the notation $U_{i,j} = \partial U_i/\partial x_j$, $U_{i,jk} = \partial^2 U_i/\partial x_j\partial x_k$, $U_{i,jkl} = \partial^3 U_i/\partial x_j\partial x_k\partial x_l$ is introduced, and the lower index "0" after the parentheses indicates that the derivatives are taken for zero arguments.

In (5) we neglected higher powers of the relative strain $\partial\hat{U}_i/\partial x_j$. The tensor $\hat{\mathbf{g}}^\alpha$ can be expanded similarly. However, experience shows [2, 14] that for a system of spins $S > 1/2$ the spin–phonon interaction caused by modulation of the tensor $\hat{\mathbf{R}}$ is two orders of magnitude stronger than the similar interaction as a result of modulation of the Landé tensor. Thus, we shall subsequently neglect linear spin–phonon interaction with respect to $\hat{\mathbf{S}}^\alpha$ [see (4)] which is significant for an $S = 1/2$ spin system [2, 14] and was studied in detail for transverse acoustic pulses in [13].

Spin–phonon interaction is introduced in (4) by replacing $\hat{\mathbf{R}}(0) \longrightarrow \hat{\mathbf{R}}$. We then obtain the expansion for f

$$f(\hat{\mathbf{S}}^\alpha\hat{\mathbf{R}}\hat{\mathbf{S}}^\alpha) = f(\hat{\mathbf{S}}^{\alpha 2}) + f'(\hat{\mathbf{S}}^{\alpha 2})\hat{\mathbf{S}}^\alpha \times\left[\left(\frac{\partial\hat{\mathbf{R}}}{\partial U_{i,j}}\right)\frac{\partial\hat{U}_i}{\partial x_j} + \left(\frac{\partial\hat{\mathbf{R}}}{\partial U_{i,jk}}\right)\frac{\partial^2\hat{U}_i}{\partial x_j\partial x_k} + \left(\frac{\partial\hat{\mathbf{R}}}{\partial U_{i,jkl}}\right)\frac{\partial^3\hat{U}_i}{\partial x_j\partial x_k\partial x_l}\right]\hat{\mathbf{S}}^\alpha + \ldots. \quad (6)$$

For the function $f(\hat{\mathbf{S}}^{\alpha 2})$ it is sufficient to assume that this can be expanded as a Taylor series. Then f is a multiple of the unit matrix. Substituting (6) into (4) taking into account $\hat{\mathbf{R}}(0) \longrightarrow \hat{\mathbf{R}}$, omitting $f(\hat{\mathbf{S}}^{\alpha 2}) = f(S(S+1))\hat{I}$, and going over to index notation, we find the Hamiltonian of the spin–phonon interaction

$$\hat{H}_{\rm int} = \sum_{\alpha=1}^{N}\left(G_{ijkl}\frac{\partial\hat{U}_k}{\partial x_l} + \beta_{ijklm}\frac{\partial^2\hat{U}_k}{\partial x_m\partial x_l} + \sigma_{ijklmp}\frac{\partial^3\hat{U}_k}{\partial x_l\partial x_m\partial x_p}\right)\hat{S}^\alpha_{ij}, \quad (7)$$

where

$$\hat{S}^\alpha_{ij} = \frac{1}{2}(\hat{S}^\alpha_i\hat{S}^\alpha_j + \hat{S}^\alpha_j\hat{S}^\alpha_i),$$

$$G_{ijkl} = f'(S(S+1))\left(\frac{\partial R_{ij}}{\partial U_{k,l}}\right)_0,$$

$$\beta_{ijklm} = f'(S(S+1))\left(\frac{\partial R_{ij}}{\partial U_{k,lm}}\right)_0,$$

$$\sigma_{ijklmp} = f'(S(S+1))\left(\frac{\partial R_{ij}}{\partial U_{k,lmp}}\right)_0.$$

The components G_{ijkl} of the local spin–phonon interaction tensor have been measured experimentally for very many crystals [14], which cannot be said of the components β_{ijklm} and σ_{ijklmp} of the first- and second-order spin–phonon nonlocality tensors introduced by us. Quite clearly, in order of magnitude $\beta_{ijklm} \sim h_m G_{ijkl}$, $\sigma_{ijklmp} \sim h_m h_p G_{ijkl}$, where h_m is the distance between the nearest paramagnetic centers along the x_m-axis. The second term in parentheses in (7) is invariant with respect to spatial inversion. Thus, $\beta_{ijklm} \neq 0$ is only found in crystals possessing no center of symmetry in the presence of paramagnetic impurities. Consequently this tensor, like the tensor ξ_{ijkls} in (3), makes a contribution to the acoustic activity which will be shown directly below.

The reasoning put forward above can only arbitrarily be called a derivation of $\hat{H}^{\rm int}$ since expression (7) was obtained without using any specific spin–phonon coupling mechanism. Hence nothing can be said about the values of the parameters characterizing this interaction. However, this reasoning shows that the Hamiltonian (7) is the most general linear Hamiltonian of the spin–phonon interaction for $S \geq 1$ with respect to the relative strain and its spatial derivatives. The most commonly used mechanism for this interaction is the Van Vleck mechanism [2, 10, 14] according to which the acoustic wave modulates the intracrystalline field near the paramagnetic ion. Interaction between the gradient of this field and the quadrupole moment of the effective spin induces quantum transitions between Zeeman sublevels described by $\hat{H}_s$. Note that in the case $S = 1/2$ we have $\hat{S}^\alpha_{ij} = 0$ [2] because of the properties of Pauli operators. Thus, for $S = 1/2$ the Hamiltonian $\hat{H}_{\rm int}$ is linear with respect to $\hat{S}^\alpha_{ij}$ ($i = x, y, z$) and is attributed to modulation by the elastic field of the Landé tensor [12, 13].

We shall use a semiclassical approach to obtain the equations of motion describing the self-consistent dynamics of the paramagnetic subsystem and the field of elastic pulses. On account of the high power of the acoustic pulses, the elastic field will be described classically. However, the spins of the paramagnetic ions,

being strictly quantum objects, should obey the laws of quantum mechanics. From the formal point of view, the essence of the semiclassical approach is as follows. The operators $\hat{U}_k$ and $\hat{p}_k$ ($k = x, y, z$) in (3) and (7) are replaced by the C-number functions U_k and p_k, respectively, which form the classical vectors $\mathbf{U} = (U_x, U_y, U_z)$, $\mathbf{p} = (p_x, p_y, p_z)$, for which we write the classical Hamiltonian equations

$$\frac{\partial \mathbf{U}}{\partial t} = \frac{\delta H}{\delta \mathbf{p}}, \quad \frac{\partial \mathbf{p}}{\partial t} = -\frac{\delta H}{\delta \mathbf{U}}. \tag{8}$$

Here H is obtained from $\hat{H}$ by substituting

$$\hat{U}_k \longrightarrow U_k, \quad \hat{S}_{ij} \longrightarrow \langle \hat{S}_{ij} \rangle \equiv \mathrm{Tr}(\hat{\rho}\hat{S}_{ij}),$$

$\hat{\rho}$ is the density operator of the spin subsystem. Thus, the quantum averaging of the spin and field variables is performed independently, which corresponds to neglecting the quantum correlations between these variables.

The dynamics of the spin subsystem can be described by the equation for the density operator:

$$i\hbar \frac{\partial \hat{\rho}}{\partial t} = [\hat{\tilde{H}}, \hat{\rho}], \tag{9}$$

where

$$\hat{\tilde{H}} = \hat{H}_\mathrm{s} + \hat{\tilde{H}}_\mathrm{int}, \tag{10}$$

and $\hat{\tilde{H}}_\mathrm{int}$ is obtained from $\hat{H}_\mathrm{int}$ by using the substitution $\hat{U}_k \longrightarrow U_k$.

Using (1)–(3) and (7)–(10), we can write self-consistent equations for the spin–elastic ($S > 1/2$) dynamics in a crystal of arbitrary geometry when a picosecond acoustic pulse propagates in this crystal.

The acoustic activity is expressed most clearly, and can be observed most easily, when an acoustic signal propagates along the symmetric acoustic axes [1]. In these cases the effect is observed as rotation of the plane of polarization of the transverse component of the elastic strain. Let us assume that an elastic pulse propagates along an external magnetic field $\mathbf{B}_0$ parallel to one of the fourfold axes of a cubic crystal (z-axis). In this case, the Landé tensor has a diagonal form and

$$\hat{H}_\mathrm{s} = \sum_{\alpha=1}^{N} \hbar\omega_0 \hat{S}_z^\alpha, \tag{11}$$

where $\omega_0 = \mu_B g_{zz} B_0/\hbar$ is the frequency spacing between the neighboring levels of the Zeeman triplet and the spin operators may be represented by matrices of the form [2]

$$\hat{S}_x^\alpha = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$

$$\hat{S}_y^\alpha = \frac{i}{\sqrt{2}}\begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \tag{12}$$

$$\hat{S}_z^\alpha = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

We shall consider the one-dimensional case when all the dynamic variables depend on z and t. Accordingly we only retain derivatives with respect to the single spatial variable z in (3) and (7).

Rotations by 90° about the z-axis ($x \longrightarrow y, y \longrightarrow -x, z \longrightarrow z$) and by 180° about the x-axis ($x \longrightarrow x, y \longrightarrow -y, z \longrightarrow -z$) are symmetry transformations for a cubic crystal, i.e., the coefficients in (3) and (7) remain constant for these transformations. Applying a standard procedure to find the nonzero components of the tensors using crystal symmetry transformations (see, for example [15–17]) we obtain from (3) for our case

$$\begin{aligned}\hat{H}_\mathrm{ph} = \frac{1}{2}\int\Bigg\{&\frac{\hat{p}_x^2 + \hat{p}_y^2 + \hat{p}_z^2}{\rho} + \lambda_{11}\left(\frac{\partial \hat{U}_z}{\partial z}\right)^2 \\ &+ \lambda_{44}\left[\left(\frac{\partial \hat{U}_x}{\partial z}\right)^2 + \left(\frac{\partial \hat{U}_y}{\partial z}\right)^2\right] \\ &+ (\lambda_{44} + \lambda_{166})\left[\left(\frac{\partial \hat{U}_x}{\partial z}\right)^2 + \left(\frac{\partial \hat{U}_y}{\partial z}\right)^2\right]^2 \frac{\partial \hat{U}_z}{\partial z} \\ &+ \left(\lambda_{11} + \frac{\lambda_{111}}{3}\right)\left(\frac{\partial \hat{U}_z}{\partial z}\right)^3 \\ &+ \xi_{54z}\left(\frac{\partial \hat{U}_x}{\partial z}\frac{\partial^2 \hat{U}_y}{\partial z^2} - \frac{\partial \hat{U}_x}{\partial z}\frac{\partial^2 \hat{U}_x}{\partial z^2}\right) - \nu_{111}\left(\frac{\partial^2 \hat{U}_z}{\partial z^2}\right)^2 \\ &- \nu_{166}\left[\left(\frac{\partial^2 \hat{U}_x}{\partial z^2}\right)^2 + \left(\frac{\partial^2 \hat{U}_y}{\partial z^2}\right)^2\right]\Bigg\} d\mathbf{r}.\end{aligned} \tag{13}$$

Here we use the Voigt notation [1, 2]: $xx \longrightarrow 1, yy \longrightarrow 2, zz \longrightarrow 3, yz \longrightarrow 4, xz \longrightarrow 5, xy \longrightarrow 6$.

We now write the interaction Hamiltonian (7) under these assumptions. We first note that the spin operators are transformed as follows as a result of successive

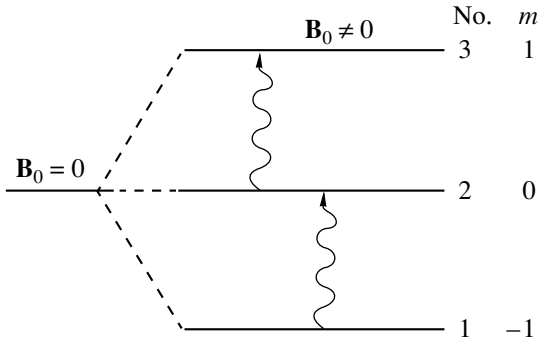


Fig. 1. Diagram showing quantum spin–phonon transitions (wavy arrows) when an acoustic wave propagates along an external magnetic field $\mathbf{B}_0$ (Faraday geometry) in an $S = 1$ spin system; No. is the number of the quantum level measured from the ground state of the Zeeman triplet, m are the magnetic quantum numbers of the Zeeman sublevels of a paramagnetic ion which appear as the eigenvalues of the operator S_z.

inversions of the coordinate axes: $x \longrightarrow -x$; $\hat{S}_x \longrightarrow \hat{S}_x$; $\hat{S}_y \longrightarrow -\hat{S}_y$; $\hat{S}_z \longrightarrow -\hat{S}_z$; $y \longrightarrow -y$; $\hat{S}_x \longrightarrow -\hat{S}_x$; $\hat{S}_y \longrightarrow \hat{S}_y$; $\hat{S}_z \longrightarrow -\hat{S}_z$; $z \longrightarrow -z$; $\hat{S}_x \longrightarrow -\hat{S}_x$; $\hat{S}_y \longrightarrow \hat{S}_y$; $\hat{S}_z \longrightarrow \hat{S}_z$. It then follows that the operator $\hat{S}_{ij}$ only changes sign in these operations if the inverted coordinate is the same as one of its indices.

The expression for (7) may be simplified because the combination $\hat{S}_x^{\alpha 2} + \hat{S}_y^{\alpha 2} + \hat{S}_z^{\alpha 2} = S(S+1)$ is an integral of motion. Consequently, adding to (7) a term proportional to this combination merely displaces all the quantum levels of the paramagnetic system by the same quantity and is thus unobservable. This provides a basis for assuming $\mathrm{Tr}\,\hat{H}_{\mathrm{int}} = 0$ [14, 18]. We then obtain the relationships

$$G_{13} + G_{23} + G_{33} = 0, \quad \sigma_{133} + \sigma_{233} + \sigma_{333} = 0, \quad (14)$$

the first of which was given in [14, 18]. Using (14) and also the fact that because of the cubic symmetry $G_{23} = G_{13}$, $\sigma_{233} = \sigma_{133}$, $G_{33} = G_{22} = G_{11}$, $\sigma_{333} = \sigma_{222} = \sigma_{111}$, we rewrite (7) in the form

$$\hat{H}_{\mathrm{int}} = \int n\Big[\frac{3}{2}\Big(G_{11}\frac{\partial \hat{U}_z}{\partial z} + \sigma_{111}\frac{\partial^3 \hat{U}_z}{\partial z^3}\Big)\hat{S}_z^2 + \Big(G_{44}\frac{\partial \hat{U}_x}{\partial z} + \beta_{54z}\frac{\partial^2 \hat{U}_y}{\partial z^2} + \sigma_{443}\frac{\partial^3 \hat{U}_x}{\partial z^3}\Big)\hat{S}_{xz} + \Big(G_{44}\frac{\partial \hat{U}_y}{\partial z} - \beta_{54z}\frac{\partial^2 \hat{U}_x}{\partial z^2} + \sigma_{443}\frac{\partial^3 \hat{U}_y}{\partial z^3}\Big)\hat{S}_{yz}\Big]d\mathbf{r}, \quad (15)$$

where $n = \sum_{\alpha=1}^{n}\delta(\mathbf{r} - \mathbf{r}_\alpha)$ is the concentration of paramagnetic centers and $\delta(\mathbf{r} - \mathbf{r}_\alpha)$ is the Dirac delta function.

Using (10), (12), and (15), we express the semiclassical interaction Hamiltonian $\hat{\tilde{H}}_{\mathrm{int}}$ as follows:

$$\hat{\tilde{H}}_{\mathrm{int}} = \sum_{\alpha=1}^{N}\hat{\tilde{H}}_{\mathrm{int}}^{\alpha},$$

where the Hamiltonian of the interaction of an α spin and the elastic field is

$$\tilde{H}_{\mathrm{int}}^{\alpha} = \begin{pmatrix} V_\parallel & V_\perp^* & 0 \\ V_\perp & 0 & -V_\perp^* \\ 0 & -V_\perp & V_\parallel \end{pmatrix}, \quad (16)$$

$$V_\parallel = \frac{3}{2}\Big(G_{11}u + \sigma_{111}\frac{\partial^2 u}{\partial z^2}\Big), \quad (17)$$

$$V_\perp = \frac{1}{2\sqrt{2}}\Big(G_{44}\psi - i\beta_{54z}\frac{\partial \psi}{\partial z} + \sigma_{443}\frac{\partial^2 \psi}{\partial z^2}\Big), \quad (18)$$

$u = \mathscr{E}_{zz}$, $\psi = \mathscr{E}_{xz} + i\mathscr{E}_{yz}$.

It follows from (16)–(18) that as the pulse propagates in Faraday geometry, the transverse component of the strain induces $1 \rightleftarrows 2$ and $2 \rightleftarrows 3$ quantum transitions within the Zeeman triplet (Fig. 1). These transitions respectively induce two circular polarizations rotating in opposite directions. As a result of the difference between the equilibrium populations of levels 1 and 2 which are determined by the Boltzmann factors, the phase velocities of these polarizations are different which leads to the acoustic Faraday effect (rotation of the plane of polarization of the transverse component of the elastic field in the magnetic field). The longitudinal component of the strain field changes the frequencies of these transitions (acoustic dynamic Stark effect).

Replacing in (13) and (15) all the field operators by C-number functions and the spin operators by their quantum averages following the semiclassical approach, we obtain after using (8), (12), (15), and the identity $\rho_{11} + \rho_{22} + \rho_{33} = 1$

$$\frac{\partial^2 \psi}{\partial t^2} - a_\perp^2\frac{\partial^2 \psi}{\partial z^2} + i\frac{\xi_{54z}}{\rho}\frac{\partial^3 \psi}{\partial z^3} - \frac{\nu_{166}}{\rho}\frac{\partial^4 \psi}{\partial z^4} - \frac{\lambda_{44} + \lambda_{166}}{\rho}\frac{\partial^2}{\partial z^2}(u\psi) = \frac{n}{\sqrt{2}p}\Big(G_{44}\frac{\partial^2}{\partial z^2} - i\beta_{54z}\frac{\partial^3}{\partial z^3} + \sigma_{443}\frac{\partial^4}{\partial z^4}\Big)(\rho_{23} - \rho_{12}), \quad (19)$$

$$\frac{\partial^2 u}{\partial t^2} - a_\parallel^2\frac{\partial^2 u}{\partial z^2} - \frac{\nu_{111}}{\rho}\frac{\partial^4 u}{\partial z^4} - \frac{3\lambda_{11} + \lambda_{111}}{\rho}\frac{\partial^2}{\partial z^2}(u^2) - \frac{\lambda_{44} + \lambda_{166}}{2\rho}\frac{\partial^2}{\partial z^2}(|\psi|^2) \quad (20)$$

$$= -\frac{3n}{2\rho}\left(G_{11}\frac{\partial^2}{\partial z^2} + \sigma_{111}\frac{\partial^4}{\partial z^4}\right)\rho_{22},$$

where $a_\perp = \sqrt{\lambda_{44}/\rho}$, $a_\parallel = \sqrt{\lambda_{11}/\rho}$ are the velocities of the transverse and longitudinal sound, respectively.

Using (9)–(12) and (15), we supplement the wave equations by constitutive equations for the elements of the density matrix of the paramagnetic center:

$$\frac{\partial \rho_{33}}{\partial t} = \frac{i}{\hbar}(V_\perp \rho_{32} - V_\perp^* \rho_{32}^*), \tag{21}$$

$$\frac{\partial \rho_{32}}{\partial t} = -i\left(\omega_0 + \frac{V_\parallel}{\hbar}\right)\rho_{32} + \frac{i}{\hbar}V_\perp^*(\rho_{33} - \rho_{22}) - \frac{i}{\hbar}V_\perp \rho_{31}, \tag{22}$$

$$\frac{\partial \rho_{31}}{\partial t} = -2i\omega_0 \rho_{31} - \frac{i}{\hbar}V_\perp^*(\rho_{32} + \rho_{21}), \tag{23}$$

$$\frac{\partial \rho_{21}}{\partial t} = -i\left(\omega_0 - \frac{V_\parallel}{\hbar}\right)\rho_{21} - \frac{i}{\hbar}V_\perp^*(\rho_{22} - \rho_{11}) - \frac{i}{\hbar}V_\perp \rho_{31}, \tag{24}$$

$$\frac{\partial \rho_{11}}{\partial t} = -\frac{i}{\hbar}(V_\perp^* \rho_{21}^* + V_\perp \rho_{21}). \tag{25}$$

The system of equations (19)–(25) describes the self-consistent dynamics of the paramagnetic centers and picosecond elastic pulses as the latter propagate along one of the fourfold axes of a cubic crystal parallel to the external magnetic field. In this case, the acoustic activity effect has three mechanisms: (1) a natural mechanism attributable to the spatial dispersion of the crystal in the absence of paramagnetic impurities and described by the tensor ξ_{54z} [see the third term on the left-hand side of (19)] which keeps this equation invariant with respect to the operation $z \longrightarrow -z$; (2) a magnetic (Faraday) mechanism caused by the time dispersion for the $1 \rightleftarrows 2$ and $2 \rightleftarrows 3$ quantum transitions [see second terms on the right-hand sides of (22) and (24) and also the first term in parenthesis in (18) and the right-hand side of (19)]; (3) a cross mechanism which occurs as a result of the space–time nonlocality of the spin–phonon coupling described by the tensor component β_{54z} [see (18), (19), (22), and (24)]. All three mechanisms contribute to the linear acoustic activity but in the nonlinear pulse regime they are responsible for a pulse structure where the plane of polarization of the transverse component is rotated inside the pulse.

3. MODIFIED SYSTEM OF EQUATIONS FOR LONG–SHORT-WAVELENGTH RESONANCE

We shall analyze the system (19)–(25) in the wave transparency approximation, i.e. when the spectrum of the elastic pulse does not overlap the quantum transitions within the Zeeman triplet. Formally, this approximation may be written as follows [19, 20]

$$\omega_0 \tau_p^{\min} \gg 1, \tag{26}$$

where $\tau_p^{\min}$ is the characteristic minimum time scale of the pulse.

In order to achieve fairly efficient interaction between the paramagnetic centers and the elastic strain field the Zeeman energy must exceed the energy of the thermal motion, i.e., $\hbar\omega_0 > k_B T$, where k_B is the Boltzmann constant. This inequality ensures an appreciable thermodynamically equilibrium population difference between the Zeeman sublevels of the paramagnetic ion which leads to efficient absorption of acoustic waves by the paramagnetic subsystem. These experiments are usually carried out at liquid helium temperatures $T \simeq 4$ K [2, 8, 10, 14]. The last inequality then gives $\omega_0 > 5 \times 10^{11}$ s^{-1}. Magnetic fields which can give Zeeman splittings $\omega_0 \sim 10^{12}$ s^{-1} have now been achieved in laboratories. Condition (26) can then be satisfied by elastic pulses of duration $\tau_p \sim 10$ ps.

Under condition (26) the pulse interacts weakly with the medium, exciting it only negligibly [11, 19, 20]. Hence, the spin–phonon interaction energy is negligible compared with the intrinsic magnetic energy of the paramagnetic center. Thus, we can supplement (26) with

$$V_\parallel \ll \hbar\omega_0, \quad |V_\perp| \ll \hbar\omega_0. \tag{27}$$

It can be seen that $|V_\perp|/\hbar$ plays the role of the characteristic frequency of the quantum transitions (Rabi acoustic frequency) or the characteristic reciprocal time τ_p^{-1} of the dynamic spin–elastic interaction process. It therefore follows that conditions (26) and (27) show good agreement.

Equations (22)–(24) can now be solved using the method of successive approximations with respect to the derivatives on the left-hand sides. In the zeroth approximation the left-hand sides of these equations can be neglected, and then taken into account in the first, second, and so on, approximations. In the zeroth approximation equations (22)–(24) yield a system of algebraic linear equations for ρ_{32}, ρ_{21}, and ρ_{31}. Solving this and expanding in terms of the small dynamic

parameters $V_\parallel/\hbar\omega_0$ and $V_\perp/\hbar\omega_0$ as far as powers no higher than the third, we obtain

$$\rho_{32}^{(0)} = \frac{V_\perp^*}{\hbar\omega_0}(\rho_{33}-\rho_{22}) - \frac{V_\parallel V_\perp^*}{\hbar^2\omega_0^2}(W_3-W_2) + \frac{V_\parallel^2 V_\perp^*}{\hbar^3\omega_0^3}(W_3-W_2) + \frac{|V_\perp|^2 V_\perp^*}{2\hbar^3\omega_0^3}(W_1-2W_2+W_3), \quad (28)$$

$$\rho_{21}^{(0)} = -\frac{V_\perp^*}{\hbar\omega_0}(\rho_{22}-\rho_{11}) - \frac{V_\parallel V_\perp^*}{\hbar^2\omega_0^2}(W_2-W_1) + \frac{V_\parallel^2 V_\perp^*}{\hbar^3\omega_0^3}(W_2-W_1) + \frac{|V_\perp|^2 V_\perp^*}{2\hbar^3\omega_0^3}(W_1-2W_2+W_3), \quad (29)$$

where W_j (j = 1, 2, 3) is the initial population of the jth level of the Zeeman triplet (it is assumed that the initial density matrix only contains diagonal elements determining the populations of the corresponding quantum levels) and the superscript (0) indicates the zeroth approximation with respect to the time derivative.

In the second and third terms on the right-hand sides of (28) and (29) we made the substitutions $\rho_{33} \longrightarrow W_3$, $\rho_{22} \longrightarrow W_2$, $\rho_{11} \longrightarrow W_1$ because the diagonal elements $\hat{\rho}$ are energy characteristics and thus their changes are proportional to $|V_\perp|^2$ (see below). Consequently, allowing for these changes in the second and third terms would generate fourth and higher powers of the strain components which we shall neglect.

Systematically substituting (28) and (29) into the left-hand sides of equations (22) and (24), respectively, in the first order of the approximation with respect to the derivatives we find

$$\rho_{32} = \rho_{32}^{(0)} + i\frac{W_3-W_2}{\hbar\omega_0^2}\frac{\partial V_\perp^*}{\partial t}, \quad (30)$$

$$\rho_{21} = \rho_{21}^{(0)} - i\frac{W_2-W_1}{\hbar\omega_0^2}\frac{\partial V_\perp^*}{\partial t}. \quad (31)$$

Here for simplicity the index (1) corresponding to the first approximation is omitted from ρ_{32} and ρ_{21}.

A contribution to the corrections to the zeroth approximations $\rho_{32}^{(0)}$ and $\rho_{21}^{(0)}$ was made by their first terms linearized with respect to $V_\perp^*$ in (28) and (29), respectively. This circumstance is consistent with (26) and (27) and as a result the nonlinearity and time nonlocality (dispersion) appear additively in (30) and (31). This statement is general for wave packets propagating in media under conditions of wave transparency [11, 19, 20]. Since in our case, the spatial nonlinearity also appears as small corrections, in this section the nonlinearity and dispersion will be taken into account additively.

Substituting (30) and (31) into (21) and (25), respectively, after integrating over t we find

$$\rho_{33} = W_3 + \frac{W_2-W_3}{\omega_0^2}|V_\perp|^2,$$

$$\rho_{11} = W_1 - \frac{W_1-W_2}{\omega_0^2}|V_\perp|^2, \quad (32)$$

$$\rho_{22} = W_2 + \frac{W_1-2W_2+W_3}{\omega_0^2}|V_\perp|^2.$$

Strictly speaking, an expansion as far as the second order in $\partial/\partial t$ should be performed in (30) and (31) because in (19) the spatial dispersion of the transverse strain component is taken into account as far as the second order in $\partial/\partial z$. We shall subsequently only allow for a minimal first-order spatial and temporal dispersion since this order provides the main contribution to the acoustic activity effect. We thus assume $\nu_{166} = \sigma_{443} = 0$. At the same time we retain the term $\sim\partial^4 u/\partial z^4$ in (20) because this gives a minimal allowance for the dispersion of the longitudinal strain component.

Taking this into account, from (28)–(32) and (17)–(20) after using the approximation of unidirectional propagation [11, 20] (frequently called the approximation of slowly varying profile in comoving coordinates [21]), we obtain

$$i\frac{\partial\psi}{\partial t} + \beta_0\frac{\partial^2\psi}{\partial\zeta^2} + i\eta\frac{\partial}{\partial\zeta}(|\psi|^2\psi) = i\mu\frac{\partial}{\partial\zeta}(u\psi), \quad (33)$$

$$\frac{\partial u}{\partial t} + \alpha_0\frac{\partial u}{\partial\zeta} + \nu\frac{\partial^3 u}{\partial\zeta^3} + qu\frac{\partial u}{\partial\zeta} = \frac{\mu}{2}\frac{\partial}{\partial\zeta}(|\psi|^2), \quad (34)$$

where $\zeta = z - a_{\perp 0}t$,

$$\beta_0 = \frac{1}{2}\left[\frac{n(W_1-W_3)}{4\rho\hbar\omega_0^2}\left(G_{44}^2 + 2\frac{\beta_{54z}G_{44}\omega_0}{G_{44}a_{\perp 0}} - \frac{\xi_{54z}}{a_{\perp 0}}\right)\right],$$

$$\nu = \frac{\nu_{111}}{2\rho a_{\perp 0}},$$

$$\eta = G_{44}^2\frac{n(W_1-W_2)}{64\rho a_{\perp 0}\hbar^3\omega_0^3},$$

$$\mu = \frac{1}{2\rho a_{\perp 0}} \times\left[\lambda_{44} + \lambda_{166} - 3G_{11}G_{44}^2\frac{n(1-3W_2)}{8\hbar\omega_0^2}\right], \quad (35)$$

$$q = \frac{3\lambda_{11}+\lambda_{111}}{4\rho a_{\perp 0}}, \quad \alpha_0 = \frac{a_\parallel^2 - a_{\perp 0}^2}{2a_{\perp 0}},$$

$$a_{\perp 0} = a_\perp\left[1 - \frac{nG_{44}^2(W_1 - W_3)}{4\rho a_\perp^2 \hbar\omega_0}\right].$$

The system (33), (34) describes the nonlinear propagation of a longitudinal–transverse acoustic pulse in a cubic crystal parallel to an external magnetic field in the wave transparency regime.

It can be seen from (35) that the velocity of the transverse sound is renormalized as a result of the spin–phonon interaction (in a medium at thermodynamic equilibrium $W_1 > W_3$ this velocity decreases). If $R \equiv nG_{44}^2/(4\rho a_\perp^2 \hbar\omega_0) > 1$, at temperature T below a certain critical level T_c we have $a_{\perp 0}^2 < 0$ which corresponds to an equilibrium phase transition accompanied by the formation of static strains and magnetization [22]. In the present study we shall not touch on this aspect, assuming $R < 1$. Having taken the values $G_{44} \sim 10^{-13}$ erg [2, 14, 18], $\rho \simeq 5$ g/cm^3, $a_\perp \simeq 5 \times 10^5$ cm/s, $\omega_0 \sim 10^{12}$ s^{-1} for Fe^{2+} ions in an MgO crystal matrix, we find that the condition $R < 1$ can be satisfied for concentrations of paramagnetic centers $n \sim 10^{23}$ cm^{-3} which is four or five orders of magnitude higher than the real concentrations of Fe^{2+} ions. For real concentrations we have $a_{\perp 0} \approx a_\perp$.

The acoustic activity effect is contained in the coefficient β_0 in equation (33). In order to show this, we shall first linearize the system (33), (34). Then substituting $\psi = \mathscr{E}_{xz} + i\mathscr{E}_{yz}$, after separating the real and imaginary parts we obtain

$$\frac{\partial \mathscr{E}_{xz}}{\partial t} + \beta_0 \frac{\partial^2 \mathscr{E}_{yz}}{\partial \zeta^2} = 0, \quad \frac{\partial \mathscr{E}_{yz}}{\partial t} - \beta_0 \frac{\partial^2 \mathscr{E}_{xz}}{\partial \zeta^2} = 0.$$

The corresponding dispersion equation in laboratory coordinates has the form

$$\omega = a_{\perp 0}k \pm \beta_0 k^2.$$

Bearing in mind that the second term is only a correction, we approximately obtain

$$k_\pm = \omega/a_{\perp 0} \mp \beta_0 \omega^2/a_{\perp 0}^3.$$

The different signs in this last expression correspond to two natural modes rotating in opposite directions, having different phase velocities $v_+ = \omega/k_+$ and $v_- = \omega/k_-$. The angle of rotation of the plane of polarization per unit length (instantaneous angle) for a linear monochromatic wave is given by

$$\kappa_l = \frac{1}{2}(k_+ - k_-) = -\frac{\beta_0 \omega^2}{a_{\perp 0}^3}. \tag{36}$$

This wave may be expressed in the form

$$\psi \propto \exp[i\Omega_l t + ik(z - a_{\perp 0}t)], \tag{37}$$

where $\Omega_l = \kappa_l a_{\perp 0}$.

It follows from the expression for β_0 [see (35)] that all three acoustic activity mechanisms noted above make additive contributions to this effect in the wave transparency approximation.

Before studying characteristics of the pulsed nonlinear acoustic activity using (33) and (34), we shall first examine the degree of generality of this system. We first note that for nanosecond pulses with rf filling the following representation holds

$$\psi = \chi \exp[-i(\Omega t - k\zeta)],$$

where

$$\frac{\partial \chi}{\partial t} \ll \Omega\chi, \quad \frac{\partial \chi}{\partial \zeta} \ll k\chi, \quad \Omega = \beta_0 k^2.$$

Then, neglecting the derivatives with respect to χ in the nonlinear terms, we arrive at one of the systems describing long–short-wavelength resonance [23]:

$$i\frac{\partial \chi}{\partial t} + \beta_0 \frac{\partial^2 \chi}{\partial \xi^2} + \eta_0 |\chi|^2 \chi = \mu_0 u \chi, \tag{38}$$

$$\frac{\partial u}{\partial t} + \alpha_1 \frac{\partial u}{\partial \xi} + \nu \frac{\partial^3 u}{\partial \xi^3} + qu\frac{\partial u}{\partial \xi} = \frac{\mu_0}{2}\frac{\partial}{\partial \xi}(|\chi|^2). \tag{39}$$

Here we have $\xi = \zeta + 2\beta_0 kt$, $\eta_0 = \eta k$, $\mu_0 = \mu k$, and $\alpha_1 = \alpha_0 + 2\beta_0 k$. The role of the short-wavelength component is played by the transverse component of the elastic pulse and that of the long-wavelength component is played by the longitudinal component.

The system (38), (39) is not integrable [24, 25]. The spatial scale of the nanosecond elastic pulse is $l_s \sim 10^{-4}$ cm $\gg h$. In this case, the term $\partial^3 u/\partial\xi^3$ can be neglected in (39). In addition, for some cubic crystals the third-order elastic modulus λ_{166} is several times higher than λ_{111} [15]. For germanium, for example, we find $\lambda_{166} \approx 60$ din/cm^2, $\lambda_{111} \approx 12$ din/cm^2. In these cases we can assume $q = 0$ with a high degree of accuracy in (39). If the crystal contains no paramagnetic impurities, then $\eta_0 = 0$. The acoustic activity is then determined only by the natural contribution and the dynamics of a nanosecond elastic pulse with rf filling for the transverse component is described by an integrable Yajima–Oikawa system [23] which is a reduced variant of the Zakharov system [26].

To sum up, we can say that the system (33) and (34) modifies the well-known long–short-wavelength resonance systems to the case when the slowly varying envelope approximation does not hold for the short-wavelength component. In this case, the component ψ can sometimes only arbitrarily be considered to be short-wave since the time and spatial scales of this component may be comparable with the corresponding scales of the long-wavelength component u. Subsequently equations (33) and (34) will sometimes be called a modified long–short-wavelength resonance system. We note that this is a coupled system compris-

ing a "differentiated" nonlinear Schrödinger equation and the Korteweg–de Vries equation which can be integrated using the inverse scattering method [27–30].

The system (33), (34) can be represented as the variational principle for the functional $\int_{-\infty}^{\infty} L\,d\zeta$, where the "Lagrangian" density is

$$\begin{aligned} L &= \frac{1}{2}\left(\frac{\partial\theta^*}{\partial\zeta}\frac{\partial\theta}{\partial t}+\frac{\partial\theta}{\partial\zeta}\frac{\partial\theta^*}{\partial t}\right)\\ &+\frac{i}{2}\beta_0\left(\frac{\partial^2\theta^*}{\partial\zeta^2}\frac{\partial\theta}{\partial\zeta}-\frac{\partial^2\theta}{\partial\zeta^2}\frac{\partial\theta^*}{\partial\zeta}\right)\\ &+\frac{\eta}{2}\left|\frac{\partial\theta}{\partial\zeta}\right|^4+\frac{\partial Q}{\partial\zeta}\frac{\partial Q}{\partial t}+\alpha_0\left(\frac{\partial Q}{\partial\zeta}\right)^2-\nu\left(\frac{\partial^2 Q}{\partial\zeta^2}\right)^2\\ &+\frac{q}{3}\left(\frac{\partial Q}{\partial\zeta}\right)^3-\mu\left|\frac{\partial\theta}{\partial\zeta}\right|^2\frac{\partial Q}{\partial\zeta}. \end{aligned} \tag{40}$$

Here the "potentials" θ and Q are introduced using the relationships $\psi = \partial\theta/\partial\zeta$ and $u = \partial Q/\partial\zeta$.

Knowing L, we can easily construct the density

$$\mathcal{H}_{\text{eff}} = \Pi_\theta\frac{\partial\theta}{\partial t}+\Pi_{\theta^*}\frac{\partial\theta^*}{\partial t}+\Pi_Q\frac{\partial Q}{\partial t}-L$$

of the "effective Hamiltonian"

$$H_{\text{eff}}\equiv\int_{-\infty}^{\infty}\mathcal{H}_{\text{eff}}d\zeta,$$

where Π_θ, Π_{θ^*}, and Π_Q are canonical pulses given respectively by

$$\Pi_\theta = \frac{\partial L}{\partial(\partial\theta/\partial t)},\quad \Pi_{\theta^*} = \frac{\partial L}{\partial(\partial\theta^*/\partial t)},$$

$$\Pi_Q = \frac{\partial L}{\partial(\partial Q/\partial t)}.$$

Then, turning from the "potentials" θ and Q to the dynamic variables ψ and u, we write the system (33) and (34) in the Hamiltonian representation:

$$\frac{\partial\psi}{\partial t} = \frac{\partial}{\partial\zeta}\frac{\delta H_{\text{eff}}}{\delta\psi^*},\quad \frac{\partial u}{\partial t} = \frac{1}{2}\frac{\partial}{\partial\zeta}\frac{\delta H_{\text{eff}}}{\delta u}, \tag{41}$$

where

$$\begin{aligned} H_{\text{eff}} &= \int_{-\infty}^{\infty}\left[\frac{i}{2}\beta_0\left(\psi^*\frac{\partial\psi}{\partial\zeta}-\psi\frac{\partial\psi^*}{\partial\zeta}\right)-\frac{\eta}{2}|\psi|^4\right.\\ &\left.-\alpha_0 u^2-\frac{q}{3}u^3+\nu\left(\frac{\partial u}{\partial\xi}\right)^2+\mu|\psi|^2 u\right]d\xi \end{aligned} \tag{42}$$

is a conserved quantity.

It is easy to see that this system has at least another three integrals of motion:

$$\begin{aligned} &\int_{-\infty}^{\infty}u\,d\zeta = \text{const},\quad \int_{-\infty}^{\infty}\psi\,d\zeta = \text{const},\\ &\int_{-\infty}^{\infty}(u^2+|\psi|^2)d\zeta = \text{const}, \end{aligned} \tag{43}$$

which clarify the dynamics of the acoustic pulse and may make it slightly easier to study the system (33), (34). The existence of the variational principle may help us to find approximate solutions of the modified long–short-wave resonance system using the Ritz–Whitha "averaged Lagrangian" method [31].

4. STRUCTURE OF AN ELASTIC SOLITON IN AN ACOUSTICALLY ACTIVE MEDIUM

A general analysis of the system (33), (34) is extremely complex. Without claiming to be general, we shall attempt to find an approximate solution in the form of a localized traveling pulse. We shall subsequently call this pulse a soliton, giving this concept a generalized meaning, and not necessarily implying that the elastic self-interaction with itself is similar. Accordingly, we shall seek a solution in the form

$$\begin{aligned} u &= u(\zeta-v_0 t),\\ \psi &= F(\zeta-v_0 t)\exp[i(\Omega_s t+\Phi(\zeta-v_0 t))]. \end{aligned} \tag{44}$$

Quite clearly the parameter v_0 is related to the pulse velocity v in the laboratory coordinates by $v_0 = v - a_{\perp 0}$. The meaning of the second free parameter Ω_s will become clear from the following. As a result of integrating, equation (34) has the form

$$u'' = \frac{v_0-\alpha_0}{\nu}u-\frac{q}{2\nu}u^2+\frac{\mu}{2\nu}F^2, \tag{45}$$

where the prime denotes a derivative with respect to $\zeta - v_0 t$.

Substituting the second expression (44) into (33) after separating the real and imaginary parts yields

$$-v_0F'+\beta_0(F\Phi''+2F'\Phi')+3\eta F^2F' = \mu(uF)', \tag{46}$$

$$F'' = \frac{\Omega_s-v_0\Phi'}{\beta_0}F+\Phi'^2F+\frac{\Phi'}{\beta_0}F(\eta F^2-\mu u). \tag{47}$$

Multiplying (46) by F and integrating, we find an expression for Φ':

$$\Phi' = \frac{v_0}{2\beta_0}-\frac{3\eta}{4\beta_0}F^2+\frac{\mu}{\beta_0}\left(u-\frac{1}{F^2}\int_0^F uf\,df\right). \tag{48}$$

The system (45), (47), (48) is invariant with respect to the transformations $F \longrightarrow -F$, $u \longrightarrow u$. We shall

assume that there exists a functional dependence $u = u(F)$ and that this function can be expanded as a Maclaurin series in powers of F. In accordance with this invariance, the series will only contain even powers of F:

$$u = AF^2 + BF^4 + \ldots, \tag{49}$$

where A and B are coefficients to be determined. Here it is assumed that $u = 0$ if $F = 0$, i.e., no static longitudinal strain is present.

Substituting (49) into (45), (48), and (47), we obtain

$$2A(FF'' + F'^2) = \frac{2(v_0 - \alpha_0)A + \mu}{2\nu}F^2 + \frac{2(v_0 - \alpha_0)B - qA^2}{2\nu}F^4 + \ldots, \tag{50}$$

$$\Phi' = \frac{v_0}{2\beta_0} + \frac{3}{4\beta_0}(\mu A - \eta)F^2 + \ldots, \tag{51}$$

$$F'' = \left[\frac{\Omega_s}{\beta_0} - \left(\frac{v_0}{2\beta_0}\right)^2\right]F - \frac{v_0(\mu A - \eta)}{2\beta_0^2}F^3 + \ldots. \tag{52}$$

The dots in (50) and (51) imply further expansion in higher even powers of F and those in (52) in terms of odd powers of F.

The first integral (52) has the form

$$F'^2 = \left[\frac{\Omega_s}{\beta_0} - \left(\frac{v_0}{2\beta_0}\right)^2\right]F^2 - \frac{v_0(\mu A - \eta)}{4\beta_0^2}F^4 + \ldots. \tag{53}$$

Here the integration constant is assumed to be zero because outside the pulse propagation region (for $\zeta - v_0 t \longrightarrow \pm\infty$) the strain with all its derivatives vanishes.

After substituting (53) into (52) and (50) and equating the coefficients of F^2, F^4, and so on, on the left- and right-hand sides, we can find A and B. We give an expression for A:

$$A = \frac{\mu}{2}\left\{\alpha_0 - v_0 + 4\nu\left[\frac{\Omega_s}{\beta_0} - \left(\frac{v_0}{2\beta_0}\right)^2\right]\right\}^{-1}. \tag{54}$$

Note that for $\nu = 0$ (54) yields the expression for A obtained by direct integration of (34) when $\nu = q = 0$, $u = u(\zeta - v_0 t)$. The expression for B is fairly cumbersome but contains the parameter q of the acoustic anharmonicity of especially longitudinal strain. In this context we note that the lowest power of F for which the coefficient q appears in (53) is six. At the same time the elastic moduli of the longitudinal–transverse anharmonicity contained in μ are to the fourth power relative to the transverse strain of F in (50). Thus, the contribution of q to the nonlinear dynamics is of a higher order of smallness than the contribution of μ. Below we shall call the coefficients η and μ the constants of the transverse and longitudinal–transverse nonlinearities, respectively.

Confining ourselves to the third power of F on the right-hand side of (52) (a similar approach proved fairly useful to analyze the nonlinear propagation of ultrashort optical pulses [4]), after using (54), (44), and (49), integrating, and converting to laboratory coordinates, we find an approximate solution of the modified long–short-wavelength resonance system:

$$\psi = \psi_m e^{i\phi}\operatorname{sech}\left(\frac{z - vt}{\Delta}\right), \tag{55}$$

$$\phi = \Omega_s t + k(z - vt) + \frac{3}{2k\Delta}\tanh(z - vt), \tag{56}$$

$$\psi_m = \frac{4|\beta_0|}{\Delta} \times \left[4\beta_0 k\left(\frac{\mu^2}{\alpha_0 - 2\beta_0 k + 4\nu/\Delta^2} - 2\eta\right)\right]^{-1/2}, \tag{57}$$

$$\Omega_s = \beta_0(k^2 + \Delta^{-2}), \tag{58}$$

$$v = a_{\perp 0} + 2\beta_0 k, \tag{59}$$

$$u = u_m\operatorname{sech}^2\left(\frac{z - vt}{\Delta}\right), \tag{60}$$

$$u_m = \frac{2\mu\beta_0}{k}\{[\mu^2 - 2\eta(\alpha_0 - 2\beta_0 k)]\Delta^2 - 8\nu\eta\}^{-1}. \tag{61}$$

The solution (59)–(61) is a two-parameter one. The soliton spatial width Δ and the wave number k were taken as the free parameters. We shall subsequently analyze this solution. From the condition that the radicand in (57) is nonnegative we conclude that the solution (55)–(61) allows the existence of "supersonic" ($v > a_{\perp 0}$) and "subsonic" ($v < a_{\perp 0}$) solitons. These terms are slightly arbitrary and suggest that the soliton velocity is only related to the velocity of the transverse sound. However, the velocity of a supersonic soliton may well be lower than the velocity of the longitudinal sound. As a result of the approximation of slowly varying pulse profile in comoving coordinates used to derive (33) and (34), we find $2|\beta_0 k| \ll a_{\perp 0}$ [see (59)] and thus the soliton velocity cannot differ substantially from $a_{\perp 0}$. The constraints imposed on the free parameters k and Δ in both cases may be written in the form

$$v > a_{\perp 0}\colon\ \beta_0 k > 0,$$
$$\alpha_0 - \frac{\mu^2}{2\eta} + \frac{4\nu}{\Delta^2} < 2\beta_0 k < \min\left(\alpha_0 + 4\frac{\nu}{\Delta^2}, 2\pi\frac{|\beta_0|}{h}\right), \tag{62}$$

$$v < a_{\perp 0}\colon\ \beta_0 k < 0,$$
$$\frac{\mu^2}{2\eta} - \alpha_0 - \frac{4\nu}{\Delta^2} < 2|\beta_0 k| < 2\pi\frac{|\beta_0|}{h}. \tag{63}$$

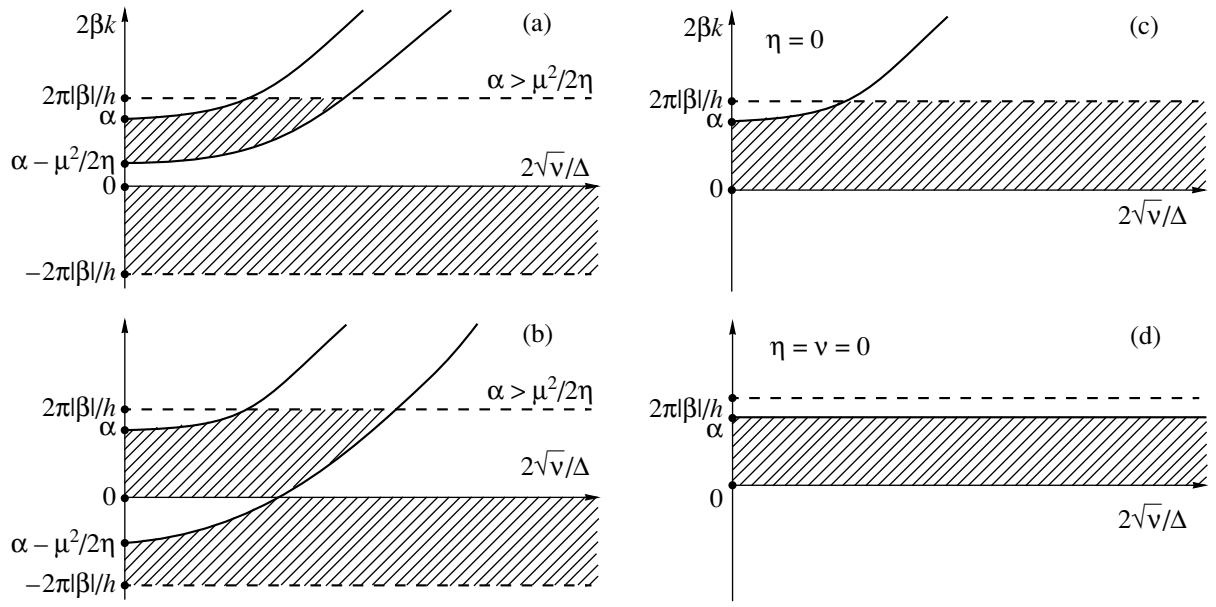


Fig. 2. Regions of free parameters of a two-component soliton of the system (33), (34) on the half-plane ($2\sqrt{\nu}/\Delta > 0$) for various relationships between the coefficients of this system (parameters of the medium). The upper quadrants ($2\beta k > 0$) correspond to the free parameters of a supersonic ($v > a_{\perp 0}$) soliton and the lower quadrants ($2\beta k < 0$) correspond to the free parameters of a subsonic ($v < a_{\perp 0}$) soliton.

Here it is assumed that the absolute value of k cannot exceed π/h (horizontal dashed line in Fig. 2) [32] (or the length of an elastic wave $\lambda < 2h$). Quite clearly for the case of a nonlinear wave (in particular, an acoustic pulse) this constraint is approximate and possibly an estimate. We shall subsequently assume that $\nu > 0$ (in particular, in cubic one-dimensional crystals $\tilde{\nu} = 2a_{\perp 0}\nu = h^2 a_{\parallel}^2/12$ [33, 34]). In addition, during propagation along symmetric acoustic axes in cubic crystals we usually find $a_{\parallel} > a_{\perp}$ [32] and thus $\alpha_0 > 0$. Since in a thermodynamically equilibrium crystal $W_1 > W_3$ we find $\eta > 0$ [see (35)].

We note that

$$v_{\perp} = a_{\perp 0} + 2\beta_0 k \quad \text{and} \quad v_{\parallel} = a_{\perp 0} + 4\nu/\Delta^2$$

are the velocities of differentiated nonlinear Schrödinger and Korteweg–de Vries solitons in the laboratory coordinates. It can be seen from (57) and (61) that $\psi_m = 0$ for $v_{\perp} - v_{\parallel} = \alpha_0$ whereas $u_m \neq 0$, i.e., at this resonance all the energy of the transverse component is transferred to the longitudinal one.

In Fig. 2 the regions of free parameters on the plane (k, Δ^{-1}) corresponding to supersonic (upper quadrants, $2\beta k > 0$) and subsonic (lower quadrants, $2\beta k < 0$) solitons are shown hatched. In cases of strong transverse nonlinearity when $\alpha_0 > \mu^2/2\eta$ the values of $|k|$ and Δ for a subsonic soliton may have arbitrary values ($|k| < \pi/h$) whereas for a supersonic soliton the region of permissible values of k and Δ on the plane ($2\beta_0 k$: $2\sqrt{v}\,\Delta^{-1}$) is a strip of width $\mu^2/2\eta$ along the vertical bounded by two parabolas and the horizontal dashed curve $2\beta_0 k = 2\pi|\beta_0|/h$ (Fig. 2a). As the coefficient $\mu^2/2\eta$ increases, the lower parabola moves downward parallel to the abscissa and for $\mu^2/2\eta > \alpha_0$ falls partly within the lower quadrant, decreasing the range of free parameters for the subsonic soliton (Fig. 2b). In this case, the upper parabola remains fixed. For $\mu^2/2\eta \longrightarrow \infty$ ($\eta = 0$) the lower parabola is displaced in the lower quadrant toward $-\infty$, reducing to zero the region of free parameters for the subsonic soliton (Fig. 2c). Thus, only supersonic solitons can form in the absence of spin–phonon interaction ($\eta = 0$). If $\nu = 0$, both parabolas straighten out to form horizontal lines. In other respects the analysis is similar to that just made. In particular, for $\eta = 0$ and $\nu = 0$ the region of free parameters of a supersonic ($v > a_{\perp 0}$) soliton degenerates into a horizontal strip of width α_0 (Fig. 2d). It is important to note that for $\eta = \nu = q = 0$ and also for $k\Delta \gg 1$ the solution (55)–(61) yields the solution of the Yajima–Oikawa system [see (38), (39) for $\nu = q = 0$]. It follows from Fig. 2d and (59) and also from the expression for α_0 that the velocity of a nanosecond elastic soliton in a crystal containing no paramagnetic impurities lies in the range $a_{\perp 0} < v < (a_{\parallel}^2 - a_{\perp 0}^2)/2a_{\perp 0} + a_{\perp 0}$. If $\mu = 0$, the lower and upper parabolas in Fig. 2a are the same, narrowing to zero the permissible range of free parameters for supersonic solitons, which as a result cannot form in the absence of longitudinal–transverse nonlinearity. Thus, for $\eta = 0$, $\mu \neq 0$ only supersonic solitons can form whereas for

$\mu = 0$, $\eta \neq 0$ only subsonic solitons form. In the general case when $\mu \neq 0$, $\eta \neq 0$ the regions of free parameters of a two-component longitudinal–transverse soliton comprise two nonintersecting subsonic and supersonic zones separated by a forbidden zone ("gap") where if $\alpha_0 < \mu^2/2\eta$ both zones touch at the point $2\sqrt{\lambda}\,\Delta^{-1} = \sqrt{\mu^2/2\eta - \alpha_0}$ on the abscissa. In all cases, phase self-modulation [see (56)] leads to more frequent oscillations from the soliton tails toward their center.

It is easy to see from (62), (63), and (61) that both for supersonic and subsonic solitons the sign of u_m is the same as μ. Consequently, for $\mu < 0$ the longitudinal component comprises a local traveling compressive strain whereas for $\mu > 0$ it comprises a tensile strain. We also note that as Δ decreases, the amplitude of the transverse strain ψ_m increases for $v > a_{\perp 0}$ and $v < a_{\perp 0}$. The situation with the longitudinal strain is more complex: for a supersonic soliton the value of u_m increases with decreasing Δ whereas for a subsonic soliton it decreases.

It is important to determine the initial conditions for which super- or subsonic two-component solitons can form. In general, this question can be answered by solving the Cauchy problem for a modified long–short-wavelength resonance system which is of separate interest.

The parameter Ω_s has the meaning of the angle of rotation of the plane of polarization of a soliton per unit time. It follows from (58) that the direction of this rotation is determined by the sign of β_0. For practical purposes the parameter κ_s which determines the angle of rotation of the plane of polarization of a soliton per unit length is more convenient. Quite clearly we have $\kappa_s = \Omega_s/v \approx \Omega_s/a_{\perp 0}$, $k \approx \omega/a_{\perp 0}$, and $\Delta \approx \tau_p a_{\perp 0}$. From (58) and (36) and comparing (55), (56) with (37) we then obtain

$$\kappa_s = -\kappa_l[1 + (\omega\tau_p)^{-2}]. \tag{64}$$

Following [35] we call the dimensionless parameter $(k\Delta)^{-2} = (\omega\tau_p)^{-2}$ the video parameter since for $(\omega\tau_p)^{-2} \ll 1$ we have an envelope soliton and for $(\omega\tau_p)^{-2} \geq 1$ we have a video soliton (i.e., a soliton which contains of the order of a single wavelength of the oscillations of the transverse component). The "minus" sign in front of κ_l indicates that the direction of rotation of the plane of polarization of an acoustic soliton is opposite to the direction of rotation of a monochromatic linear wave. The shorter the soliton for fixed ω, the larger the difference between κ_s and κ_l. For envelope solitons when $(\omega\tau_p)^{-2} \ll 1$ we have $|\kappa_s| \approx \kappa_l$.

Expression (64) is also obtained in analyses of the optical Faraday effect [11]. The soliton solution under discussion together with expression (64) does not allow a passage to the limit of solutions of linear equations when $\eta = \mu = 0$. This can be explained by the fact that solitons are essentially nonlinear structures whose amplitudes tend to infinity when the coefficients of the nonlinear terms tend to zero. Quite clearly, when an initial strain pulse enters a medium, this pulse begins to undergo a certain evolution as a result of which one or more solitons may form together with "soliton tails" comprising small-amplitude (linear) longitudinal and transverse waves. This analysis suggests that longitudinal linear waves should appear as precursors of two-component solitons since $v < a_{\parallel}$. In addition, when $\beta_0 k > 0$ ($v > a_{\perp 0}$) solitons should be recorded behind longitudinal linear waves, followed by transverse small-amplitude waves. In accordance with (64), the direction of rotation of the plane of polarization of these waves is predicted to be the opposite of this rotation for solitons. If $\beta_0 k$ ($v < a_{\perp 0}$) the soliton precursors are initially longitudinal and then transverse small-amplitude waves.

We shall now give some estimates. We shall take as an example an MgO crystal containing Fe^{2+} paramagnetic centers at liquid helium temperature $T = 4$ K. Let us assume that the magnetic field creates Zeeman splitting $\omega_0 = 10^{12}$ s^{-1}. We then have $b \equiv \hbar\omega_0/k_B T = 2$, $W_1 = e^b/Z = 0.87$, $W_2 = 1/Z = 0.12$, $W_3 = e^{-b}/Z = 0.01$ (here $Z \equiv 1 + 2\sinh b$ is the partition function). In this case, condition (26) can be satisfied, assuming $\tau_p \sim 10$ ps. Then taking [2, 10, 14, 15, 31, 36] $G_{11} = 1.3 \times 10^{-13}$ erg, $G_{44} = 0.8 \times 10^{-13}$ erg, $n \sim 10^{19}$ cm^{-3}, $\rho \simeq 3.5$ g/cm^3, $h \simeq 5 \times 10^{-8}$ cm, $a_{\perp 0} \simeq 6 \times 10^5$ cm/s, $a_{\parallel} \simeq 9 \times 10^5$ cm/s, $\lambda_{166} \simeq -10^{12}$ din/cm^2, $\lambda_{44} \simeq 5 \times 10^{11}$ din/cm^2, we find from (35) $\eta = 6.4 \times 10^3$ cm/s, $\mu = 2.5 \times 10^5$ cm/s, $\beta_0 = 4 \times 10^{-6}$ cm^2/s, $\nu = 1.3 \times 10^{-10}$ cm^3/s, $\alpha_0 = 3.9 \times 10^5$ cm/s, and $\Delta = 5 \times 10^{-6}$ cm. Here we assumed that $\xi_{54z} = \beta_{54z} = 0$ for MgO : Fe^{2+}. In this example, the main contribution to the longitudinal–transverse nonlinearity (to the parameter μ) is made by the acoustic anharmonicity whereas the transverse nonlinearity (the parameter η) is determined by the spin–phonon interaction. Substituting these values of the parameters into (59) and (60), we conclude that in this particular case only supersonic solitons can form with constraints for the wave number $0 < k < \pi/h = 6 \times 10^7$ cm^{-1}. For $k = \Delta^{-1} = 2 \times 10^5$ cm^{-1} (videopulse) we obtain from (57), (61), (64), and (36) $\psi_m = 5 \times 10^{-3}$, $u_m \simeq -7 \times 10^{-6}$, and $\kappa_s = 0.6$ rad/cm. Thus, the soliton deformation consists mainly of the transverse component whereas the longitudinal component is negligible. The pressure P_s inside this soliton reaches $P_s \simeq \rho a_{\perp 0}^2 \psi_m \simeq 6$ kbar. The signal intensity is $I \simeq \rho a_{\perp 0}^3 \psi_m^2 \simeq 2 \times 10^6$ W/cm^2. If the magnetic field is now reduced to $\omega_0 \simeq 10^{11}$ c^{-1} and the duration increased to $\tau_p \simeq 100$ ps, a similar analysis shows that only subsonic solitons with $0 < |k| < \pi/h$ can form in this case. Then for $k = \Delta^{-1} = 2 \times 10^4$ cm^{-1} we have $\psi_m \approx 2 \times 10^{-3}$, $u_m \simeq 10^{-6}$, $\kappa_s = 0.08$ rad/cm, $P_s \simeq 0.1$ kbar, and $I \simeq 3 \times 10^5$ W/cm^2. Note that in this case the acoustic anharmo-

nicity and the spin–phonon interaction make comparable contributions to the longitudinal–transverse nonlinearity. We again note that $\psi_m \gg u_m$. This factor justifies neglect of the longitudinal component in [13]. In both examples the acoustic dispersion of the longitudinal component plays a negligible role since $4\nu/\Delta^2 \ll 2|\beta_0 k|$, α_0, $\mu^2/2\eta$ [see (57), (61)]. However, the situation changes radically if $a_{\perp 0} = a_{\parallel}$ ($\alpha_0 = 0$). This condition is satisfied for elasto-isotropic ionic crystals with central forces of interaction between the ions in the absence of natural acoustic activity [31]. In this case, the long–short-wavelength resonance condition is strictly satisfied: the linear velocities of both components are equal which ensures efficient energy exchange between them. We shall illustrate this by assuming $a_{\parallel} = a_{\perp 0} = 6 \times 10^5$ cm^{-1} in the two examples given above. For $\omega_0 = 10^{12}$ c^{-1}. $\tau_p = 10$ ps only supersonic solitons can form with $0 < k < 10^6$ cm^{-1}. Taking $k = \Delta^{-1} = 2 \times 10^5$ cm^{-1} we obtain $\psi_m \approx 2 \times 10^{-5}$, $u_m \approx -0.6 \times 10^{-5}$, $\kappa_s = 0.64$ rad/cm. Thus, for $\alpha_0 = 0$ the longitudinal and transverse strains of the two-component soliton are of comparable order of magnitude. Here we note the fundamental role of the acoustic dispersion of the longitudinal component which determines the upper limit of the wave number k [see (62)].

For $\omega_0 = 10^{11}$ s^{-1}, $\tau_p = 100$ ps, $\alpha_0 = 0$ it is only possible to have supersonic solitons for which $0 < k < 0.08$ cm^{-1}. The minimum value of the videoparameter is $(k_{\max}\Delta)^{-2} = 10^4$. Thus, in this case only videosolitons should be formed whereas envelope solitons are eliminated. The very high value of the videoparameter suggests that the transverse pulse component is a unipolar videosoliton with a rotating plane of polarization (in all cases the longitudinal component is unipolar). Calculations for $k = 0.02$ cm^{-1} give $\psi_m = 2 \times 10^{-5}$, $u_m = -8 \times 10^{-4}$, $\kappa_s = 0.04$ rad/cm. Thus here we have $|u_m| \gg \psi_m$. Since $(k\Delta)^{-2} \gg 1$, the value of κ_l for a plane linear wave with $k < 0.08$ cm^{-1} is 10^4 times lower than the similar value of κ_s for a videosoliton. As a result, the magnetic rotation of the plane of polarization for a linear wave with $k < 0.08$ cm^{-1} cannot be observed experimentally whereas for a two-component videosoliton of duration $\tau_p \sim 100$ ps this effect may be very appreciable: up to several degrees per centimeter. On the basis of these numerical examples we can also conclude that for $\alpha_0 = 0$ the soliton amplitudes are 10–100 times smaller than those obtained when the linear velocities are mismatched. In addition, for $\alpha = 0$ the longitudinal strain component plays an increased role in the soliton formation.

5. CONCLUSIONS

The wave transparency approximation (26) has been used to obtain a system of nonlinear equations (33), (34) for the coupled dynamics of the longitudinal and transverse components of an acoustic picosecond ($\tau_p \sim 10$ ps) pulse in a crystal possessing the property of acoustic activity. On the basis of an approximate quasisoliton solution of this system, it is concluded that the plane of polarization of the transverse component of an elastic soliton rotates in the opposite direction to the rotation for a plane linear wave. Both cases are provided for by this system which generalizes the well-known nonlinear long–short-wave resonance systems to pulses containing down to a single period of the vibrations when the standard approximation of a slowly varying envelope ceases to hold. Strictly speaking this resonance occurs when the linear velocities of the longitudinal and transverse sound are similar [$\alpha_0 = 0$, see (34)]. The numerical estimates given here are consistent with this statement: in this case energy exchange between the two strain components has the highest intensity. It is shown that for such short pulses the term "long–short-wavelength resonance" is to a considerable extent arbitrary since the wavelength of the transverse component may be comparable with its spatial duration Δ and with the duration of the longitudinal component. A more appropriate term could well be "longitudinal–transverse resonance".

From the mathematical point of view the system (33), (34) may be curious in that it contains two integrable equations, the differentiated nonlinear Schrödinger equation and the Korteweg–de Vries equation. Studies of the deep mathematical structure of this system, in particular its integrability, may well be of some value. This interest may be partly stimulated by the fact that this system has Lagrangian and Hamiltonian structures and also nontrivial conservation laws [see (42), (43)]. The most far-reaching results may well be achieved by numerical integration of a modified long–short-wavelength resonance system. This will allow us to study the stages of formation of isolated two-component solitons for which the transverse component has a rotating plane of polarization and also processes of intersoliton interaction.

At this point the question of the stability of these approximate quasisoliton solutions with respect to transverse perturbations arises (in particular, with respect to self-focusing). For single-component solitons this question can be solved reasonably well using a qualitative analysis and the method of moments [37, 38]. In this case, it is sufficient to know the dependence of the soliton velocity on its amplitude. For two-component solitons this problem cannot be solved so easily at least because in many cases qualitatively different dependences of the soliton velocity on the amplitudes of the longitudinal and transverse strains are observed.

The spatial nonlocality of the spin–phonon interaction is taken into account minimally here using the tensor β_{ijklm} which is nonzero in crystals without a center of symmetry in the presence of paramagnetic impurities (a pure crystal containing no such impurities can

then have a center of symmetry). This nonlocality contributes to the acoustic activity. As the duration of the acoustic pulse decreases to $\tau_p \sim 1$ ps allowance should be made for higher terms of the expansion in terms of the spatial derivatives of the crystal lattice site displacements, containing the tensor σ_{ijklmp} [see (7)]. In this case, the terms $\sim\partial^2 V_\perp/\partial t^2$ must be added on the right-hand sides of expressions (30) and (31). As a result for pulses with $\tau_p \sim 1$ ps the term $\sim\partial^3\psi/\partial\zeta^3$ must be added to the left-hand side of (46).

It is interesting to study the magnetic acoustic activity effect over a wider frequency range, not only for $\omega \ll \omega_0$. Of particular interest is the case where the pulse spectrum overlaps the quantum transitions in the subsystem: $\omega_0\tau_p \ll 1$ [11, 12, 19, 20]. This situation may be achieved by reducing the external magnetic field to levels when $\omega_0 \sim 10^{11}$ s^{-1} for $\tau_p \sim 1$–10 ps. Under conditions of spectral overlap the excitation of paramagnetic quantum impurities is considerably stronger than that in the wave transparency regime (26) considered here.

For $\omega \simeq \omega_0$ a two-component elastic pulse can propagate under conditions of self-induced transparency. In [8] self-induced transparency was studied for longitudinal rf strain pulses propagating at a certain angle to $\mathbf{B}_0$ in an $S = 1$ spin system. In [9] a similar effect was studied for transverse pulses propagating parallel to $\mathbf{B}_0$ in an $S = 1/2$ system. In this context it may be interesting to study the characteristics of self-induced transparency when the frequency ω of the transverse pulse component is the same as ω_0 and the low-frequency longitudinal component propagates with the transverse component in the long–short-wave resonance regime.

The question of the acoustic activity effect over a broad frequency range between $\omega \ll \omega_0$ and $\omega \gg \omega_0$ may be answered using the method of analytic continuation of the dispersion parameters to the complex plane proposed in [11].

There is thus reason to assume that on entering the picosecond range of durations the traditional effects of physical acoustics (including acoustic activity), which are observed for envelope pulses and also for plane linear waves, will acquire significant characteristics.

ACKNOWLEDGMENTS

This work was partly supported by the Russian Foundation for Basic Research (project no. 00-02-17436).

REFERENCES

1. V. A. Krasil'nikov and V. V. Krylov, *Introduction to Physical Acoustics* (Nauka, Moscow, 1984).
2. J. W. Tucker and V. W. Rampton, *Microwave Ultrasonics in Solid State Physics* (North-Holland, Amsterdam, 1972; Mir, Moscow, 1975).
3. J. T. Darrow, B. B. Hu, X. C. Chang, and D. H. Auston, Opt. Lett. **15**, 323 (1990).
4. S. A. Akhmanov, V. A. Vysloukh, and A. S. Chirkin, *The Optics of Femtosecond Pulses* (Nauka, Moscow, 1998).
5. S. A. Akhmanov and V. É. Gusev, Usp. Fiz. Nauk **162** (3), 3 (1992) [Sov. Phys. Usp. **35**, 153 (1992)].
6. A. V. Kim, M. Yu. Ryabikin, and A. M. Sergeev, Usp. Fiz. Nauk **169** (1), 58 (1999).
7. S. L. McCall and E. L. Hahn, Phys. Rev. Lett. **18**, 908 (1967).
8. N. S. Shiren, Phys. Rev. B **2**, 2471 (1970).
9. G. A. Denisenko, Zh. Éksp. Teor. Fiz. **60**, 2269 (1971) [Sov. Phys. JETP **33**, 1220 (1971)].
10. V. A. Golenishchev-Kutuzov, V. V. Samartsev, N. K. Solovarov, and B. M. Khabibulin, *Magnetic Quantum Acoustics* (Nauka, Moscow, 1997).
11. S. V. Sazonov, Zh. Éksp. Teor. Fiz. **107**, 20 (1995) [JETP **80**, 10 (1995)].
12. S. V. Sazonov, J. Phys.: Condens. Matter **4**, 6485 (1992).
13. S. V. Sazonov, J. Phys.: Condens. Matter **6**, 6295 (1994).
14. S. A. Altshuler and B. M. Kozyrev, *Electron Paramagnetic Resonance in Compounds of Transition Elements* (Nauka, Moscow, 1972; Halsted, New York, 1975).
15. L. K. Zarembo and V. A. Krasil'nikov, *Introduction to Nonlinear Physical Acoustics* (Nauka, Moscow, 1966).
16. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 7: *Theory of Elasticity* (Nauka, Moscow, 1987; Pergamon, New York, 1986).
17. R. H. Pantell and H. E. Puthoff, *Fundamentals of Quantum Electronics* (Wiley, New York, 1969; Mir, Moscow, 1972).
18. E. Taker, in *Physical Acoustics: Principles and Methods*, Vol. 4, Part A: *Applications to Quantum and Solid State Physics*, Ed. by W. P. Mason (Academic, New York, 1966; Mir, Moscow, 1969).
19. É. M. Belenov and A. V. Nazarkin, Pis'ma Zh. Éksp. Teor. Fiz. **51**, 252 (1990) [JETP Lett. **51**, 288 (1990)].
20. É. M. Belenov, A. V. Nazarkin, and V. A. Ushchapovskiĭ, Zh. Éksp. Teor. Fiz. **100**, 762 (1991) [Sov. Phys. JETP **73**, 422 (1991)].
21. M. B. Vinogradova, O. V. Rudenko, and A. P. Sukhorukov, *The Theory of Waves* (Nauka, Moscow, 1990).
22. N. S. Buĭnov, V. R. Nagibarov, and N. K. Solovarov, Ukr. Fiz. Zh. **22**, 151 (1977).
23. R. K. Dodd, J. C. Eilbeck, J. Gibbon, and H. C. Morris, *Solitons and Nonlinear Wave Equations* (Academic, New York, 1982; Mir, Moscow, 1988).
24. E. S. Benilov and S. P. Burtzev, Phys. Lett. A **98**, 256 (1983).
25. V. G. Makhan'kov, Yu. P. Rybakov, and V. I. Sanyuk, Usp. Fiz. Nauk **162** (2), 121 (1992) [Sov. Phys. Usp. **35**, 55 (1992)].
26. V. E. Zakharov, Zh. Éksp. Teor. Fiz. **62**, 1745 (1972) [Sov. Phys. JETP **35**, 908 (1972)].

27. D. N. Kaup and A. C. Newell, J. Math. Phys. **19**, 798 (1978).
28. T. Kawata, N. Kobayashi, and H. Inoue, J. Phys. Soc. Jpn. **46**, 1008 (1978).
29. V. A. Vysloukh and I. V. Cherednik, Teor. Mat. Fiz. **78**, 35 (1989).
30. S. Novikov, S. V. Manakov, L. P. Pitaevskii, and V. E. Zakharov, *Theory of Solitons: the Inverse Scattering Method* (Nauka, Moscow, 1980; Consultants Bureau, New York, 1984).
31. D. Anderson, Phys. Rev. A **27**, 3135 (1983).
32. C. Kittel, *Introduction to Solid State Physics* (Wiley, New York, 1976, 5th ed.; Fizmatlit, Moscow, 1963).
33. V. É. Gusev and A. A. Karabutov, *Laser Optical Acoustics* (Nauka, Moscow, 1991).
34. A. M. Kosevich and A. S. Kovalev, *Introduction to Nonlinear Physical Mechanics* (Naukova Dumka, Kiev, 1989).
35. S. V. Sazonov, Fiz. Tverd. Tela (St. Petersburg) **37**, 1612 (1995) [Phys. Solid State **37**, 875 (1995)].
36. O. Anderson, in *Physical Acoustics: Principles and Methods*, Vol. 3, Part B: *Lattice Dynamics*, Ed. by W. P. Mason (Academic, New York, 1965; Mir, Moscow, 1968).
37. V. S. L'vov, *Nonlinear Spin Waves* (Nauka, Moscow, 1987).
38. V. I. Vlasov, V. I. Talanov, and V. A. Petrishchev, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **14**, 1353 (1971).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 1, 2000, pp. 167–169.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 1, 2000, pp. 189–192.

SOLIDS
Structure

Spiral Magnetic Configuration in a Thin Film with Biaxial Anisotropy

Yu. I. Gorobets*, Yu. I. Dzhezherya*, and L. P. Mironenko**
**Institute of Magnetism, Ukrainian National Academy of Sciences, Kiev, 252680 Ukraine*
***Kiev Polytechnical Institute, Kiev, 252056 Ukraine*
Received November 5, 1999

Abstract—The conditions for the existence of a spiral magnetic configuration are determined for a thin ferromagnetic film with biaxial anisotropy. *© 2000 MAIK "Nauka/Interperiodica".*

The Landau–Lifshitz equations for magnetic systems can have a wide spectrum of solutions, which in turn correspond to a local minimum of the energy. The configurations corresponding to such solutions do not arise spontaneously, but, being created artificially, they remain stable.

Examples of such structures are a spiral magnetic configuration, which is stable in a thin film with biaxial anisotropy.

Let us consider a thin ferromagnetic film in the *yx* plane with biaxial anisotropy. Let us assume that the easy magnetization axis is directed along the *z*-axis orthogonal to the plane of the film, while the orthorhombic anisotropy axis is directed along the *x*-axis.

We shall show that for a definite ratio of the magnetic parameters a spiral (helicoidal) configuration, whose magnetization is modulated in the *x* direction according to a periodic law, can be stable in such a system.

To solve this problem we proceed from the energy functional taking into account the basic types of magnetic interactions of the system:

$$E = \int d\mathbf{v} W(\mathbf{M}),$$
$$W(\mathbf{M}) = \left\{\frac{\alpha}{2}(\nabla\mathbf{M})^2 - \frac{\beta}{2}M_z^2 + \frac{\rho}{2}M_x^2\right\} \quad (1)$$
$$+ \frac{1}{2}M_i(\mathbf{r})\int d\mathbf{v}' M_j(\mathbf{r}')\frac{\partial^2}{\partial x_i \partial x_j'}\frac{1}{|\mathbf{r}-\mathbf{r}'|},$$

where α and β are the exchange interaction and easy-axis anisotropy constants, ρ is the orthorhombic anisotropy constant, and **M** is the magnetization of the material. The integration extends over the volume occupied by the magnetic field, and *W* is the magnetic energy density. The second term in the expression for the energy density corresponds to the magnetostatic interaction, $x_i - x$, y, z, $M_i - M_x$, M_y, M_z.

We represent the magnetization vector in terms of the angular variables in a spherical system with polar axis *x*:

$$\mathbf{M} = M_0(\cos\theta, \sin\theta\sin\varphi, \sin\theta\cos\varphi), \quad (2)$$

where θ and φ are polar and azimuthal angles, and M_0 is the saturation magnetization of the film material.

It is shown in [1–3] that in thin magnetic films the magnetization distribution can be treated as uniform over the thickness. Adopting this assumption, we shall determine one variant of a one-dimensional periodic structure whose magnetization varies in the *x* direction.

To this end we make in Eq. (1) the substitution of variables according to Eq. (2) and integrate it with respect to y, y', z, z'.

As a result of these transformations the energy of the system assumes the form

$$E = \frac{E_0}{2L_x}\left\{\int dx\left(\left(\frac{\partial\theta}{\partial x}\right)^2 + \alpha\sin^2\theta\left(\frac{\partial\varphi}{\partial x}\right)^2\right.\right.$$
$$\left. - \beta\sin^2\theta\cos^2\varphi + (\rho + 4\pi)\cos^2\theta\right) \quad (3)$$
$$+ \iint dx' dx J(x - x')(\cos\varphi(x)\cos\varphi(x')$$
$$\left.\times \sin\theta(x)\sin\theta(x') - \cos\theta(x)\cos\theta(x'))\right\},$$

where

$$E_0 = M_0^2 L_x L_y L_z, \quad J(x) = \frac{2}{L_z}\ln\left(\frac{x^2 + L_z^2}{x^2}\right).$$

Setting the variation of the energy functional (3) to zero we obtain an equation for the angular variables of

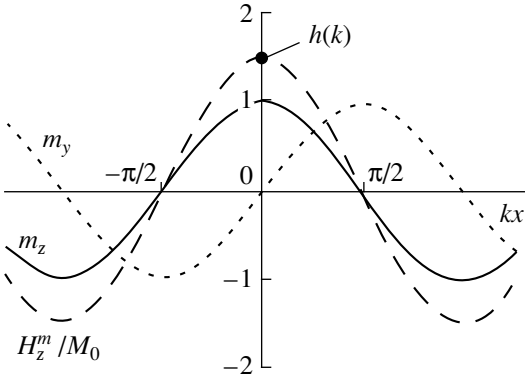


Fig. 1. Distribution of the magnetization and the intrinsic magnetostatic field in a ferromagnetic film.

the magnetization field:

$$\begin{aligned} &-\alpha\frac{\partial^2\theta}{\partial x^2}+\left(\alpha\left(\frac{\partial\varphi}{\partial x}\right)^2-\beta\cos^2\varphi-\rho\right)\sin\theta\cos\theta\\ &-h_x^m(x)\sin\theta+h_z^{(m)}(x)\cos\theta\cos\varphi\;=\;0,\\ &-\alpha\frac{\partial}{\partial x}\sin^2\theta\frac{\partial\varphi}{\partial x}+\beta\sin^2\theta\sin\varphi\cos\varphi\\ &-h_z^m(x)\sin\theta\sin\varphi\;=\;0. \end{aligned} \tag{4}$$

In Eqs. (4) the following notation is used for determining the characteristic magnetostatic fields of the system:

$$\begin{aligned} h_z^m(x)\;&=\;\int dx'\,J(x-x')\sin\theta(x')\cos\varphi(x'),\\ h_x^m(x)\;&=\;4\pi\cos\theta-\int dx'\,J(x-x')\cos\theta(x'). \end{aligned} \tag{5}$$

It is easy to show that the system of equations (4) can possess a particular solution that describes a helicoidal magnetic structure:

$$\theta\;=\;\pi/2,\quad \varphi(x)\;=\;kx, \tag{6}$$

with the following relation between the wave number k of the spiral and the anisotropy constant β:

$$\beta\;=\;h(k). \tag{7}$$

where

$$\begin{aligned} h(k)\;&=\;4\int_0^\infty dt\ln\left(\frac{1+t^2}{t^2}\right)\cos(kLt)\\ &=\;\frac{4\pi}{|kL|}(1-\exp(-|kL|)) \end{aligned}$$

is the amplitude of the magnetostatic field in the film in relative units M_0.

Figure 1 shows schematically the distribution of the magnetic field in the film along the x axis:

$$H_z^m(x)\;=\;M_0h(k)\cos kx$$

and the components of the magnetization of the system. It follows from the relation (7) that the solution (6) exists if the anisotropy constant satisfies the condition $\beta < 4\pi$.

The solution (6) of the variational equations (4) determines a helicoidal configuration corresponding to an extremum of the magnetic energy of a thin film with biaxial anisotropy. Consequently, to answer the question of the stability of the structure we shall consider a second variation of the energy near the ground state (6):

$$\begin{aligned} E_2\;&=\;\frac{E_0}{2L_x}\{E_2(\delta\varphi)+E_2(\delta\theta)\},\\ E_2(\delta\varphi)\;&=\;\int dx\left(\alpha\left(\frac{\partial\delta\varphi}{\partial x}\right)^2-\beta\delta\varphi^2\sin^2kx\right)\\ &+\iint dx'dx\,J(x-x')\delta\varphi(x')\delta\varphi(x)\sin kx\sin kx', \end{aligned} \tag{8}$$

$$\begin{aligned} E_2(\delta\theta)\;&=\;\int dx\left(\alpha\left(\frac{\partial\delta\theta}{\partial x}\right)^2+(\rho+4\pi-\alpha k^2)\delta\theta^2\right)\\ &-\iint dx'dx\,J(x-x')\delta\varphi(x')\delta\varphi(x). \end{aligned}$$

To eliminate ambiguity in the Fourier expansion of the variations, we shall confine the variations in the values of the wave number q to the region $[-k, k]$, which is essentially the analog of the first Brillouin zone, and we shall represent $\delta\varphi$ and $\delta\theta$ in the form

$$\begin{pmatrix}\delta\varphi(x)\\ \delta\theta(x)\end{pmatrix}\;=\;\frac{1}{\sqrt{L_x}}\sum_n\sum_q\begin{pmatrix}a_n(q)\\ b_n(q)\end{pmatrix}\exp(i(q+2kn)x). \tag{9}$$

The summation in Eq. (9) extends over the admissible values of the wave numbers q and the reciprocal lattice sites n.

On the basis of the obvious condition

$$\frac{1}{L_x}\int_{-L_x/2}^{L_x/2}dx\exp(i(q-q'+(n-m)2k)x)\underset{L_x\to\infty}{=}\delta_{qq'}\delta_{nm}, \tag{10}$$

where δ_{nm} and $\delta_{qq'}$ are the Kronecker delta functions, it is easy to show that the values of the expansion coefficients have the form

$$\begin{pmatrix}a_n(q)\\ b_n(q)\end{pmatrix}\;=\;\frac{1}{\sqrt{L_x}}\int_{-L_x/2}^{L_x/2}dx\begin{pmatrix}\delta\varphi(x)\\ \delta\theta(x)\end{pmatrix}\exp(i(q+2kn)x). \tag{11}$$

Substituting the expression (9) for the variations of the magnetization field into the quadratic form (8) and performing the integration, we obtain the following result:

$$E_2(\delta\varphi) = \sum_q \sum_n \{\mathscr{E}_1(q, n) a_n^*(q) a_n(q) + d(q, n)(a_{n+1}^*(q) a_n(q) + a_n^*(q) a_{n+1}(q))\}, \quad (12)$$

$$E_2(\delta\theta) = \sum_q \sum_n \mathscr{E}_2(q, n) b_n^*(q) b_n(q),$$

where

$$\mathscr{E}_1(q, n) = \alpha(q + kn)^2 - \frac{1}{2}h(k) + \frac{1}{4}h(q + 2kn - k) + \frac{1}{4}h(q + 2kn + k),$$

$$\mathscr{E}_2(q, n) = \alpha(q + 2kn)^2 + \rho - \alpha k^2 + 4\pi - h(q + 2kn),$$

$$d(q, n) = \frac{1}{4}(h(k) - h(q + k + 2kn)).$$

The helicoidal structure described by the relations (6) corresponds to a minimum of the energy and is stable if the quadratic forms (9) are positive-definite. The quadratic form $E_2(\delta\theta)$ satisfies this requirement if

$$\mathscr{E}_2(0, 0) = \rho - \alpha k^2 > 0. \quad (13)$$

Thus, the relation (13) is the first necessary condition for stability of the spiral structure, which, as we shall see below, can exist only in the presence of orthorhombic anisotropy in the system. The orthorhombic anisotropy stabilizes the structure, preventing a change in orientation of the magnetization vector parallel to the plane of the film.

The quadratic form $E_2(\delta\varphi)$ possesses a nondiagonal form and is difficult to analyze precisely. However, the condition for the quadratic form to be nonnegative and for a range of parameters of the magnetic films for which the spiral configuration is stable to exist is

$$\mathscr{E}_1(q, n) - d(q, n)^2/\mathscr{E}_1(q, n + 1) - d(q, n - 1)^2/\mathscr{E}_1(q, n - 1) \geq 0. \quad (14)$$

We note first that the expression (14) is strictly zero for $q = 0$ and $n = 0$. This obvious result is due to the invariance of the system with respect to uniform translations.

The inequality (14) determines the limits of the region of parameters of the system where the configuration under study is reliably stable within the framework of the proposed model. Analysis of the relation (14) shows that for fixed k and values of Λ/L below a certain limit this inequality breaks down for $n = 0$, $q = \pm k$. This condition determines the limits of stability of the structure under study.

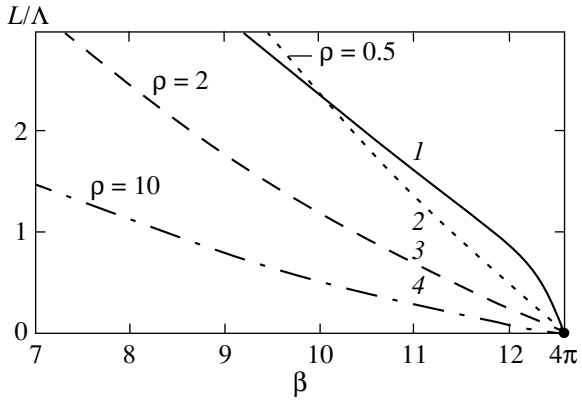


Fig. 2. Region of stable existence of a spiral magnetic structure.

Figure 2 shows in the $(\beta - L/\Lambda)$ plane the range of parameters of magnetic films for which the stability condition of a spiral configuration is satisfied. This region is bounded above by the curve (*1*) and below by one of the curves (*2*, *3*, *4*), each of which corresponds to a definite value of the orthorhombic anisotropy constant ρ and is determined by the solution of the inequality (13).

In summary, the results of these investigations have shown that a stable spiral magnetic configuration can exist in thin magnetic films for a definite ratio of the magnetic parameters. Additional investigations also show that this structure is stable with respect to bending deformations.

REFERENCES

1. A. P. Malozemoff and J. C. Slonczewski, *Magnetic Domain Walls in Bubble Materials* (Academic, New York, 1979; Mir, Moscow, 1982).
2. Yu. I. Dzhezherya, Zh. Éksp. Teor. Fiz. **115**, 1315 (1999) [JETP **88**, 726 (1999)].
3. V. G. Bar'yakhtar and Yu. I. Gorobets, *Cylindrical Magnetic Domains and Their Lattices* (Naukova Dumka, Kiev, 1988).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 1, 2000, pp. 170–181.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 1, 2000, pp. 193–206.

SOLIDS
Structure

Investigation of the Percolation Transition in a Nonwetting Liquid–Nanoporous Medium System

V. D. Borman, A. M. Grekhov*, and V. I. Troyan

Moscow State Engineering Physics Institute, Moscow, 115409 Russia
**e-mail: a_grekhov@mail.ru*

Received December 21, 1999

Abstract—The flows of liquid into and out of a nanoporous medium are studied as processes leading to the fluctuation formation and the growth of fractal clusters of filled and empty pores, respectively. The conditions for stable growth of such fluctuations are analyzed as a function of the interfacial energy between the liquid and the porous medium and the surface energy of the liquid. Expressions are obtained for the pressure at which the barrier for fluctuation filling and emptying of the pores vanishes. In general, it is shown for porous media with a pore-size distribution that these processes can be interpreted as a percolation phase transition. The volume and susceptibility of a liquid–porous medium system near the transition points with inflow and outflow of the liquid are calculated. The phenomenon of nonoutflow of a nonwetting liquid from a porous medium and hysteresis of the flow of liquid into and out of a porous medium are explained on the basis of the mechanism considered. The results of an experimental investigation of these processes in the system liquid Wood's alloy–silochrome 80 and silochrome 120 are presented. The experimental data obtained can be described on the basis of the proposed mechanism. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Despite the long history of investigations of the phenomena associated with the properties of the interface between a liquid and a solid, many questions remain unclear. Examples are capillary phenomena in ultrathin channels, the dynamics of the interphase interaction, and the nonwettability (wettability) of porous media. The transport of a nonwetting liquid in porous media with nanometer pores, which is eliciting great interest from the standpoint of fundamental science and on account of the many applications of nanotechnologies and membrane technology [1–9], is associated with these phenomena.

The most suitable theory for describing the filling of porous media with a nonwetting liquid is percolation theory [10–13]. In percolation theory the filling of a porous medium with a nonwetting liquid can be represented as a process of the formation and growth of fractal clusters and the formation of an infinite cluster from pores which are filled at the Laplace pressure. The external pressure compensates the surface tension forces. According to these ideas, as the pressure decreases below the Laplace pressure, the state of a nonwetting liquid in the pores should be unstable and liquid should leave the pores in the matrix. However, the existing experimental data [14–17] show that effects which cannot be explained on the basis of these ideas without assuming a special geometric structure for the pores are observed in a nonwetting liquid—porous medium system. Thus, the experiments of [14] on the filling of porous glass, containing pores with 3.5 nm radii, with liquid mercury showed that liquid mercury does not emerge from the pores at atmospheric pressure and room temperature. The mass and properties of samples of the porous medium with liquid mercury, which were filled at a pressure of 5×10^3 atm, remained unchanged for months under normal conditions.

The observed difference of the onset pressures for filling and outflow of a nonwetting liquid (hysteresis) is attributed [15–17] to the fact that the contact angles when the liquid flows into and out of the pores are different. These angles are empirical parameters. No physical model explaining why the contact angles are different and making it possible to calculate their values or associate them qualitatively with the parameters of the liquid—porous medium system has been published.

The present work is devoted to an investigation of these phenomena. The results of an experimental investigation of the hysteresis and the phenomenon of nonoutflow of a nonwetting liquid in the system liquid Wood's alloy—silochrome are presented in Section 4. A physical mechanism that makes it possible to explain the experimental data is proposed in Sections 2 and 3.

A nonwetting liquid (liquid metal) filling a porous medium containing nanometer size pores changes from a bulk phase into a state which is characterized by a large specific surface area of the interface between the liquid and the porous medium (dispersed state). Estimates show that the surface energy of the metal in a nanopore in a dielectric matrix can be comparable to the volume energy of a metal. Since pores are filled only when a definite pressure is reached [5–9], the transition of a liquid from a bulk phase into a dispersed state is a threshold phenomenon. Consequently, near

threshold the fluctuations in the system [18, 19] become decisive, and the questions concerning the possibility of the filling of a porous medium and the stability of a nonwetting liquid in a porous medium can be studied naturally on the basis of the theory of fluctuations, analyzing at various pressures the probability of the fill fluctuations increasing (Section 2). Since at a pressure near the threshold p_I the adiabatic formation work δA of the fill fluctuations is close to the temperature, a transition of the system into a state with filled pores under the action of thermal fluctuations is possible. Thus, the condition $\delta A \sim T$ makes it possible to determine only the pressure range $p > p_I$ where the formation of pores accessible to a nonwetting liquid is possible in a porous medium. At the same time it follows from percolation theory that the filling of a porous medium should be determined by the fraction of pores through which a nonwetting liquid can pass. Consequently, there arises the question of whether or not the condition $\delta A \sim T$ is a condition for filling of a porous medium or a condition for the formation of an infinite cluster of filled pores.

The volume $v(p)$ of pores filled at a given pressure and the static susceptibility $\varepsilon(p)$ of the system can be calculated if the distribution function $F(N)$ of clusters over the number N of pores in them is found for a bounded porous medium with a distribution of pores over the pore radius R (Section 3). However, since filling occurs for $|\vartheta - \vartheta_c| \ll 1$ (ϑ is defined below in Section 3) and the characteristic size (correlation length) of a cluster is much greater than the radius, $\xi \gg R$, in the present work the distribution function $F(N)$ for an infinite medium, well-known in percolation theory, is used to calculate $v(p)$ and $\varepsilon(p)$ [13, 20]. The dependence of the filled volume $v(p)$ near a percolation transition, where the correlation length is limited by the size of the porous medium, is calculated in order to describe the experimental data.

The phenomenon of nonoutflow of a nonwetting liquid and hysteresis of filling and emergence of liquid can be explained at the same time on the basis of the mechanism considered. It turned out that these phenomena can be attributed to the different dependence of the work δA with percolation transitions of filling and emergence of a liquid on the ratio of the interfacial energy between a liquid and a porous medium to the surface energy of the liquid.

Hysteresis and nonoutflow of a nonwetting liquid were investigated in the system Wood's alloy (T_{melt} = 340 K)—porous medium (silochromes S80 and S120) with the same framework material (SiO_2) but with different pore radii and relative width of the pore size distribution (Section 4). The well-known method of preliminary heating of a porous medium, in which the composition of the adsorbed layer changes [21], was used to change the surface energy of the interface between the liquid and the porous medium. It was found that the experimentally obtained pressure dependences of the volume of the system can be described by theoretical relations for the neighborhood of percolation transition points. The pressures at these points were determined according to the maximum of the static susceptibility of the system in the processes of filling with and emergence of the liquid. The measured volume of the metal remaining in the pores and the characteristic pressures agree, to within the limits of error, with the computed values.

2. THRESHOLD FILLING AND EMERGENCE PRESSURES FOR A NONWETTING LIQUID IN A POROUS MEDIUM

We shall calculate the threshold pressure at which filling of a porous medium with metal is possible and the pressure at which the metal can emerge from the porous medium. The filling and stability of a nonwetting liquid are studied below for porous media whose framework consists of spheres which have grown together at the points of contact [22]. The pores in this model are the free space between the spheres. Examples of such porous media are opals [8] and silochromes [23] with pore diameter ~10 nm. For S120 and S80 silochromes the fraction of the pore volume (porosity) is $\varphi \approx 0.65$, and there are $\sim 10^{17}$ pores per gram of the porous medium. For simplicity, we shall assume first that the pores are spheres with the same radius R.

We shall assume that as a result of thermal fluctuations at pressure p macroscopically small regions in the form of clusters consisting of N metal-filled pores arise near the filling threshold of a porous medium. Each cluster starts at the boundary between the porous medium and the metal and, on account of the limited volume of the pores, grows by the attachment of other filled pores to it. It can be assumed that at the onset of growth each cluster consists of branched chains of filled pores. We shall neglect the interaction between the clusters. The external pressure performs work when the porous medium is filled with metal. In the process, energetically unfavorable meniscus surfaces of metal in the pores and the interface between the metal and the porous medium are formed. The elastic state of the porous medium also changes. If the adiabatic work of the formation of fill fluctuations is $\delta A(N)$ and energy dissipation due to friction can be neglected (see below), then the probability for fluctuations to appear is $w = w_0 \exp(\delta S)$ [18], where $\delta S = -\delta A/T$ is the entropy fluctuation. Consequently, if δA ($\delta A > 0$) increases as the number N of pores in a cluster increases, the probability of a fluctuation decreases. This corresponds to thermodynamic stability of the initial state of the system. If $\delta A \sim T$ the fill fluctuation can increase. Then the system becomes unstable and liquid starts to fill the porous medium.

We shall calculate the components of the adiabatic work $\delta A(N)$ of formation of a fill fluctuation of N pores. The work performed as the metal expands in a porous medium can be written in the form

$$\delta A_{\text{ex}} = -p\frac{4}{3}\pi R^3 N. \quad (1)$$

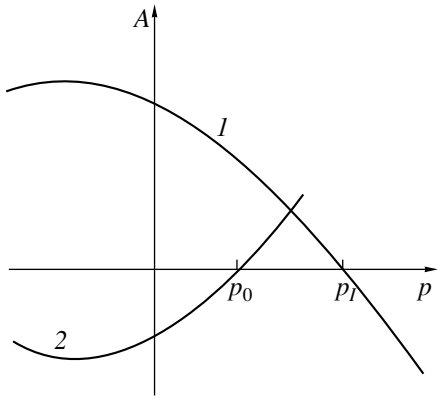


Fig. 1. Pressure dependence of the formation energies A_I and A_0 of inflow and outflow fluctuations, respectively, of a nonwetting liquid in a porous medium (curves *1* and *2*, respectively); p_I and p_0 are the onset pressures for inflow and outflow of liquid.

The work of formation of an interface between the metal and the porous medium is

$$\delta A_{s1} = \delta\sigma 4\pi R^2(1-\eta)N, \tag{2}$$

where $\delta\sigma = \sigma_{ms} - \sigma_{sg}$, σ_{ms}, and σ_{sg} are the surface energy of the metal–porous medium interface and the porous medium–gas interface, and η is the interconnectedness constant of the pores, equal to the ratio of the meniscus surfaces formed to the total surface area of the pores. In each filled pore miniscuses of metal form at the boundary with the pores. Assuming, for simplicity, the miniscuses to be flat, the expression for the work δA_{s2} of formation of miniscuses can be written in the form

$$\delta A_{s2} = \sigma_m 4\pi R^2 \eta N. \tag{3}$$

Here σ_m is the surface energy of the metal. Taking account of the shape of the miniscuses only leads to the appearance of a new coefficient in this expression and does not affect the character of the dependence $\delta A(N)$.

The change in the elastic energy of a porous medium compressed by external pressure in the presence of fill fluctuations can be comparable to the surface energy, since the compressibility of the porous medium is several orders of magnitude greater than the compressibility of the framework material [22]. Before fill fluctuations appear the porous medium is compressed by a pressure p and the pressure inside the pores is zero (the pressure p is measured from the atmospheric pressure). After fill fluctuations appear the elastic energy partially relaxes, since the pressure inside the filled pores becomes equal to the external pressure. Consequently, the component of the work of formation of fill fluctuations that is associated with the relaxation of the elastic energy is a negative quantity ($\delta A_{el} < 0$). Using the linear relation between the elastic stress and strain tensors for porous media and keeping in mind the boundary conditions for the initial and final elastic states, the following expression can be obtained, in agreement with [24], for the change in the elastic energy accompanying the formation of a fill fluctuation of N pores:

$$\delta A_{el} \simeq -(\chi - \chi_s)p^2\frac{4}{3}\pi R^3 N, \tag{4}$$

where χ and χ_s are the compressibility values of the porous medium and the framework material of the porous medium. For porous media $\chi \gg \chi_s$, according to [22], and $\chi \sim (1-\varphi)^{-\alpha}$ and $\alpha = 3.7$. The derivation of the expression (4) assumed that the elastic properties are the same over the entire volume of the porous medium.

Summing Eqs. (1)–(4), the expression for the work $\delta A_I(N)$ of formation of fill fluctuations of N pores can be written in the form

$$\begin{aligned}\delta A_I(N) &= A_I(p)\frac{4}{3}\pi R^3 N,\\ A_I(p) &= -\chi p^2 - p + \frac{3\delta\sigma}{R}\left[1 + \eta\left(\frac{\sigma_m}{\delta\sigma} - 1\right)\right].\end{aligned} \tag{5}$$

It follows from Eq. (5) that $\delta A_I(N)$ is a linear function of the number N of filled pores. We note that the number N in Eq. (5) can be macroscopically large. The sign of $\delta A_I(N)$ depends on the pressure (see Fig. 1). It also follows from Eq. (5) that if w_0 is independent of N, then the probability of formation of a fluctuation consisting of N pores is determined by the probability of each pore being filled raised to the power N. Such a fluctuation grows as a result of successive fluctuation filling of individual pores. The function $\delta A_I(p)$ changes sign at pressure

$$p_I = -\left\{1 - \sqrt{1 + 12\chi\frac{\delta\sigma}{R}\left[1 + \eta\left(\frac{\sigma_m}{\delta\sigma} - 1\right)\right]}\right\}\frac{1}{2\chi}. \tag{6}$$

At pressures $p < p_I$ the quantity $\delta A(N)$ is greater than zero for any value of N and the probability of a fluctuation decreases with increasing N. Consequently, it equals zero for any macroscopically large number of pores. Fluctuations of finite length arise, but filling of the porous medium does not occur. For pressures near p_I the work δA is of the order of T and thermal fluctuations in the system can lead to the formation of clusters consisting of N pores. At pressures $p > p_I$ filling of individual pores becomes energetically favorable, since $\delta A(N)$ is negative. Since $\delta A \sim N$, for $p > p_I$ the probability w is approximately one. The pressure difference $p - p_I$ gives rise to motion of the liquid in the porous medium. Estimates show that for characteristic values of the parameters of the porous medium and metal $\sigma_m \sim \delta\sigma \approx 0.5$ J/m^2, $R \sim 10$ nm, $\chi \sim 10^{-4}$ atm^{-1}, and $\eta \approx 0.7$ with the pressure deviating from p_I by the amount $\delta p = -10^{-2}p_I$, the work changes from $\delta A \sim T$ to ~1 eV, and the probability of a fill fluctuation at the temperature $T = 400$ K changes by ten orders of magnitude. Consequently, for systems

with such values of the characteristic parameters the probability w/w_0 changes abruptly from zero to one in a narrow pressure range ($\delta p/p \sim 10^{-2}$) near the pressure p_I. A temperature dependence of the filled volume should be observed in this pressure range. It follows from Eq. (6) that the threshold pressure increases as the pore radius decreases. Consequently, in a porous medium with a pore-radius distribution the state of the liquid in pores with a large radius is energetically more favorable at this pressure. As pressure increases, filling of smaller pores becomes possible. Hence it follows that filled pores should form fractal clusters, and subsequent flow of liquid and a change in the cluster shape are energetically unfavorable.

If it is assumed that fractal clusters form from filled pores, then the dependence of the work δA_I on N will once again be linear. In accordance with [13] if the size of a fractal cluster is a, then for 3D systems the number of pores in a cluster is $N \sim a^{2.5}$. As this cluster increases in size, miniscuses appear only in pores belonging to the shell of the cluster. For 3D systems, because of the multiconnectedness of the space, the dependence of the number N_0 of pores in the shell on the cluster size is the same as for the number of pores in a cluster: $N_0 \sim a^{2.5}$. Consequently, all components (1)–(3) and (5) of the work of formation of a fill fluctuation are proportional to N ($\sim a^{2.5}$), and the fill onset pressure p_I will differ from the expression (6) only by a different value of the coefficient η.

The filling of pores is accompanied by energy dissipation due to friction forces. A nonwetting liquid above the threshold pressure moves in a porous medium under the action of the pressure difference $\delta p = p - p_I$. Consequently, the ratio of the dissipation energy to the energy of the metal in pores can be estimated as $\delta p/p_I$. The quantity δp determines the pore filling time τ_0, whose upper bound can be estimated, in accordance with [18, 20], as $\eta_1 N/\delta p$. For a liquid metal (mercury) the viscosity is $\eta_1 \sim 10^{-3}$ Pa s and for $N \sim 10^4$ (the size of the granules of the porous medium is $L \sim 10^{-2}$ cm and the radius is $R \sim 10$ nm) and $\delta p = 10^4$ Pa the pore filling time is $\tau_0 \sim 10^{-3}$ s. If the characteristic time of a pressure variation in the experiment is $t \gg \tau_0$, then the filling can be interpreted as a process due to thermal fluctuations and energy dissipation due to friction can be neglected compared with the adiabatic work of filling of the pores with metal ($\delta p/p_I \sim 10^{-3}$). In the experiments described above the time is $t \sim 1$ s and the inequality $t \gg \tau_0$ holds for $\delta p \ll 10$ atm.

We shall now investigate the stability of a nonwetting liquid in a porous medium at pressure p. A cluster consisting of N empty pores can form in a liquid-filled porous medium as a result of fluctuations. The work $\delta A_0(N)$ of formation of such a cluster can also be written as a sum of expressions of the type (1)–(4). However, The signs of the terms in $\delta A_0(N)$ change. Only the formation of an interface between a porous medium and the gas in the empty pores ($\delta A_{s1} < 0$) is energetically favorable. The metal flowing out of the pores performs work, and consequently $\delta A_{ex} > 0$. The work performed when the elastic state of the system ($\delta A_{el} > 0$) changes and the work of formation of the miniscuses of metal in pores located next to the empty pores formed ($\delta A_{s2} > 0$) becomes positive. Ultimately, the expression for the work of formation of a cluster of empty pores can be written in the form

$$\delta A_0(N) = A_0(p)\frac{4}{3}\pi R^3 N,$$
$$A_0(p) = \chi p^2 + p - \frac{3\delta\sigma}{R}\left[1 - \eta\left(\frac{\sigma_m}{\delta\sigma} + 1\right)\right]. \quad (7)$$

In accordance with Eqs. (7) the energy barrier for the formation of a cluster consisting of N empty pores either grows linearly or decreases as N increases (see Fig. 1), depending on the sign of $A_0(p)$. The sign of the function $A_0(p)$ changes at the pressure

$$p_0 = -\left\{1 - \sqrt{1 + 12\chi\frac{\delta\sigma}{R}\left[1 - \eta\left(\frac{\sigma_m}{\delta\sigma} + 1\right)\right]}\right\}\frac{1}{2\chi}. \quad (8)$$

For $p < p_0$ the quantity $\delta A_0(N)$ is less than zero for any N and the growth of clusters of empty pores becomes energetically favorable. The relation $\delta A_0 \sim T$ is satisfied at $p = p_0$, and therefore the pressure $p = p_0$ is the pressure at which the nonwetting liquid starts to flow out of the porous medium.

Comparing (6) and (8) shows that the filling pressure is greater than the pressure at which the liquid leaves the pores, and hysteresis of filling and outflow of the liquid should be observed. The liquid can remain in the porous medium if $p_0 < 0$. In accordance with Eq. (8) this should occur for sufficiently large values of the ratio $\sigma_m/\delta\sigma$, when

$$\eta > \left(1 + \frac{\sigma_m}{\delta\sigma}\right)^{-1}. \quad (9)$$

It follows from the relation (9) that the opposite inequality can be obtained by increasing $\delta\sigma$, for example, through changing the composition of the adsorbed layer on the surface of the porous medium. In this case, the nonwetting liquid should flow out of the medium at pressure p_0 (8) with a definite value of the interconnectedness constant η of the pores.

3. VOLUME AND SUSCEPTIBILITY OF A POROUS MEDIUM–NONWETTING LIQUID SYSTEM

Let us consider the filling of a porous medium containing pores of different size. The total number of pores per unit volume of the porous medium that can be filled in a fluctuation manner increases with pressure and can reach the percolation threshold. This raises the

question of whether or not the condition $\delta A \sim T$ is the condition for filling of the porous medium or the condition for the formation of an infinite cluster of filled pores. In accordance with the estimate made above of the change in the energy barrier for the system metal–porous medium with $R \sim 10$ nm pores in the present analysis, the exponential pressure dependence of the probability of formation of a fill fluctuation (the dependence $w = w_0 \exp[-\delta A(p)/T]$) is replaced by a step function $w = w_0\theta(p - p_I)$. This means that the probability of fluctuation filling of pores with radius R does not depend on temperature and for $t \gg \tau_0$ it is equal to zero at pressures $p < p_I(R)$ and 1 at pressures $p > p_I(R)$. The dependence of the work of fluctuation formation of a fractal cluster of filled pores on the size of the cluster and the pressure is analyzed below, after which the specific volume of the filled pores and the static susceptibility (compressibility) of the liquid–porous medium system for pressures near the percolation threshold are calculated.

We shall assume that the pores are spherical, the size distribution of the pores $f(R)$ is spatially uniform, and the width of the distribution is small enough ($\Delta R \ll R$) so that percolation effects due to the strong variance of the pore radii can be neglected [25]. If the interaction between fluctuations is neglected once again, then the adiabatic work of formation of a cluster of N filled pores at pressure p can be written in the form

$$\delta A(N) = \sum_{J=1}^{N} A_I(p, R_J)\frac{4}{3}\pi R_J^3. \qquad (10)$$

Here $A_I(p, R_J)$ is identical to the expression (5) with $R_J = R$.

Since a cluster can form only by successive fluctuation filling of connected pores, the condition

$$A_I(p, R_J) \sim T \qquad (11)$$

should be satisfied for each filled pore in Eq. (10). The probability of fluctuation filling of pores with radius R_J that do not satisfy the condition (11) is exponentially small at pressures $p < p_I$ and is assumed to be zero. In accordance with the percolation theory, the number N of filled pores in such a cluster in the expression (10) is limited and is determined by the relative fraction (ϑ) of pores for which $\delta A \sim T$, and consequently they can be filled by the nonwetting liquid at pressure p (accessible pores). If the pore-size distribution is normalized as $\int f(R)dR = n$, where n is the number of pores per unit volume ($\sim 10^{17}$ cm^{-3}), then $\vartheta(p)$ at pressure p is

$$\vartheta(p) = \frac{1}{n}\int_{R(p)}^{\infty} f(R)dR. \qquad (12)$$

Here the lower limit of integration $R(p)$ should be determined from the condition (11) at pressure p. Consequently, the probability of fluctuation filling for all accessible pores is close to 1 ($w/w_0 \sim 1$). The quantity $\vartheta(p)$ determines the average probability of finding an accessible pore at a given location in the porous medium. It is analogous to the fraction of lattice sites, which determines percolation and is usually introduced in percolation theory. In our case this is the fraction of lattice sites which are occupied by accessible pores, which depends on the external parameter—the pressure. It follows from Eq. (12) that in accordance with Eq. (6), as pressure increases, $R(p)$ decreases and the fraction of accessible pores $\vartheta(p)$ increases. Using this analogy, we can introduce the critical fraction of accessible pores ϑ_c and the critical pressure p_c ($\vartheta(p_c) = \vartheta_c$) for which an infinite cluster consisting of accessible pores arises. It is convenient to use the distribution function $F(N)$ of clusters over the number of pores in a cluster to describe the possible values of the number N of pores in clusters and the number of clusters containing N pores at a given pressure and with a given value of $\vartheta(p)$. The function $F(N)$ can be normalized to the total number of clusters per unit volume. The analogy noted above between the fraction $\vartheta(p)$ of pores and the fraction of lattice sites determining percolation makes it possible to use in the problem being discussed the following well-known [13, 20] mass distribution function of the clusters (in our case the distribution over the number of pores N in a cluster, since the mass is proportional to the number of pores in it):

$$F(N) = cN^{-\tau}. \qquad (13)$$

For 3D systems $\tau = 2.2$, and c is a normalization factor. The dependence (13) has been obtained in [13] from scaling considerations for clusters containing a large number of sites ($N \gg 1$) (in our case a large number of pores), and values of $\vartheta(p)$ close to the percolation threshold, where $|\vartheta - \vartheta_c| \ll 1$. The number of pores in clusters in Eq. (13) is bounded from above by the value of $N(\xi)$ for a cluster whose geometric size is ξ:

$$N(\xi) \sim |\vartheta - \vartheta_c|^{-1/\sigma}, \qquad \xi = R|\vartheta - \vartheta_c|^{-\nu}. \qquad (14)$$

Here $\sigma = 0.44$ and $\nu = 0.89$ for 3D systems. The quantity ξ in the percolation theory is called the correlation length [13]. It determines the characteristic range of possible values of N in the distribution (13). The number of clusters with $N > N(\xi)$ near the percolation threshold is exponentially small [20]. In accordance with Eqs. (6), (12), and (14) the number of pores $N(\xi)$ in a cluster of size ξ and the correlation length increase with pressure and diverge at pressure where the fraction $\vartheta(p)$ of accessible pores reaches the value ϑ_c. The value of ϑ_c according to [13] depends on the structure of the porous medium (by analogy with the dependence of the critical fraction of lattice sites on the number of nearest neighbors). In 3D systems percolation arises when the relative volume of accessible pores is $\vartheta_c = 0.16$, and the corresponding critical fraction ϑ_c of accessible pores should be determined from Eq. (12). It follows from

Eqs. (13) and (14) that for $\vartheta(p) < \vartheta_c$ the number of pores in clusters is limited.

Filling of a porous medium submerged in a liquid as pressure increases starts in a region with a thickness of the order of ξ near the surface of the medium. Clusters of filled pores consist of pores for which the condition (11) is satisfied. Analogously, pores that are accessible but are located outside a near-surface region of thickness ξ do not become filled with liquid.

The conversion of clusters of accessible pores into clusters of filled pores can be interpreted as an interaction of the clusters. A cluster consisting of n accessible pores of radii $R > R(p)$, which is located at the boundary (shell) of a cluster consisting of m filled pores becomes filled with liquid as pressure increases up to p, and a new cluster consisting of $n + m$ filled pores is formed. For such an interaction of fractal clusters the preexponential factor w_0 in the probability of fluctuation filling will depend on the numbers n and m: $w_0 \sim n^{-q_1} m^{q_2}$. The exponents q_1 and q_2 take account of the fraction of pores on a fractal surface through which the liquid flows. As an estimate, it can be assumed that the values of q_1 and q_2 are close and, according to [20], are equal to 0.93. The dependence $w_0(n, m)$ leads to a new estimate of the characteristic pore filling time with $\tau \sim \tau_0 n^{q_1} m^{-q_2}$. For this dependence of the time τ on the number of pores contained in clusters, the kinetics of the formation of the function $F(N)$ for clusters of filled pores changes. However, if the characteristic pressure variation time t in an experiment is much greater than τ, the stationary distribution functions of clusters of filled and accessible pores are the same and are described by the expression (13). In the experiments described below, this inequality holds for a relative pressure excess $\delta p/p_I \ll 10^{-3}$ above the threshold value and the time t is about 1 s.

Knowing cluster distribution function $F(N)$, the relative volume $\nu(p)$of all filled pores per unit volume of a porous medium in a layer thickness ξ at pressure p can be calculated as

$$\nu(p) = \int_1^{N(\xi)} NF(N)\frac{4}{3}\pi\bar{R}^3 dR. \qquad (15)$$

Here $\bar{R}$ is the average radius of the pores in clusters at pressure p. Assuming that the distribution function $F(N)$ of clusters of filled pores can be described by the expressions (13) and (14), it can be shown that for ϑ near the critical value ϑ_c the integral (15) has no singularities. As pressure and $\vartheta(p)$ increase, the thickness ξ of the filling region increases. For $\vartheta \sim \vartheta_c$ a cluster of filled pores with geometric size of the order of the size of a porous medium (infinite cluster) forms, and a further increase of the filled volume with increasing pressure will be due to attachment of clusters and filled pores to other clusters of filled pores and to an infinite cluster.

Differentiating the integrals (12) and (15) with respect to the limits of integration and using the functions (13), (14), and (6), the following expression can be obtained for the susceptibility ε of a liquid–porous medium system:

$$\varepsilon = \frac{\partial}{\partial p}\nu(p) \sim |\vartheta - \vartheta_c|^{(\tau-\sigma-2)/\sigma} f(R)\frac{dR}{dp_I}. \qquad (16)$$

Using the well-known [13] dependences of the exponents τ and σ on the basic critical exponents ν, β ($\tau = (2\nu d - \beta)(\nu d - \beta))^{-1}$, $\sigma = (\nu d - \beta)^{-1}$, d is the dimension of the space), determining the dependence of the correlation length and the formation probability of an infinite cluster near the percolation threshold on $|\vartheta - \vartheta_c|$, we obtain $(\tau - \sigma - 2)\sigma^{-1} = \beta - 1$ and $\varepsilon \sim |\vartheta - \vartheta_c|^{\beta-1}$. Hence it follows that the critical exponent of the susceptibility is determined only by the critical exponent of the formation probability of an infinite cluster. For 3D systems the critical exponent of an infinite cluster is $\beta = 0.42$ [13]. Consequently, as $\vartheta \longrightarrow \vartheta_c$ the expression (16) diverges. This divergence is related with the formation of an finite cluster of filled pores in a porous medium at the pressure p_c where $\vartheta(p_c) \simeq \vartheta_c$. It follows from Eq. (16) that the maximum value of the pressure derivative of the volume of the system should be observed at a pressure where in accordance with Eq. (12) ϑ is close to ϑ_c. According to Eq. (16), the value of this derivative will be greater for porous media with a narrower size distribution of pores. It also follows from Eq. (16) that the condition for filling of a porous medium is $\delta A_I \sim T$ for pores with radius such that the fraction of accessible pores (12) is ϑ_c.

The results of percolation theory, which were used here and were obtained on the basis of scaling, do not depend on the type of lattice: the lattices can be arbitrary, including irregular, and there need not be a lattice at all [13]. In the calculations performed, a cluster distribution function $F(N)$ for a spatially unbounded uniform porous medium with pores of the same size was used. At the same time, it is known [2] that the critical exponents of the mass of a fractal cluster decrease at the boundary of the medium. It can be assumed that for the case at hand a uniform porous medium filling a half-space and containing a narrow distribution of pores, the macroscopic characteristics of the system determined by an integral of the function $F(N)$ can be described qualitatively correctly for the experimentally accessible region $\xi \gg R$ in measurements of $\nu(p)$, where ξ is comparable to the size of the porous medium. In this case $|\vartheta - \vartheta_c| \ll 1$ and distortions of the function $F(N)$ in the near-surface region of the porous medium can be neglected. We note that for small values of $\Delta R/R$ the dependence of the percolation threshold on the concentration of large spheres (pores) is weak [25]. On the other hand the fraction of the filled volume of pores in the near-surface layer of thickness ξ can be estimated from above as the fraction of all accessible pores ϑ, since by virtue of the multiconnectedness of 3D systems, for small $\Delta R/R$ all accessible pores in a layer ξ will be filled with liquid (see Section 4). This estimate leads to a different value of

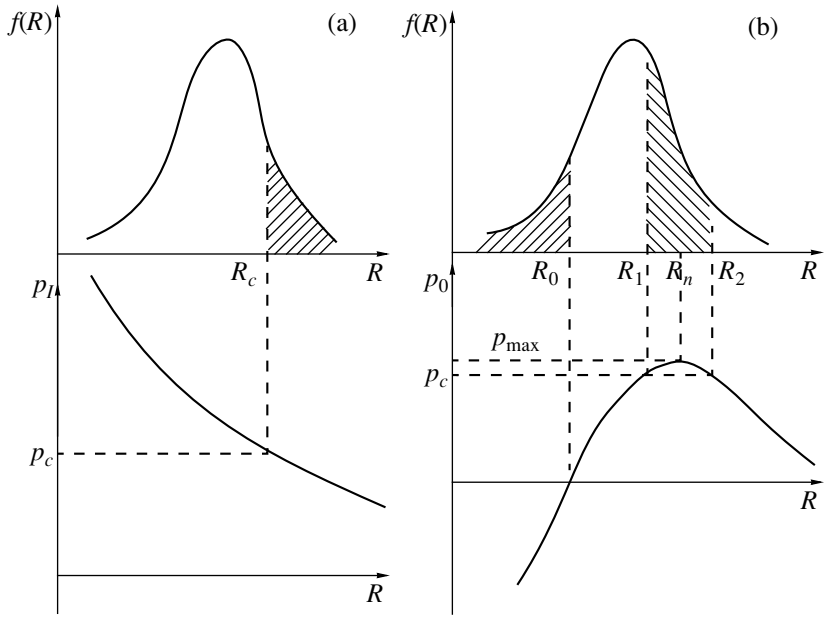


Fig. 2. (a) Pore filling pressure versus the pore radius. The pressure p_c corresponds to the appearance of an "infinite" cluster of filled pores, when $\vartheta(p_c) = \vartheta_c$. The hatched region corresponds to the critical fraction ϑ_c of accessible pores. (b) Pressure of liquid outflow from pores versus the pore radius. The pressure $p_{\max}$ corresponds to the onset of liquid outflow from a porous medium. The hatched region of radii $[R_1, R_2]$ corresponds to the condition $\vartheta(p_c)$; pores with $R < R_n$ remain filled.

the critical exponent for the susceptibility of the system without changing the qualitative results.

We shall now discuss in greater detail the pore-size dependence of the pressure at which the energy barrier for filling vanishes. As a simplification we shall assume that the compressibility of the porous medium is small and $\chi\delta\sigma\,\bar{R}^{-1} \ll 1$. Then the first term in the expression (5) can be neglected. For a porous medium with a pore-size distribution the coefficient η depends on the radius of the pores, since for any two neighboring connected pores with different radii the fraction of the meniscus surfaces formed during filling with respect to the surface of the pores ($4\pi R^2$) is smaller for larger pores. Hence it follows that the coefficient η can be written in the form

$$\eta \simeq q(\varphi)\left(\frac{R_{\min}}{R}\right)^2 \qquad (17)$$

Here $R_{\min}$ is the minimum observable pore size in a porous medium, and the coefficient $q(\varphi)$ (<1) depends on the fraction φ of the volume of the pores in the porous medium and can be estimated as $\varphi^{2/3}$. Then $\eta < 1$. As the exposition below will make clear, the results obtained do not change qualitatively, if $\eta \sim R^{-\alpha}$, $0 < \alpha < 2$. Using Eqs. (6) and (17) the condition for vanishing of the energy barrier for filling of pores with radius R can be written as

$$p_I = \frac{3\delta\sigma}{R}\left[1 + q(\varphi)\left(\frac{R_{\min}}{R}\right)^2\left(\frac{\sigma_m}{\delta\sigma} - 1\right)\right]. \qquad (18)$$

It follows from Eq. (18) (see Fig. 2a) that the pressure p_I increases monotonically as the pore radius decreases. When the susceptibility ε of the system reaches its maximum value, in accordance with the expression (18) the value of the radius R of the pores in Eq. (16) should equal $R(p_c) = R_c$ in Eq. (12) for which the fraction of admissible pores is $\vartheta = \vartheta_c$. Then all pores with radii $R > R_c$ become accessible for filling.

We now return to the question of the stability of a nonwetting liquid in a porous medium. The main effect of the pore-size distribution is the possibility that part of the nonwetting liquid will remain in the porous medium. The work of formation of a cluster of N empty pores can be written in the form of the expression (10) where the function $A_I(p, R_j)$ must be replaced by $A_0(p, R)$ (7). Repeating the arguments leading to the expression (18), we can obtain the condition under which the energy barrier for liquid to flow out of pores with radius R vanishes ($A_0(p, R) = 0$):

$$p_0 = \frac{3\delta\sigma}{R}\left[1 - q(\varphi)\left(\frac{R_{\min}}{R}\right)^2\left(\frac{\sigma_m}{\delta\sigma} + 1\right)\right]. \qquad (19)$$

It follows from Eq. (19) (see Fig. 2b) that the pressure p is maximum ($p_{\max} = 2\delta\sigma/R_n$) for pore radius

$$R_n = R_{\min}\sqrt{3q(\varphi)}\sqrt{\frac{\sigma_m}{\delta\sigma} + 1}\,. \qquad (20)$$

For pressures $p > p_{\max}$ the energy barrier $A_0(p, R)$ is greater than 0 for all pores and a nonwetting liquid should remain in the porous medium. For pressures $p < p_{\max}$ the radii of pores freed of liquid fall in the range $[R_1, R_2]$ (see Fig. 2b).

The pressure p becomes negative for pores with radius less than R_0, which is equal to

$$R_0 = R_{\min}\sqrt{q(\varphi)}\sqrt{\frac{\sigma_m}{\delta\sigma} + 1}. \quad (21)$$

This means that the energy barrier remains positive for all pores with $R < R_0$ at any pressure. Consequently, growing clusters of empty pores cannot contain pores with $R < R_0$ at any pressure. Calculations similar to those performed in the derivation of (16) yield the following expression for the susceptibility of a system with a liquid flowing out of a prefilled porous medium:

$$\varepsilon(p) \sim |\vartheta(p) - \vartheta_c|^{-(\tau-\sigma-2)/\sigma} \times\left[f(R_1(p))\frac{\partial R_1}{\partial p_0} - f(R_2(p))\frac{\partial R_2}{\partial p_0}\right]. \quad (22)$$

Here $f(R)$ is the pore distribution function for radii R_1 and R_2 for which $p(R_1) = p(R_2)$. The fraction $\vartheta(p)$ of pores which can become freed of liquid, i.e., for which $A_0(p, R) \sim T$, should be

$$\vartheta = \frac{1}{n}\int_{R_1}^{R_2} f(R)dR. \quad (23)$$

According to Eq. (23), as the pressure decreases, the range $[R_1, R_2]$ and the fraction of pores ϑ for which $A_0(p, R) \sim T$ increase. Since $(\tau - \sigma - 2)/\sigma \sim 0.5$, the derivative of the expression (22) increases as $\vartheta \longrightarrow \vartheta_c$. We note that just as Eq. (16) the expression (22) becomes inapplicable for $\vartheta > \vartheta_c$, since an infinite cluster of empty pores is formed and the distribution function $F(N)$ of clusters over the number of pores in a cluster changes. According to the expression (22), the susceptibility ε is maximum when the fraction of pores with radii such that $p(R) < p(R_n)$ is ϑ_c. It follows from Eq. (22) that the susceptibility will be maximum if the value of R_n is close to the average value of the pore radius for which the distribution function $f(R)$ is maximum. In contrast to filling of a porous medium, when a liquid flows out of the medium, clusters of empty pores are formed not in a layer of thickness of the order of the correlation length at the surface of the medium but rather over the entire volume of the medium. Comparing the expressions (18) and (19) shows that $p_0 < p_I$ and hysteresis should be observed for filling of and outflow of a nonwetting liquid from a porous medium.

In accordance with Eq. (19), pores with small radii, $R < R_0$, can remain filled as pressure decreases to zero. If R_0 lies outside the range $[R_{\min}, R_{\max}]$ of the pore-size distribution and $R_0 < R_{\min}$, then all of the liquid should flow out of the porous medium. For $R_0 > R_{\max}$ the quantities $f(R_1)$ and $f(R_2)$ in Eq. (22) are zero and the nonwetting liquid should remain entirely in the porous medium. The phenomenon of nonwetting mercury not flowing out of porous glass was observed in [14]. In the intermediate case part of a nonwetting liquid should remain in the medium. We note that in accordance with Eq. (21) the ratio of R_0 and $R_{\max}$, $R_{\min}$ can be changed by varying σ_m and $\delta\sigma$.

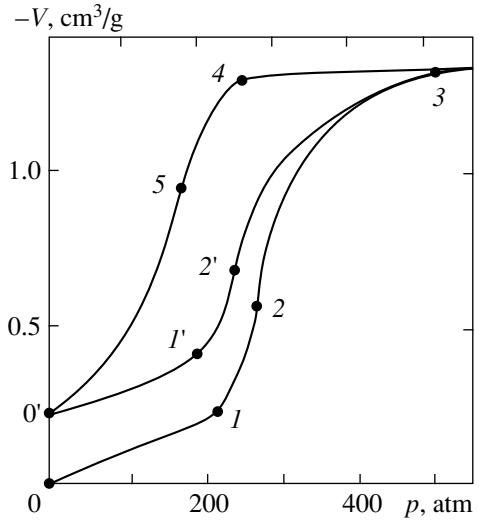


Fig. 3. Experimental dependence (for 1 g of silochrome) of the change in volume of the system silochrome 80–Wood's alloy versus pressure: 0–*3*, decrease of volume with increasing pressure; *3*–0', increase of volume with decreasing pressure after filling of a porous medium; 0'–*3*, decrease of volume with repeated increase of pressure.

4. EXPERIMENT. DISCUSSION OF RESULTS

Two types of porous media were used in the experiments:[1] silochrome 120 (S120) and silochrome 80 (S80), whose framework consists of silicon dioxide SiO_2. According to published data, the minimum $R_{\min}$ and maximum $R_{\max}$ pore radii for S120 and S80 are, respectively, 10 nm, 15 nm and 20 nm, 25 nm [19]. The specific volume and the specific surface area $V_p \approx 1.3$ cm^3/g, $S_p = 120$ m^2/g for S120 and $V_p = 1.3$ cm^2/g, $S_p = 80$ m^2/g for S80. Porous media consist of powders with average granule size $L \sim 10^{-4}$ m. Liquid Wood's alloy, whose melting temperature is $T = 340$ K, was used as the nonwetting liquid. The experiments were performed in the temperature range $T > 400$ K. To change the composition of the layer adsorbed on the surface of the silochromes, the silochromes were heated in a special chamber up to temperature $T < 500$ K in 1.5 h.

Silochrome powder with mass 2.00 ± 0.05 g was placed in a high-pressure chamber. The chamber was filled with liquid Wood's alloy. The construction of the chamber made it possible to perform experiments up to 10^3 atm pressure. The excess pressure (above atmospheric pressure) and the volume of the liquid–porous medium system were measured to within 3% using a strain-gauge pressure sensor and a displacement sensor.

Figure 3 shows the experimental pressure dependences of the change in the volume V of the system

[1] A. A. Geidarov participated at the initial stage of the experimental investigations performed in this work.

Table

Type of silochrome	p_1, atm		p_2, atm		p_4, atm		p_5, atm		V_0	
	exp.	theor.	exp.	theor.	exp.	theor.	exp.	theor.	exp.	theor.
S120	280	285	420	430	380	390	310	315	0.14	0.14
S80	200	220	240	265	240	240	180	170	0.22	0.23
Annealed S120	250	260	400	410	320	320	280	280	0.16	0.16
Annealed S80	150	150	210	270	235	220	150	140	0.55	0.56

Note: The annealing was performed at $T = 473$ K for $t = 1.5$ h.

Wood's's alloy–silochrome 80 (for 1 g of silochrome). These dependences were corrected taking into account the premeasured compressibility values of the chamber and the liquid Wood's alloy. Consequently, the change in volume on the linear section from zero up to the point *1* is due to elastic deformation of an unfilled porous medium with increasing pressure. The measured value of the compressibility $\chi = V_s^{-1}\partial V/\partial p$ (where V_s is the volume of a sample of the porous medium) on this section is $\chi \approx 4 \times 10^{-4}$ atm^{-1}. For $p > p_1$ a substantial decrease in the volume of the system is observed at the point *1* with increasing pressure. The section *1–3* corresponds to filling of the pores with liquid. The change in the volume of the system due to elastic compression of the porous medium accompanying an increase in pressure from p_1 to p_3 is ≤3%. On the section *1–2* the compressibility of the system increases monotonically up to a maximum value at the point *2*. This point is the point of inflection in the function $V(p)$. The volume of the system remains unchanged in the pressure range $p > p_3$. This means that the liquid has filled all pores in the sample and the compressibility of such a porous medium does not exceed the measurement accuracy. The specific volume of the pores, determined according to the section *1–3* taking into account the compressibility of the porous medium, is $V_p \approx 1.3 \pm 0.07$ cm^3/g and agrees with the tabulated value for silochrome 80 [23].

As pressure decreases, the volume of the system increases at the point *4*. This increase corresponds to the onset of flow of liquid out of the porous medium. On the section *4–5* the compressibility increases, reaching its highest value at the point *5* for flow of liquid out of the pores. At the point 0' the pressure becomes zero. As pressure is increased once again, the volume of the system starts to decrease at the point 0', so that the volume determined by the segment 0–0' corresponds to the volume V_0 of the liquid remaining in the pores with zero excess pressure. Refilling of the pores with liquid starts at the pressure $p_1' = 210$ atm close to the pressure $p_1 = 200$ atm. The function $V(p)$ does not change for repeated outflow of the liquid. After a repeated "fill–outflow" cycle the volume of the liquid remaining in the porous medium likewise remains unchanged. Similar dependences have been obtained for the system silochrome 120–Wood's alloy. The pressures for liquid to enter (p_1) and leave (p_4) the medium and the pressures for which the maximum compressibility is reached with liquid flowing into and out of the pores (p_2 and p_5)were determined from the dependences $V(p)$ obtained. These data are presented in the table. The table also gives the values of the relative volume of the metal remaining in the pores.

It follows from the data presented in the table that the volume of the metal remaining in the pores with the same state of the surface is greater for a porous medium with a larger average pore radius (S80). All characteristic pressures are higher for a porous medium with a smaller average pore radius (S120). After the porous medium is heated, the volume of the metal remaining in the pores increases and the characteristic pressures decrease.

We shall now compare the experimental data with the theoretical results. We note that the measured values of the compressibility of an unfilled porous medium ($\chi = 4 \times 10^{-4}$ atm^{-1}) and filling onset pressures ($p_1 = 2 \times 10^2$ atm) are such that $\chi p_1 \sim 10^{-2} \ll 1$. Consequently, the elastic compressibility of a porous medium in the expressions (6) and (8) can be neglected.

According to the preceding section, the expression for the specific volume $V(p)$ of a metal flowing under pressure p into the pores in the powder granules of a porous medium can be written in the form of a product:

$$V(p) = \frac{1}{\rho}\nu(p)\Phi\left(\frac{\xi}{L}\right), \tag{24}$$

where ρ is the density of the porous medium, $\rho = \rho_0(1 - \varphi)$, ρ_0 is the density of the framework (SiO_2) of the porous medium, φ is the porosity, and $\nu(p)$ is the relative volume (determined by the expression (15)) of a porous medium in the near-surface layer of the granules, the thickness of the layer being equal to the correlation length ξ. The function $\Phi(\xi/L)$ describes the ξ dependence of the ratio of the volume of the layer ξ to the volume of a granule. Assuming that the granules are spheres with radii L, the function $\Phi(\xi/L)$ can be written in the form

$$\Phi\left(\frac{\xi}{L}\right) = 3\left[\frac{\xi}{L} - \left(\frac{\xi}{L}\right)^2 + \frac{1}{3}\left(\frac{\xi}{L}\right)^3\right], \tag{25}$$

$$\xi = \frac{\bar{R}}{|\vartheta - \vartheta_c|^{\nu}}.$$

Here $\bar{R}$ is the average pore radius and $\vartheta(p)$ is the fraction of pores (12) accessible at pressure p. In accordance with the expressions (12), (18), (24), and (25), $\vartheta(p)$ increases with pressure and at $|\vartheta(p) - \vartheta_c| \ll 1$, i.e., near the threshold of a percolation transition the correlation length ξ is much greater than $\bar{R}$ and increases on a macroscopic scale L. For $\xi = L$ a cluster whose geometric size is equal to the size of a granule is formed. Further filling of accessible pores with increasing pressure should occur by attachment of other clusters and individual filled pores over the entire volume of the granules to this cluster. Consequently, for $\xi > L$ we have $\Phi(\xi/L) = 1$. For $\xi = L$, in accordance with Eqs. (16), (24), and (25), the susceptibility of the liquid–porous medium system reaches its maximum value. Consequently, the experimentally established pressure p_2 can be calculated from Eqs. (12) and (18) and the equation

$$\xi = \frac{\bar{R}}{|\vartheta(p) - \vartheta_c|^{\nu}} = L. \qquad (26)$$

It follows from Eq. (26) that for this pressure $|\vartheta - \vartheta_c| = (\bar{R}/L)^{1/\nu}$, and for $\nu = 0.89$ the quantity $\vartheta(p_2)$ is equal to ϑ_c $(1–10^{-4.5})$ for the sample S80 ($R = 20$ nm, $L = 2 \times 10^{-2}$ cm).

The quantity $V(p)$ in Eq. (24) is limited from above by the value of the relative volume of all accessible (see Eq. (12)) pores in the layer ξ. Their equality means that all accessible pores become filled with liquid. At the same time, in the layer ξ with $|\vartheta(p) - \vartheta_c| \ll 1$ most of the pores (but not all) belong to the cluster with the maximum size $N(\xi)$. Some of the pores which are not connected with the surface of the granules via accessible pores cannot be filled. However, in accordance with Eqs. (13) and (15) most of the filled pores should belong to large clusters whose is size close to the thickness of the layer ξ and which are connected with the surface. Consequently, for a neighborhood of the percolation transition point and a layer of thickness ξ with $|\vartheta(p) - \vartheta_c| \ll 1$ we can write

$$V(p) \simeq \int_{R(p_c)}^{\infty} f(R)\frac{4}{3}\pi R^3 dR = V_c(p). \qquad (27)$$

Here the lower limit of integration is determined from Eq. (18) for $R(p)$ at pressure satisfying the equation $\vartheta(p_c) = \vartheta_c$.

For $\vartheta < \vartheta_c$, far from the percolation transition point, the geometric size of clusters of filled pores changes with increasing pressure in the "microscale" of pore radii and pores in the near-surface layer of thickness $\sim R$ in the granules become filled, and the filled volume is $V \sim \vartheta(R/L)(1/\rho)$. For the experimental S80 and S120 samples we have $R/L \sim 10^{-4}$ and therefore the filled volume is small: $V \sim 10^{-4}\vartheta/\rho$.

For threshold sensitivity of volume measurements in the experiment equal to 7×10^{-2} cm^3/g, filling of a porous medium is observed only near the percolation transition point. Indeed, it follows from Eqs. (25), (26), and (27) for small ξ/L that

$$V(p) \simeq \frac{1}{\rho} V_c(p)\frac{\xi}{L}$$

and for $V = 10^{-2}$ we have $\xi/L = 10^{-2} V_c\rho$ and $\vartheta - \vartheta_c = 10^{-2.4}$. This last relation makes it possible to calculate the measured value of the pressure p_1 for onset of filling. We note that as pressure changes from p_1 to p_2 the quantity $|\vartheta_c - \vartheta|$ and the correlation length change by two orders of magnitude.

The expressions (12) and (27) contain the distribution function of the pore radii $f(R)$. Several methods of determining the function $f(R)$ are described in the literature [3]. One method is based on measuring the function $V(p)$ with filling of a porous medium with a nonwetting liquid (for example, mercury). The function $f(R)$ is obtained by differentiating the function $V(p)$ assuming that the change in volume of the system starts in accordance with the Laplace pressure when pores of maximum size are filled in the entire volume of the porous medium and not under the condition $\vartheta = \vartheta_c$. The functions $f(R)$ obtained by different methods differ from one another [3]. Consequently, in what follows, the Gaussian distribution with the average pore radius $\bar{R}$ is used to analyze the experimental data; this radius lies in the range between the tabulated values of $R_{\min}$ and $R_{\max}$. The quantity $\bar{R}$ and the variance δ of the radii in the Gaussian distribution were treated as free parameters, whose values were determined by comparing with the experimental data on the pressures p_1, p_2, p_4, and p_5 and the volume of the liquid remaining in the porous medium. The expression (18) for the pressure p at which the energy threshold for filling vanishes contains unknown parameters: $\delta\sigma$—the change in the surface energy of a porous medium when the medium is filled with liquid—and the coefficient q which determines the relative fraction of the surface area of the miniscuses of the liquid. In the calculations it was assumed that $\delta\sigma$ is a free parameter and is the same for silochromes S80 and S120 with the same surface preparation, since their framework material is the same. The coefficient q was assumed to equal $\varphi^{2/3}$ (φ is the fraction of the pore volume, i.e., the porosity).

The volumes of the metal entering and remaining in the pores also depend on these parameters. When the condition (19) for the development of a pore-emptying fluctuation to be energetically favorable is satisfied, liquid can flow out of any pore into the volume of a prefilled porous medium. Consequently, the integral (15) of the distribution function of clusters for the volume of

empty pores should be equal to the integral over pore sizes (see Fig. 2b):

$$V(p) = \frac{1}{\rho}\int_{R_1(p)}^{R_2(p)} f(R)\frac{4}{3}\pi R^3 dR. \quad (28)$$

Here the radii $R_1(p)$ and $R_2(p)$ are solutions of Eq. (19) for $p < p_{\max}$. The maximum of the derivative of the function $V(p)$ in accordance with Eqs. (22) and (26) should be observed for $\xi = L$.

According to Eqs. (19) and (21), liquid should remain in pores with radii $R < R_0$ under excess pressure $p = 0$. The specific volume of such pores is

$$V_0 = \frac{1}{\rho}\int_0^{R_n} f(R)\times\frac{4}{3}\pi R^3 dR. \quad (29)$$

The dependences (24) and (28) were calculated for the system silochrome 80–Wood's alloy for the chosen values of the average radius $R = 22$ nm, variance $\delta/R = 0.1$, and $q = 0.7$ ($\varphi = 0.6$ for S80), and $\delta\sigma = 0.15$. The characteristic points $p_{1,\,2,\,4,\,5}$ of the given dependence are presented in the table. Within the limits of experimental error, they should agree with the experimental results. These same parameters were used to calculate the volume V_0 of the metal remaining in the silochrome S80.

As one can see from the table, the computed values agree with the experimentally obtained values within a 10% error range. Similar dependences for the system S120–Wood's alloy describe the experimental data within the limits of error for the parameters $R = 11$ nm, $\delta/R = 0.2$, $q = 0.7$ ($\varphi = 0.65$ for S120), and $\delta\sigma = 0.15$. The values of the characteristic pressures and the residual volume V_0 (see table) also agree.

The observed decrease of the characteristic pressures with preheating of silochromes can be explained using the expressions (12), (18), (19), (24), (28), and (29) by a decrease in the surface energy $\delta\sigma$. Then, according to Eq. (21), the volume of the remaining metal should increase as $\delta\sigma$ decreases. This is in fact observed in experiments with the silochromes S80 and S120. The computed values of the pressures and the volumes of the remaining metal V_0 agree to within the error limits with the experimentally established values for both experimental systems after preliminary heating with the new value $\delta\sigma = 0.14$ (see table). We note that in additional experiments with other values of the heating temperature of silochromes and with sorption of various molecules, substantial changes in V_0 were observed from $V_0 = 0$ to $V_0 = 1/\rho$ and the corresponding dependences (18) and (19) of the changes in the characteristic pressures.

The phenomenon of nonoutflow of the nonwetting liquid is attributed to the structure of the pores in a porous medium [26]. It is assumed that large pores surrounded by smaller pores exist in a porous medium. Then they can be filled at the Laplace pressure for small pores. As pressure subsequently decreases, first small pores are freed of liquid, while liquid should remain in the large pores. It is obvious that assuming such a pore structure makes it impossible to explain the observed strong change (from 0 to $1/\rho$) in the volume of the metal remaining in the porous medium accompanying a change in the surface energy.

The dependences presented in Fig. 3 and the data in the table were obtained at temperature $T = 450$ K. Additional experiments at temperature $T = 400$ K have shown that the dependences $V(p)$ are the same as those obtained at $T = 450$ K. This corresponds to previous estimates of the change in the energy barrier and probability of a fill (outflow) liquid for the metal–porous medium system studied with pore size $R \sim 10$ nm. For systems with $\delta\sigma \sim 0.01$ J/m^2 and $R \sim 10$ nm, with a pressure change $\delta p/p \sim \delta/R$ (δ is the variance of the pore radii) the quantity δA is approximately 0.01 eV and is comparable to the temperature. In this case the volume of the system should depend on the temperature in the range $T = 100$–1000 K.

As follows from Eqs. (7), (17), and (19), the work of formation of a cluster of empty pores is positive for pores with radius $R < R_0$, since for pores with small radius an increase of the fluctuation energy accompanying the formation of miniscuses of liquid is not compensated by a decrease in this energy when the interface between the liquid and the porous medium vanishes. This is why a nonwetting liquid does not flow out. The pressure at which the conditions for percolation transitions for filling and outflow of liquid are satisfied also depends on the ratio between σ_m and $\delta\sigma$. However, these dependences are different, since the work of formation of a fill fluctuation is determined by the sum and not the difference of the interfacial energy and the energy of the miniscuses of liquid. Consequently, analysis of the work of formation of fluctuations without any assumptions about the special geometric structure of pores is sufficient to explain the hysteresis.

We note that although the above analysis of fluctuations agrees with the conventional approach to the description of the kinetics of a first-order phase transition, the nature of the hysteresis is fundamentally different. Hysteresis in a first-order phase transition is attributed to the nonlinear dependence of δA on the number of particles in a nucleus and the metastable states which arise as a consequence of this. In the system under study, because the pore space is limited, fill fluctuations grow by attachment of other pores to a cluster (fractal). This results in a linear dependence of the work δA on N. The above analysis of the stability of the fluctuations makes it possible to establish only the critical pressure at which $\delta A \sim T$, and the hysteresis of filling–outflow of the liquid is due to the different work δA of formation of fluctuations in different initial states of the system.

In summary, the experiments performed with porous media with different pore sizes and surface energy as a result of heat treatment have shown that the proposed model of the inflow and outflow of a nonwetting liquid satisfactorily describes all experimental data obtained. Specifically, it is possible to describe in a consistent manner the hysteresis of inflow and outflow of liquid and the nonoutflow a nonwetting liquid. In accordance with this model the instability of inflow and outflow fluctuations of a nonwetting liquid in a porous medium with a pore-size distribution is manifested under conditions of a percolation transition, where the size of clusters of filled pores is spatially limited by the correlation length ξ, which approaches infinity as the fraction of pores which are either accessible to the liquid or can be freed of the liquid increases up to the critical value ϑ_c.

ACKNOWLEDGMENTS

In conclusion, it is our duty to thank A.A. Fetisov for formulating the problem. A.P. Alekhin who participated in the experiments with modified surfaces, E.V. Gribanov, V.V. Konyukov, V.N. Tronin, and L.A. Maksimov for a helpful discussion of the results, and D.N. Voskresenskii for helpful remarks. We thank J.J.M. Beenakker with whom the formulation of the problem of describing the investigated phenomena on the basis of the theory of fluctuations was discussed.

This work was partially supported by INTAS (project no. 99-01744).

REFERENCES

1. V. N. Bogomolov, Phys. Rev. B **51**, 17040 (1995).
2. R. Cafiero, G. Caldarelh, and A. Gabrielli, Phys. Rev. E **56**, 1291 (1997).
3. M. Mulder, *Basic Principles of Membrane Technology* (Kluwer, Dordrecht, 1996; Mir, Moscow, 1999).
4. T. Naheiri, K. A. Ludwig, M. Anand, *et al.*, Sep. Sci. Technol. **32**, 1589 (1997).
5. M. B. Rao and S. Sircar, J. Membr. Sci. **85**, 253 (1993).
6. Yu. A. Alekseev, V. N. Bogomolov, and T. B. Zhukova, Izv. Akad. Nauk SSSR, Ser. Fiz. **50**, 418 (1986).
7. V. N. Bogomolov, Usp. Fiz. Nauk **124**, 171 (1978) [Sov. Phys. Usp. **21**, 77 (1978)].
8. V. G. Balakirev, V. N. Bogomolov, V. V. Zhuravlev, *et al.*, Kristallografiya **38** (3), 111 (1993) [Crystallogr. Rep. **38**, 348 (1993)].
9. V. N. Bogomolov, Poverkhnost **9**, 136 (1992).
10. J. Feder, *Fractals* (Plenum, New York, 1988; Mir, Moscow, 1990).
11. B. M. Smirnov, Usp. Fiz. Nauk **149**, 177 (1986) [Sov. Phys. Usp. **29**, 481 (1986)].
12. I. M. Sokolov, Usp. Fiz. Nauk **150**, 221 (1986) [Sov. Phys. Usp. **29**, 506 (1986)].
13. M. B. Isichenko, Rev. Mod. Phys. **64**, 961 (1992).
14. Yu. A. Kumzerov, A. A. Nabereznov, and S. V. Vakhrushev, Phys. Rev. B **52**, 4772 (1995).
15. P. G. de Gennes, Usp. Fiz. Nauk **151**, 620 (1987).
16. A. Yu. Fadeev and V. A. Eroshenko, Ross. Khim. Zh. **39** (6), 93 (1995).
17. V. A. Eroshenko and A. Yu. Fadeev, Zh. Fiz. Khim. **70**, 1482 (1996).
18. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 6: *Fluid Mechanics* (Nauka, Moscow, 1964; Pergamon, New York, 1987).
19. H. Haken, *Synergetics: an Introduction* (Springer-Verlag, Berlin, 1977; Mir, Moscow, 1979).
20. A. A. Abrikosov, Pis'ma Zh. Éksp. Teor. Fiz. **29**, 72 (1979) [JETP Lett. **29**, 65 (1979)].
21. *Physical and Chemical Aspects of Adsorbents and Catalysts*, Ed. by B. G. Linsen (Academic, London, 1970; Mir, Moscow, 1978).
22. B. M. Smirnov, Usp. Fiz. Nauk **152**, 133 (1987) [Sov. Phys. Usp. **30**, 420 (1987)].
23. V. A. Dzis'ko, A. P. Karnaukhov, and D. V. Tarasov, *Phisicochemical Fundamentals of Oxide Catalyst Synthesis* (Nauka, Novosibirsk 1978).
24. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 7: *Theory of Elasticity* (Nauka, Moscow, 1965; Pergamon, New York, 1986).
25. M. I. Ozhovan and K. N. Semenov, Zh. Éksp. Teor. Fiz. **101**, 1286 (1992) [Sov. Phys. JETP **75**, 696 (1992)].
26. Churl-Young Park and Son-Ki Ihm, AIChE J. **36**, 1641 (1990).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 1, 2000, pp. 182–187.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 1, 2000, pp. 207–212.

Excitation of Coupled TM–TE Focus Wave Modes in Biisotropic Media¶

P. Hillion
Institut Henri Poincare 86 Bis Route de Croissy, 78110 Le Vésinet, France
Received December 21, 1999

Abstract—We prove that two kinds of coupled TM–TE focus wave modes can propagate in a biisotropic medium, and we consider what happens when a TM or TE focus wave mode impinges from a vacuum on a slab made of a biisotropic material. We get the amplitudes of the reflected and refracted waves in a chiral slab.

1. INTRODUCTION

Some years ago, new solutions for Maxwell's equations called focus wave modes (FWM) were discovered [1, 2] with the property to be distortion-free. It was expected that these waves, able in theory to carry electromagnetic energy at large distances, could replace lasers or radars in some situations. However, at the same time, new chiral materials were developed [3, 4] that improved the furtivity of targets. Thus, a natural question arises: what happens when an FWM impinges on a chiral target? Part of the incident wave is reflected and part is refracted inside the target, but in what proportion?

To answer this question is a difficult task, and in this work we assume that the incident field is a TM or TE monochromatic FWM since in this case Maxwell's equations do not depend on one coordinate, making calculations easier. In a vacuum, TM and TE fields propagate separately; however, in a biisotropic medium, one may expect that chirality will couple these two components of the electromagnetic field. We first need to discuss this point before looking for the amplitudes of the reflected and refracted waves.

For an incident monochromatic field, one may neglect the dependence of ε, μ, β, permittivity, permeability, and chirality on frequency so that the constitutive relations in a chiral medium are [5–7]

$$\begin{aligned}\mathbf{D} &= \varepsilon\mathbf{E} + i\beta\mathbf{B} = q\mathbf{E} + i\mu\beta\mathbf{H}, \quad i = \sqrt{-1},\\ \mathbf{H} &= \mu^{-1}\mathbf{B} + i\beta\mathbf{E} = n^{-2}(q\mathbf{B} + i\mu\beta\mathbf{D}),\end{aligned} \tag{1}$$

$$n^2 = \varepsilon\mu, \quad q = \varepsilon + \mu\beta^2. \tag{1a}$$

Here, **E**, **D**, **B**, **H** are the usual components of the electromagnetic field.

¶This article was submitted by the authors in English.

2. COUPLED TM–TE FIELDS IN BIISOTROPIC MEDIA

TM and TE fields do not depend on one coordinate (say y), and we can write Maxwell's equations

$$\begin{aligned}\partial_z H_y &= -c^{-1}\partial_t D_x,\\ \partial_x H_y &= c^{-1}\partial_t D_z,\\ c^{-1}\partial_t B_y &= \partial_x E_z - \partial_z E_x,\end{aligned} \tag{2a}$$

$$\begin{aligned}\partial_z E_y &= c^{-1}\partial_t B_x,\\ \partial_x E_y &= -c^{-1}\partial_t B_x,\\ c^{-1}\partial_t D_y &= \partial_z H_x - \partial_x H_z\end{aligned} \tag{2b}$$

with the additional divergence equations

$$\partial_x D_x + \partial_z D_z = 0, \quad \partial_x B_x + \partial_z B_z = 0. \tag{3}$$

We start with the harmonic plane wave solutions of Eqs. (2) and (3).

2.1. Coupled TM–TE Plane Waves

Leaving aside the harmonic term $\exp(ikct)$ and substituting Eq. (1) into Eq. (2a) and (2b) give

$$\begin{aligned}\partial_z(H_y + i\beta E_y) &= -ik(D_x - i\beta B_x) = -ik\varepsilon E_x,\\ \partial_x(H_y + i\beta E_y) &= ik(D_z - i\beta B_z) = ik\varepsilon E_z,\\ ik\mu(H_y - i\beta E_y) &= \partial_x E_z - \partial_z E_x,\end{aligned} \tag{4a}$$

and

$$\begin{aligned}\partial_z(qE_y - i\mu\beta H_y) &= ik(qB_x + i\mu D_x) = ikn^2 H_x,\\ \partial_x(qE_y - i\mu\beta H_y) &= -ik(qB_z + i\mu\beta D_z) = -ikn^2 H_z,\\ ik(qE_y + i\mu\beta H_y) &= \partial_z H_x - \partial_x H_z,\end{aligned} \tag{4b}$$

while the divergence equations become

$$\begin{aligned} q(\partial_x E_x + \partial_z E_z) + i\mu\beta(\partial_x H_x + \partial_z H_z) &= 0, \\ i\mu\beta(\partial_x E_x + \partial_z E_z) - \mu(\partial_x H_x + \partial_z H_z) &= 0, \end{aligned} \tag{5}$$

which implies

$$\partial_x E_x + \partial_z E_z = 0, \quad \partial_x H_x + \partial_z H_z = 0. \tag{5a}$$

From (4a) and (5a), we get for the component E_x two equations in which $\Delta = \partial_x^2 + \partial_z^2$,

$$\begin{aligned} ik\mu\partial_z(H_y + i\beta E_y) &= k^2 n^2 E_x, \\ ik\mu\partial_z(H_y - i\beta E_y) &= \partial_x\partial_z E_z - \partial_z^2 E_x = -\Delta E_x. \end{aligned} \tag{6}$$

Successively eliminating $i\beta E_y$ and H_y from these relations, we obtain

$$\begin{aligned} 2ik\mu\partial_z H_y &= -(\Delta - n^2 k^2)E_x, \\ 2k\mu\beta\partial_z E_y &= -(\Delta + n^2 k^2)E_x. \end{aligned} \tag{7}$$

Similarly, we deduce from (4b) and (5a)

$$\begin{aligned} ik\partial_z(qE_y - i\mu\beta H_y) &= -k^2 n^2 H_x, \\ ik\partial_z(qE_y + i\mu\beta H_y) &= \partial_z^2 H_x - \partial_x\partial_z H_z = \Delta H_x \end{aligned} \tag{8}$$

and the successive elimination of qE_y and $\mu\beta H_y$ gives

$$\begin{aligned} 2ikq\partial_z E_y &= (\Delta - k^2 n^2)H_x, \\ 2k\mu\beta\partial_z H_y &= -(\Delta + k^2 n^2)H_x. \end{aligned} \tag{9}$$

Then, eliminating successively $\partial_z H_y$ and $\partial_z E_y$ from (7) and (9), we get two equations:

$$\begin{aligned} (\Delta + k^2 n^2)H_x + i\beta(\Delta - n^2 k^2)E_x &= 0, \\ i\mu\beta(\Delta - k^2 n^2)H_x - q(\Delta + n^2 k^2)E_x &= 0. \end{aligned} \tag{10}$$

Changing ∂_z into ∂_x and E_x, H_x into $-E_x$, eE_z would supply similar relations for the pair (E_z, H_z) in relations (6)–(10).

We now assume that H_x and E_x are the solutions to the scalar Helmholtz equation in which m is to be determined

$$(\Delta + k^2 m^2)E_x, H_x = 0, \tag{11}$$

and substituting (11) into (10) gives the homogeneous system of equations

$$\begin{aligned} (n^2 - m^2)H_x - i\beta(n^2 + m^2)E_x &= 0, \\ i\mu\beta(n^2 + m^2)H_x + q(n^2 - m^2)E_x &= 0 \end{aligned} \tag{12}$$

with a nontrivial solution if its determinant is zero, which requires

$$g(n^2 - m^2)^2 - \mu\beta^2(n^2 + m^2)^2 = 0 \tag{13}$$

or using the definition (1a) of q

$$\varepsilon(n^2 - m^2)^2 - 4\mu\beta^2 n^2 m^2 = 0 \tag{3a}$$

with the solutions

$$\begin{aligned} \varepsilon m^2 &= (\varepsilon + 2\mu\beta^2)n^2 \pm 2n^2\beta(\mu q)^{1/2} \\ \text{or} \\ m^2 &= [\mu\beta \pm (\mu q)^{1/2}]^2. \end{aligned} \tag{14}$$

This implies, as expected [6], that two kinds of coupled TM–TE harmonic plane waves can propagate in a biisotropic medium.

We can derive explicitly the plane wave solutions of Eqs. (4a) and (4b). One of the components $E_{x,z}$ or $H_{x,z}$ (say E_x) is an arbitrary solution of the Helmholtz equation (11) so that, for a harmonic plane wave with amplitude A propagating in a direction making the angle u with oz, E_x is

$$\begin{aligned} E_x &= A\exp(ik\Omega_p), \\ \Omega_p &= ct - m(x\sin u + z\cos u), \end{aligned} \tag{15}$$

where m is one of the two solutions (14). The phase Ω_p satisfies the characteristic equation

$$(\partial_z\Omega_p)^2 + (\partial_z\Omega_p)^2 - m^2 c^{-2}(\partial_t\Omega_p)^2 = 0. \tag{16}$$

Then, H_x is obtained from (12)

$$H_x = i\beta(n^2 + m^2)(n^2 - m^2)^{-1}E_x, \tag{17}$$

while from the first equation of the systems (7) and (9) we obtain

$$\begin{aligned} 2\mu m H_y \cos u &= (m^2 + n^2)E_x, \\ 2qmE_y \cos u &= (m^2 + n^2)H_x. \end{aligned} \tag{18}$$

Finally, E_z and H_z are deduced from the divergence equations (5a):

$$E_z = -E_x \tan u, \quad H_z = -H_x \tan u. \tag{18a}$$

The relations (15), (17), (18), and (18a) give the components of the coupled TM–TE electromagnetic plane waves in a biisotropic medium.

2.2. Coupled TM–TE Focus Wave Modes

With the subscript j taking the values 1, 2, 3 for the respective coordinates x, y, z, the TM–TE FWMs have the form [8]

$$\begin{aligned} E_j &= e_j\exp(ik\Omega), \quad H_j = h_j\exp(ik\Omega), \\ \Omega &= ct - Z - gX^2, \end{aligned} \tag{19}$$

in which the amplitudes e_j, h_j depend on x, z, t, and

$$\begin{aligned} Z &= m(x\sin u + z\cos u),\\ X &= m(x\cos u - z\sin u),\\ g &= (a + ct + Z)^{-1}, \end{aligned} \tag{19a}$$

where a is a pure imaginary parameter, while m is still given by (14) since the phase Ω must be a solution [9] to the characteristic equation (16). Simple calculation gives

$$\begin{aligned} w_x \equiv \partial_x\Omega &= -m(1-g^2X^2)\sin u - 2mgX\cos u,\\ w_z \equiv \partial_z\Omega &= -m(1-g^2X^2)\cos u + 2mgX\sin u,\\ w_0 \equiv c^{-1}\partial_t\Omega &= 1 + g^2X^2. \end{aligned} \tag{20}$$

Substituting Eq. (20) into Eq. (16) proves the statement.

We now assume the frequency kc in Eq. (19) is large enough to make the derivatives of e_j and h_j negligible with respect to the derivatives of $k\Omega$ (high-frequency approximation). Then, using Eq. (20), we find that

$$\begin{aligned} \partial_l E_j &= ikw_l E_j, \quad \partial_l H_j = ikw_l H_j,\\ \partial_l &= \partial/\partial_l, \quad l = 1, 3,\\ c^{-1}\partial_t E_j &= ikw_0 E_j, \quad c^{-1}\partial_t H_j = ikw_0 H_j. \end{aligned} \tag{21}$$

Substituting Eq. (19) into Eqs. (2) and (3); using Eqs. (1) and (21); and deleting $\exp(ikQ)$, the Maxwell equations become

$$\begin{aligned} w_z(h_y + i\beta e_y) &= -\varepsilon w_0 e_z,\\ w_x(h_y + i\beta e_y) &= \varepsilon w_0 e_z,\\ w_0\mu(h_y - i\beta e_y) &= w_x e_z - w_z e_x, \end{aligned} \tag{22a}$$

$$\begin{aligned} w_z(qe_y - i\mu\beta h_y) &= n^2 w_0 h_x,\\ w_x(qe_y - i\mu\beta h_y) &= -n^2 w_0 h_z,\\ w_0(qe_y + i\mu\beta h_y) &= w_z h_x - w_x h_z, \end{aligned} \tag{22b}$$

while we obtain

$$\begin{aligned} q(w_x e_x + w_z e_z) + i\mu\beta(w_x h_x + w_z h_z) &= 0,\\ i\mu\beta(w_x e_x + w_z e_z) - \mu(w_x h_x + w_z h_z) &= 0 \end{aligned} \tag{23}$$

for the divergence equations, implying $w_x e_x + w_z e_z = 0$, $w_x h_x + w_z h_z = 0$.

According to Eq. (15), the components of the wave vector for plane waves are $w_x = -m\sin u$ and $w_z = -m\cos u$ with $w_0 = 1$; then, Eqs. (22a), (22b) reduce to Eqs. (4a), (4b) so that changing $-m\sin u$, $-m\cos u$ into w_x/w_0, w_z/w_0 in the expressions of the TM–TE plane waves supplies the coupled TM–TE FWMs. Thus, from Eqs. (18), (18a), we find that

$$E_z = -w_x E_x/w_z, \quad H_z = -w_x H_x/w_z, \tag{24a}$$

$$\begin{aligned} 2\mu w_z H_y &= -(m^2 + n^2)w_0 E_x,\\ 2q w_z E_y &= (m^2 + n^2)w_0 H_x. \end{aligned} \tag{24b}$$

However, the arbitrary component E_x is now a solution to the wave equation (that reduces to Eq. (11) for plane waves)

$$(\partial_x^2 + \partial_z^2 - m^2c^2\partial_t^2)E_x = 0, \tag{25}$$

and we choose for E_x a scalar FWM that is [8]

$$E_x = Ag^{1/2}\exp(ik\Omega), \tag{26}$$

in which the amplitude A is multiplied by $g^{1/2}$ [where g is the function (19a)]. Expression (17) for H_x is formally unchanged, leading to

$$\begin{aligned} H_x &= i\beta(n^2 + m^2)(n^2 - m^2)^{-1}E_x\\ &= iq(n^2 - m^2)(n^2 + m^2)^{-1}E_x/\mu\beta. \end{aligned} \tag{26a}$$

Successively substituting the two solutions (14) m_+ and m_- into Eqs. (24), (26), and (26a) gives the expressions for the two kinds of coupled TM–TE FWMs propagating in a biisotropic medium.

3. EXCITATION OF COUPLED TM–TE FOCUS WAVE MODES

From now on, the index s (used as subscript or superscript for convenience) takes the values 0 and 2 (for the incident and reflected fields in vacuum) and 1_+ and 1_-) for the refracted fields in the chiral medium, the plus and minus signs corresponding to the two solutions (14)).

We first have to give the expression of the incident TM–TE FWMs. In a vacuum, these fields propagate separately and for the TM field, assuming $\varepsilon = \mu = 1$, $\beta = 0$, the Maxwell equations (2a) become

$$\begin{aligned} \partial_z H_y^0 &= -c^{-1}\partial_t E_x^0,\\ \partial_x H_y^0 &= c^{-1}\partial_t E_z^0,\\ c^{-1}\partial_t H_y^0 &= \partial_x E_z^0 - \partial_z E_x^0. \end{aligned} \tag{27}$$

For H_y^0, we take a scalar FWM with amplitude A_0 and angle of incidence u_0

$$H_y^0 = A_0 g_0^{1/2}\exp(ik_0\Omega_0), \quad \Omega_0 = ct - Z_0 - g_0X_0^2, \tag{28}$$

$$\begin{aligned} Z_0 &= x\sin u_0 + z\cos u_0,\\ X_0 &= x\cos u_0 - z\sin u_0,\\ g_0 &= (a + ct + Z_0)^{-1}. \end{aligned} \tag{28a}$$

Substituting Eq. (28) into (27) and using the high-frequency approximation immediately gives

$$E_x^0 = -w_z^0 H_y^0/w_0^0, \quad E_z^0 = w_x^0 H_y^0/w_0^0 \tag{29}$$

with

$$w_x^0 = -(1-g_0^2X_0^2)\sin u_0 - 2g_0X_0\cos u_0,$$
$$w_z^0 = -(1-g_0^2X_0^2)\cos u_0 + 2g_0X_0\sin u_0, \qquad (29a)$$
$$w_0^0 = 1 + g_0^2X_0^2.$$

Similarly, for a TE FWM with amplitude B_0,

$$E_y^0 = B_0 g_0^{1/2}\exp(ik_0\Omega_0),$$
$$H_x^0 = w_z^0E_y^0/w_0^0, \quad H_z^0 = -w_x^0E_y^0/w_0^0. \qquad (30)$$

We now assume that one these two FWMs impinges on the face of a chiral slab located in the $z = 0$ plane. Our first task is to get the conditions for reflection and refraction.

3.1. Conditions for Reflection and Refraction

The conditions for reflection and refraction are obtained by imposing the continuity of the transverse component $k_x = k\partial_x\Omega$ of the wave vector [10]. Since the parameter a in Ω is arbitrary, we make the calculations easier by assuming $a = 0$ so that, according to Eqs. (19) and (19a),

$$\Omega = ct - Z - X^2(ct+Z)^{-1}$$
$$= [c^2t^2 - m^2(x^2+z^2)][ct + mx\sin u + mx\cos u]^{-1}. \qquad (31)$$

Simple calculation gives

$$(\partial_x\Omega)_{z=0} = -2xm^2(ct+mx\sin u)^{-1}$$
$$-m(c^2t^2 - m^2x^2)(ct+mx\sin u)^{-2}\sin u. \qquad (31a)$$

Let us now look at conditions for reflection: in a vacuum, $m = 1$ and, with the previous convention on the index s, we have $u = u_0$ and $u = u_2$ for the incident and reflected fields. In addition, we do not assume *a priori* the equality of the frequencies k_0c and k_2c. Then, using Eq. (31), the continuity of $(k\partial_x\Omega)_{z=0}$ gives

$$2k_0x(ct + x\sin u_0)^{-1}$$
$$+ k_0(c^2t^2 - x^2)(ct + x\sin u_0)^{-2}\sin u_0$$
$$= 2k_2x(ct + x\sin u_2)^{-1} \qquad (32)$$
$$+ k_2(c^2t^2 - x^2)(ct + x\sin u_2)^{-2}\sin u_2,$$

implying the usual Descartes–Snell relations:

$$k_2 = k_0, \quad \sin u_2 = \sin u_0 \quad (u_2 = \pi - u_0). \qquad (32a)$$

The situation is somewhat different for refraction, and we first shift the coordinate x to x' inside the slab. At this point, it is not necessary to show a difference between the two solutions m_+ and m_-, so we note k_1c and u_1 (the frequency and the angle of refraction). Then, still using Eq. (31a), the continuity of $(k\partial_x\Omega)_{z=0}$ supplies the relation

$$2k_0x(ct + x\sin u_0)^{-1}$$
$$+ k_0(c^2t^2 - x^2)(ct + x\sin u_0)^{-2}\sin u_0$$
$$= 2k_1m^2x'(ct + mx'\sin u_1)^{-1} \qquad (33)$$
$$+ mk_1(c^2t^2 - m^2x'^2)(ct + mx'\sin u_1)^{-2}\sin u_1.$$

This relation seems very intricate, but one can easily check that Eq. (33) yields conditions

$$u_1 = u_0, \quad mk_1 = k_0, \quad (mx' - x)_{z=0} = 0 \qquad (33a)$$

that are very different from the Descartes–Snell law: the refracted FWM undergoes a jump frequency $k_0 \longrightarrow k_1$ (a Göos–Hanken-like shift $x \longrightarrow x'$), but it propagates in the direction of the incident FWM. Note that the refractive index m generally depends on frequency, so the jump condition may be written as $\omega_0 = \omega_1 m(\omega_1)$ and total reflection happens when this equation has no solution. In theory, it would be possible to choose the frequency of the incident field in a way that none, one, or both of the FWMs can propagate in a chiral slab.

We need the expressions on the interface $z = 0$ of the phase Ω, its derivatives, and the form of the attenuation factor g. Using conditions (32a) and (33a) for reflection and refraction and taking into account that the phase Ω_1 is defined with respect to the x' coordinate, we easily get

$$(k_0\Omega_0)_{z=0} = (k_2\Omega_2)_{z=0} = (mk_1\Omega_1)_{z=0} = \Sigma,$$
$$(g_0 = g_2 = g_1)_{z=0} = G \qquad (34)$$

from Eqs. (19) and (19a) with the parameters of the incident field and restoring a

$$\Sigma = k_0(ct - x\sin u_0 - Gx^2\cos u_0),$$
$$G = (a + ct + x\sin u_0)^{-1}. \qquad (34a)$$

Similarly, we get

$$(w_x^0)_{z=0} = (w_x^2)_{z=0} = (w_x^1)_{z=0} = V_x,$$
$$(w_z^0)_{z=0} = -(w_z^2)_{z=0} = (w_z^1)_{z=0} = V_z, \qquad (35)$$
$$(w_0^0)_{z=0} = (w_0^2)_{z=0} = (w_0^1)_{z=0} = V_0$$

from Eq. (20a) with

$$V_x = -(1 - G^2x^2\cos^2 u_0)\sin u_0 - 2Gx\cos^2 u_0,$$
$$V_z = -(1 - G^2x^2\cos^2 u_0 + 2Gx\sin u_0)\cos u_0, \qquad (35a)$$
$$V_0 = 1 + G^2x^2\cos u_0.$$

We have now all the elements to discuss the excitation of coupled TM–TE FWMs in a chiral slab.

3.2. Excitation by an Incident TM Focus Wave Mode

The amplitudes of the reflected and refracted FWMs (Fresnel coefficients) are obtained by imposing the continuity of the tangential components of **E** and **H** on the interface $z = 0$, that is, with the previous convention on the index s:

$$\begin{aligned}[(H^0 + H^2 - H^{1_+} - H^{1_-})_{x,y}]_{z=0} &= 0,\\ [(E^0 + E^2 - H^{1_+} - H^{1_-})_{x,y}]_{z=0} &= 0.\end{aligned} \tag{36}$$

For an incident TM FWM, these conditions become

$$[(H^0 + H^2 - H^{1_+} - H^{1_-})_y]_{z=0} = 0, \tag{37a}$$

$$[(H^2 - H^{1_+} - H^{1_-})_x]_{z=0} = 0, \tag{37b}$$

$$[(E^2 - E^{1_+} - E^{1_-})_y]_{z=0} = 0, \tag{37c}$$

$$[(E^0 + E^2 - E^{1_+} - E^{1_-})_x]_{z=0} = 0. \tag{37d}$$

So, an incident TM FWM generates TM- and TE-reflected FWMs. Now, to satisfy Eq. (37), we need the expressions of the tangential components of the electromagnetic field on $z = 0$; for sake of simplicity, we introduce the functions with Σ and G given by (34a)

$$\begin{aligned}P &= G^{1/2}\exp(i\Sigma),\\ Q^{\pm} &= G^{1/2}\exp[i\Sigma(m_{\pm})^{-1}],\end{aligned} \tag{38}$$

in which m_+ and m_- are the two solutions (14) for m.

Then, taking into account Eqs. (34), (35), and (38) for the incident and reflected fields, we get

$$\begin{aligned}(H_y^s)_{z=0} &= A_sP, \quad (E_y^s)_{z=0} = B_sP, \quad s = 0, 2,\\ (E_x^0)_{z=0} &= -V_zA_0P/V_0,\\ (E_x^2)_{z=0} &= V_zA_2P/V_0,\\ (H_x^0)_{z=0} &= V_zB_0P/V_0,\\ (H_x^2)_{z=0} &= -V_zB_2P/V_0\end{aligned} \tag{39}$$

from Eqs. (28), (29), and (30).

Taking $A = V_zC/V_0$ as the amplitude in Eq. (26) and still using Eqs. (34), (35), and (38), we get from (24), (26), and (26a)

$$\begin{aligned}2\mu[(H^{1_+})_y]_{z=0} &= -(n^2 + m_+^2)C_{1_+}Q^+,\\ 2\mu\beta[(E^{1_+})_y]_{z=0} &= i(n^2 - m_+^2)C_{1_+}Q^+,\\ [(E^{1_+})_x]_{z=0} &= V_zC_{1_+}Q^+/V_0,\\ \mu\beta[(H^{1_+})_x]_{z=0} &= iqr_+V_zC_{1_+}Q^+/V_0\end{aligned} \tag{40}$$

with

$$r_+ = (n^2 - m_+^2)(n^2 + m_+^2)^{-1}. \tag{40a}$$

Changing the plus sign into a minus sign in Eqs. (40) and (40a) gives the expressions on the interface $z = 0$ of the second refracted field. The relations (39) and (40) depend on four unknown amplitudes (A_2, B_2, C_{1_+}, C_{1_-}), while A_0 is the amplitude of the incident TM FWM and $B_0 = 0$. Substituting Eqs. (39) and (40) into (37) and respecting the lexicographic order, we obtain

$$\begin{aligned}2\mu(A_0 + A_2)P + (n^2 + m_+^2)C_{1_+}Q^+\\ + (n^2 + m_-^2)C_{1_-}Q^- = 0,\end{aligned} \tag{41a}$$

$$i\mu\beta q^{-1}B_2P - r_+C_{1_+}Q^+ - r_-C_{1_-}Q^- = 0, \tag{41b}$$

$$\begin{aligned}2i\mu\beta B_2P + (n^2 - m_+^2)C_{1_+}Q^+\\ + (n^2 - m_-^2)C_{1_-}Q^- = 0,\end{aligned} \tag{41c}$$

$$(A_0 - A_2)P + C_{1_+}Q^+ + C_{1_-}Q^- = 0. \tag{41d}$$

Summing and subtracting Eqs. (41a) and (41c) gives (due to the fact that $n^2 = \varepsilon\mu$)

$$(A_0 + A_2 + i\beta B_2)P + \varepsilon(C_{1_+}Q^+ + C_{1_-}Q^-) = 0, \tag{42a}$$

$$\mu(A_0 + A_2 - i\beta B_2)P + m_+^2C_{1_+}Q^+ + m_-^2C_{1_-}Q^- = 0; \tag{42b}$$

substituting Eq. (41a) into (42a) gives

$$A_0(1 - \varepsilon) + A_2(1 + \varepsilon) + i\beta B_2 = 0; \tag{43a}$$

while from Eqs. (41d) and (42b), we obtain

$$\begin{aligned}(m_+^2 - m_-^2)C_{1_-}Q^-\\ = [A_0(\mu - m_+^2) + A_2(\mu + m_+^2) + i\mu\beta B_2]P,\\ (m_-^2 - m_+^2)C_{1_+}Q^+\\ = [A_0(\mu - m_-^2) + A_2(\mu + m_-^2) + i\mu\beta B_2]P;\end{aligned} \tag{43b}$$

finally, substituting Eq. (43b) into (41b) gives

$$\begin{aligned}i\mu\beta[r_- - r_+ + q^{-1}(m_-^2 - m_+^2)]B_2\\ = A_0[r_+(\mu - m_-^2) - r_-(\mu - m_+)^2]\\ + A_2[r_+(\mu + m_-^2) - r_-(\mu + m_+)^2].\end{aligned} \tag{44}$$

From Eqs. (43) and (44), we get in terms of A_0 the amplitudes A_2 and B_2 of the reflected TM–TE FWMs; as expected, they do not depend on the functions P or $Q^{\pm}$. Then, the amplitudes C_{1_+} and C_{1_-} of the refracted coupled TM–TE FWMs are obtained from Eq. (43b) as function of the phases $P/Q^{\pm}$, that is, according to Eq. (38) of $\exp[i\Sigma(1 - 1/m_{\pm})]$ with Σ given by Eq. (34). We refrain from giving the rather intricate expressions of A_2, B_2, and $C_{1_{\pm}}$; however, it is clear that the Fresnel coefficients for

the excitation of coupled TM–TE FWMs in a chiral slab are amenable to analytical calculations.

Exchanging the previous results H_y and E_y; H_x, H_z and $-E_x$, $-E_z$; A_0, A_2 and B_0, B_2 gives the reflected and refracted TM–TE FWMs excited by an incident TE FWM.

4. DISCUSSION

Some of the results obtained in this work for TM and TE FWMs are easily extended to arbitrary electromagnetic FWMs, for instance, the propagation of two kinds of FWMs in a biisotropic medium. Now, the phase of FWMs is a nonlinear function of space and time. This implies refraction conditions very different from Descartes–Snell conditions making things easier since FWMs undergo only a jump frequency and a lateral shift with propagation in the direction of the incident field; these properties hold for any incident FWM.

To get the amplitudes of the waves excited inside a chiral medium by an incident FWM, we had to use the high-frequency approximation so that Maxwell's equations formally take the same aspect as that for harmonic plane waves, except that now the wave vector is a function of space and time. The same approximation also works for a general FWM, but the calculations are a bit more intricate.

FWMs and plane waves share the same status: both propagate in all space for $-\infty < t < +\infty$, and *stricto sensu* no physical FWM exists (no more than physical plane waves). In both cases, one has to be satisfied with practical realizations giving a good approximation of these mathematical solutions in some bounded region of space–time. Many works have been devoted to the generation of approximate FWMs [11–15] with experimental settings in acoustics [16, 17] (but apparently not in electromagnetism yet) of beams having a less dispersion than usual Gaussian beams [18]. The author believes that, in the future, approximate electromagnetic FWMs will be generated, leading to practical applications.

REFERENCES

1. J. M. Brittingham, Appl. Phys. **54**, 1179 (1983).
2. A. P. Kiselev, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **24**, 1014 (1983).
3. D. L. Jaggard, N. Engheta, and J. C. Liu, Electron. Lett. **26**, 1332 (1990).
4. P. L. E. Uslenghi, Electromagnetics **10**, 201 (1990).
5. D. L. Jaggard, A. R. Mickelson, and C. H. Papas, Appl. Phys. **18**, 211 (1979).
6. N. Engheta and D. L. Jaggard, IEEE Trans. Antennas Propag. **30**, 6 (1988).
7. M. P. Silverman, J. Opt. Soc. Am. A **5**, 1952 (1988).
8. P. Hillion, J. Opt. (Paris) **29**, 345 (1998).
9. R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1962).
10. J. A. Arnaud, *Beam and Fiber Optics* (Academic, New York, 1976).
11. R. Donnelly and R. Ziolkowski, Proc. R. Soc. London, Ser. A **437**, 637 (1992).
12. J. Hernandez, R. Ziolkowski, and S. Parker, J. Acous. Soc. Am. **92**, 550 (1992).
13. P. Hillion, J. Opt. Soc. Am. A **9**, 137 (1992).
14. V. Borisov and A. Utkins, J. Phys. A **26**, 406 (1993).
15. R. Ziolkowski, IEEE Trans. Antennas Propag. **40**, 888 (1992).
16. R. Ziolkowski, K. Lewis, and B. Cook, Phys. Rev. Lett. **62**, 147 (1989).
17. R. Ziolkowski and K. Lewis, J. Appl. Phys. **68**, 6083 (1990).
18. D. Palmer, R. Donnelly, and R. MacIsaac, Phys. Rev. E **48**, 1410 (1993).

Journal of Experimental and Theoretical Physics, Vol. 91, No. 1, 2000, pp. 188–196.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 1, 2000, pp. 213–222.

SOLIDS
Structure

Total Reflection of Ultrasound from a Ferromagnetic Plate with Surface Pinning of the Spins

Yu. G. Peĭsakhovich* and A. A. Shtygashev
Novosibirsk State Pedagogical University, Novosibirsk, 630068 Russia
**e-mail: ygp@nspu.nsu.ru*
Received February 24, 2000

Abstract—Using the effective transfer matrix for coupled elastic and spin waves, formulas are derived for the amplitudes of reflection and transmission of right-hand polarized elastic waves incident, in a direction along the normal, on a transversely magnetized ferromagnetic film with homogeneous conditions for pinning of the spins on the surfaces of the film. It is shown that a series of lines with reflectance maxima appears, to the extent of the magnetoelastic coupling, near the magnetoacoustic resonance. The shapes of the spectral line contours and the effect of the thickness of the film and damping are analyzed. The possibility of a similar effect in other polariton systems is discussed. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The resonance properties of thin magnetic films and multilayer structures are substantially determined by the boundary conditions for resonant fields at an interfacial surface. The electrodynamic and exchange boundary conditions, including surface anisotropy effects, determine the magnetostatic spectrum and the spin-wave resonance spectrum of films [1–3]. The mechanical conditions at the surface are important for magnetoacoustic resonance [1, 5].

The resonance properties of films are especially striking in the reflection and transmission amplitude spectra for different types of waves interacting with the magnetic subsystem of a film. A theoretical calculation of the reflection and transmission coefficients is difficult to perform in general because of many geometric and structural factors, such as the angle of incidence, direction and magnitude of the external magnetic field, the magnetization, the values of the anisotropy constants, the existence of several sublattices, different branches of the excitation spectrum, dissipative processes and so on, each of which leads to unique effects and is manifested individually most strongly only in a definite simplified situation.

Narrow—with respect to frequency, magnitude, and orientation of the external magnetic field—lines of nondissipative tenfold reduction of the transmission amplitude of longitudinal elastic waves through a plate of the antiferromagnet $KMnF_3$ in the nuclear magnetoacoustic resonance frequency range have recently been observed in [4, 5]. In [4, 5] this effect was attributed to the mutual interference quenching of two characteristic magnetoacoustic waves, if their phase difference at the exit from the plate is a multiple of π. The theory of resonance between elastic and nuclear spin waves in two sublattice antiferromagnets requires a complicated symmetry and quantitative analysis. In [4, 5] such an analysis was performed only for one frequency of intersection of the resonating branches.

On the other hand, in our view, the total nondissipative reflection under conditions where an externally incident wave is coupled with the wave of characteristic elementary excitations of the film is to some extent physically a general effect. Consequently, in the present paper a simpler and well-studied [1, 3] related system will be examined—the problem of the reflection of an elastic circularly polarized wave, incident in a direction along the normal on a ferromagnetic film magnetized in a direction perpendicular to the surface, is presented.

In such a geometry the right- and left-hand spiral circular variables separate in the equations of motion of the magnetoelastic waves as well as in the boundary conditions describing the partial or total surface pinning of the spins. Depolarization phenomena such as the acoustic Faraday effect are absent. As a result, expressions for the reflection and transmission coefficients will be obtained analytically and analyzed numerically. We shall show that a series of narrow total resonance reflection lines, which become smeared when the damping of waves in the film is taken into account, appears, to the extent of the magnetoelastic coupling, near the magnetoacoustic resonance. These lines are superposed on an ordinary half-wave resonance transmission comb [6]. The total-reflection resonance conditions include, besides the obvious trigonometric factors, the parameters of coherent coupling of partial waves, which depend strongly on the frequency and other physical quantities, and they do not reduce to a resonance condition presented in [4, 5]. This is because we take into account successively the existence of reflected waves in the film and the conditions of surface pinning of the spins on the incident and transmitted wave sides. In our case all partial waves are plane monochromatic waves. We shall discuss only

1063-7761/00/9101-0188$20.00

briefly, and qualitatively, the effects which can appear with the use of short pulses such as those used in the experiment of [4, 5].

As will be clear from the derivation, the significance of the results obtained is not limited by the specific phonon-magnon system mentioned in the title. The formulas derived below for the reflection and transmission coefficients have a general physical meaning, and they substantially generalize the well-known formulas for the reflection of a wave, not necessarily of an elastic nature, incident in a direction along the normal on a plate to the case where polariton coupling of the externally incident wave with a branch of characteristic wave oscillations occurs in the plate under conditions of total or partial pinning of the internal wave field on the surface of the plate.

2. REFLECTION AND TRANSMISSION COEFFICIENTS

Let us consider the standard one-dimensional scattering problem. Let the x axis be perpendicular to a uniform plate, which occupies the region $0 \le x \le a$. Incident and reflected monochromatic plane waves with frequency ω and wave number k are present to the left ($x < 0$) of the plate:

$$u = e^{\kappa x} + re^{-\kappa x}, \quad \kappa = ik \tag{1}$$

the time factor $e^{i\omega t}$ is omitted, and a transmitted wave

$$u = te^{\kappa(x-a)} \tag{2}$$

is present to the right of the plate ($x > a$).

For acoustic elastic waves u is the solution of the wave equation from the theory of elasticity:

$$\left(v^2\frac{d^2}{dx^2} + \omega\right)u = 0, \tag{3}$$

where v is the speed of sound in the medium outside the plate, i.e., $k = \omega/v$. The boundary conditions on the plate surfaces $x = 0$ and $x = a$ require continuity of u and its derivative $u' = du/dx$. It is convenient to use the transfer matrix (propagator) M to describe the procedure for matching the solutions. This matrix relates the values of the column vector of the state

$$\Psi(x) = \begin{pmatrix} u(x) \\ u'(x) \end{pmatrix}$$

at the boundaries of the layer [7–9]

$$\Psi(a) = M\Psi(0). \tag{4}$$

The transmittance $T = |t|^2$ and reflectance $R = |r|^2$ can be expressed in terms of the elements of an effective two-dimensional matrix of a transition through the plate:

$$M_{\text{eff}} = L_0^{-1}M^{-1}L_0 \equiv \begin{pmatrix} (M_{\text{eff}})_{11} & (M_{\text{eff}})_{12} \\ (M_{\text{eff}})_{21} & (M_{\text{eff}})_{22} \end{pmatrix},$$
$$L_0 = \begin{pmatrix} 1 & 1 \\ \kappa & -\kappa \end{pmatrix} \tag{5}$$

using formulas for the amplitudes

$$t = \frac{1}{(M_{\text{eff}})_{11}}, \quad r = \frac{(M_{\text{eff}})_{22}}{(M_{\text{eff}})_{11}}. \tag{6}$$

If the incident wave arrives on the right-hand side from the region $x > a$, then the amplitudes are given by the expressions

$$t_r = \frac{\det(M_{\text{eff}})}{(M_{\text{eff}})_{11}}, \quad r_r = -\frac{(M_{\text{eff}})_{12}}{(M_{\text{eff}})_{11}}. \tag{6a}$$

3. TRANSFER MATRIX

To obtain a two-dimensional transfer matrix M for passage through the plate it is necessary to take account of the fact that the propagation of the field u inside the plate is described by a system of coupled wave equations which originate from the equations of motion of the internal degrees of freedom (for a ferromagnetic plate these are the equations of the theory of elasticity, coupled with the Landau–Lifshitz equations [1, 3]). The dimension of the transfer matrix in a plate is determined by the general order of the system of differential equations and by the number of coupled partial waves. However, for all additional internal degrees of freedom pinning boundary conditions must be imposed at the surfaces. They give relations between the partial amplitudes and make it possible to construct a transfer matrix of lower degree, sufficient for matching at the boundaries the solutions corresponding only to an external scattered field. This technique was used in [7], which concerned the propagation of elastic waves in a system of piezoelectric crystal layers with surface pinning of the electric or magnetic field. Here we shall also consider the very simple case where an interaction of the scattered wave with only one branch of the characteristic oscillations is present in the plate, but in contrast to [7] these oscillations propagate along the x axis and are not surface waves. This situation can occur in a uniaxial ferromagnetic plate magnetized in a direction perpendicular to the surface. In this case the following system of equations describes the propagation of right-hand spiral circular monochromatic plane wave components

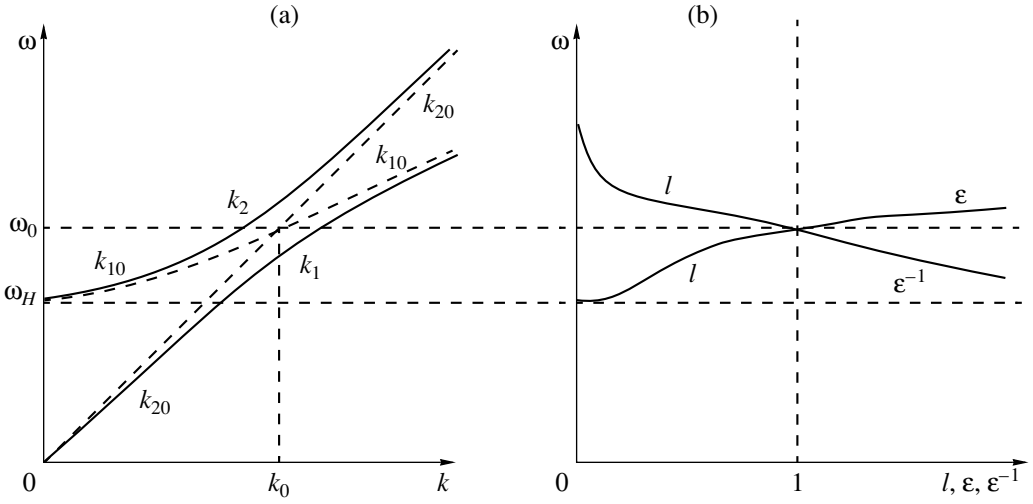


Fig. 1. (a) Spectrum of magnetoacoustic waves near the bottom magnetoacoustic resonance point; (b) relative coherent coupling parameter.

of the elasticity field u and magnetization m along the X axis [1, 3]:

$$\begin{pmatrix} Z_1 & \Delta_1 \\ \Delta_2 & Z_2 \end{pmatrix}\begin{pmatrix} u \\ m \end{pmatrix} = 0, \tag{7}$$

$$Z_1(k) = \omega^2 - v_t^2 k^2, \quad \Delta_1(k) = -ik\frac{B}{\rho M_0},$$
$$Z_2(k) = \omega - (\omega_H + \eta k^2), \quad \Delta_2(k) = ikgB, \tag{8}$$

where v_t is the speed of sound in the plate, $\omega_H = gH_0$ is the ferromagnetic resonance frequency, H_0 is the magnetic field intensity, g is the magnetomechanical ratio, η is the effective exchange constant, B is the magnetoelastic coupling constant, M_0 is the magnetization, and ρ is the density of the plate. The homogeneous exchange boundary conditions for surface pinning have the form [1–3]

$$(m + hm')_{x=0} = 0, \quad (m + \tilde{h}m')_{x=a} = 0, \tag{9}$$

where $m' = dm/dx$, and h and $\tilde{h}$ are the pinning parameters. The characteristic equation

$$Z_1Z_2 - \Delta_1\Delta_2 \equiv (k^2 - k_{10}^2)(k^2 - k_{20}^2) - k^2b^2 = 0,$$
$$k_{10}^2 = \frac{\omega - \omega_H}{\eta}, \quad k_{20}^2 = \left(\frac{\omega}{v_t}\right)^2, \quad b^2 = \frac{gB^2}{\rho M_0 \eta v_t^2} \tag{10}$$

gives four roots describing the repulsion of the branches (Fig. 1a):

$$k_{1,2} = (0.5B_1 \mp \sqrt{(0.5B_1)^2 - B_2})^{1/2},$$
$$k_{3,4} = -k_{1,2},$$
$$B_1 = k_{10}^2 + k_{20}^2 + b^2 = k_1^2 + k_2^2,$$
$$B_2 = k_{10}^2k_{20}^2 = k_1^2k_2^2. \tag{11}$$

The general solution of the system (7) has the form

$$u = \sum_{j=1}^{4} A_j e^{\kappa_j x}, \quad m = \sum_{j=1}^{4} l_j A_j e^{\kappa_j x}, \tag{12}$$

where $\kappa_j = ik_j$, A_j are partial amplitudes, and the coherent coupling coefficients l_j ($j = 1, 2, 3, 4$) of m and u waves are

$$l_j \equiv \lambda_j^{-1} = -\frac{Z_1(k_j)}{\Delta_1(k_j)} = -\frac{\Delta_2(k_j)}{Z_2(k_j)}, \quad l_{3,4} = -l_{1,2}. \tag{13}$$

The state vector in the plate and the transfer matrix are four-dimensional. It is convenient to work with two-dimensional blocks:

$$\begin{pmatrix} \Psi \\ \Psi' \end{pmatrix}_x = LM_{\text{diag}}(x)\begin{pmatrix} A_{12} \\ A_{34} \end{pmatrix},$$
$$A_{ij} = \begin{pmatrix} A_i \\ A_j \end{pmatrix}, \quad i, j = 1, 2, 3, 4, \tag{14}$$

where the transfer matrix in the diagonal representation is

$$M_{\text{diag}}(x) = \begin{pmatrix} M_{d1}(x) & 0 \\ 0 & M_{d2}(x) \end{pmatrix},$$
$$M_{dj}(x) = \begin{pmatrix} e^{\kappa_j x} & 0 \\ 0 & e^{-\kappa_j x} \end{pmatrix}, \quad j = 1, 2, \tag{15}$$

and the diagonalization matrix L is the Van der Monde matrix

$$L = \begin{pmatrix} L_1 & L_2 \\ \kappa_1^2 L_1 & \kappa_2^2 L_2 \end{pmatrix}, \qquad L_j = \begin{pmatrix} 1 & 1 \\ \kappa_j & -\kappa_j \end{pmatrix}, \quad j = 1, 2. \tag{16}$$

It is also convenient to introduce the matrices C_j and $S_j = M_{dj}(a)C_j$ of the form

$$C_j = d_j \begin{pmatrix} \bar{f}_j e^{\kappa_j a} - \bar{g}_j \\ f_j e^{\kappa_j a} - g_j \end{pmatrix}, \quad S_j = d_j \begin{pmatrix} \bar{f}_j - \bar{g}_j e^{\kappa_j a} \\ f_j - g_j e^{\kappa_j a} \end{pmatrix}, \tag{17}$$

$$d_j^{-1} = g_j \bar{f}_j e^{-\kappa_j a} - \bar{g}_j f_j e^{\kappa_j a}, \quad j = 1, 2,$$

where $g_j = 1 + h\kappa_j$, $\bar{g}_j = 1 - h\kappa_j$, $f_j = 1 + \tilde{h}\kappa_j$ and $\bar{f}_j = 1 - \tilde{h}\kappa_j$. Substituting the expressions (12) into the boundary conditions (9), we obtain pairwise coupling of the partial amplitudes

$$A_{34} = CA_{12}, \quad C = -\frac{l_1}{l_2} C_2 C_1^{-1}. \tag{18}$$

Separating the block of the first row in Eq. (14), we have

$$\Psi(x) = (L_1 M_{d1}(x) + L_2 M_{d2}(x)C)A_{12}. \tag{19}$$

Hence, by definitions (4) and (5), we find the desired effective two-dimensional transfer matrix

$$M_{\text{eff}} = \Gamma_c \Gamma_s^{-1},$$

$$\Gamma_c = \lambda_1 \tilde{L}_1 C_1 - \lambda_2 \tilde{L}_2 C_2,$$

$$\Gamma_s = \lambda_1 \tilde{L}_1 S_1 - \lambda_2 \tilde{L}_2 S_2, \tag{20}$$

$$\tilde{L}_j = L_0^{-1} L_j = \frac{1}{2} \begin{pmatrix} 1 + \chi_j & 1 - \chi_j \\ 1 - \chi_j & 1 + \chi_j \end{pmatrix},$$

$$\chi_j = \frac{\kappa_j}{\kappa} = \frac{k_j}{k}, \quad j = 1, 2,$$

whose determinant is

$$\det(M_{\text{eff}}) = \det(M) = \frac{\tilde{\delta}}{\delta},$$

$$\delta = (\lambda_1 d_1 - \lambda_2 d_2)(\lambda_1 d_1 k_1 s_1 - \lambda_2 d_2 k_2 s_2) \tag{21}$$

$$+ \tilde{h}(\lambda_1 d_1 k_1^2 - \lambda_2 d_2 k_2^2)(\lambda_1 d_1 c_1 - \lambda_2 d_2 c_2),$$

where $s_j = \sin k_j a - h k_j \cos k_j a$, $c_j = \cos k_j a + h k_j \sin k_j a$, $j = 1, 2$; $\tilde{\delta}$ is obtained from δ by the substitution $h \rightleftarrows -\tilde{h}$. The transfer matrix is unimodular if $\tilde{h} = -h$ (specifically, for complete pinning of spins $h = \tilde{h} = 0$) and also with completely free spins, when $h \longrightarrow \infty$ and $\tilde{h} \longrightarrow \infty$ simultaneously.

4. SPECTRUM OF TRANSMISSION AND REFLECTION AMPLITUDES

In the absence of coupling of u and m waves, i.e., if $B \longrightarrow 0$, the frequencies ω_0 and wave vectors k_0 of a magnetoacoustic resonance are determined by the equation $k_0^2 = k_{10}^2 = k_{20}^2$ and are equal to

$$(\omega_0)_{1,2} = \bar{\omega}(1 \mp \sqrt{1 - 2\omega/\bar{\omega}}), \qquad (k_0)_{1,2} = (\omega_0)_{1,2}/v_t, \tag{22}$$

where $\bar{\omega} = v_t^2/2\eta$. Near the lower resonance point $(k_0, \omega_0)_1$ the group velocity v_m of spin waves is ordinarily low compared with the sound velocity: $v_m/v_t \simeq 2\eta\omega_H/v_t^2 \ll 1$. Near the upper point $(k_0, \omega_0)_2$ they are of the same order of magnitude, $v_m/v_t \simeq 2\eta(k_0)_2/v_t \leq 2$. Between the two resonance frequencies $(\omega_0)_{1,2}$ a sound wave corresponds, according to Eq. (11), to the upper branch of the roots $k_2 = k_{20}(\lambda_1 \longrightarrow 0, \lambda_2 \longrightarrow \infty)$; below the bottom and above the upper resonance frequencies the sound wave corresponds to the lower branch of the roots $k_1 = k_{20}(\lambda_1 \longrightarrow \infty, \lambda_2 \longrightarrow 0)$, and Eqs. (20) and (6) give well-known formulas for the transmission and reflection amplitudes of a u wave with wave number k for a plate

$$t = \tau_j, \quad r = i\rho_j \tau_j, \quad j = 1, 2,$$

$$\tau_j^{-1} = \cos k_j a - i\frac{1 + \chi_j^2}{2\chi_j} \sin k_j a, \tag{23}$$

$$\rho_j = \frac{1 - \chi_j^2}{2\chi_j} \sin k_j a,$$

i.e., there is no reflection, if $k = k_{20}$ or $k_{20}a = n\pi$ (n is an integer, half-wave plate).

In the presence of appreciable coupling between u and m waves and for arbitrary values of the pinning parameters h and $\tilde{h}$ Eqs. (20) and (6) can be used to obtain the following formulas, which generalize Eq. (23), for t and r:

$$t = \delta\tau, \quad r = i\rho\tau. \tag{24}$$

Without writing out the general, complicated expressions for the factors τ and ρ, here we shall consider only

the important limiting cases (slightly overdetermining δ in Eq. (21) on account of obvious cancellations):

(i) $h = \tilde{h} = 0$ (long-wavelength Kittel limit [1, 10] of rigid pinning of spins on both boundaries):

$$\delta = (\lambda_1\chi_1 - \lambda_2\chi_2)(\lambda_1 \sin k_2 a - \lambda_2 \sin k_1 a),$$

$$\tau^{-1} = \lambda_1^2\chi_1\tau_1^{-1}\sin k_2 a + \lambda_2^2\chi_2\tau_2^{-1}\sin k_1 a - \lambda_1\lambda_2\Phi_0,$$

$$\rho = -\lambda_1^2\chi_1\rho_1\sin k_2 a - \lambda_2^2\chi_2\rho_2\sin k_1 a + \lambda_1\lambda_2 F_0, \quad (25)$$

$$\Phi_0 = 1 - \alpha_1^*\alpha_2^*, \quad F_0 = 1 - \frac{1}{2}(\alpha_1^*\alpha_2 + \alpha_1\alpha_2^*),$$

$$\alpha_j = \cos k_j a + i\chi_j \sin k_j a, \quad j = 1, 2;$$

(ii) $h = \tilde{h} = \infty$ (short-wavelength Ament–Rado limit [1, 11], no pinning of spins on both surfaces)

$$\delta = (\lambda_1\chi_1^{-1} - \lambda_2\chi_2^{-1})(\lambda_1 \sin k_2 a - \lambda_2 \sin k_1 a),$$

$$\tau^{-1} = \lambda_1^2\chi_1^{-1}\tau_1^{-1}\sin k_2 a + \lambda_2^2\chi_2^{-1}\tau_2^{-1} + \sin k_1 a + i\lambda_1\lambda_2\Phi_\infty,$$

$$\rho = -\lambda_1^2\chi_1^{-1}\rho_1\sin k_2 a - \lambda_2^2\chi_2^{-1}\rho_2\sin k_1 a - \lambda_1\lambda_2 F_\infty, \quad (26)$$

$$\Phi_\infty = 1 - \beta_1^*\beta_2^*, \quad F_\infty = 1 - \frac{1}{2}(\beta_1\beta_2^* + \beta_1^*\beta_2),$$

$$\beta_j = \cos k_j a + i\chi_j^{-1}\sin k_j a, \quad j = 1, 2;$$

(iii) $h = 0$, $\tilde{h} = \infty$ (on the left-hand boundary, on the side of incidence of the u wave, rigid pinning; on the right-hand boundary—total freedom of the m field):

$$\delta = (\lambda_1\chi_2 - \lambda_2\chi_1)(\lambda_1\chi_1 \cos k_2 a - \lambda_2\chi_2 \cos k_1 a),$$

$$\tau^{-1} = \chi_1\chi_2(\lambda_1^2\tau_1^{-1}\cos k_2 a + \lambda_2^2\tau_2^{-1}\cos k_1 a) - \frac{1}{2}\lambda_1\lambda_2\Phi,$$

$$\rho = -\chi_1\chi_2(\lambda_1^2\rho_1\cos k_2 a + \lambda_2^2\rho_2\cos k_1 a) + \frac{i}{2}\lambda_1\lambda_2 F, \quad (27)$$

$$\Phi = \chi_1^2(1 + \beta_1^*\alpha_2^*) + \chi_2^2(1 + \alpha_1^*\beta_2^*),$$

$$F = \chi_1^2(1 - \beta_1\alpha_2^*) + \chi_2^2(1 - \alpha_1^*\beta_2);$$

(iv) $h = \infty$, $\tilde{h} = 0$ (on the left-hand boundary, on the side of incidence of the u wave, complete freedom, on the right-hand boundary—rigid pinning of the m field); in this case τ^{-1} is given by the same expression and ρ is given by the expression which is the complex-conjugate of Eq. (27), and

$$\delta = (\lambda_1\chi_1 - \lambda_2\chi_2) \times (\lambda_1\chi_2 \cos k_2 a - \lambda_2\chi_1 \cos k_1 a). \quad (28)$$

This difference in δ between Eqs. (27) and (28) is explained by comparing the left-hand (6) and right-hand (6a) amplitudes.

5. TOTAL REFLECTION

The transmission coefficient T vanishes in accordance with Eq. (6) when $(M_{\text{eff}})_{11}$ possesses poles. The existence of poles in the matrix elements of M_{eff} is due to the vanishing of the determinant Γ_s and $L_0 = -2\kappa$ in the denominators of the elements of the inverse matrices in Eq. (20). It is evident from the expressions (25)–(28) that in the absence of wave damping a series of such poles (zeros of the parameter δ) exists on the real axis ω.

In the absence of magnetoelastic coupling, $B = 0$, the condition $\delta = 0$ gives in addition to Eq. (23) the well-known equation for the spin-wave resonance spectrum of a plate [2]

$$\tan k_{10} a = \frac{k(h - \tilde{h})}{1 + k^2 h\tilde{h}}. \quad (29)$$

For $B \neq 0$, neglecting damping, the condition $\delta = 0$ gives a discrete spectrum of resonance frequencies of total magnetoelastic reflection. This spectrum is, evidently, different from the discrete magnetoelastic resonance spectrum in a plate for boundary conditions with total or partial pinning of the elastic field u on the surfaces. For example, in the case (i) $h = \tilde{h} = 0$ of rigid pinning of the field m it follows from Eq. (25) that total reflection occurs for

$$(\lambda_1 k_1 - \lambda_2 k_2)(\lambda_1 \sin k_2 a - \lambda_2 \sin k_1 a)\tau = 0, \quad (30)$$

and the spectrum of standing magnetoelastic waves under the condition of rigid clamping of the boundaries of the plate $u(0) = u(a) = 0$ is given by the equation $M_{12} = 0$, i.e.,

$$[(\lambda_1^2 + \lambda_2^2)\sin k_1 a \sin k_2 a - 2\lambda_1\lambda_2(1 - \cos k_1 a \cos k_2 a)]\delta^{-1} = 0. \quad (31)$$

The other pinning conditions change the form but not the structure of these equations.

The resonance lines of total magnetoelastic reflection are very narrow, and because of damping they are localized in a narrow range near the magnetoelastic resonance frequencies ω_0 [roots (22)]. The width of the region of strong magnetoelastic reflection is determined by the ω width of a single maximum near ω_0 of

the parameter l of the relative coherent coupling of the branches

$$l = \begin{cases} \varepsilon, & \varepsilon < 1 \\ \varepsilon^{-1}, & \varepsilon > 1, \end{cases} \qquad \varepsilon = \left|\frac{l_1}{l_2}\right| = \frac{k_1 k_2 b^2}{(k_{10}^2 - k_1^2)(k_2^2 - k_{20}^2)}. \tag{32}$$

Near magnetoacoustic resonance points $(k_0, \omega_0)_{1,2}$ the parameter $l \sim 1$ for $|(k_1 - k_0)(k_2 - k_0)| \leq b^2$, and away from these points for $|(k_1 - k_0)(k_2 - k_0)| \gg b^2$ (i.e., $|k_{10} - k_{20}| \gg b$) we have

$$l = \frac{k_{10} k_{20} b^2}{(k_{10}^2 - k_{20}^2)^2}. \tag{33}$$

This means that for $k_{10} \sim k_{20} \gg b$

$$l \simeq \left(\frac{\Delta\omega}{\omega - \omega_0}\right)^2, \quad \Delta\omega = \frac{1}{2}\gamma b, \qquad \gamma^{-1} = \left|v_t^{-1} - v_m^{-1}\right|, \tag{34}$$

where $\Delta\omega$ is the characteristic width of the curve $l(\omega)$, estimated from the conditions $|k_1 - k_{10}| \sim |k_2 - k_{20}| \sim |k_{10} - k_{20}| \sim b/2$. This is the width of the region near ω_0 where the magnetoelastic coupling is strong enough so that the corresponding terms in δ, τ, and ρ could compete and the resonances of the reflection maxima could appear. It is obvious that near the bottom resonance $\gamma \simeq v_m \ll v_t$. Away from the bottom resonance point, toward the left, where $k_{10} \ll k_{20}$ and $\omega_H \leq \omega < \omega_0$, we have $l \simeq k_{10} b^2 / k_{20}^3 = \alpha_1 b^2 \sqrt{\omega - \omega_H} \ll 1$ ($\alpha_1 = (v_t/\omega_H)^3/\sqrt{\eta}$) and to the right, where $k_{10} \gg k_{20}$ and $\omega_H \leq \omega_0 \ll \omega$, we have $l \simeq k_{20} b^2 / k_{10}^3 = \beta_1 b^2/\sqrt{\omega} \ll 1$ ($\beta_1 = \eta^{3/2}/v_t$). Away from the upper resonance point, to the left, where $k_{10} \simeq k_{20}$ and $\omega_H \ll \omega \leq \omega_0$, we have the estimate (34) and to the right, where $k_{20} \gg k_{10}$ and $\omega_H \ll \omega_0 \ll \omega$, we have $l \simeq k_{10} b^2 / k_{20}^3 = \beta_2 b^2/\omega^{5/2}$ ($\beta_2 = v_t^3/\sqrt{\eta}$). The qualitative curves $l(\omega)$, $\varepsilon(\omega)$, and $\varepsilon^{-1}(\omega)$, near the bottom magnetoacoustic resonance point are shown in Fig. 1b.

The parameter l also determines the form of the contour of individual spectral lines of reflection maxima, since it appears in the coefficients in front of the trigonometric functions in the expressions for T and R. Near the upper magnetoacoustic resonance point these functions have close periods of oscillations with respect to frequency. Moreover, the upper resonance point can be absent completely because of the downward bend of the curve of the spectrum of short-wavelength magnons in a crystal. The bottom magnetoacoustic resonance point can be easily observed, so that further estimates and calculations will be presented only for it.

It follows even from Eq. (30) that if the unlikely cases where k_1 and k_2 are comparable and also $k_1 l_2 = k_2 l_1$ are ignored, then because of the substantial difference in the slopes of the branches $\omega = \omega(k_1)$ and $\omega = \omega(k_2)$ small periods $\Delta\omega_{10} \simeq \pi v_m/a$ (from $k_{10} a = n\pi$) of total magnetoelastic reflection should be observed against the background of large periods $\Delta\omega_{20} \simeq \pi v_t/a$ (from $k_{20} a = n\pi$, n is an integer) of the resonance acoustic transparency near the bottom magnetoacoustic resonance point ω_0 in the frequency band $\Delta\omega$. Two of the most interesting regimes can be singled out, depending on the thickness a of the film.

(i) Thin film, $\Delta\omega \leq \Delta\omega_{10} \ll \Delta\omega_{20}$. In the region $\Delta\omega$ the contour of the spectrum R is determined by individual lines of reflection maxima. The width of these maxima in terms of the wave number is $\delta k_{10} \simeq \pi/a$ near ω_0 (where $l \sim 1$) and decreases sharply as $\delta k_{10} \simeq l(\pi/a)$ away from ω_0 (where $l \ll 1$); the frequency width of the line is $\delta\omega \sim v_m \delta k_{10} \sim v_m l/a$, i.e., according to Eq. (34) it is proportional to the magnetoelastic coupling constant B^2 and inversely proportional to $(\omega - \omega_0)^2$.

Figure 2a shows the computational results for the spectrum of the reflection coefficient R of transverse elastic waves for a thin film in the case of rigid pinning of spins at the boundaries of the film $h = \tilde{h} = 0$. The thickness of the film $a = 2.295 \times 10^{-4}$ cm; the remaining parameters of the system are characteristic for yttrium iron garnet [1] and are: $B = 6.96 \times 10^6$ ergs/cm^3, $H_0 = 870$ Oe, $M_0 = 140$ G, $v_t = 3.873 \times 10^5$ cm/s, $\eta = 0.1$ cm^2/s, i.e., $\omega_0 = 1.546 \times 10^{10}$ s^{-1}, $k_0 = 4.021 \times 10^4$ cm^{-1}, and $k_0 a = 9.227$.

The possibility of resolving spectroscopically the resonance magnetoelastic reflection lines in an experiment is evidently determined by the condition that their spectral width be greater than the characteristic smearing parameter associated with the dissipative damping of magnetoelastic waves. The damping constant of elastic waves is comparatively small, and the damping of spin waves depends strongly on frequency and temperature [1, 12]. Without going into the details of the physical mechanisms leading to the dissipation of wave energy [1], we shall take this into account, as usual, phenomenologically, adding to the quasispin wave and quasisound wave numbers small positive imaginary parts $k_{10} \longrightarrow k_{10} + i\gamma_1$ and $k_{20} \longrightarrow k_{20} + i\gamma_2$. This gives damping of the partial waves, a shift of the poles of the matrix elements off the real axis ω, and smearing right up to vanishing of the resonance reflection maxima, if $\delta k_{10} \leq \gamma_1$.

Figure 2b shows the reflection spectrum for the same values of the parameters as in Fig. 2a, but taking into account the dissipative processes for characteristic

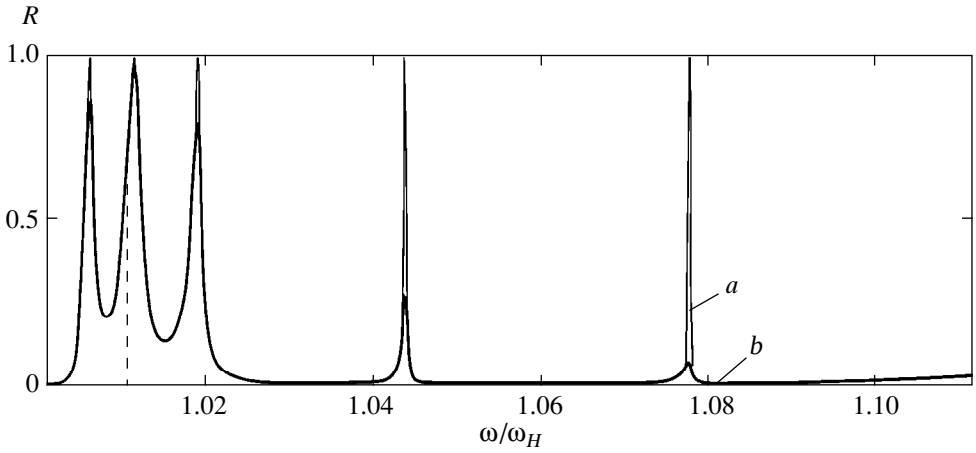


Fig. 2. Spectrum of the reflection coefficient R for a thin film: (*a*) no damping (fine line), (*b*) taking account of damping (heavy line).

values of the damping coefficients for yttrium iron garnet [1]: $\gamma_1 = 100$ cm^{-1} and $\gamma_2 = 2$ cm^{-1}.

(ii) Thick film, $\Delta\omega_{10} \ll \Delta\omega \sim \Delta\omega_{20}$. The region $\Delta\omega$ of magnetoelastic coupling contains very many extremely narrow reflection resonances $\delta k_{10} \sim l\pi/a$, which are strongly smeared by dissipative damping of the spin waves and the instrumental function of the apparatus. However, since $\Delta\omega$ is comparable to the acoustic transmission period $\Delta\omega_{20}$ near the magnetoelastic resonance point, there arises a unique structure of the reflection and transmission spectra which is strongly different from the structure described by Eqs. (23).

Smoothing the dependence $|\delta|^2$, $|\rho|^2$, and $|\tau^{-1}|^2$ on the high-frequency component $\langle\cos^2 k_{10}a\rangle = \langle\sin^2 k_{10}a\rangle = 1/2$ by averaging over a large number of periods of the component (indicated by the brackets), it is easy to obtain expressions for T and R that depend only on the low-frequency component. For example, for rigid pinning of spins $h = \tilde{h} = 0$, replacing the low-frequency wave number $k_{1,2}$ by k_{20}, we obtain

$$T = \frac{(1+\Delta_1)^2(1+\Delta_2\sin^2(k_{20}a))}{1+\Delta_3+\Delta_4\sin(k_{20}a)\cos(k_{20}a)+\Delta_6\sin^2(k_{20}a)},$$
$$\Delta_6 = \Delta_5 + \frac{k_{20}^2 - k^2}{4k_{20}^2k^2}, \tag{35}$$

where Δ_i ($i = 1, \ldots, 5$) are smooth algebraic functions of k_1, k_2, and k, consisting of terms which are proportional to the powers l^n ($n = 1, \ldots, 4$), so that $\Delta_i \longrightarrow 0$ as $l \longrightarrow 0$. It is evident that several spikes of T and R can occur near $\Delta\omega$, where $l \sim 1$. Depending on the values of a and k_0, both a maximum close to 1 and a dip close to 0 can appear near the center of the spectral line.

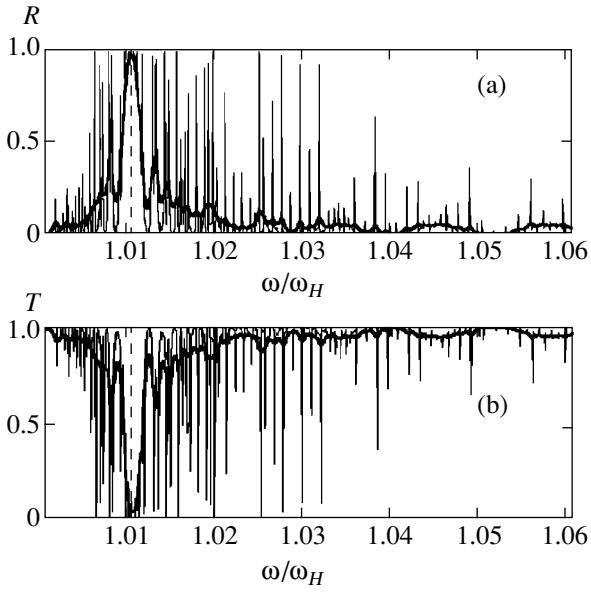


Fig. 3. a) Spectrum of the reflection coefficient R for a thick film with no damping (fine line), taking account of smoothing of the high-frequency component (heavy line); b) the same for the transmission coefficient T.

The fine line in Fig. 3a corresponds to the computational results for the spectrum of the reflection coefficient R, and the fine line in Fig. 3b corresponds to the spectrum of the transmission coefficient T for transverse elastic waves without damping ($\gamma_1 = \gamma_2 = 0$) for a thick film ($a = 5.29 \times 10^{-3}$ cm, $k_0a = 212.7$); all other parameters are the same as in the case a. The heavy line shows the smoothing effect of the measuring process. A Gaussian function with spectral width $\Delta\omega_{\text{gaus}} = 1.4 \times 10^7$ s^{-1} was taken as the instrumental function.

Figures 4*a* and 4*b* show the same spectra taking into account the damping $\gamma_1 = 100$ cm^{-1} and $\gamma_2 = 2$ cm^{-1}. It is evident that near the magnetoacoustic resonance ω_0, in the band $\Delta\omega$, reflection is intensified and transmission T, conversely, decreases to very small values. Away from ω_0 all curves transform into a half-wave sinusoid of $2k_{20}a$.

All this refers to right-hand circularly polarized waves. As is well known, the interaction of left-hand circular components of elastic and spin waves in the system under study has virtually no effect on their spectrum [1, 3]; this is ordinarily manifested in the Farady rotation of the polarization plane for linearly polarized elastic wave passing through a plate. However, it follows from our analysis that near total magnetoelastic reflection lines at the exit we should obtain only a right-hand circularly polarized wave.

6. OTHER SYSTEMS

Our approach can be applied with little modification to other quasi-one-dimensional stationary scattering problems with participation of two or more pairs of coupled waves of different physical nature.

The generalization for a more complicated experimental geometry and more complex magnets, specifically, for antiferromagnets, can be done using an internal irreducible transfer matrix with dimension equal to the order of the system of coupled wave equations in a medium [8]. The required conditions for pinning of the spins once again make it possible to obtain a two-dimensional transfer matrix, whose poles will give the lines of the reflection maxima.

It is obvious that this method can be used to solve the system of equations in [5] for nuclear magnetoelastic branches, engendered by monochromatic waves of nuclear ferromagnetic and antiferromagnetic vectors in $KMnF_3$. If the frequencies of the points of intersection of an elastic branch with two magnetic branches are substantially different, as proposed in [5], then near each of them formulas of the type presented above (thick-film regime) can be used with an appropriate interpretation of the parameters (see Eqs. (20)–(32) from [5] and the expressions (7)–(9)).

We note that at the magnetoelastic resonance point $l_1 = l_2$, but none of our formulas (21) and (25)–(28) for $\delta = 0$ with different pinning conditions gives the condition of nontransmission $\cos[(k_1 - k_2)a/2] = 0$ presented in [4, 5]. In addition, the change observed in [5] with increasing magnetic field in the signal transmission time through the sample can be explained by the nonmonochromaticity of this signal, since the duration of the sound pulses formed by the generator in the experiment of [5] is only two orders of magnitude greater than the main period of the resonance oscillations. Usually, when the magnetic field intensity H increases, the characteristic resonance frequencies increase, raising the entire pattern (Fig. 1) together with the nontransmission pole upwards in frequency. This means that as H increases, the wave packet at the exit from the sample at first becomes enriched with high-frequency quasielastic components $\sim k_{20}$ of the branch k_2 with a high group velocity and then, after passing through the region of the resonance nontransmission maximum, it becomes enriched with low-frequency components $\sim k_{10}$ with a low group velocity.

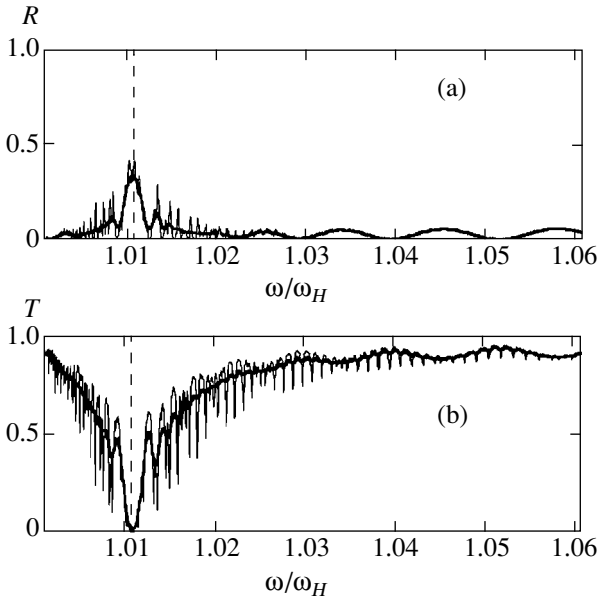


Fig. 4. (a) Same as in Fig. 3a taking account of damping of the waves; (b) same as in Fig. 3b taking account of damping.

The possibility of these effects being manifested in physical systems such as electric polaritons in piezoelectric crystals with interaction of electromagnetic waves with optical branches of elastic oscillations [13] and so on requires a special investigation. Especially tempting is the case of magnetic polaritons for right-hand circularly polarized electromagnetic waves, incident in the direction of the normal on a ferromagnetic plate, similar to that studied in Section 3. The corresponding system of equations, describing the interaction of the magnetic field of an electromagnetic wave **b** and a right-hand spin wave in a plate, is identical to the system (7), where the elastic field u must be replaced by the high-frequency magnetic field **b**, the speed of sound v_t must be replaced by the phase velocity of light v_f in the medium, and the coupling parameters do not depend on the wave vector k and equal $\Delta_1 = 4\pi\omega^2$ and $\Delta_2 = \gamma M_0$ [1, 3], i.e., the magnetostatic coupling in such polaritons is much stronger than in magnetoelastic waves. The characteristic equation has the form

$$(k^2 - k_{10}^2)(k^2 - k_{20}^2) - \tilde{k}^4 = 0,$$
$$\tilde{k}^4 = 4\pi\omega^2\gamma M_0/\eta v_f^2, \tag{36}$$

the coherent coupling coefficients (13) do not change sign for reflected waves, i.e., $l_{3,4} = l_{1,2}$. This means that in the matrices (17) the sign of the elements in the bottom rows must be changed, $f_j \longrightarrow -f_j$ and $g_j \longrightarrow -g_j$.

Ultimately, we obtain from Eqs. (18)–(21) for the parameter δ in Eq. (21) the expression

$$\delta = (\lambda_1 d_1 k_1 - \lambda_2 d_2 k_2) \times [(\lambda_1 d_1 s_1 - \lambda_2 d_2 s_2) + \tilde{h}(\lambda_1 d_1 k_1 c_1 - \lambda_2 d_2 k_2 c_2)], \quad (37)$$

which, just as in Eqs. (25)–(28), gives for different limiting cases of surface pinning of the spins

$$\begin{aligned} &\text{(i)}\ \delta = (\lambda_1 - \lambda_2)(\lambda_1 k_1 \sin k_2 a - \lambda_2 k_2 \sin k_1 a), \quad h - \tilde{h} = 0, \\ &\text{(ii)}\ \delta = (\lambda_1 - \lambda_2)(\lambda_1 k_2 \sin k_2 a - \lambda_2 k_1 \sin k_1 a), \quad h = \tilde{h} = \infty, \\ &\text{(iii, iv)}\ \delta = (\lambda_1 - \lambda_2)(\lambda_1 k_2 \cos k_2 a - \lambda_2 k_1 \cos k_1 a), \quad h = 0, \quad \tilde{h} = \infty. \end{aligned} \quad (38)$$

However, in contrast to Eq. (33) the parameter increasing monotonically with ω

$$\varepsilon = \left|\frac{l_1}{l_2}\right| = \frac{\tilde{k}^4}{|k_2^2 - k_{20}^2|} \quad (39)$$

is different from zero at the threshold of propagation of waves of the upper branch $k_2 = 0$ and $\omega_n = \omega_H + \omega_M$, where $\omega_M = 4\pi\gamma M_0$, and assumes the value $\varepsilon_{\min} = \omega_M/\eta k_H^2$ and $k_H = \omega_H/v_f$. Unfortunately, ordinarily $\varepsilon_{\min} \gg 1$, and total reflection is noticeable for $\varepsilon \leq 1$. To reach this region the phase velocity v_f of light in the medium can be decreased using retardation systems, the ratio M_0/H_0 can be decreased, or a material can be chosen with strong spatial dispersion of magnons with a large value of η. In any case, the effect should be more noticeable for films with a resonance thickness a on the branch k_1, i.e., for (i) and (ii) near $\sin k_1 a = 0$ and for (iii) and (iv) near $\cos k_1 a = 0$.

7. CONCLUSIONS

We have derived formulas for the coefficients of reflection and transmission for right-hand polarized monochromatic elastic plane waves, incident in a normal direction on a transversely magnetized uniaxial ferromagnetic film with uniform boundary conditions for surface pinning of the spins. A nonunimodular effective transfer matrix relating the elastic field in the medium on both sides of the plate was obtained, and the zeros and poles of this transfer matrix in the absence of damping of the waves were studied. It was shown that in addition to the half-wave maxima of the transmission coefficient of the film for elastic waves, a series of lines with reflection coefficient maxima for hypersound, whose position is close to the half-wave conditions for spin waves, can appear, depending on the magnetoelastic coupling constant, near the magnetoelastic resonance. The width and number of potentially observable lines are determined by the thickness of the film, the magnitude of the magnetoelastic coupling, the damping parameters for partial waves in a plate, and the frequency distance from the magnetoelastic resonance point. A numerical analysis of the results was performed for values of the parameters close to the real values. The possibility of the existence of a similar effect in other polariton systems is discussed.

It would be interesting to study a multilayer system and follow the behavior of the reflection resonances of a single layer with diffraction-grating Bragg resonances superposed on them [7].

REFERENCES

1. A. G. Gurevich and G. A. Melkov, *Magnetic Oscillations and Waves* (Nauka, Moscow, 1994).
2. N. M. Salanskiĭ and M. Sh. Erukhimov, *Physical Properties and Application of Magnetic Films* (Nauka, Novosibirsk, 1975).
3. A. I. Akhiezer, V. G. Bar'yakhtar, and S. V. Peletminskii, *Spin Waves* (Nauka, Moscow, 1967; North-Holland, Amsterdam, 1968).
4. Kh. G. Bogdanova, V. A. Golenishchev-Kutuzov, M. I. Kurkin, *et al.*, Zh. Éksp. Teor. Fiz. **112**, 1830 (1997) [JETP **85**, 1001 (1997)].
5. Kh. G. Bogdanova, V. A. Golenishchev-Kutuzov, M. I. Kurkin, *et al.*, Zh. Éksp. Teor. Fiz. **115**, 1727 (1999) [JETP **88**, 947 (1999)].
6. M. Born and E. Wolf, *Principles of Optics* (Pergamon, Oxford, 1969; Nauka, Moscow 1973).
7. V. I. Al'shits and A. L. Shuvalov, Zh. Éksp. Teor. Fiz. **103**, 1356 (1993) [JETP **76**, 663 (1993)].
8. Yu. G. Peisakhovich, J. Phys. A **29**, 5103 (1996); **32**, 3133 (1999).
9. Yu. G. Peĭsakhovich and A. A. Shtygashev, Fiz. Nizk. Temp. **25**, 455 (1999) [Low Temp. Phys. **25**, 332 (1999)].
10. C. Kittel, Phys. Rev. **110**, 1295 (1958).
11. W. S. Ament and G. T. Rado, Phys. Rev. **97**, 1558 (1955).
12. M. I. Kaganov, N. B. Pustyl'nik, and T. I. Shalaeva, Usp. Fiz. Nauk **167**, 191 (1997) [Phys. Usp. **40**, 181 (1997)].
13. M. K. Balakirev and I. A. Gilinskiĭ, *Waves in Piezoelectric Crystals* (Nauka, Novosibirsk, 1982).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 1, 2000, pp. 197–205.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 1, 2000, pp. 223–231.

SOLIDS
Electronic Properties

Quasiballistic Electron Transport in a Three-Dimensional Microconstriction

N. G. Galkin, V. A. Geyler, and V. A. Margulis*

Mordovian State University, Saransk, 430000 Russia
**e-mail: margulis@mrsu.ru; theorphysics@mrsu.ru*

Received February 21, 2000

Abstract—Quasiballistic electron transport in three-dimensional microconstrictions in a longitudinal magnetic field is studied. The case where the quantization of the conductance is destroyed by an impurity located in the microconstriction is investigated. The transmission coefficients and a formula for the conductance of such a microconstriction are found. It is shown that the dependence of the conductance $G(E)$ on the electron energy has kinks at the threshold of the steps for which impurity scattering is strong. These kinks contain a pronounced vertical segment of the curve $G(E)$. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Electron transport in nanostructures of various dimensions, such as two-dimensional quantum channels and constrictions as well as three-dimensional quantum wires and microconstrictions, is an object of intense theoretical [1–9] and experimental [10–12] study. This is due to the discovery of the quantization of conductance observed in such nanostructures in the ballistic transport regime. In ballistic electron transport the electron mean-free path is greater than the characteristic dimensions of the nanostructure, so that scattering of propagating electronic modes is due only to the geometry of the samples (specifically, their length and the shape of the transverse cross section). In such a transport regime a characteristic step dependence of the conductance G on the electron energy E arises in the propagating modes, i.e., the presence of a plateau between the thresholds of the conductance quantization steps. The temperature of the sample also plays an important role in transport, since it changes the form of the curve $G(E)$, specifically, it smoothes the conductance quantization steps, decreasing the plateau region [4, 9].

We note that a magnetic field applied to the nanostructure influences the transport regime. This is due to the fact that a magnetic field can intensify electronic confinement in a nanostructure and lead to hybridization of the dimensional and magnetic confinement as a result of coupling of the motion of electrons parallel and perpendicular to the magnetic field $\mathbf{B}$ [4, 9]. In addition, in a strong magnetic field the dependence $G(\mathbf{B})$ is oscillatory (Shubnikov–de Haas and Aharonov–Bohm oscillations) [4, 9].

Impurities present in the propagation path of electronic modes destroy ballistic transport regimes [13]. Even a single impurity present in a two-dimensional channel [14–17] or in a three-dimensional quantum wire [13, 18] changes the transport regime from ballistic to quasiballistic. In the process conductance quantization is restructured because of the appearance of peaks near the threshold of the steps due to reflection of electronic modes from impurities [13]. When two impurities are present in a wire an entire series of peaks appears near the thresholds; several impurities have an even larger effect because of the complicated character of the reflection and transmission of electronic modes. Similar effects have also been observed in two-dimensional constrictions [14–17].

The confinement potential, characterizing the extent of the constriction, near a saddle point can be simulated by an inverted harmonic oscillator potential [20]. Quantum transport in two-dimensional microconstrictions was investigated in [19, 20], and a relation was obtained between the form of the conductance quantization steps and the geometry of the constriction. The theoretical and experimental result, obtained prior to 1991, on quantum transport in two-dimensional microconstrictions are contained in the review [21]. In a three-dimensional microconstriction the characteristic frequency of an inverted oscillator, as shown in [9], plays the same role as temperature, i.e., it smoothes the thresholds of the steps, increasing their slope and decreasing the size of the plateau.

Our objective in the present work is to investigate the quasiballistic regime of a three-dimensional quantum constriction in a longitudinal magnetic field in the presence of a single impurity in the microconstriction. We note that the conductance of a three-dimensional constriction in the ballistic transport regime has been studied in [4, 9].

1063-7761/00/9101-0197$20.00

2. ELECTRONIC STATES AND THE GREEN'S FUNCTION OF THE UNPERTURBED HAMILTONIAN

Let us consider the state of an electron in a quantum three-dimensional constriction which is not perturbed by an impurity potential. We shall simulate the potential of the transverse section of such a constriction by a two-dimensional harmonic oscillator potential [4, 9]. Then the Hamiltonian of the single-electron spinless states in a longitudinal magnetic field $B = B_z$ is

$$H^0 = \frac{1}{2m^*}\left(\mathbf{p} - \frac{e}{c}\mathbf{A}\right)^2 + \frac{m^*}{2}(\omega_x^2 x^2 + \omega_y^2 y^2 - \omega_z^2 z^2), \quad (1)$$

where $\mathbf{A} = (Bx, -By, 0)/2$ and ω_j are the frequencies of the harmonic potential. We introduce the hybrid frequencies according to the formula

$$\omega_{1,2} = \frac{1}{2}[\sqrt{(\omega_x + \omega_y)^2 + \omega_c^2} \pm \sqrt{(\omega_x - \omega_y)^2 + \omega_c^2}], \quad (2)$$

and then the spectrum of the Hamiltonian (1) has the form [9]

$$E^{\pm}_{n_1, n_2, \lambda} = \hbar\omega_1\left(n_1 + \frac{1}{2}\right) + \hbar\omega_2\left(n_2 + \frac{1}{2}\right) + \hbar\omega_z\lambda, \quad (3)$$

where $n_1, n_2 \in \mathbb{N}$, and $\lambda \in \mathbb{R}$. The eigenvalues of H^0 correspond to the wave functions $\Psi^{0\pm}_{n_1, n_2, \lambda}(x, y, z) = \Psi_{n_1, n_2}(x, y)\Psi^{\pm}_{\lambda}(z)$. Here Ψ_{n_1, n_2} is the eigenfunction of the operator

$$H^0_{x,y} = \frac{1}{2m^*}\left(\mathbf{p} - \frac{e}{c}\mathbf{A}\right)^2 + \frac{m^*}{2}(\omega_x^2 x^2 + \omega_y^2 y^2),$$

whose eigenvalues are $\hbar\omega_1(n_1 + 1/2) + \hbar\omega_2(n_2 + 1/2)$, and $\Psi^{\pm}_{\lambda}(z)$ are the eigenfunctions of the operator

$$H^0_z = -\frac{\hbar^2}{2m^*}\frac{\partial^2}{\partial z^2} - \frac{m^*}{2}\omega_z^2 z^2,$$

which correspond to the point of the continuous spectrum $\hbar\omega_z\lambda$. The function $\Psi^{\pm}_{\lambda}(z)$ has the form

$$\Psi^{\pm}_{\lambda}(z) = \left(\frac{2m^*}{\hbar^3\omega_z}\right)^{1/4} \times [2\pi(1 + e^{-2\pi\lambda})]^{-1/2}\mathbf{E}\left(-\lambda, \pm\sqrt{\frac{2m^*\omega_z}{\hbar}}z\right), \quad (4)$$

where $\mathbf{E}(\lambda, x)$ is a modified parabolic cylinder function [22].

The function $\Psi_{n_1, n_2}(x, y)$ has a simple form when the cross section of the constriction is circular ($\omega_x = \omega_y = \omega_0$). In this important case it is convenient to switch to a different parameterization of the spectrum. Specifically, let $\Omega = \sqrt{4\omega_0^2 + \omega_c^2}$ and then the spectrum H^0_{xy} consists of the eigenvalues

$$E_{mn} = \frac{\hbar\omega_c}{2}m + \frac{\hbar\Omega}{2}(2n + |m| + 1),$$

where $m \in \mathbb{Z}$ and $n \in \mathbb{N}$. It is convenient to write the eigenfunction corresponding to E_{mn} in polar coordinates [18]

$$\Psi_{mn}(\rho, \varphi) = \frac{\exp(im\varphi)}{\sqrt{2\pi}}R_{mn}(\rho), \quad (5)$$

where

$$R_{mn}(\rho) = c_{mn}|\rho|^m \exp\left(-\frac{\rho^2}{4l_\Omega^2}\right)L_n^{|m|}\left(\frac{\rho^2}{2l_\Omega^2}\right), \quad (6)$$

$l_\Omega = \sqrt{\hbar/m^*\Omega}$, and $L_n^{|m|}(x)$ are generalized Laguerre polynomials.

The normalization factor is

$$c_{mn} = \frac{1}{l_\Omega^{|m|+1}}\left(\frac{n!}{2^{|m|}(n + |m|)!}\right)^{1/2}. \quad (7)$$

To find the Green's function of the Hamiltonian H^0 we shall write first an expression for the kernel of the propagator of the operator $\exp(-itH_0/\hbar)$. Using the results of [23, 24], after some transformations we obtain for the kernel of the propagator $K_0(\mathbf{r}, \mathbf{r}'; t)$ the expression

$$K_0(\mathbf{r}, \mathbf{r}'; t) = \left(\frac{m^*}{2\pi i\hbar}\right)^{3/2}\frac{\Omega_1\Omega_2(\omega_x\omega_y\omega_z)^{1/2}}{(p^2(t) - q^2(t))^{1/2}\sinh\omega_z t}$$

$$\times \exp\left\{\left(-\frac{m^*}{2i\hbar}\right)[M_1(t)(x^2 + x'^2) + M_2(t)(y^2 + y'^2) + M_3(t)(z^2 + z'^2) + M_4(t)xx' + M_5(t)yy' + M_6(t)zz' + M_7(t)(xy - x'y') + M_8(t)(xy' - yx')]\right\}. \quad (8)$$

Here

$$p(t) = \Omega_2(\omega_x + \omega_y)\sin\frac{\Omega_1 t}{2},$$

$$q(t) = \Omega_1(\omega_y - \omega_x)\sin\frac{\Omega_2 t}{2},$$

$$F_{1,2}(t) = \left[p(t)\cos\frac{\Omega_1 t}{2} + q(t)\cos\frac{\Omega_2 t}{2}\right],$$

$$M_1(t) = \frac{\Omega_1\Omega_2\omega_x F_{1,2}(t)}{p^2(t)-q^2(t)},$$

$$M_2(t) = \frac{\Omega_1\Omega_2\omega_y F_{1,2}(t)}{p^2(t)-q^2(t)}, \quad M_3(t) = \omega_z\cosh\omega_z t,$$

$$M_4(t) = -\frac{2\Omega_1\Omega_2\omega_x F_{2,1}(t)}{p^2(t)-q^2(t)},$$

$$M_5(t) = -\frac{2\Omega_1\Omega_2\omega_y F_{2,1}(t)}{p^2(t)-q^2(t)}, \quad M_6(t) = -\frac{2\omega_z}{\sinh\omega_z t}, \tag{9}$$

$$M_7(t) = \frac{\omega_c}{p^2(t)-q^2(t)}\left[\frac{\omega_y-\omega_x}{\omega_x+\omega_y}p^2(y) - \frac{\omega_x+\omega_y}{\omega_y-\omega_x}q^2(t)\right],$$

$$M_8(t) = \frac{\omega_c\omega_x\omega_y p(t)q(t)}{[p^2(t)-q^2(t)](\omega_x^2-\omega_y^2)},$$

where $\Omega_1 = \sqrt{(\omega_x+\omega_y)^2+\omega_c^2}$, $\Omega_2 = \sqrt{(\omega_x-\omega_y)^2+\omega_c^2}$, and $F_{2,1}(t)$ is obtained from $F_{1,2}(t)$ by interchanging the indices.

For a circular cross section ($w_x = w_y$) the propagator (8) becomes

$$K_0(\mathbf{r},\mathbf{r}';t) = \left(\frac{m^*}{2\pi i\hbar}\right)^{3/2}\frac{\Omega}{2\sin(\Omega t/2)}\left(\frac{\omega_z}{\sinh\omega_z t}\right)^{1/2}$$
$$\times\exp\left\{-\frac{m^*}{2i\hbar}\left[\frac{\Omega}{\sin\Omega t/2}\left((x'y-xy')\sin\frac{\omega_c t}{2}\right.\right.\right.$$
$$\left.-(xx'+yy')\cos\frac{\omega_c t}{2}+\frac{1}{2}(p^2+p'^2)\cos\frac{\Omega t}{2}\right)$$
$$\left.\left.+\frac{\omega_z}{\sinh\omega_z t}((z^2+z'^2)\cosh\omega_z t-2zz')\right]\right\}. \tag{10}$$

The resolvent $(H^0-E)^{-1}$ can be written in terms of the propagator as

$$(H^0-E)^{-1} = \frac{i}{\hbar}$$
$$\times\lim_{\varepsilon\to 0}\int_0^\infty\exp\{-it[H^0-(E+i\varepsilon)]/\hbar\}dt. \tag{11}$$

As follows from Eq. (11), the kernel of the resolvent, i.e., the Green's function $G^0(\mathbf{r},\mathbf{r}';E)$, is given by the formula

$$G^0(\mathbf{r},\mathbf{r}';E) = \frac{i}{\hbar}$$
$$\times\lim_{\varepsilon\to 0}\int_0^\infty K^0(\mathbf{r},\mathbf{r}';t)\exp\{i(E+i\varepsilon)t/\hbar\}dt. \tag{12}$$

For a constriction with a circular cross section the following expression can be obtained for $G^0(\mathbf{r},\mathbf{r}';E)$ in the form of a series for $\mathrm{Im}\,\zeta > 0$:

$$G^0(\mathbf{r},\mathbf{r}';\zeta) = \frac{e^{i\pi/4}}{2\pi\hbar}\left(\frac{m^*}{\pi\hbar\omega_z}\right)^{1/2}$$
$$\times\sum_{m=-\infty}^{\infty}\sum_{n=0}^{\infty}\exp[im(\varphi-\varphi')]R_{mn}(\rho)R_{mn}(\rho')$$
$$\times\Gamma\left(\frac{1}{2}-i\frac{\zeta-E_{mn}}{\hbar\omega_z}\right)$$
$$\times U\left(-i\frac{\zeta-E_{mn}}{\hbar\omega_z}, e^{-i\pi/4}\left(\frac{2m^*\omega_z}{\hbar}\right)^{1/2}\max(z,z')\right)$$
$$\times U\left(-i\frac{\zeta-E_{mn}}{\hbar\omega_z}, e^{-i\pi/4}\left(\frac{2m^*\omega_z}{\hbar}\right)^{1/2}\min(z,z')\right). \tag{13}$$

Here $U(a,x)$ is a parabolic cylinder function [22].

For $\mathrm{Im}\,\zeta < 0$ the Green's function is the complex-conjugate of the expression (13), $G^0(\mathbf{r},\mathbf{r}';\zeta) = \overline{G^0(\mathbf{r}',\mathbf{r};\zeta)}$. On the real axis

$$G^0_+(\mathbf{r},\mathbf{r}';E) = \lim_{\varepsilon\to+0}G^0(\mathbf{r},\mathbf{r}';E+i\varepsilon), \quad E\in\mathbb{R}.$$

From Eq. (13) we obtain

$$G^0_+(\mathbf{r},\mathbf{r}';E) = \frac{1}{4\pi\hbar}\left(\frac{2m^*}{\hbar\omega_z}\right)^{1/2}\sum_{m=-\infty}^{\infty}\sum_{n=0}^{\infty}\exp[im(\varphi-\varphi')]$$
$$+\frac{R_{mn}(\rho)R_{mn}(\rho)}{\{1+\exp[2\pi(E_{mn}-E)/\hbar\omega_z]\}^{1/2}}$$
$$\times\mathbf{E}\left(\frac{E_{mn}-E}{\hbar\omega_z}, \left(\frac{2m^*\omega_z}{\hbar}\right)^{1/2}\max(z,z')\right)$$
$$\times\mathbf{E}\left(\frac{E_{mn}-E}{\hbar\omega_z}, -\left(\frac{2m^*\omega_z}{\hbar}\right)^{1/2}\min(z,z')\right). \tag{14}$$

For $G^0_-(\mathbf{r},\mathbf{r}';E) = \lim_{\varepsilon\to+0}G^0(\mathbf{r},\mathbf{r}';E-i\varepsilon)$ we have the relation $G^0_-(\mathbf{r},\mathbf{r}';E) = \overline{G^0_+(\mathbf{r}',\mathbf{r};E)}$. We shall use the expression (14) in the next section to find the states of an electron scattered by an impurity.

3. GREEN'S FUNCTION AND STATES OF THE HAMILTONIAN WITH AN IMPURITY

Let us consider the Hamiltonian of an electron perturbed by a single impurity located in a constriction. We shall simulate the impurity potential by a point potential (zero-radius potential). The main difference of such a potential from a more realistic one with a finite range, as will be shown below, is that only states with magnetic quantum number $m = 0$ (s states) are scattered by a point

potential. The point perturbation potential $V(\mathbf{r})$ can be written correctly in the form of a pseudopotential [25] as

$$V(\mathbf{r}) = \lambda\delta(\mathbf{r}-\mathbf{a})[1+(\mathbf{r}-\mathbf{a})\nabla], \tag{15}$$

where (a_x, a_y, a_z) are the coordinates of the impurity.

The Green's function of the operator $H^0 + V$ can be found using the formula [26]

$$G(\mathbf{r}, \mathbf{r}'; \zeta) = G^0(\mathbf{r}, \mathbf{r}'; \zeta) - [Q(\zeta, \mathbf{a}) + \lambda]^{-1} G^0(\mathbf{r}, \mathbf{a}; \zeta) G^0(\mathbf{a}, \mathbf{r}'; \zeta), \tag{16}$$

where $Q(\zeta, \mathbf{a})$ is the Crane function, determined from the relation

$$Q(\zeta, \mathbf{a}) = \lim_{\mathbf{r}\to\mathbf{a}}\left[G^0(\mathbf{r}, \mathbf{a}; \zeta) - \frac{m^*}{2\pi\hbar^2}\frac{1}{|\mathbf{r}-\mathbf{a}|}\right]. \tag{17}$$

The limit in Eq. (17) can be calculated conveniently as follows. Let us consider the Green's function $G^1(\mathbf{r}, \mathbf{r}'; \zeta)$ of the operator

$$H^0 = \frac{1}{2m^*}\left(\mathbf{p} - \frac{e}{c}\mathbf{A}\right)^2 + \frac{m^*}{2}(\omega_x^2 x^2 + \omega_y^2 y^2) \tag{18}$$

with $\omega_x = \omega_y \equiv \omega_0$. This function has the form [18]

$$G^1(\mathbf{r}, \mathbf{r}'; \zeta) = \frac{m^*}{2\pi i\hbar}\sum_{m=-\infty}^{\infty}\sum_{n=0}^{\infty}\exp[im(\varphi-\varphi')] \times R_{mn}(\rho)R_{mn}(\rho')\frac{\exp(ip_{mn}|z-z'|/\hbar)}{p_{mn}}, \tag{19}$$

where

$$p_{mn} = \left\{2m^*\left[\zeta - \frac{\hbar\omega_c}{2}m - \frac{\hbar\Omega}{2}(2n+|m|+1)\right]\right\}^{1/2}.$$

Then $Q(\zeta, \mathbf{a}) = Q_1(\zeta, \mathbf{a}) + Q_2(\zeta, \mathbf{a})$, where

$$Q_1(\zeta, \mathbf{a}) = \lim_{\mathbf{r}\to\mathbf{a}}\left[G^1(\mathbf{r}, \mathbf{a}; \zeta) - \frac{m^*}{2\pi\hbar^2}\frac{1}{|\mathbf{r}-\mathbf{a}|}\right], \tag{20}$$

and $Q_2(\zeta, \mathbf{a})$ is determined by the expression

$$Q_2(\zeta, \mathbf{a}) = \lim_{\mathbf{r}\to\mathbf{a}}[G^0(\mathbf{r}, \mathbf{a}; \zeta) - G^1(\mathbf{r}, \mathbf{a}; \zeta)]. \tag{21}$$

The function $Q_1(\zeta, \mathbf{a})$ is used in [18]. To calculate $Q_2(\zeta, \mathbf{a})$ we shall confine our attention to the important case where the impurity lies on the symmetry axis of a nanostructure, i.e., when $a_x = a_y = 0$. It is obvious from Eq. (6) that the wave function of an electron decreases exponentially, $\propto \exp(-\rho^2/4l_\Omega^2)$, away from the symmetry axis. Therefore only an impurity located virtually on the symmetry axis scatters most efficiently. We note that for $m \neq 0$ we have $R_{mn}(0) = 0$, and for $m = 0$

$$R_{0n}(0) = \sqrt{\frac{m^*\Omega}{\hbar}} = \frac{1}{l_\Omega}.$$

An expression for $Q_2(E, a)$ can be obtained from Eqs. (19) and (20) in the form

$$Q_2(E, \mathbf{a}) = \frac{1}{4\pi\hbar\omega_z}\frac{1}{l_\Omega^2 l_z} \times \sum_{n=0}^{\infty}\left\{\frac{\mathbf{E}(\varepsilon_n; a_z/l_z)\mathbf{E}(\varepsilon_n; -a_z/l_z)}{(1+\exp 2\pi\varepsilon_n)^{1/2}} - \varepsilon_n^{-1/2}\right\}, \tag{22}$$

where

$$l_z = \left(\frac{\hbar}{2m^*\omega_z}\right)^{1/2},$$

$$\varepsilon_n = \frac{\Omega}{\omega_z}\left[\left(n+\frac{1}{2}\right) - \frac{E}{\hbar\Omega}\right].$$

The quantity $Q_1(E, \mathbf{a})$ is calculated in [18] and is given by

$$Q_1(E, \mathbf{a}) = \frac{m^*}{2\pi\hbar^2}\sqrt{\frac{m^*\Omega}{2\hbar}}Z\left(\frac{1}{2}, \frac{1}{2} - \frac{E}{\hbar\Omega}\right), \tag{23}$$

where $Z(s, x)$ is the generalized zeta function (the Hurwitz function). Convergence of the series in Eq. (22) can be proved by studying separately the real and imaginary parts of the series. For the real part the general term in the series can be estimated as $\sim Cn^{-5/2}$, and for the imaginary part as $\sim Ce^{-\pi\varepsilon_n}$, where C = const. Therefore the series in Eq. (22) converges. Finally, we obtain for the Q function

$$Q(E, \mathbf{a}) = \frac{1}{4\pi\hbar\omega_z}\frac{1}{l_\Omega^2 l_z}\left\{\left(\frac{\omega_z}{\Omega}\right)^{1/2}Z\left(\frac{1}{2}, \frac{1}{2} - \frac{E}{\hbar\Omega}\right) + \sum_{n=0}^{\infty}\left[\frac{\mathbf{E}(\varepsilon_n, a_z/l_z)\mathbf{E}(\varepsilon_n, -a_z/l_z)}{[1+\exp(2\pi\varepsilon_n)]^{1/2}} - \varepsilon_n^{-1/2}\right]\right\}. \tag{24}$$

In the case $\mathbf{a} = 0$ the real and imaginary parts can be easily separated in the expression (24):

$$Q(E) \equiv Q(E, 0) = \frac{1}{4\pi\hbar\omega_z}\frac{1}{l_\Omega^2 l_z}\left\{\left(\frac{\omega_z}{\Omega}\right)^{1/2}Z\left(\frac{1}{2}, \frac{1}{2} - \frac{E}{\hbar\Omega}\right) + \sum_{n=0}^{\infty}\left[\frac{1}{\sqrt{2}}\left|\frac{\Gamma(1/4 + i\varepsilon_n/2)}{\Gamma(3/4 + i\varepsilon_n/2)}\right| \times [1+\exp(-2\pi\varepsilon_n)]^{-1/2} - \varepsilon_n^{-1/2}\right] + i\sum_{n=0}^{\infty}\frac{1}{\sqrt{2}}\left|\frac{\Gamma(1/4 + i\varepsilon_n/2)}{\Gamma(3/4 + i\varepsilon_n/2)}\right|[1+\exp(2\pi\varepsilon_n)]^{-1/2}\right\}. \tag{25}$$

Next we analyze in detail the case where the point impurity lies at the center of the cross section of the

waist of the constriction ($\mathbf{a} = 0$). As already mentioned above, the role of scattering will be strongest for an impurity located on the symmetry axis.

Let us consider a wave function of H^0 of the form $\Psi^{0\pm}_{mn\lambda}(r, \varphi, z) = \Psi_{mn}(\rho, \varphi)\Psi^{\pm}_{\lambda}(z)$, which corresponds to energy

$$E = \frac{\hbar\omega_c}{2}m + \frac{\hbar\Omega}{2}(2n + |m| + 1) + \hbar\omega_z\lambda.$$

This value of the energy E corresponds to a wave function of the perturbed Hamiltonian H described by the formula

$$\Psi^{\pm}_{mn\lambda}(r, \varphi, z) = \Psi^{0\pm}_{mn\lambda}(r, \varphi, z) + [Q(E) + \alpha]^{-1}\Psi^{0\pm}_{mn\lambda}(0)G^0_+(r, \rho, z; 0, 0, 0; E), \quad (26)$$

where the parameter α is related with the scattering length a_0 by the formula $\alpha = m^*/2\pi\hbar a_0$.

As noted above, in the case at hand only states with $m = 0$ are scattered by a point potential. As follows from Eq. (26), $m = 0$ for the scattered mode also. Thus, for the analysis below it is sufficient to study a transition from the mode $(0, n_0)$ into the mode $(0, n_1)$.

4. TRANSMISSION COEFFICIENTS

To eliminate the ambiguity associated with the difference between G^0_+ and G^0_-, it is convenient to use directly the formula (13) in the upper half-plane ($\mathrm{Im}\,\zeta \geq 0$), writing it in the form

$$G^0(\mathbf{r}, \mathbf{r}'; \zeta) = -\frac{1}{(2\pi)^{3/2}e^{i3\pi/4}}\frac{1}{\hbar\omega_z l_z l_\Omega^2} \times \sum_{m=-\infty}^{0}\sum_{n=0}^{\infty}\exp[im(\varphi - \varphi')]R_{mn}(\rho)R_{mn}(\rho') \times \Gamma\left(\frac{1}{2} - i\lambda_{mn}\right)U\left(-i\lambda_{mn}, \frac{\max(z, z')}{l_z e^{i\pi/4}}\right) \times U\left(-i\lambda_{mn}, -\frac{\min(z, z')}{l_z e^{i\pi/4}}\right), \quad (27)$$

where $\lambda_{mn} = (\zeta - E_{mn})/\hbar\omega_z$. We take the eigenfunction $\Psi^{0\pm}_{mn\lambda}(\rho, \varphi, z)$ in the form

$$\Psi^{0\pm}_{mn\lambda}(\rho, \varphi, z) = R_{mn}(\rho)e^{im\varphi}U(-i\lambda, \pm ze^{-i\pi/4}/l_z). \quad (28)$$

Then the eigenfunction of the perturbed Hamiltonian is

$$\Psi^{\pm}_{mn\lambda}(\rho, \varphi, z) = \Psi^{0\pm}_{mn\lambda}(\rho, \varphi, z) - [Q(E) + \alpha]^{-1}\Psi^{0\pm}_{mn\lambda}(0)G^0(r, \rho, z; 0, 0, 0; E). \quad (29)$$

To find the transmission coefficient $T_{(0, n_0)\to(0, n_1)}$ from the mode $(0, n_0)$ into the mode $(0, n_1)$ we write the function (29) with $m = 0$ in the form

$$\Psi^{+}_{0n_0}(\rho, \varphi, z) = R_{0n_0}(\rho)U\left(-i\lambda_{0n_0}, \frac{z}{l_z e^{i\pi/4}}\right) + \frac{1}{(2\pi)^{3/2}e^{i3\pi/4}}\frac{1}{\hbar\omega_z l_z l_\Omega^2}U(-i\lambda_{0n_0}, 0)[Q(E) + \alpha]^{-1} \times \sum_{n=0}^{\infty}R_{0n}(\rho)\Gamma\left(\frac{1}{2} - i\lambda_{0n}\right)U\left(-i\lambda_{0n}, \frac{\max(z, 0)}{l_z e^{i\pi/4}}\right) \times U\left(-i\lambda_{0n}, -\frac{\min(z, 0)}{l_z e^{i\pi/4}}\right). \quad (30)$$

It follows from Eq. (30) that for $z > 0$

$$\Psi^{+}_{0n_0\lambda}(\rho, \varphi, z) = R_{0n_0}(\rho)U(-i\lambda_{0n_0}, ke^{-i\pi/4}) + \frac{1}{(2\pi)^{3/2}e^{i3\pi/4}}\frac{1}{\hbar\omega_z l_z l_\Omega^2}U(-i\lambda_{0n_0}, 0)[Q(E) + \alpha]^{-1} \times \sum_{n=0}^{\infty}R_{0n}(\rho)\Gamma\left(\frac{1}{2} - i\lambda_{0n}\right)U(-i\lambda_{0n}, 0)U(-i\lambda_{0n}, ke^{-i\pi/4}), \quad (31)$$

and for $z < 0$

$$\Psi^{+}_{0n_0\lambda}(\rho, \varphi, z) = R_{0n_0}(\rho)U(-i\lambda_{0n_0}, ke^{-i\pi/4}) + \frac{1}{(2\pi)^{3/2}e^{i3\pi/4}}\frac{1}{\hbar\omega_z l_z l_\Omega^2}U(-i\lambda_{0n_0}, 0)[Q(E) + \alpha]^{-1} \times \sum_{n=0}^{\infty}R_{0n}(\rho)\Gamma\left(\frac{1}{2} - i\lambda_{0n}\right)U(-i\lambda_{0n}, 0)U(-i\lambda_{0n}, ke^{-i\pi/4}), \quad (32)$$

where $k = z/l_z$.

We note the asymptotic expansions, required for the analysis below, of the parabolic cylinder function in the limit $z \longrightarrow +\infty$ [22]:

$$U\left(-i\lambda_{0n}, \frac{k}{e^{i\pi/4}}\right) \sim \exp\left(\frac{\pi\lambda_{0n}}{4} + \frac{i\pi}{8}\right)|k|^{i\lambda_{0n} - 1/2}\exp\frac{ik^2}{4},$$

$$U\left(-i\lambda_{0n}, -\frac{k}{e^{i\pi/4}}\right) \sim \exp\left(-\frac{3\pi\lambda_{0n}}{4} - \frac{3i\pi}{8}\right) \times |k|^{i\lambda_{0n} - 1/2}\exp\frac{ik^2}{4} - \frac{\sqrt{2\pi}}{\Gamma(1/2 - i\lambda_{0n})} \times \exp\left(-\frac{\pi\lambda_{0n}}{4} - \frac{7i\pi}{8}\right)|k|^{-i\lambda_{0n} - 1/2}\exp\left(-\frac{ik^2}{4}\right). \quad (33)$$

Substituting the asymptotic expansions (33) into Eqs. (31) and (32) we find the asymptotic expansion of $\Psi^{+}_{0n_0\lambda}$ in the limit $z \longrightarrow \pm\infty$. We note that the factor $\exp(iz^2/4l_z^2)$ in $\Psi^{+}_{0n_0\lambda}$ corresponds to a state in which the current flows from the origin of the coordinates in both directions, and the factor $\exp(-iz^2/4l_z^2)$ corresponds to the state in which the current flows toward the origin of coordinates, likewise from both directions.

Using the asymptotic expressions for $\Psi^{+}_{0n_0\lambda}$ which follow from Eq. (33), we find the coefficients of R_{0n_0} for $z < 0$, which represent a wave moving from left to right:

$$\Psi_{in}(z) = -\frac{\sqrt{2\pi}}{\Gamma(1/2 - i\lambda_{0n_0})}\exp\left(-\frac{\pi\lambda_{0n_0}}{4} - \frac{7i\pi}{8}\right) \times |k|^{-i\lambda_{0n_0} - 1/2}\exp\left(-\frac{ik^2}{4}\right). \quad (34)$$

The coefficient of R_{0n_0} for $z > 0$ represents a wave moving in the same direction:

$$\Psi_{\text{out}}(z) = \left[\delta_{n_0n_1} + \frac{1}{(2\pi)^{3/2}e^{i3\pi/4}}\frac{1}{\hbar\omega_z l_z l_\Omega^2}[Q(E) + \alpha]^{-1} \times \Gamma\left(\frac{1}{2} - i\lambda_{0n_1}\right)U(-i\lambda_{0n_0}, 0)U(-i\lambda_{0n_1}, 0)\right] \times \exp\left(\frac{\pi\lambda_{0n_1}}{4} + \frac{i\pi}{8}\right)|k|^{i\lambda_{0n_1} - 1/2}\exp\frac{ik^2}{4}. \quad (35)$$

The transmission coefficient $T_{(0n_0)\to(0n_1)}(E)$ has the form

$$T_{(0n_0)\to(0n_1)}(E) = \lim_{z\to\infty}\left|\frac{\Psi_{\text{out}}(z)}{\Psi_{\text{in}}(-z)}\right|^2. \quad (36)$$

Substituting the expression (34) and (35) into Eq. (36) and using [22]

$$U(-i\lambda, 0) = \frac{\sqrt{\pi}2^{i\lambda/2 - 1/4}}{\Gamma(3/4 - i\lambda/2)}, \quad (37)$$

we obtain for the transmission coefficient

$$T_{(0n_0)\to(0n_1)}(E) = \frac{\exp[\pi(\lambda_{0n_1} - \lambda_{0n_0})/2]}{1 + \exp(-2\pi\lambda_{0n_0})} \times \left|\delta_{n_0n_1} - e^{i\pi/4}\frac{2^{-i(\lambda_{0n_1} - \lambda_{0n_0})}}{4\sqrt{2}\pi\hbar\omega_z l_z l_\Omega^2[Q(E) + \alpha]}\frac{\Gamma(1/4 - i\lambda_{0n_1}/2)}{\Gamma(3/4 - i\lambda_{0n_0}/2)}\right|^2. \quad (38)$$

We note that $\lambda_{0n_1} - \lambda_{0n_0} = \Omega/\omega_z(n_0 - n_1)$ does not depend on the particle energy E. The limit $\alpha \longrightarrow \infty$ corresponds to the absence of an impurity. Then the Hamiltonian H converts into H^0. The expression (38) in this limit becomes

$$T_{(0n_0)\to(0n_1)}(E) \longrightarrow \delta_{n_0n_1}\frac{1}{1 + \exp(-2\pi\lambda_{0n_0})}. \quad (39)$$

The expression (39) is identical to the well-known Büttiker formula for the transmission coefficient of a mode with $m = 0$ in a three-dimensional constriction [4]. As noted above, a point impurity does not scatter states with $m \neq 0$. Consequently, Büttiker's formula is valid for these states as well:

$$T_{(m_0n_0)\to(m_1n_1)}(E) = \delta_{m_0m_1}\delta_{n_0n_1}\frac{1}{1 + \exp(-2\pi\lambda_{m_0n_0})}. \quad (40)$$

On this basis the general expression for the transmission coefficient has the form

$$T_{(m_0n_0)\to(m_1n_1)}(E) = \frac{\delta_{m_0m_1}\exp[\pi\Omega(n_0 - n_1)/2\omega_z]}{1 + \exp[-2\pi\lambda_{0n_0}(E)]} \times \left|\delta_{n_0n_1} - e^{i\pi/4}\delta_{0m_0}\frac{2^{i\Omega(n_1 - n_0)/2\omega_z}}{4\sqrt{2}\pi\hbar\omega_z l_z l_\Omega^2[Q(E) + \alpha]} \times \frac{\Gamma(1/4 - i\lambda_{0n_1}/2)}{\Gamma(3/4 - i\lambda_{0n_0}/2)}\right|^2. \quad (41)$$

5. CONDUCTANCE OF A QUANTUM CONSTRICTION

We now introduce the quantity γ according to the formula $\gamma = a_0 l_z/2l_\Omega^2$. Then it is convenient to write $q(E) = Q(E)/\alpha\gamma$ and the expression for the transmission coefficient (41) in the form

$$T_{(m_0n_0)\to(m_1n_1)}(E) = T^{\text{geom}}(E) + T^{\text{imp}}(E), \quad (42)$$

where the ballistic part of the transmission coefficient is

$$T^{\text{geom}}(E) = \frac{\delta_{m_0m_1}\exp[\pi\Omega(n_0 - n_1)/2\omega_z]}{1 + \exp[-2\pi\lambda_{0n_0}(E)]}\delta_{n_0n_1}, \quad (43)$$

and the part of the transmission coefficient that is due to scattering by the impurity and the geometry of the constriction is

$$T^{\mathrm{imp}}(E) = \frac{\delta_{m_0 m_1}\exp[\pi\Omega(n_0-n_1)/2\omega_z]}{1+\exp[-2\pi\lambda_{0n_0}(E)]}$$

$$\times\left\{-\delta_{n_0 n_1}\frac{\gamma}{\pi\sqrt{2}}\frac{\left|\Gamma(1/4+i\varepsilon_{n_0}/2)\right|^2}{\left|1+\gamma q(E)\right|^2}\right. \qquad (44)$$

$$\times[(1+\gamma\mathrm{Re}q)\exp(\pi\varepsilon_{n_0}/2)+\gamma\mathrm{Im}q\exp(-\pi\varepsilon_{n_0}/2)]$$

$$\left.+\frac{\gamma^2}{2}\frac{1}{\left|1+\gamma q(E)\right|^2}\frac{\left|\Gamma(1/4+i\varepsilon_{n_1}/2)\right|^2}{\left|\Gamma(3/4+i\varepsilon_{n_0}/2)\right|^2}\right\}.$$

We shall now calculate the conductance of a constriction with an impurity at temperature $T = 0$ using the Büttiker–Landauer formula:

$$\frac{G}{G_0} = \sum_{m_0 m_1 n_0 n_1} T_{m_0 n_0 \to m_1 n_1}. \qquad (45)$$

In Eq. (45) it is convenient to single out the term that is determined only by the geometry of the constriction and does not depend on the scattering by the impurity $G^{\mathrm{geom}} = G_1$. Then $G = G_1 + G_2$, where G_2 depends on the scattering parameter γ. As follows from what has been said above, we obtain for G_1/G_0 the expression

$$\frac{G_1(E)}{G_0} = \sum_{mn}\frac{1}{1+\exp[-2\pi(E-E_{mn})/\hbar\omega_z]}, \qquad (46)$$

where $G_0 = 2e^2/h$ is the conductance quantum.

The term $G_1(E)/G_0$ in the conductance has been studied in detail in our preceding work [9]. The term that depends on the scattering, $G_2(E)/G_0$, has the form, in accordance with Eqs. (42) and (45),

$$\frac{G_2(E)}{G_0} = -\frac{\gamma}{\pi\left|1+\gamma q(E)\right|^2}\sum_{n=0}^{\infty}[1+\exp(2\pi\varepsilon_n)]^{-1}$$

$$\times\left|\Gamma\left(\frac{1}{4}+\frac{i\varepsilon_n}{2}\right)\right|^2\left[(1+\gamma\mathrm{Re}q(E))\exp\frac{\pi\varepsilon_n}{2}\right.$$

$$\left.+\gamma\mathrm{Im}q(E)\exp(-\pi\varepsilon_n/2)\right]+\frac{\gamma^2}{8\pi^2\left|1+\gamma q(E)\right|^2} \qquad (47)$$

$$\times\sum_{n=0}^{\infty}\left|\Gamma\left(\frac{1}{4}+\frac{i\varepsilon_n}{2}\right)\right|^2\exp\left(-\frac{\pi\varepsilon_n}{2}\right)$$

$$\times\sum_{l=0}^{\infty}\left|\Gamma\left(\frac{1}{4}+\frac{i\varepsilon_l}{2}\right)\right|^2\exp\left(-\frac{\pi\varepsilon_l}{2}\right).$$

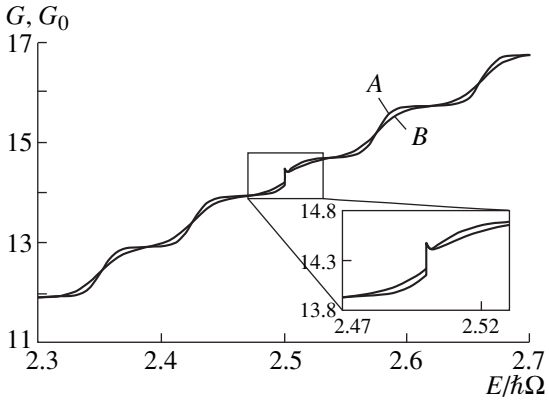


Fig. 1. Plot of the conductance G (in units of G_0) versus energy: curve A—$T = 0$ K, curve B—$T = 2$ K; $\omega_x = \omega_y = 1.61\times10^{13}$ s^{-1}, $\omega_x/\omega_z = 13$, $B = 5$ T.

For nonzero temperature the expression for the conductance can be written in the form

$$\frac{G(E,T)}{G_0} = \int_0^{\infty} G(E)\left(-\frac{\partial f_0}{\partial E}\right)dE, \qquad (48)$$

where f_0 is the Fermi function and $G(E)$ is given by the expression (43). The temperature dependence of the conductance of a constriction with no impurity was investigated in detail in [9], where it was shown that an increase of temperature results in smearing of the steps and in a small inclination of the plateau in $G_1(E, T)$. The effect of the temperature on the term $G_2(E, T)$ of the conductance is somewhat weaker than on the term $G_1(E, T)$. A plot of $G(E)$ at different temperatures is shown in Fig. 1.

6. DISCUSSION

As shown in the preceding section the conductance of a constriction under conditions of a quasiballistic transport regime can be represented as a sum of two terms. The first one is due to the scattering processes which depend on the geometric parameters of the three-dimensional constriction and do not depend on the scattering by an impurity. This term, $G^{\mathrm{geom}}(E, T)$, is identical to the formula for the conductance in a ballistic transport regime, as obtained in [9]. The dependences of this term on the electron energy, magnetic field, and temperature have been investigated in detail in [9]. We note that this term makes the main contribution to the conductance in the quasiballistic regime as well, since $G_2 \ll G_1$. Scattering by a point impurity increases the resistance of the system and correspondingly results in a small general downward displacement of the entire conductance curve (see Fig. 2). Since only states with $m = 0$ are scattered by an impurity located at the center

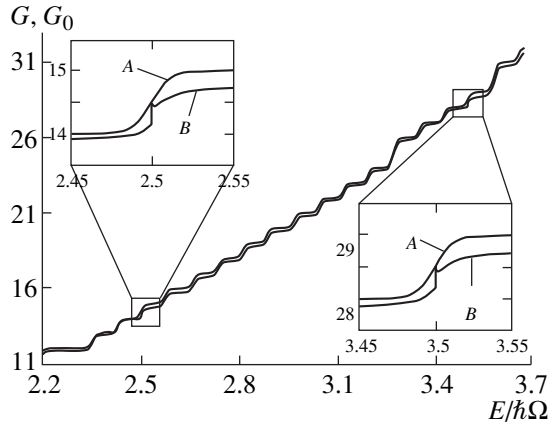


Fig. 2. Plot of the conductance versus energy at temperature $T = 0$ K: curve *A*—constriction with no impurity, curve *B*—constriction with an impurity; $\omega_x = \omega_y = 1.61 \times 10^{13}$ s^{-1}, $\omega_x/\omega_z = 13$, $B = 5$ T.

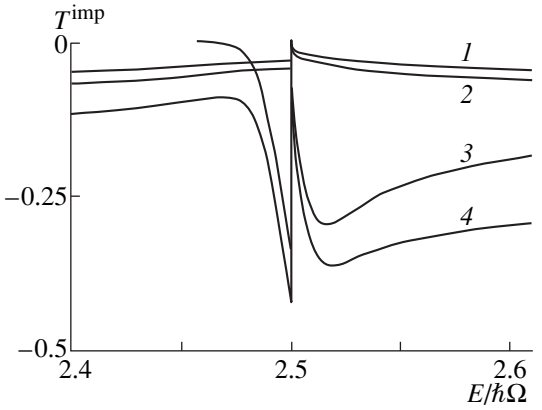


Fig. 3. Plots of the transmission coefficients $T^{\mathrm{imp}}(E) \equiv T^{\mathrm{imp}}_{n_0 n_1}(E)$ versus energy at temperature $T = 0$ K: curve *1*—T^{imp}_{00}, curve *2*—T^{imp}_{11}, curve *3*—T^{imp}_{22}, curve *4*—$\sum_{n=0}^{2} T^{\mathrm{imp}}_{nn}$; $\omega_x = \omega_y = 1.61 \times 10^{13}$ s^{-1}, $\omega_x/\omega_z = 13$, $B = 5$ T.

of the cross section of the waist of the constriction, the corresponding conductance steps shift downwards by a greater amount than the others (Fig. 2). Of greatest interest is the nonmonotonic behavior of the conductance in a small neighborhood of the thresholds of the steps. It is shown in the insets (Fig. 2) that at the points where $E/\hbar\Omega - 1/2$ is an integer G_2 changes abruptly to zero. At this point the curve $G(E)$ possesses a kink, and a pronounced vertical segment of $G(E)$ forms on the threshold. This dependence of $G(E)$ is due to the corresponding behavior of $G_2(E)$, which in turn is due to the behavior of the transmission coefficients $T^{\mathrm{imp}}(E)$ near the points where $E/\hbar\Omega - 1/2$ is equal to an integer N. Plots of the energy dependences of the transmission coefficients $T^{\mathrm{imp}}(E)$, contributing to $G_2(E)$ near the point where $N = 2$, are presented in Fig. 3. Numerical analysis showed that the second term in $T^{\mathrm{imp}}(E)$, containing γ^2, is much smaller in modulus than the negative first term, containing γ. The sum of these terms is much smaller than the transmission coefficient $T^{\mathrm{geom}}(E)$. However, there exists a small region where the first term in $T^{\mathrm{imp}}(E)$ is not too small. This is a neighborhood of the point where $1 + \gamma \mathrm{Re}\, q(E) \approx 0$. At this point the small factors containing γ cancel in both terms due to the geometry of the constriction and the presence of the impurity. Consequently, the sum of both terms in $T^{\mathrm{imp}}(E)$ to the left of the point where $E/\hbar\Omega - 1/2 = N$ (see Fig. 3, for $T_{n_0 \to n_1}$, $n_0 = n_1 = 2$) becomes a relatively large negative quantity (approximately –0.4 in Fig. 3). However, at the point where $E/\hbar\Omega - 1/2 = N$, the function $Z(1/2, 1/2 - E/\hbar\Omega)$ entering into $q(E)$ takes infinite value and, therefore, all impurity-induced terms in the transmission coefficients vanish. For this reason, the decrease of these terms is replaced by even sharper, almost vertical, increase up to zero. Then, to the right from the point $E/\hbar\Omega - 1/2 = N$, the terms considered give a nonmonotonic dependence of the transmission coefficients $T^{\mathrm{imp}}(E)$ (Fig. 3). Since the negative first term is larger in modulus than the second term, their sum is once again negative to the right of the point where $E/\hbar\Omega - 1/2 = N$. This is responsible for the stronger decrease in the conductance $G(E)$ than $G_1(E)$ at these steps after the point where $G_2(E)$ vanishes. We note that, as follows from the numerical analysis, only the transmission coefficients from $G_2(E)$ for which $n_0 = n_1 = n \le N$ make an appreciable contribution to the conductance; the transmission coefficient with $n = N$ makes the main contribution.

It is interesting to note that the situation is completely different for a three-dimensional wire with a circular cross section [18]. This is because only a positive term $\sim\gamma^2$ appears in the expression for the transmission coefficients $T^{\mathrm{imp}}(E)$ of a wire, in contrast to the constriction studied here (there is no analog of the negative term in Eq. (44) for the case of a wire). It is this term that gives for a wire the narrow and large in magnitude peaks at the thresholds of steps with $m = 0$, and outside the region of the threshold the scattering by an impurity is negligibly small.

Finally, we note that the resonance condition $1 + \gamma \mathrm{Re}\, q(E) \approx 0$ corresponds to the condition for the existence of a bound state $Q(E) = -\alpha$. Hence it follows that a bound state engenders the resonance shown in Figs. 1 and 2.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research and the Ministry of Education of the Russian Federation.

REFERENCES

1. E. N. Bogachek, A. M. Zagoskin, and I. O. Kulik, Fiz. Nizk. Temp. **16**, 1404 (1990) [Sov. J. Low Temp. Phys. **16**, 796 (1990)].
2. J. A. Torres, J. I. Pascual, and J. J. Saenz, Phys. Rev. B **49**, 16581 (1994).
3. E. N. Bogachek, M. Jonson, R. I. Shekhter, *et al.*, Phys. Rev. B **47**, 16635 (1993); **50**, 18341 (1994).
4. A. G. Scherbakov, E. N. Bogachek, and Uzi Landman, Phys. Rev. B **53**, 4054 (1996).
5. E. N. Bogachek, A. G. Scherbakov, and Uzi Landman, Phys. Rev. B **53**, R13246 (1996).
6. E. N. Bogachek, A. G. Scherbakov, and Uzi Landman, Phys. Rev. B **54**, R11094 (1996).
7. E. N. Bogachek, A. G. Scherbakov, and Uzi Landman, Phys. Rev. B **56**, 14917 (1997).
8. E. N. Bogachek, A. G. Scherbakov, and Uzi Landman, Phys. Rev. B **56**, 1065 (1997).
9. N. G. Galkin, V. A. Geĭler, and V. A. Margulis, Zh. Éksp. Teor. Fiz. **117**, 593 (2000) [JETP **90**, 517 (2000)].
10. J. I. Pascual, J. Mendez, J. Gomez-Herrero, *et al.*, Phys. Rev. Lett. **71**, 1852 (1993).
11. J. I. Pascual, J. Mendez, J. Gomez-Herrero, *et al.*, Science **267**, 1793 (1995); J. Vac. Sci. Technol. B **13**, 1280 (1995).
12. L. Olesen, E. Laegsgaard, I. Stensgaard, *et al.*, Phys. Rev. Lett. **72**, 2251 (1994).
13. V. A. Geĭler, V. A. Margulis, and L. I. Filina, Zh. Éksp. Teor. Fiz. **113**, 1376 (1998) [JETP *86*, 751 (1998)].
14. Y. B. Levinson, M. I. Lubin, and E. V. Sukhorukov, Phys. Rev. B **45**, 11976 (1992).
15. M. B. Levinson, M. I. Lyubin, and E. V. Sukhorukov, Pis'ma Zh. Éksp. Teor. Fiz. **54**, 405 (1991) [JETP Lett. **54**, 401 (1991)].
16. M. I. Lyubin, Pis'ma Zh. Éksp. Teor. Fiz. **57**, 346 (1993) [JETP Lett. **57**, 361 (1993)].
17. V. A. Geĭler, V. A. Margulis, and I. I. Chuchaev, Pis'ma Zh. Éksp. Teor. Fiz. **58**, 666 (1993) [JETP Lett. **58**, 648 (1993)].
18. V. A. Geĭler and V. A. Margulis, Zh. Éksp. Teor. Fiz. **111**, 2215 (1997) [JETP **84**, 1209 (1997)].
19. L. I. Glazman, G. B. Lesovik, D. E. Khmel'nitskiĭ, and R. I. Shekhter, Pis'ma Zh. Éksp. Teor. Fiz. **48**, 218 (1988) [JETP Lett. **48**, 238 (1988)].
20. L. I. Glazman and A. V. Khaetskiĭ, Pis'ma Zh. Éksp. Teor. Fiz. **48**, 546 (1988) [JETP Lett. **48**, 591 (1988)].
21. C. W. J. Beenakker and H. van Houten, in *Solid State Physics*, Ed. by H. Ehrenrech and D. Turnbull (Academic, New York, 1991), Vol. 44, p. 83.
22. *Handbook of Mathematical Functions*, Ed. by M. Abramowitz and I. A. Stegun (Dover, New York, 1971; Nauka, Moscow, 1979).
23. N. Kokiantonis and D. P. L. Castrigiano, J. Phys. A **18**, 45 (1985).
24. L. F. Urrutia and C. Manterola, Int. J. Theor. Phys. **25**, 75 (1985).
25. A. I. Baz', Ya. B. Zel'dovich, and A. M. Perelomov, *Scattering, Reactions and Decays in Nonrelativistic Quantum Mechanics* (Nauka, Moscow, 1971; Israel Program for Scientific Translations, Jerusalem, 1966).
26. V. A. Geĭler and V. A. Margulis, Teor. Mat. Fiz. **58**, 461 (1984).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 1, 2000, pp. 206–225.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 1, 2000, pp. 232–252.

MISCELLANEOUS

Directed Self-Avoiding Walks and Statistics of Rigid-Chain Polymer Molecules

A. E. Arinshtein

Semenov Institute of Chemical Physics, Russian Academy of Sciences, Moscow, 117977 Russia
e-mail: arinst@polymer.chph.ras.ru

Received November 18, 1999

Abstract—The statistics of rigid-chain polymer conformations is described on the basis of a model of directed self-avoiding walks. The generating functions for the distribution function of a chain in one-, two-, and three-dimensional spaces are constructed. It is shown that the statistics of the conformational states of chains with finite interunit flexural stiffness can differ strongly from Gaussian statistics. If the chain length is comparable to the Kuhn segment length, then the molecule is strongly anisotropic (almost rectilinear), but as the chain length increases, the molecule starts to bend and ultimately coils up. However, since a coil contains extended, almost rectilinear, chain sections, the coil is not truly Gaussian, even though the squared average size of the coil is directly proportional to the chain length. It is shown that under certain conditions the existence of almost rectilinear chain sections results in the appearance of orientational order in the system. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Many physical properties of polymer systems are determined by the conformational properties of macromolecules, i.e., the set of their possible conformations and their conformational mobility. The progress achieved in the description of the statistics of the conformations of linear polymer chains is largely due to the fact that the conformational properties of polymer chains as a whole depend very little on the characteristic features of their chemical structure, i.e., they reflect certain general, fundamental properties of polymers. In addition, the fact that the number n of links in a polymer chain is large makes it possible to pass, when necessary, to the asymptotic limit ($N \longrightarrow \infty$).

The simplest model that makes it possible to take into account the conformational properties of long linear molecules is the model of an ideal (phantom) flexible polymer chain consisting of a chain of freely-jointed weightless links. Even though the range of applicability is relatively narrow (dilute solutions in θ solvents and polymer melts, where the interaction of the links of one chain is compensated by their interaction with the environment), this model is extremely helpful for understanding many features of polymer systems [1, Chapter 1, Section 1.1; 2, Chapter 1, Section 1; 3, Chapter 7, Sections 7.3–7.4].

The orientations of two neighboring links in an ideal freely-jointed chain are independent of one another. However, if the polymer molecule possesses some interlink stiffness and the orientations of two neighboring units are correlated, then the concept of a Kuhn segment (or persistence length) [1, Introduction; 2, Chapter 1, Section 3] is used to describe such a chain. A Kuhn segment is introduced as a chain section of length Δl in which the memory of the orientation of its starting section is completely lost, so that the orientations of neighboring Kuhn segments are uncorrelated, making it possible to describe the statistics of the conformational states of such a molecule using the description of a freely-jointed chain with a link equal to the Kuhn segment. In more general (idealized) models a Kuhn segment is treated as a structureless object, since it is assumed that if a large number of Kuhn segments fits within the length of a molecule, then the structure of the segments, as determined by the characteristic features of the chemical structure of specific polymer molecules, essentially has no effect on the conformational properties of a long polymer chain as a whole. Thus, a molecule with a certain interunit stiffness can be treated as freely-jointed only if a large number of Kuhn segments fits within the length of the molecule.

If the stiffness of a polymer molecule is large, i.e., the number of Kuhn segments in the chain is small, the asymptotic representation of a polymer molecule as a freely-jointed chain is unsuitable and more complicated models must be used to describe its statistical properties. It is certainly of interest to study such rigid-chain systems as well as systems consisting of chains of limited length theoretically, since such a study will make it possible to explain a number of experimentally observed unusual properties of so-called oligomer systems, which fall between low- and high-molecular chemical compounds.

But even if a polymer molecule contains a sufficiently large number of Kuhn segments, in some cases treating each Kuhn segment as a structureless (isotropic) object is too coarse an approximation. Obviously,

the structure of a Kuhn segment of a polymer chain is determined by the characteristic features of its chemical structure, and at first glance it appears that taking account of the internal structure of a Kuhn segment will result in different consequences for different homological series. However, for all types of linear polymers the Kuhn segments possess a very important property in common: since, by definition, a Kuhn segment is an almost rectilinear section of a chain, it is, first and foremost, a strongly anisotropic object. This means that a rigid-chain polymer consists of strongly anisotropic elements, and it is this local anisotropy (the presence of quite extended rectilinear sections) of a polymer molecule that can result in the appearance of supramolecular structures, influencing the macroscopic properties of the polymer system and in some cases completely determining its state. It turns out that the anisotropy of Kuhn segments results in the appearance of a local anisotropy of a rigid-chain polymer melt, and under certain conditions it can lead to liquid-crystal ordering of the entire system as a whole. For example, anomalies in the rheological properties are observed in liquid oligomeric systems [4], a liquid-crystal state arises in melts of rigid-chain polymers (polyethylene type) [5], and so on. Scaling analysis of the appearance of anisotropy in a rigid-chain polymeric system for various ratios of the lengths of a chain and its Kuhn segment is presented in Appendix 1.

Thus, under certain conditions the properties of rigid-chain polymers are determined by their small-scale structure (on scales less than or the order of the persistence length). It is obvious that when these scales are taken into account the statistics of the conformational states of rigid-chain molecules is different from that of long flexible polymer chains, and the asymptotic description employed for the latter is inapplicable for rigid-chain molecules, since it is too coarse.

One of the simplest methods for obtaining an idealized description of a freely-jointed phantom chain is to represent it as a trajectory of random walks on a regular periodic lattice [6] (see also [1, Chapter 1, Section 1.1; 2, Chapter 1, Section 6; 3, Chapter 7, Sections 7.3 and 7.4]). Here the theory of random processes [7–9, Chapter 11, Sections 11.1–11.3], which is well-developed for lattices and in the continuous case, is used. The trajectory of a random walk describing a linear polymer chain consisting of N links is a sequence of N steps. The first step is taken from a point $\mathbf{R}_0$ and the last step arrives at the point $\mathbf{R}_N$. At each step the next hop can occur with equal probability in the direction of any of the nearest lattice sites.

On this basis, to describe the statistics of the conformational properties of an oligomeric molecule in terms of random walks it is necessary to use a modification of the random-walk model—a variant of directed self-avoiding walks [10–12]. Apparently, the problem of directed self-avoiding walks was first formulated by Feynman [13, Chapter 2, Section 4] in a calculation of the relativistic quantum-mechanical propagator in two-dimensional space-time ((1+1) dimensions). Later this problem was discussed in connection with an investigation of exactly solvable models in statistical physics, such as the one-dimensional Ising model, the Potts model in a magnetic field, and others. It should be noted that the possibility of describing polymer chains with finite interunit stiffness in terms of directed self-avoiding walks was also studied in [14, 15]. However, these works were concerned with only a one-parameter model, in which the probabilities of various relative orientations of two neighboring links assume only two values. The more general situation where the probabilities of various relative orientations of two neighboring links are different from one another is not described in the works mentioned. Moreover, the spatial distribution function of a polymer chain was constructed only in a one-dimensional space, where the problem reduces to a one-dimensional Ising chain. In spaces with a larger number of dimensions, the works [14, 15] were limited to a calculation of only the average squared displacement $\langle \mathbf{R}^2 \rangle$.

In the present paper some variants of the description of the conformational statistics of oligomeric molecules in the language of directed self-avoiding walks, making it possible to take into account more fully all of the above-enumerated features of the statistics of the conformational states of oligomeric chains and to construct the generating function for the spatial distribution function of a polymeric chain are proposed.

2. DIRECTED SELF-AVOIDING WALKS ON REGULAR LATTICES AS A MODEL OF A LINEAR CHAIN WITH FINITE STIFFNESS

We shall use the random-walk model to describe the statistics of the conformations of chains possessing a finite interunit bending stiffness. We shall consider random walks of a particle on a regular lattice for which the probability of the direction of each subsequent hop depends on the direction of the hop at the preceding step. We shall call such walks directed self-avoiding and the trajectories of such a random walk correspond to the conformations of a linear chain with finite stiffness.

Let the vector $\mathbf{e}(n)$ correspond to an edge in the lattice along which a hop occurs at the n step of the walks. This same vector $\mathbf{e}(n) \equiv \mathbf{e}(l)$ corresponds to the nth link in the chain ($l = l_0 n$, l_0 is the length of a lattice edge, corresponding to the length of one link in the chain). We shall represent the vector connecting the beginning and end of the chain using the vectors $\mathbf{e}(n)$ introduced:

$$R(N) = l_0 \sum_{n}^{N} \mathbf{e}(n), \tag{1}$$

The complete information about the state of a chain is contained in the distribution function $G(\mathbf{R}_n, n)$, which

gives the probability that the nth link of the chain (more accurately, the end of the chain), which starts at the origin of the coordinates, falls at a lattice site with the coordinates $\mathbf{R}_n \equiv \{R_{nx} = l_0 m_x, R_{ny} = l_0 m_y, R_{nz} = l_0 m_z\}$. For $n = N$ the distribution function $G(\mathbf{R}_N, N)$ gives the probability that the end of the chain has moved relative to its beginning by the vector $\mathbf{R}_N$.

In the random-walk model the probability of a wandering particle reaching some point at a particular step can depend only on the location of the particle at the preceding steps and is completely independent of the future location of the particle. Consequently, the distribution function $G(\mathbf{R}_n, n)$, giving the probability that a wandering particle has moved by the vector $\mathbf{R}_n$ in n steps, does not depend on the total number of steps. For a linear molecule this means that the distribution function $G(\mathbf{R}_n, n)$, determining, specifically, the probability that the distance between the n links of a chain is $\mathbf{R}_n$, does not depend on the length of the molecule. Knowing the distribution function $G(\mathbf{R}_n, n)$ gives us a complete picture of the spatial distribution of the mass of a molecule.

In the walk of a particle (both random and directed self-avoiding) $(2d)^N$ different trajectories can be realized in N steps. Determining which fraction of these trajectories corresponds to the displacement $\mathbf{R}_n$, we find the desired distribution function $G(\mathbf{R}_N, N)$. Therefore, singling out the corresponding trajectories, which we give using the expression (1), and averaging over all possible realizations of the walk of a particle, the distribution function $G(\mathbf{R}_N, N)$ can be written as

$$G(\mathbf{R}_N, N) = \left\langle \delta\left(\mathbf{R}_N - l_0 \sum_n^N \mathbf{e}(n)\right)\right\rangle$$
$$= \left\langle \delta\left(m_{Nx} - \sum_n^N e_x(n)\right)\delta\left(m_{Ny} - \sum_n^N e_u(n)\right)\right.$$
$$\left. \times \delta\left(m_{Nz} - \sum_n^N e_z(n)\right)\right\rangle, \qquad (2)$$

where the brackets signify averaging of all possible trajectories and the analog of the Dirac delta function with a vector argument is used—the Kronecker delta function with vector argument—$\delta(\mathbf{R}) \equiv \delta(m_x)\delta(m_y)\delta(m_z)$: $\delta(m) \equiv \delta_{m,0}$, $\delta_{m,0} = 0$ if $m \neq 0$ and $\delta_{m,0} = 1$ if $m = 0$.

Since the vectors $\mathbf{e}(n)$ can possess components 1, 0, or –1, the expression (2) can be written as

$$G(\mathbf{R}_N, N) = \left\langle \delta\left(m_{Nx} - \sum_{n=1}^N \sigma_n^x\right)\right.$$
$$\left. \times \delta\left(m_{Ny} - \sum_{n=1}^N \sigma_n^y\right)\delta\left(m_{Nz} - \sum_{n=1}^N \sigma_n^z\right)\right\rangle, \qquad (3)$$

where the numbers σ_n^x, σ_n^y, and σ_n^z assume the values ± 1 or 0 and averaging is performed over all possible values of σ_n^x, σ_n^y, and σ_n^z satisfying the condition $(\sigma_n^x)^2 + (\sigma_n^y)^2 + (\sigma_n^z)^2 = 1$.

In two-dimensional space this expression simplifies somewhat:

$$G(\mathbf{R}_N, N)$$
$$= \left\langle \delta\left(m_{Nx} - \sum_{n=1}^N \sigma_n^x\right)\delta\left(m_{Ny} - \sum_{n=1}^N \sigma_n^y\right)\right\rangle, \qquad (4)$$

where the numbers σ_n^x and σ_n^y assume the values ± 1 or 0, and the averaging is performed over all possible values of σ_n^x and σ_n^x satisfying the condition $(\sigma_n^x)^2 + (\sigma_n^y)^2 = 1$.

An even simpler expression is obtained in the one-dimensional case:

$$G(x_N, N) = \left\langle \delta\left(m_N - \sum_{n=1}^N \sigma_n\right)\right\rangle, \quad x_N = l_0 m_N, \qquad (5)$$

where the averaging extends only over two values, $\sigma_n = \pm 1$.

The construction of the distribution function $G(\mathbf{R}_N, N)$ in accordance with Eqs. (3)–(5) is a quite difficult problem to which we shall return below. We call attention especially to the one-dimensional variant of the problem, which is of more than methodological interest. Actually, the one-dimensional case is of little physical interest for the problem of the conformations of a rigid chain: it can describe only the folded structure of oriented (i.e., possessing liquid-crystal ordering) chains which are subjected to stretching. However, for describing the diffusion of impurities in a condensed phase even the one-dimensional variant of the problem of directed self-avoiding walks has an important physical meaning [16].

3. RMS SIZE OF A MOLECULE WITH FINITE INTERLINK STIFFNESS

When it is difficult to find the complete distribution function, the moments of the distribution function can yield information about the properties of the system. One of the most informative and consequently most widely used is the second moment, which characterizes the size of a molecule. It can be calculated quite simply in terms of the distribution function:

$$\langle R^2(N)\rangle = \int \mathbf{R}^2 G(\mathbf{R}, N) d\mathbf{R}. \qquad (6)$$

However, in some cases it can also be determined from other considerations. Indeed, using Eq. (1) we can write the following expression for the squared size of a mol-

ecule (the squared distance between the start and end of the chain):

$$\langle \mathbf{R}^2(N)\rangle = l_0^2 \sum_{n_1, n_2}^{N} \langle \mathbf{e}(n_1)\cdot\mathbf{e}(n_2)\rangle = l_0^2\left\{ N + 2\sum_{n_1=1}^{N}\sum_{n_2=1}^{N-n_1} \langle \mathbf{e}(n_1)\cdot\mathbf{e}(n_1+n_2)\rangle \right\}. \quad (7)$$

The expression (7) contains the correlation function

$$g(\Delta l) = \langle \mathbf{e}(l)\cdot\mathbf{e}(l+\Delta l)\rangle = \langle \mathbf{e}(n)\cdot\mathbf{e}(n+\Delta n)\rangle, \quad \Delta l = l_0\Delta n, \quad (8)$$

which takes into account the correlation of the orientations of the various links of a linear molecule, which is due to the interlink flexural stiffness of the chain. The calculation of the function $g(\Delta l)$ is the main difficulty arising in this problem.

We note that the length Δl of a Kuhn segment for the chain under consideration is ordinarily determined from the condition $g(\Delta l) \approx 0$.

3.1. Two-Dimensional Problem. Square Lattice

We shall consider the simplest variant of the problem first: a random walk on a two-dimensional square lattice. The stiffness of the chain (the probability of choosing a direction of the walk at the next step in the lattice as a function of the direction of the preceding step) is taken into account as follows:

the probability that the motion is in the same direction is maximum and equals $\alpha_+(T)$;

the probability that motion occurs in the perpendicular direction is $\alpha_\perp(T)$;

the probability of a 180° turn is $\alpha_-(T)$.

It is obvious that the probabilities α_i introduced in this manner satisfy $\alpha_-(T) < \alpha_\perp(T) < \alpha_+(T)$.

We shall enumerate the directions in the plane (Fig. 1) and introduce the column vector of the distribution of probabilities of the direction of motion at a step n:

$$|P(n)\rangle = \begin{pmatrix} P_1(n) \\ P_2(n) \\ P_3(n) \\ P_4(n) \end{pmatrix}. \quad (9)$$

These column vectors of the probability distribution at steps n and $n+1$ are related by a transition matrix:

$$|P(n+1)\rangle = \mathbf{T}_2|P(n)\rangle, \quad (10)$$

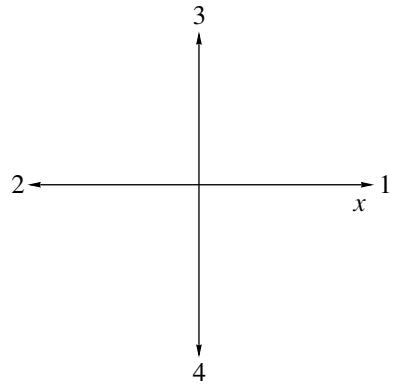


Fig. 1.

where

$$\mathbf{T}_2 = \begin{pmatrix} \alpha_+ & \alpha_- & \alpha_\perp & \alpha_\perp \\ \alpha_- & \alpha_+ & \alpha_\perp & \alpha_\perp \\ \alpha_\perp & \alpha_\perp & \alpha_+ & \alpha_- \\ \alpha_\perp & \alpha_\perp & \alpha_- & \alpha_+ \end{pmatrix}.$$

After m steps the column vectors are related with one another using the same transfer matrix:

$$|P(n+m)\rangle = \mathbf{T}_2^m|P(n)\rangle. \quad (11)$$

Since the scalar product of the vectors $\mathbf{e}(n)$ and $\mathbf{e}(n+m)$ can assume the values –1, 0, and 1, and taking into account the spatial isotropy of the transfer matrix $\mathbf{T}_2$, the desired correlation function can be expressed in terms of the matrix elements of the mth power of these matrices:

$$g(m) = \langle \mathbf{e}(n)\cdot\mathbf{e}(n+m)\rangle = (\mathbf{T}_2^m)_{i,i} - (\mathbf{T}_2^m)_{i,i\pm1}, \quad (12)$$

where a plus sign is used for odd i and a minus sign for even i.

Averaging of all possible orientations the desired correlation function can be written as

$$g(m) = \langle \mathbf{e}(n)\cdot\mathbf{e}(n+m)\rangle = \frac{1}{4}\mathrm{Sp}(\mathbf{Q}_2\mathbf{T}_2^m), \quad (13)$$

where $\mathbf{Q}_2$ is a block-diagonal matrix:

$$\mathbf{Q}_2 = \begin{pmatrix} \mathbf{s} & 0 \\ 0 & \mathbf{s} \end{pmatrix}, \quad \mathbf{s} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}. \quad (14)$$

We note that $\mathbf{T}_2$ is also block matrix:

$$\mathbf{T}_2 = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{pmatrix}, \quad (15)$$

where $\mathbf{A}$ and $\mathbf{B}$ are 2×2 square matrices:

$$\mathbf{A} = \begin{pmatrix} \alpha_+ & \alpha_- \\ \alpha_- & \alpha_+ \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} \alpha_\perp & \boldsymbol{\alpha}_\perp \\ \alpha_\perp & \boldsymbol{\alpha}_\perp \end{pmatrix},$$

and the matrices $\mathbf{A}$ and $\mathbf{B}$ commute with one another and therefore can be diagonalized simultaneously:

$$\begin{aligned} \mathbf{A} &= \mathbf{O}\begin{pmatrix} \alpha_+ - \alpha_- & 0 \\ 0 & \alpha_+ + \alpha_- \end{pmatrix}\mathbf{O}^{-1}, \\ \mathbf{B} &= \mathbf{O}\begin{pmatrix} 0 & 0 \\ 0 & 2\alpha_\perp \end{pmatrix}\mathbf{O}^{-1}. \end{aligned} \tag{16}$$

Here $\mathbf{O}$ is an orthogonal matrix:

$$\mathbf{O}^{\pm 1} = \begin{pmatrix} 1/\sqrt{2} & \pm 1/\sqrt{2} \\ \mp 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}.$$

Successive diagonalization of the matrix $\mathbf{T}_2$ (taking into account its block structure) gives

$$\begin{aligned} \mathbf{T}_2 &= \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}\begin{pmatrix} \mathbf{A}-\mathbf{B} & 0 \\ 0 & \mathbf{A}+\mathbf{B} \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \\ &= \begin{pmatrix} \frac{\mathbf{O}}{\sqrt{2}} & \frac{\mathbf{O}}{\sqrt{2}} \\ \frac{-\mathbf{O}}{\sqrt{2}} & \frac{\mathbf{O}}{\sqrt{2}} \end{pmatrix}\begin{pmatrix} \mathbf{d}_{2-} & 0 \\ 0 & \mathbf{d}_{2+} \end{pmatrix}\begin{pmatrix} \frac{\mathbf{O}^{-1}}{\sqrt{2}} & \frac{-\mathbf{O}^{-1}}{\sqrt{2}} \\ \frac{\mathbf{O}^{-1}}{\sqrt{2}} & \frac{\mathbf{O}^{-1}}{\sqrt{2}} \end{pmatrix}, \end{aligned} \tag{17}$$

where

$$\mathbf{d}_{2-} = \begin{pmatrix} \alpha_+ - \alpha_- & 0 \\ 0 & \alpha_+ - \alpha_- - 2\alpha_\perp \end{pmatrix},$$

$$\mathbf{d}_{2+} = \begin{pmatrix} \alpha_+ - \alpha_- & 0 \\ 0 & \alpha_+ - \alpha_- + 2\alpha_\perp \end{pmatrix}.$$

The following matrices will be required to calculate the trace of the mth power of the matrix $\mathbf{T}_2$:

$$\begin{aligned} \mathbf{O}^{-1}\mathbf{s}\mathbf{O} &= \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \\ &= \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}, \end{aligned} \tag{18a}$$

$$\mathbf{O}^{-1}\mathbf{s}\mathbf{O}(\mathbf{d}_{2-})^m = \begin{pmatrix} 2(\alpha_+ - \alpha_-)^m & 0 \\ 0 & 0 \end{pmatrix}. \tag{18b}$$

Then, since the matrix $\mathbf{Q}_2$ is quasidiagonal, we obtain finally that the desired correlation function depends on the difference of the probability that the direction of motion is preserved and the probability that the direction changes by the angle π:

$$\begin{aligned} g(m) &= \langle \mathbf{e}(n)\mathbf{e}(n+m)\rangle \\ &= \frac{1}{4}\mathrm{Sp}(\mathbf{Q}_2\mathbf{T}_2^m) = (\alpha_+ - \alpha_-)^m. \end{aligned} \tag{19}$$

3.2. Three-Dimensional Problem. Cubic Lattice

The problem of random walk in a cubic lattice can also be solved completely in the three-dimensional case.

We shall enumerate the direction in a three-dimensional space (Fig. 2). As done for a two-dimensional square lattice, we introduce the probability distribution for the direction of motion at a step n

$$|P(n)\rangle = \begin{pmatrix} P_1(n) \\ P_2(n) \\ P_3(n) \\ P_4(n) \\ P_5(n) \\ P_6(n) \end{pmatrix}, \tag{20}$$

which are related at steps n and $n + 1$ by the transfer matrix

$$|P(n+1)\rangle = \mathbf{T}_3|P(n)\rangle, \quad \mathbf{T}_3 = \begin{pmatrix} \mathbf{A} & \mathbf{B} & \mathbf{B} \\ \mathbf{B} & \mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{B} & \mathbf{A} \end{pmatrix}, \tag{21}$$

where $\mathbf{A}$ and $\mathbf{B}$ are the 2×2 square matrices defined in Eq. (15). Next, we have

$$|P(n+m)\rangle = T_3^m|P(n)\rangle, \tag{22}$$

$$g(m) = \langle \mathbf{e}(n)\mathbf{e}(n+m)\rangle = \frac{1}{6}\mathrm{Sp}(\mathbf{Q}_3\mathbf{T}_3^m), \tag{23}$$

where the matrix $\mathbf{Q}_3$ (similarly to the matrix $\mathbf{Q}_2$) has a block-diagonal form:

$$\mathbf{Q}_3 = \begin{pmatrix} \mathbf{s} & 0 & 0 \\ 0 & \mathbf{s} & 0 \\ 0 & 0 & \mathbf{s} \end{pmatrix}, \quad \mathbf{s} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}. \tag{24}$$

The matrix $\mathbf{T}_3$ can be easily diagonalized:

$$\mathbf{T}_3 = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{-1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{-1}{\sqrt{3}} & 0 & \frac{1}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} \mathbf{A}-\mathbf{B} & 0 & 0 \\ 0 & \mathbf{A}-\mathbf{B} & 0 \\ 0 & 0 & \mathbf{A}+2\mathbf{B} \end{pmatrix}$$

$$\times \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{-2}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} & \frac{-2}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix} = \begin{pmatrix} \frac{\mathbf{O}}{\sqrt{3}} & \frac{\mathbf{O}}{\sqrt{3}} & \frac{\mathbf{O}}{\sqrt{3}} \\ 0 & \frac{-\mathbf{O}}{\sqrt{3}} & \frac{\mathbf{O}}{\sqrt{3}} \\ \frac{-\mathbf{O}}{\sqrt{3}} & 0 & \frac{\mathbf{O}}{\sqrt{3}} \end{pmatrix} \quad (25)$$

$$\times \begin{pmatrix} \mathbf{d}_{3-} & 0 & 0 \\ 0 & \mathbf{d}_{3-} & 0 \\ 0 & 0 & \mathbf{d}_{3+} \end{pmatrix} \begin{pmatrix} \frac{\mathbf{O}^{-1}}{\sqrt{3}} & \frac{\mathbf{O}^{-1}}{\sqrt{3}} & \frac{-2\mathbf{O}^{-1}}{\sqrt{3}} \\ \frac{\mathbf{O}^{-1}}{\sqrt{3}} & \frac{-2\mathbf{O}^{-1}}{\sqrt{3}} & \frac{\mathbf{O}^{-1}}{\sqrt{3}} \\ \frac{-\mathbf{O}^{-1}}{\sqrt{3}} & \frac{\mathbf{O}^{-1}}{\sqrt{3}} & \frac{\mathbf{O}^{-1}}{\sqrt{3}} \end{pmatrix}.$$

Here $\mathbf{d}_{3-} = \mathbf{d}_{2-}$, and

$$\mathbf{d}_{3+} = \begin{pmatrix} \alpha_+ - \alpha_- & 0 \\ 0 & \alpha_+ - \alpha_- + 4\alpha_\perp \end{pmatrix}.$$

Using the matrix (18a) and the matrix

$$\mathbf{O}^{-1}\mathbf{s}\mathbf{O}(\mathbf{d}_{3\pm})^m = \mathbf{O}^{-1}\mathbf{s}\mathbf{O}(\mathbf{d}_{2-})^m = \begin{pmatrix} 2\theta^m & 0 \\ 0 & 0 \end{pmatrix}, \quad (18c)$$

where $\theta = \alpha_+ - \alpha_-$, we obtain the following expression for the correlation function $g(m)$ for a three-dimensional cubic lattice:

$$g(m) = \langle \mathbf{e}(n)\mathbf{e}(n+m)\rangle = \frac{1}{6}\mathrm{Sp}(\mathbf{Q}_3\mathbf{T}_3^m) = (\alpha_+ - \alpha_-)^m. \quad (26)$$

It has turned out that the expressions for the correlation functions (19) and (26) are the same in the two- and three-dimensional cases. Therefore the rms size of a molecule in the three-dimensional case will also be described by the same expression (see Eq. (28) below). The only difference between the two- and three-dimensional problems is the normalization condition for the probabilities α_i:

$$\alpha_+ + 2\alpha_\perp + \alpha_- = 1 \text{ in the two-dimensional case,} \quad (27a)$$

$$\alpha_+ + 4\alpha_\perp + \alpha_- = 1 \text{ in the three-dimensional case.} \quad (27b)$$

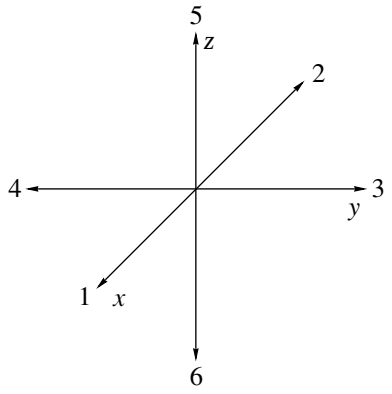


Fig. 2.

3.3. Size of an Oligomeric Chain as a Function of Its Length

Having determined the expression for the correlation function, we can also calculate the rms size of a molecule:

$$\langle \mathbf{R}^2(N)\rangle = l_0^2\left\{N + 2\sum_{n_1=1}^{N}\sum_{n_2=1}^{N-n_1}\langle \mathbf{e}(n_1)\cdot\mathbf{e}(n_1+n_2)\rangle\right\} = l_0^2\left\{\frac{1+\theta}{1-\theta}N - \frac{2\theta}{1-\theta}\frac{1-\theta^N}{1-\theta}\right\}. \quad (28)$$

Here the summation of the geometric progressions is performed twice. The parameter θ depends on the temperature (T) and the flexural stiffness of the chain (κ): $\theta \equiv \theta(T, \kappa) = \theta(\kappa/T)$.

We shall consider two limiting cases: the flexural stiffness of the chain is small, which corresponds to the restriction $\theta \ll 1$, and the flexural stiffness of the chain is large, which corresponds to the restriction $1 - \theta \equiv \varepsilon \ll 1$.

For small θ ($\theta \ll 1$, i.e., $\kappa/T \ll 1$) the dependence obtained describes the Gaussian ball:

$$\langle \mathbf{R}^2(N)\rangle \approx l_0^2 N\{1 + 2\theta(1 - 1/N)\}. \quad (29)$$

For θ close to 1 ($1 - \theta \equiv \varepsilon \ll 1$, i.e., $\kappa/T \gg 1$) and not very large N ($N \leq 1/\varepsilon$) the dependence obtained describes an essentially stiff rod:

$$\langle \mathbf{R}^2(N)\rangle \approx l_0^2 N^2\{1 - \varepsilon N/3\}. \quad (30)$$

However, even for θ close to 1 but quite large N ($1 - \theta = \varepsilon \ll 1$, $N \gg 1/\varepsilon$) the quadratic dependence of $\langle \mathbf{R}^2\rangle$ on N becomes linear in the asymptotic limit, which corresponds to a transition of the chain to the Gaussian ball.

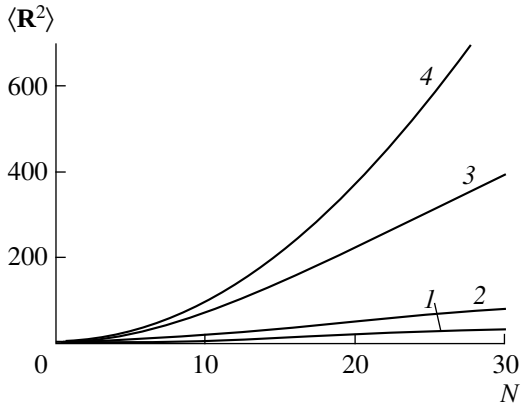


Fig. 3. Average squared size of a chain $\langle \mathbf{R}^2 \rangle$ versus the chain length N for θ = (*1*) 0.1, (*2*) 0.5, (*3*) 0.9, and (*4*) 0.99.

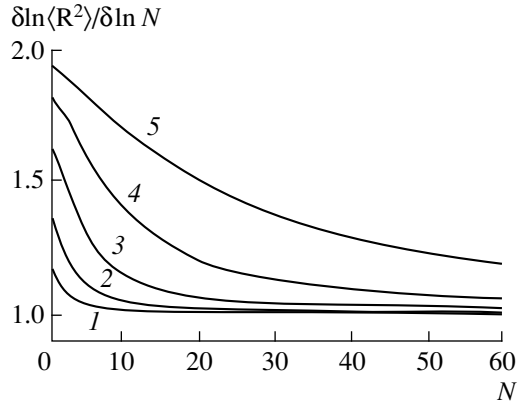


Fig. 4. Logarithmic derivative $(\partial \ln \langle \mathbf{R}^2 \rangle)/\partial \ln N$ of the average squared size of a chain with respect to the logarithm of the chain length, versus the chain length N for θ = (*1*) 0.1, (*2*) 0.25, (*3*) 0.5, (*4*) 0.75, and (*5*) 0.9.

The curves of the expression (28) for $\langle \mathbf{R}^2 \rangle$ versus N are displayed in Fig. 3 for some values of the parameter θ (θ = 0.1, 0.5, 0.9, 0.99). We note that the asymptotic expression (29) is valid for the curves *1* and *2*, and the expression (30) is valid for the curves *3* and *4*.

Having obtained the relation (19) $g(m) = \theta^m$, the number of links in a Kuhn segment can be estimated formally from the condition that the condition $g(m)$ is exponentially small: $n_K = -1/\ln\theta$. This gives $n_K \approx 1$ right up to $\theta = 0.4$ and $n_K \approx 10$ only for $\theta = 0.9$. However, if the scale on which the function $\langle \mathbf{R}^2(N) \rangle$ becomes linear is chosen as the criterion, then it is found that $n_K \approx 10$ for $\theta = 0.1$, $n_K \approx 15$ for $\theta = 0.25$, $n_K \approx 20$ for $\theta = 0.5$, and $n_K > 100$ for $\theta = 0.75$ and 0.9.

This is clearly seen in Fig. 4, which shows the logarithmic derivative $(\partial \ln \langle \mathbf{R}^2 \rangle)/\partial \ln N$ of the average squared size of a chain with respect to the logarithm of the chain length. The second criterion gives much larger values for the Kuhn segment than the first one. But it will be preferable if the fact that it is related with the scale at which the statistics of the polymer chain first becomes Gaussian is taken into account. This method for determining the Kuhn segment is close to the definition $l_0 n_K = \langle \mathbf{R}^2 \rangle / N$ presented in [2].

4. SPATIAL DISTRIBUTION OF THE LINKS IN A CHAIN

As already mentioned above, the computed dependence of the average length of a chain on its length yields information about the state of a linear molecule, but this information is far from complete. Now, we can return to the construction of the spatial distribution function of the links of a chain as defined above (see the expressions (2), (3), and (5)). This function contains complete information about the state of a chain.

In the calculations below we shall use the integral representation of the Kronecker delta function (more accurately, its Fourier transform):

$$\delta(m) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp(imq)\,dq. \tag{31}$$

Substituting this representation of the Kronecker delta function into the expression (3) we obtain

$$G(\mathbf{R}_N, N) = \frac{1}{(2\pi)^3}\int \exp(i\mathbf{m}_N \cdot \mathbf{q}) K(\mathbf{q}, N)\,d\mathbf{q}, \tag{32}$$

where the vectors $\mathbf{m}_N = \{m_{Nx}, m_{Ny}, m_{Nz}\}$ and $q = \{q_x, q_y, q_z\}$ and the function $K(\mathbf{q}, N)$ is the Fourier transform of the distribution function $G(\mathbf{R}_N, N)$ and is given by the expression

$$K(\mathbf{q}, N) = \left\langle \exp\left[i \sum_{n=1}^{N} (\sigma_n^x q_x + \sigma_n^y q_y + \sigma_n^z q_z) \right] \right\rangle, \tag{33}$$

where the brackets, just as in Section 1, signify averaging, i.e., summation over all possible values of σ_n^x, σ_n^y, $\sigma_n^z = \pm 1, 0$ satisfying the condition $(\sigma_n^x)^2 + (\sigma_n^y)^2 + (\sigma_n^z)^2 = 1$.

Since the probability that the vector of the numbers $\boldsymbol{\sigma}_{n+1}$ assumes a certain definite value depends on the value which the vector of numbers $\boldsymbol{\sigma}_n$ assumes, the averaging in the expression (33) will be performed using the functions $W(\boldsymbol{\sigma}_{n+1}, \boldsymbol{\sigma}_n)$, giving the conditional probabilities that the vector of numbers $\boldsymbol{\sigma}_{n+1}$ will assume the corresponding value provided that the vector of numbers $\boldsymbol{\sigma}_n$ possesses a prescribed value:

$$K(\mathbf{q}, N) = \sum_{\{\sigma\}} \prod_{n=2}^{N} \{ \exp(\boldsymbol{\sigma}_n \cdot \mathbf{q}) W(\boldsymbol{\sigma}_n; \boldsymbol{\sigma}_{n-1}) \} \exp(\boldsymbol{\sigma}_1 \cdot \mathbf{q}), \tag{34}$$

where $\{\boldsymbol{\sigma}\}$ means summation over all allowed sets of $\boldsymbol{\sigma}_n$ with $n = 1, 2, \ldots, N$.

Now, to each allowed vector of numbers $\boldsymbol{\sigma}_n$ we associate a direction in a cubic lattice in accordance with Fig. 2, i.e., the vector $\{1, 0, 0\}$ is associated to the direction 1, the vector $\{-1, 0, 0\}$ to the direction 2, the vector $\{0, 1, 0\}$ to the direction 3, the vector $\{0, -1, 0\}$ to the direction 4, the vector $\{0, 0, 1\}$ to the direction 5, and the vector $\{0, 0, -1\}$ to the direction 6. Then the expression (34) can be written using the matrix $\mathbf{T}_3$, introduced in Eq. (21), as

$$K_3(\mathbf{q}, N) = \sum_{j,k} [(\hat{\mathbf{q}}_3 \mathbf{T}_3)^{N-1} \hat{\mathbf{q}}_3]_{j,k}. \tag{35a}$$

Here $\hat{\mathbf{q}}_3$ is a 6×6 diagonal matrix:

$$\hat{\mathbf{q}}_3 = \begin{pmatrix} \hat{\mathbf{q}}_x & 0 & 0 \\ 0 & \hat{\mathbf{q}}_y & 0 \\ 0 & 0 & \hat{\mathbf{q}}_z \end{pmatrix},$$

$$\hat{\mathbf{q}}_x = \begin{pmatrix} e^{iq_x} & 0 \\ 0 & e^{-iq_x} \end{pmatrix}, \quad \hat{\mathbf{q}}_y = \begin{pmatrix} e^{iq_y} & 0 \\ 0 & e^{-iq_y} \end{pmatrix},$$

$$\hat{\mathbf{q}}_z = \begin{pmatrix} e^{iq_z} & 0 \\ 0 & e^{-iq_z} \end{pmatrix}.$$

In the two-dimensional case the corresponding expression has the form

$$K_2(\mathbf{q}, N) = \sum_{j,k} [(\hat{\mathbf{q}}_2 \mathbf{T}_2)^{N-1} \hat{\mathbf{q}}_2]_{j,k}. \tag{35b}$$

Here the matrix $\mathbf{T}_2$ is defined in Eq. (10), and $\hat{\mathbf{q}}_2$ is a 4×4 diagonal matrix:

$$\hat{\mathbf{q}}_2 = \begin{pmatrix} \hat{\mathbf{q}}_x & 0 \\ 0 & \hat{\mathbf{q}}_y \end{pmatrix},$$

$$\hat{\mathbf{q}}_x = \begin{pmatrix} e^{iq_x} & 0 \\ 0 & e^{-iq_x} \end{pmatrix}, \quad \hat{\mathbf{q}}_y = \begin{pmatrix} e^{iq_y} & 0 \\ 0 & e^{-iq_y} \end{pmatrix}.$$

In the one-dimensional case we have an even simpler expression:

$$K_1(\mathbf{q}, N) = \sum_{j,k} [(\hat{\mathbf{q}}\mathbf{A})^{N-1} \hat{\mathbf{q}}]_{j,k}. \tag{35c}$$

Here the matrix

$$\mathbf{A} = \begin{pmatrix} \alpha_+ & \alpha_- \\ \alpha_- & \alpha_+ \end{pmatrix}$$

was introduced in Eq. (15), and

$$\hat{\mathbf{q}} = \begin{pmatrix} e^{iq} & 0 \\ 0 & e^{-iq} \end{pmatrix}.$$

Introducing the column vectors

$$|K(N)\rangle_3 = \begin{pmatrix} K_1^{(3)}(N) \\ K_2^{(3)}(N) \\ K_3^{(3)}(N) \\ K_4^{(3)}(N) \\ K_5^{(3)}(N) \\ K_6^{(3)}(N) \end{pmatrix},$$

$$|K(N)\rangle_2 = \begin{pmatrix} K_1^{(2)}(N) \\ K_2^{(2)}(N) \\ K_3^{(2)}(N) \\ K_4^{(2)}(N) \end{pmatrix}, \quad |K(N)\rangle_1 = \begin{pmatrix} K_1^{(1)}(N) \\ K_2^{(1)}(N) \end{pmatrix},$$

the relations (35) can be written in the form

$$\begin{aligned} |K(\mathbf{q}, N)\rangle_3 &= (\hat{\mathbf{q}}_3 \mathbf{T}_3)^{N-1} |K(1)\rangle_3, \\ K_3(\mathbf{q}, N) &= \sum_{j=1}^{6} K_j^{(3)}(\mathbf{q}, N), \end{aligned} \tag{36a}$$

$$\begin{aligned} |K(\mathbf{q}, N)\rangle_2 &= (\hat{\mathbf{q}}_2 \mathbf{T}_2)^{N-1} |K(1)\rangle_2, \\ K_2(\mathbf{q}, N) &= \sum_{j=1}^{4} K_j^{(2)}(\mathbf{q}, N), \end{aligned} \tag{36b}$$

$$\begin{aligned} |K(\mathbf{q}, N)\rangle_1 &= (\hat{\mathbf{q}}_1 \mathbf{A})^{N-1} |K(1)\rangle_1, \\ K_1(\mathbf{q}, N) &= \sum_{j=1}^{2} K_j^{(1)}(\mathbf{q}, N), \end{aligned} \tag{36c}$$

where the column vectors $|K(N)\rangle_{1,2,3}$ correspond to the direction of the first step of the walk. Since the system is isotropic, without loss of generality we can set

$$K_j^{(3,2)}(1) = \delta_{j,1} \exp(iq_x), \quad K_j^{(1)}(1) = \delta_{j,1} \exp(iq).$$

Taking the inverse Fourier transform in the expressions (36), we obtain the spatial distribution function $G(\mathbf{R}_N, N)$ of the chain links in three- and one-dimensional spaces.

4.1. One-Dimensional Distribution Function

In the one-dimensional case the problem reduces to diagonalizing the 2×2 matrix $\hat{\mathbf{q}}\mathbf{A}$ defined in Eq. (35c).

Diagonalization, performed by the standard method, gives

$$\hat{\mathbf{q}}\mathbf{A} = \hat{\mathbf{O}}\begin{pmatrix} \lambda_+ & 0 \\ 0 & \lambda_- \end{pmatrix}\hat{\mathbf{O}}^{-1}, \tag{37}$$

where

$$\lambda_\pm = \alpha_+\left[\cos q \pm \sqrt{\left(\frac{\alpha_-}{\alpha_+}\right)^2 - \sin^2 q}\right],$$

$$\hat{\mathbf{O}} = \frac{1}{\sqrt{\Delta}}\begin{pmatrix} \alpha_- & \alpha_0 e^{iq} \\ -\alpha_0 e^{-iq} & \alpha_- \end{pmatrix},$$

$$\Delta = \alpha_-^2 + \alpha_0^2, \quad \alpha_0 = \alpha_+\left[i\sin q - \sqrt{\left(\frac{\alpha_-}{\alpha_+}\right)^2 - \sin^2 q}\right].$$

Substituting the expression (37) into Eqs. (36c), we obtain

$$K_1(q, N) = \frac{1}{\Delta}[(\alpha_-^2\lambda_+^{N-1} + \alpha_0^2\lambda_-^{N-1})e^{iq} + \alpha_0\alpha_-(\lambda_-^{N-1} - \lambda_+^{N-1})]. \tag{38}$$

Fourier transforming the expression (38) we obtain the desired distribution function in the one-dimensional case:

$$G_1(x_N, N) = \frac{1}{2\pi\Delta}\int_0^{2\pi}\{e^{im_N q}[(\alpha_-^2\lambda_+^{N-1} + \alpha_0^2\lambda_-^{N-1})e^{iq} + \alpha_0\alpha_-(\lambda_-^{N-1} - \lambda_+^{N-1})]dq\}, \tag{39}$$

where $x_N = l_0 m_N$. The explicit expression for the one-dimensional distribution function (39) is too complicated to present here.

We note that for isotropic walks, where $\alpha_+ = \alpha_- = 0.5$, the expression (38) becomes

$$K_1^{\text{isotr}}(q, N) = (\cos q)^{N-1}e^{iq}, \tag{40}$$

which after averaging over the direction of the first step becomes

$$\bar{K}_1^{\text{isotr}}(q, N) = (\cos q)^N. \tag{41}$$

The Fourier transform of the expression (41) gives a binomial distribution, which in the large N limit becomes a normal distribution, i.e., the ordinary Gaussian distribution function is obtained from Eq. (38) in the particular case of isotropic walks.

The expression (38) can be used to calculate the generating function for the distribution function:

$$\Gamma^{(1)}(x_N, \xi) = \frac{1}{2\pi}\int e^{im_N q}\tilde{\Gamma}^{(1)}(q, \xi)dq, \tag{42}$$

where

$$\tilde{\Gamma}^{(1)}(q, \xi) = 1 + \sum_{N=1}^{\infty}\xi^N K_1(q, N) = \frac{1 + [\alpha_- e^{iq} - \alpha_+ e^{-iq}]}{1 - 2\alpha_+\xi\cos q + (\alpha_+ - \alpha_-)\xi^2}.$$

Averaging over the direction of the first step we obtain

$$\bar{\tilde{\Gamma}}^{(1)}(q, \xi) = \frac{1 - (\alpha_+ - \alpha_-)\xi\cos q}{1 - 2\alpha_+\xi\cos q + (\alpha_+ - \alpha_-)\xi^2}, \tag{43}$$

which for $\alpha_+ = \alpha_- = 0.5$ is the same as the corresponding expression for isotopic walks [3, Chapter 7, Section 7.8]:

$$\bar{\tilde{\Gamma}}_{\text{isotr}}^{(1)}(q, \xi) = \frac{1}{1 - \xi\cos q}. \tag{44}$$

Knowing $\tilde{\Gamma}^{(1)}(q, \xi)$, the moments of the distribution function $G_1(x, N)$ can be easily calculated. Thus for $q = 0$ we obtain

$$\tilde{\Gamma}^{(1)}(0, \xi) = \frac{1}{1 - \xi},$$

i.e., all zero moments $M_0(N)$ are 1 in accordance with the normalization condition.

In contrast to isotropic walks, in the case of orientationally correlated walks the first moment of the distribution is different from zero and shows a displacement of the walking particle in the direction of the first step:

$$M_1 = \langle x(N)\rangle = \frac{l_0}{2\pi}\oint_{|\xi|=1}\frac{d\xi}{\xi^{N+1}}\frac{\partial}{\partial q}\tilde{\Gamma}^{(1)}(q, \xi)\Big|_{q=0} = \frac{l_0}{2\pi i}\oint_{|\xi|=1}\frac{d\xi}{\xi^N(1-\xi)[1-(\alpha_+ - \alpha_-)\xi]} = l_0\frac{1-\theta^N}{1-\theta}. \tag{45}$$

Here the contour integral in the complex ξ plane was used; the integration contour is a circle with radius $R = 1 - 0$ centered at the point $\xi = 0$.

The second moment can be calculated similarly:

$$\langle x^2(N)\rangle = -\frac{l_0^2}{2\pi i}\oint_{|\xi|=1}\frac{d\xi}{\xi^{N+1}}\frac{\partial^2}{\partial q^2}\tilde{\Gamma}^{(1)}(q, \xi)\Big|_{q=0} = -\frac{l_0^2}{2\pi i}\oint_{|\xi|=1}\frac{[1 + (\alpha_+ - \alpha_-)\xi]d\xi}{(1-\xi)^2[1-(\alpha_+ - \alpha_-)\xi]} = l_0^2\left[\frac{1+\theta}{1-\theta}N - \frac{2\theta}{1-\theta}\frac{1-\theta^N}{1-\theta}\right], \tag{46}$$

which is identical to the expression (28).

Analysis of the expressions (39), (42), (43), and (46) shows that for $\alpha_+ \neq \alpha_-$ the distribution function (39) will be different from a Gaussian function even in the

asymptotic limit for large N, though the dependence $\langle x^2(N)\rangle$ for large N becomes linear.

In closing this section we note that the distribution function can be expressed in terms of the partition function of a statistical system. Specifically, in the one-dimensional case we arrive at the well-studied classic Ising model in the one-dimensional variant:

$$G_1(x_N, N) = \frac{1}{2\pi} \times \int_0^{2\pi} \exp(im_N q) \sum_{\sigma = \pm 1} [Z(q, N)]_{\sigma, 1} dq, \tag{47}$$

where $Z(q, N)$ is the partition function of a one-dimensional Ising chain:

$$[Z(q, N)]_{\sigma_1\sigma_N} = \sum_{\sigma_2, \ldots, \sigma_{N-1} = \pm 1} \exp\left\{ iq \sum_{n=1}^{N} \sigma_n - V \sum_{n=1}^{N-1} \sigma_n\sigma_{n+1} \right\}, \tag{48}$$

$V = 0.5\ln(\alpha_-/\alpha_+)$. The calculation of the partition function (48) presents no difficulties, as was demonstrated above, similarly to, for example, [17].

4.2. Distribution Function in a Plane and in Space

To calculate the two-dimensional distribution function the Nth power of a 4×4 matrix must be calculated. To diagonalize this matrix it is necessary to find the roots of a quartic polynomial. In principle, the analytic expressions required can be written out using the Cardano formulas. However, these expressions are so complicated that it is essentially unrealistic to work with them. Moreover, a 6×6 matrix has to be diagonalized and the roots of a polynomial of degree six must be found in order to construct the distribution function in the three-dimensional case. In the general case this problem has no analytic solution. Consequently, we shall not use the diagonalization procedure to construct the two-dimensional distribution function (more accurately, its generating function).

The representation of the Fourier transform of the distribution function in the form of the partition function of a one-dimensional system will suggest to us how to calculate the N power of a matrix. Using the block form of the matrices $\mathbf{T}_2$ and $\hat{\mathbf{q}}_2$ we can write

$$(\mathbf{T}_2)_{ij} = (\mathbf{T}_2)_{\alpha\alpha'}^{\beta\beta'},$$
$$(\hat{\mathbf{q}}_2)_{ij} = (\hat{\mathbf{q}}_2)_{\alpha\alpha'}^{\beta\beta'} = \delta_{\alpha\alpha'}(\hat{\mathbf{q}}_2)_{\alpha}^{\beta\beta'} = \delta_{\alpha\alpha'}\delta_{\beta\beta'}(\hat{\mathbf{q}}_2)_{\alpha}^{\beta}, \tag{49}$$

where $i, j = 1, \ldots, 4$; $\alpha, \alpha', \beta, \beta' = 1, 2$,

$$(\mathbf{T}_2)_{11}^{\beta\beta'} = (\mathbf{T}_2)_{22}^{\beta\beta'} = \mathbf{A}_{\beta\beta'},$$
$$(\mathbf{T}_2)_{12}^{\beta\beta'} = (\mathbf{T}_2)_{21}^{\beta\beta'} = \mathbf{B}_{\beta\beta'},$$
$$(\hat{\mathbf{q}}_2)_1^{\beta\beta'} = (\hat{\mathbf{q}}_x)_{\beta\beta'}, \quad (\hat{\mathbf{q}}_2)_2^{\beta\beta'} = (\hat{\mathbf{q}}_y)_{\beta\beta'}.$$

Next, we represent the matrix elements of the matrices $\mathbf{T}_2$ and $\hat{\mathbf{q}}_2$ in the form

$$(\mathbf{T}_2)_{\alpha\alpha'}^{\beta\beta'} = \exp\left\{ \frac{1 + \nu\nu'}{4} \left[\ln\left(\frac{\alpha_+\alpha_-}{\alpha_\perp^2}\right) + \sigma\sigma' \ln\left(\frac{\alpha_+}{\alpha_-}\right) \right] + \ln\alpha_\perp \right\}, \tag{50}$$
$$(\hat{\mathbf{q}}_2)_\alpha^\beta = \exp\left\{ \frac{i\sigma}{2}[(q_x + q_y) + \nu(q_x - q_y)] \right\}.$$

Here the indices $\nu, \nu' = \pm 1$ correspond to the indices α, $\alpha' = 1, 2$, and the indices $\sigma, \sigma' = \pm 1$ correspond to the indices $\beta, \beta' = 1, 2$. Using the expression (50) and introducing the notation

$$p_\pm = \frac{1}{2}(q_x \pm q_y), \quad g = \frac{1}{2}\ln\left(\frac{\alpha_+\alpha_-}{\alpha_\perp^2}\right), \quad h = \frac{1}{2}\ln\left(\frac{\alpha_+}{\alpha_-}\right),$$

$K_2(\mathbf{q}, N)$ can be represented as a partition function of a one-dimensional chain of two types of "spin-1/2" particles:

$$K_2(\mathbf{q}, N) = \alpha_\perp^{N-1} \sum_{\substack{\sigma_2, \ldots, \sigma_N = \pm 1 \\ \nu_2, \ldots, \nu_N = \pm 1}} \exp\left\{ i \sum_{n=1}^{N} \sigma_n(p_+ + p_-\nu_n) + \sum_{n=1}^{N-1} \frac{1}{2}(1 + \nu_n\nu_{n+1})(g + h\sigma_n\sigma_{n+1}) \right\}. \tag{51}$$

The structure of the expression obtained for the partition function is such that if the nearest neighbors are different kinds of particles ($\nu_n \neq \nu_{n+1}$), then they do not interact. This means that the chain consists of two types of alternating noninteracting chains, and the calculation of the partition function (51) reduces to averaging over the lengths of these chains (n_j^x and n_j^y) and their number (m) under the condition that the total length of the entire chain is fixed: $\sum n_j^x + \sum n_j^y = N$.

The partition of the entire chain into a collection of two types of alternating noninteracting chains means, in the language of walk trajectories, that first n_1^x steps are taken in the direction x, n_1^y steps are taken in the direction y, n_2^x steps are taken in the direction x, n_2^y steps are taken in the direction y, and so on. The follow-

ing terms correspond to each such trajectory in the partition function (51):

$$\sum_{i=1,2}\left\{(\hat{\mathbf{q}}_y\mathbf{A})^{n_m^y}(\hat{\mathbf{q}}_y\mathbf{B})(\hat{\mathbf{q}}_x\mathbf{A})^{n_m^x-1}(\hat{\mathbf{q}}_x\mathbf{B})\ldots \times(\hat{\mathbf{q}}_y\mathbf{A})^{n_1^y-1}(\hat{\mathbf{q}}_y\mathbf{B})(\hat{\mathbf{q}}_x\mathbf{A})^{n_1^x-1}\hat{\mathbf{q}}_x\right\}_{i,1}, \tag{52a}$$

if the last steps were taken in the direction *y*, and

$$\sum_{i=1,2}\left\{(\hat{\mathbf{q}}_x\mathbf{A})^{n_{m+1}^x-1}(\hat{\mathbf{q}}_x\mathbf{B})(\hat{\mathbf{q}}_y\mathbf{A})^{n_m^y-1}(\hat{\mathbf{q}}_y\mathbf{B})(\hat{\mathbf{q}}_x\mathbf{A})^{n_m^x-1} \times(\hat{\mathbf{q}}_x\mathbf{B})\ldots(\hat{\mathbf{q}}_y\mathbf{A})^{n_1^y-1}(\hat{\mathbf{q}}_y\mathbf{B})(\hat{\mathbf{q}}_x\mathbf{A})^{n_1^x-1}\hat{\mathbf{q}}_x\right\}_{i,1}, \tag{52b}$$

if the last steps were taken in the direction *x*.

Summing the expressions (52) over all positive n_j^x, n_j^y, and *m* with the condition $\sum n_j^x + \sum n_j^x = N$, we obtain the desired expression for $K_2(\mathbf{q}, N)$:

$$K_2(\mathbf{q},N) = \sum_{\substack{m\geq 1\\ \{n_j^x\},\{n_j^y\}\geq 1}}\left\{\delta\left(\sum_{j=1}^m n_j^x+\sum_{j=1}^m n_j^y-N\right)\right.$$
$$\left.+\,\delta\left(\sum_{j=1}^{m+1} n_j^x+\sum_{j=1}^m n_j^y-N\right)\langle 1,1|(\hat{\mathbf{q}}_x\mathbf{A})^{n_{m+1}^x-1}\alpha_\perp\hat{\mathbf{q}}_x|1,1\rangle\right\}$$
$$\times\langle 1,1|(\hat{\mathbf{q}}_y\mathbf{A})^{n_m^y-1}\alpha_\perp\hat{\mathbf{q}}_y|1,1\rangle \times\langle 1,1|(\hat{\mathbf{q}}_x\mathbf{A})^{n_m^x-1}\alpha_\perp\hat{\mathbf{q}}_x|1,1\rangle\ldots \times\langle 1,1|(\hat{\mathbf{q}}_y\mathbf{A})^{n_1^y-1}\alpha_\perp\hat{\mathbf{q}}_y|1,1\rangle\langle 1,1|(\hat{\mathbf{q}}_x\mathbf{A})^{n_1^x-1}\hat{\mathbf{q}}_x|1,0\rangle. \tag{53}$$

Here $\langle 1,1|$, $|1,1\rangle$ and $|1,0\rangle$ are the row vector (11) and the column vectors $\begin{pmatrix}1\\1\end{pmatrix}$ and $\begin{pmatrix}1\\0\end{pmatrix}$.

Using the integral representation for the δ function, the expression (53) can be represented in the form

$$K_2(\mathbf{q},N) = \sum_{m,\{n_j^x\},\{n_j^y\}\geq 1}\int_0^{2\pi}\left\{\left[1+\langle 1,1|\right.\right.$$
$$\left.\times(e^{i\omega}\hat{\mathbf{q}}_x\mathbf{A})^{n_{m+1}^x-1}\alpha_\perp e^{i\omega}\hat{\mathbf{q}}_x|1,1\rangle\right]$$
$$\times\langle 1,1|(e^{i\omega}\hat{\mathbf{q}}_x\mathbf{A})^{n_m^x-1}\alpha_\perp e^{i\omega}\hat{\mathbf{q}}_x|1,1\rangle$$
$$\times\langle 1,1|(e^{i\omega}\hat{\mathbf{q}}_y\mathbf{A})^{n_m^y-1}\alpha_\perp e^{i\omega}\hat{\mathbf{q}}_y|1,1\rangle\ldots$$
$$\times\langle 1,1|(e^{i\omega}\hat{\mathbf{q}}_y\mathbf{A})^{n_1^y-1}\alpha_\perp e^{i\omega}\hat{\mathbf{q}}_y|1,1\rangle$$
$$\left.\times\langle 1,1|(e^{i\omega}\hat{\mathbf{q}}_x\mathbf{A})^{n_1^x-1}e^{i\omega}\hat{\mathbf{q}}_x|1,0\rangle e^{-iN\omega}\frac{d\omega}{2\pi}\right\}. \tag{54}$$

Going over in Eq. (54) from integration with respect to the variable ω to a contour integral with respect to the variable $\xi = \exp(i\omega)$ (the integration contour is a circle with radius $R = 1 - 0$ centered at the point $\xi = 0$), we obtain

$$K_2(\mathbf{q},N) = \frac{1}{2\pi i}\times\sum_{\substack{m\geq 1\\ \{n_j^x\},\{n_j^x\}\geq 1}}\oint_{|\xi|=1}\left\{[1+\langle 1,1|(\xi\hat{\mathbf{q}}_x\mathbf{A})^{n_{m+1}^x}\alpha_\perp\xi\hat{\mathbf{q}}_x|1,1\rangle]\right.$$
$$\times\langle 1,1|(\xi\hat{\mathbf{q}}_y\mathbf{A})^{n_m^y-1}\alpha_\perp\xi\hat{\mathbf{q}}_y|1,1\rangle \times\langle 1,1|(\xi\hat{\mathbf{q}}_x\mathbf{A})^{n_m^x-1}\alpha_\perp\xi\hat{\mathbf{q}}_x|1,1\rangle\ldots \times\langle 1,1|(\xi\hat{\mathbf{q}}_y\mathbf{A})^{n_1^y-1}\alpha_\perp\xi\hat{\mathbf{q}}_y|1,1\rangle$$
$$\left.\times\langle 1,1|(\xi\hat{\mathbf{q}}_x\mathbf{A})^{n_1^x-1}\xi\hat{\mathbf{q}}_x|1,0\rangle\right\}\xi^{-(N+1)}d\xi. \tag{55}$$

Since in Eq. (55) the restriction on the summation range in the expressions

$$\sum_{n\geq 1}\langle 1,1|(\xi\hat{\mathbf{q}}_{x,y}\mathbf{A})^{n-1}\alpha_\perp\xi\hat{\mathbf{q}}_{x,y}|1,1\rangle,$$

$$\sum_{n\geq 1}\langle 1,1|(\xi\hat{\mathbf{q}}_{x,y}\mathbf{A})^{n-1}\xi\hat{\mathbf{q}}_{x,y}|1,0\rangle$$

is lifted, these sums can be found similarly to the way that the generating functions $\tilde{\Gamma}^{(1)}(q,\xi)$ and $\bar{\tilde{\Gamma}}^{(1)}(q,\xi)$ were calculated [see Eqs. (42) and (43)]:

$$P(q,\xi) = \sum_{n=1}^{\infty}\langle 1,1|(\xi\hat{\mathbf{q}}\mathbf{A})^{n-1}\xi\hat{\mathbf{q}}|1,0\rangle = \frac{\xi[\exp(iq)-(\alpha_+-\alpha_-)\xi]}{1-2\alpha_+\xi\cos q+(\alpha_+^2-\alpha_-^2)\xi^2}. \tag{56a}$$

$$\bar{P}(q,\xi) = \sum_{n=1}^{\infty}\langle 1,1|(\xi\hat{\mathbf{q}}\mathbf{A})^{n-1}\alpha_\perp\xi\hat{\mathbf{q}}|1,1\rangle = \frac{2\alpha_\perp\xi[\cos q-(\alpha_+-\alpha_-)\xi]}{1-2\alpha_+\xi\cos q+(\alpha_+^2-\alpha_-^2)\xi^2}. \tag{56b}$$

Substituting the expressions (56) into Eq. (55) we obtain

$$K_2(\mathbf{q}, N) = \frac{1}{2\pi i}\oint\limits_{|\xi|=1}\left\{\frac{d\xi}{\xi^{N+1}}(1+\bar{P}(q_y,\xi))\right.$$
$$\left.\times\sum_{m=0}^{\infty}[\bar{P}(q_x,\xi)\bar{P}(q_y,\xi)]^m P(q_x,\xi)\right\} \quad (57)$$
$$= \frac{1}{2\pi i}\oint\limits_{|\xi|=1}\frac{(1+\bar{P}(q_y,\xi))P(q_x,\xi)}{1-\bar{P}(q_x,\xi)\bar{P}(q_y,\xi)}\frac{d\xi}{\xi^{N+1}}.$$

Here a summation over a geometric progression has been performed.

The integrand in the expression (57) determines the Fourier representation of the generating function for the distribution function:

$$\tilde{\Gamma}^{(2)}(\mathbf{q},\xi) = 1+\frac{[1+\bar{P}(q_y,\xi)]P(q_x,\xi)}{1-\bar{P}(q_x,\xi)\bar{P}(q_y,\xi)}. \quad (58)$$

Averaging over the direction of the first step we obtain

$$\bar{\tilde{\Gamma}}^{(2)}(\mathbf{q},\xi) = 1$$
$$+\frac{\bar{P}(q_x,\xi)+\bar{P}(q_y,\xi)+2\bar{P}(q_x,\xi)\bar{P}(q_y,\xi)}{4\alpha_\perp\{1-\bar{P}(q_x,\xi)\bar{P}(q_y,\xi)\}}. \quad (59)$$

For isotropic walks with $\alpha_+ = \alpha_- = \alpha_\perp = 1/4$ this expression becomes

$$\bar{\tilde{\Gamma}}^{(2)}_{\text{isotr}}(\mathbf{q},\xi) = \left[1-\frac{\xi}{2}(\cos q_x+\cos q_y)\right]^{-1}, \quad (60)$$

which is identical to the corresponding expression for isotropic walks in a plane [3, Chapter 7, Section 7.8].

Just as in the one-dimensional case, it is easy to verify that

$$\tilde{\Gamma}^{(2)}(\mathbf{q},\xi)\big|_{q=0} = 1+\frac{\xi}{1-\xi} = \frac{1}{1-\xi} = \sum_{N=0}^{N}\xi^N \quad (61)$$

(i.e., the distribution function is normalized to 1),

$$\langle x(N)\rangle = \frac{l_0}{2\pi}\oint\limits_{|\xi|=1}\frac{d\xi}{\xi^{N+1}}\frac{\partial}{\partial q_x}\tilde{\Gamma}^{(2)}(\mathbf{q},\xi)\bigg|_{\mathbf{q}=0} = \frac{l_0}{2\pi i} \quad (62a)$$
$$\times\oint\limits_{|\xi|=1}\frac{d\xi}{\xi^N(1-\xi)[1-(\alpha_+-\alpha_-)\xi]} = l_0\frac{1-\theta^N}{1-\theta},$$

$$\langle y(N)\rangle = \frac{l_0}{2\pi}\oint\limits_{|\xi|=1}\frac{d\xi}{\xi^{N+1}}\frac{\partial}{\partial q_y}\tilde{\Gamma}^{(2)}(\mathbf{q},\xi)\bigg|_{\mathbf{q}=0} = 0 \quad (62b)$$

(i.e., the first moment in the direction of the first step is different from zero),

$$\langle\mathbf{R}^2(N)\rangle = -\frac{l_0^2}{2\pi i}\oint\limits_{|\xi|=1}\frac{d\xi}{\xi^{N+1}}\left(\frac{\partial^2}{\partial q_x^2}+\frac{\partial^2}{\partial q_y^2}\right)\tilde{\Gamma}^{(2)}(\mathbf{q},\xi)\bigg|_{\mathbf{q}=0}$$
$$= -\frac{l_0^2}{2\pi i}\oint\limits_{|\xi|=1}\frac{[1+(\alpha_+-\alpha_-)\xi]d\xi}{\xi^N(1-\xi)^2[1-(\alpha_+-\alpha_-)\xi]} \quad (63)$$
$$= l_0^2\left[\frac{1+\theta}{1-\theta}N-\frac{2\theta}{1-\theta}\frac{1-\theta^N}{1-\theta}\right]$$

(i.e., the dependence of the rms displacement is identical to the one-dimensional case; see Eq. (46)).

The construction of a generating function for the distribution function in three-dimensional space is entirely analogous. The only difference is that we shall average over walk trajectories of the following type: first n_1^x steps are taken in the one-dimensional subspace x, then n_1^{yz} steps in the plane yz, then n_2^x steps once again in the one-dimensional subspace x, then n_2^{yz} steps once again in the plane yz, and so on.

Averaging over such trajectories we obtain that the generating function $\tilde{\Gamma}^{(3)}(\mathbf{q},\xi)$ for the distribution function in three-dimensional space has a form similar to the two-dimensional function $\tilde{\Gamma}^{(2)}(\mathbf{q},\xi)$:

$$\tilde{\Gamma}^{(3)}(\mathbf{q},\xi) = 1+\frac{[1+\bar{P}(q_y,q_z,\xi)]P(q_x,\xi)}{1-\bar{P}(q_x,\xi)\bar{P}(q_y,q_z,\xi)}, \quad (64)$$

where the propagator $P(q_x, q_y, \xi)$ averaged over the direction of the first step is

$$\bar{P}(q_y,q_z,\xi) = \frac{[1+\bar{P}(q_z,\xi)]\bar{P}(q_y,\xi)}{1-\bar{P}(q_y,\xi)\bar{P}(q_z,\xi)}$$
$$+\frac{[1+\bar{P}(q_y,\xi)]\bar{P}(q_z,\xi)}{1-\bar{P}(q_y,\xi)\bar{P}(q_z,\xi)}.$$

It is easy to check that the generating function (64) in the three-dimensional space possesses the same properties as the generating function (58) in the plane and (42) in the one-dimensional space:

$$\bar{\tilde{\Gamma}}^{(3)}_{\text{isotr}}(\mathbf{q},\xi) = \left[1-\frac{\xi}{6}(\cos q_x+\cos q_y+\cos q_z)\right]^{-1}, \quad (65)$$

$$\tilde{\Gamma}^{(3)}(\mathbf{q},\xi)\big|_{\mathbf{q}=0} = 1+\frac{\xi}{1-\xi} = \frac{1}{1-\xi} = \sum_{N=0}^{\infty}\xi^N, \quad (66)$$

$$\langle x(N)\rangle = \frac{l_0}{2\pi}\oint\limits_{|\xi|=1}\frac{d\xi}{\xi^{N+1}}\frac{\partial}{\partial q_x}\tilde{\Gamma}^{(3)}(\mathbf{q},\xi)\Big|_{\mathbf{q}=0} = \frac{l_0}{2\pi i} \tag{67a}$$

$$\times\oint\limits_{|\xi|=1}\frac{d\xi}{\xi^N(1-\xi)[1-(\alpha_+-\alpha_-)\xi]} = l_0\frac{1-\theta^N}{1-\theta},$$

$$\langle y, z(N)\rangle = \frac{l_0}{2\pi}\oint\limits_{|\xi|=1}\frac{d\xi}{\xi^{N+1}}\frac{\partial}{\partial q_{y,z}}\tilde{\Gamma}^{(3)}(\mathbf{q},\xi)\Big|_{\mathbf{q}=0} = 0, \tag{67b}$$

$$\langle \mathbf{R}^2(N)\rangle = -\frac{l_0^2}{2\pi i}$$

$$\times\oint\limits_{|\xi|=1}\frac{d\xi}{\xi^{N+1}}\left(\frac{\partial^2}{\partial q_x^2}+\frac{\partial^2}{\partial q_y^2}+\frac{\partial^2}{\partial q_z^2}\right)\tilde{\Gamma}^{(3)}(\mathbf{q},\xi)\Big|_{\mathbf{q}=0} \tag{68}$$

$$= -\frac{l_0^2}{2\pi i}\oint\limits_{|\xi|=1}\frac{[1+(\alpha_+-\alpha_-)\xi]d\xi}{\xi^N(1-\xi)^2[1-(\alpha_+-\alpha_-)\xi]}$$

$$= l_0^2\left[\frac{1+\theta}{1-\theta}N-\frac{2\theta}{1-\theta}\frac{1-\theta^N}{1-\theta}\right].$$

5. ORIENTATIONAL SELF-ORGANIZATION OF A RIGID-CHAIN POLYMERIC MELT

As already mentioned above, the non-Gaussian character of the spatial distribution of chain links on small scales (the existence of extended almost rectilinear sections) can result in the appearance of orientational ordering in a concentrated rigid-chain polymeric system—a melt. The variant, presented in the present paper, of the model of orientationally correlated walks makes it possible to describe similar effects. Without giving an exhaustive description of this phenomenon, we shall show as an example the conditions under which a liquid-crystal state arises in a two-dimensional rigid-chain polymeric system and how transitions occur between ordered and disordered states.

Let us consider the statistics of the conformations of a rigid chain in a polymeric melt taking account of the influence of the environment in a two-dimensional space. Since we shall study chains with finite flexural stiffness, each chain can be divided, in accordance with generally accepted ideas, into almost rectilinear sections—Kuhn segments, whose orientations are noncollinear. This means that the system under study consists of anisotropic elements. If orientational ordering of these anisotropic elements (Kuhn segments of chains) is observed in a polymer melt, then for the packing of the trial chain both the orientation of neighboring links along the chain and the orientation of links of surrounding chains will influence the orientation of each link.

The degree of orientational ordering of the system can be characterized by an order parameter which we shall define as the difference of the fraction of the links $n_x = N_x/N$ oriented in the direction x and the fraction of the links $n_y = N_y/N$ ($N = N_x + N_y$) oriented in the direction y:

$$\eta = \eta_x - \eta_y = \frac{N_x - N_y}{N} = \frac{N_x - N_y}{N_x + N_y}. \tag{69}$$

If all links of polymer chains are oriented in the direction x, then the order parameter $\eta = 1$; if they are oriented in the direction y, then $\eta = -1$; and, if the system is isotropic as a whole, then $\eta = 0$. We note that with this definition of orientational ordering of a system (when each link is not a vector but rather a director), a high degree of ordering of the system could mean complete rectification of the chains or formation of a folded structure.

To describe the statistics of the conformations of the trial chain, taking account of the orientational effect of the environment, we introduce the transfer matrix which relates the column vectors of the probability distribution of the orientations of the nth and $(n+1)$st links taking account of the orientation of the environment:

$$\hat{\mathbf{T}}_2(\eta) = \mathbf{T}_2 + \mathbf{T}_2^{(1)}(\eta) = \begin{pmatrix} \mathbf{A}(\eta) & \mathbf{B}(-\eta) \\ \mathbf{B}(\eta) & \mathbf{A}(-\eta) \end{pmatrix} \tag{70}$$

$$= \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{pmatrix} + \begin{pmatrix} \mathbf{A}_1(\eta) & \mathbf{B}_1(\eta) \\ \mathbf{B}_1(\eta) & \mathbf{A}_1(-\eta) \end{pmatrix}.$$

Here $\mathbf{T}_2$ is a block matrix, defined in Eq. (10), which relates the column vectors of the probability distribution of the orientations of the nth and $(n+1)$st links in the absence of an orienting effect due to the environment,

$$\mathbf{A} = \begin{pmatrix} \alpha_+ & \alpha_- \\ \alpha_- & \alpha_+ \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} \alpha_\perp & \alpha_\perp \\ \alpha_\perp & \alpha_\perp \end{pmatrix},$$

α_+, $\alpha_\perp$, and α_- are, respectively, the probabilities of the orientation being preserved and of rotations by 90° and 180°, respectively; the block matrix $\mathbf{T}_2^{(1)}(\eta)$ takes account of the orienting effect of the environment,

$$\mathbf{A}_1(\eta) = \begin{pmatrix} \delta_+(\eta) & \delta_-(\eta) \\ \delta_-(\eta) & \delta_+(\eta) \end{pmatrix},$$

$\delta_+(\eta)$ and $\delta_-(\eta)$ are, respectively, the changes in the probability of the orientation being preserved and of a rotation by 180° as a result of the orienting influence of

the environment in the case where a preceding link was oriented in the positive x direction;

$$\mathbf{B}_1(\eta) = -\frac{\delta_+(\eta) + \delta_-(\eta)}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

It is obvious that the functions $\delta_\pm(\eta)$ introduced above should grow monotonically in the range (–1, 1), and at the boundaries of this range and at $\eta = 0$ they should satisfy the conditions

$$\delta_\pm(\eta) = \begin{cases} -\alpha_\pm, & \eta = -1, \\ 0, & \eta = 0, \\ \alpha_\perp \pm \gamma, & \eta = 1, \end{cases} \tag{71}$$

the first of which signifies impossibility of maintaining an orientation in the direction x, if the entire environment is oriented in the direction y, and the latter means impossibility of rotation by 90°, if the entire environment is oriented in the direction x. Since $\alpha_\pm + \delta_\pm(\eta)$ is a probability, the condition $0 \le \alpha_\pm + \delta_\pm(\eta) \le 1$ is satisfied for any η. Consequently, the parameter γ, determining the relation between the rectilinear and folded conformations of the chain in a highly oriented state, should satisfy the inequality $-(\alpha_+ + \alpha_\perp) \le \gamma \le \alpha_- + \alpha_\perp$. Further calculations show that the parameter γ does not enter in the equation for determining the degree of orientational ordering of a system. This is because these two structures make the same contribution to the degree of ordering of a system determined in the manner chosen.

The transfer matrix $\hat{\mathbf{T}}_2(\eta)$ (70) can be used to find the probability distribution for the orientation of all chain links as a function of the orientational ordering of the environment. The value of the order parameter η (orientational ordering of the environment) can be calculated using the same matrix $\hat{\mathbf{T}}_2(\eta)$. The result is a relation that expresses the order parameter in terms of the functions $\delta_\pm(\eta)$, and the value of the order parameter can be determined from this relation in the self-consistent field approximation.

Indeed, since the environment of the trial chain consists of similar chains, the orientational ordering of the environment can be calculated by considering an infinitely long chain whose links have orientations which satisfy the same probability distribution as in the trial chain. Then, assuming for definiteness that the first link of an infinitely long chain corresponding to the environment of the trial chain is directed along the positive x direction, we obtain that the order parameter η can be expressed in terms of the matrix elements of the matrix $\hat{\mathbf{T}}_2(\eta)$ as follows:

$$\eta = \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\{[(\hat{\mathbf{T}}_2(\eta))^n]_{11} + [(\hat{\mathbf{T}}_2(\eta))^n]_{21} - [(\hat{\mathbf{T}}_2(\eta))^n]_{31} - [(\hat{\mathbf{T}}_2(\eta))^n]_{41}\}. \tag{72}$$

To calculate the infinite sum in Eq. (72) it is necessary to diagonalize the matrix $\hat{\mathbf{T}}_2(\eta)$. It turns out that the required diagonalization can be performed for an arbitrary value of the parameter η similarly to the way this was done for the matrix $\mathbf{T}_2$:

$$\hat{\mathbf{T}}_2(\eta) = \mathbf{U}\begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & \lambda_4 \end{pmatrix}\mathbf{U}^{-1}, \tag{73}$$

where

$$\mathbf{U}^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \dfrac{2\alpha_\perp - \beta_+(\eta)}{\rho} & 0 & \dfrac{-[2\alpha_\perp - \beta_+(-\eta)]}{\rho} \\ 0 & 0 & 1 & 0 \\ 0 & 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{pmatrix},$$

$$\mathbf{U} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/\sqrt{2} & 0 & \dfrac{2\alpha_\perp - \beta_+(-\eta)}{\rho} \\ 0 & 0 & 1 & 0 \\ 0 & -1/\sqrt{2} & 0 & \dfrac{2\alpha_\perp - \beta_+(\eta)}{\rho} \end{pmatrix}$$

are the diagonalizing matrices, $\rho \equiv \sqrt{2}\{2\alpha_\perp - [\beta_+(\eta) + \beta_+(-\eta)]/2\}$, and $\lambda_1 = \alpha_+ - \alpha_- + \beta_-(\eta)$, $\lambda_2 = 1 - 4\alpha_\perp + \beta_+(\eta) + \beta_+(-\eta)$, $\lambda_3 = \alpha_+ - \alpha_- + \beta_-(-\eta)$, and $\lambda_4 = \alpha_+ + \alpha_- + 2\alpha_\perp = 1$ are the eigenvalues of the matrix $\hat{\mathbf{T}}_2(\eta)$.

For diagonalization of the matrix $\hat{\mathbf{T}}_2(\eta)$ the calculation of the infinite system (72) reduces to summation of geometric progressions with exponents which are eigenvalues of the matrix $\hat{\mathbf{T}}_2(\eta)$. Since $|\lambda_1|, |\lambda_2|, |\lambda_3| < 1$, and $\lambda_4 = 1$, taking the limit we obtain

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}(\lambda_{1,2,3})^n = \lim_{N\to\infty}\frac{1}{N}\frac{\lambda_{1,2,3}}{1-\lambda_{1,2,3}} = 0,$$

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}(\lambda_4)^n = 1.$$

Simple calculations give the following self-consistency condition for the order parameter η:

$$\eta = \frac{\beta_+(\eta) - \beta_+(\eta)}{4\alpha_\perp - [\beta_+(\eta) + \beta_+(-\eta)]}. \quad (74)$$

Here $\beta_+(\eta) = \delta_+(\eta) + \delta_-(\eta)$ is a monotonically increasing function on the interval (–1, 1); at the ends of this interval and at $\eta = 0$ the function $\beta_+(\eta)$ satisfies the conditions

$$\beta_+(\eta) = \begin{cases} -(1-2\alpha_\perp), & n = -1, \\ 0, & \eta = 0, \\ 2\alpha_\perp, & \eta = 1. \end{cases} \quad (75)$$

Since the explicit form of the function $\beta_+(\eta)$ cannot be determined on the basis of the phenomenological analysis performed above, we shall confine our attention to an expansion of the function in a power series in η. In order that the even (in the denominator) and odd (in the numerator) parts of the functions $\beta_+(\eta)$ have the same order, we shall retain an even number of terms in the expansion. It turned out that if the parameter $\alpha_\perp$ is not too small, the analysis can be limited to the minimum nontrivial number of terms in the expansion, equal to four, i.e., the function $\beta_+(\eta)$ can be approximated by a quartic polynomial:

$$\beta_+(\eta) = \frac{1}{2}[(1-a)\eta - (1-4\alpha_\perp)(1-b)\eta^2 + a\eta^3 - (1-4\alpha_\perp)b\eta^4], \quad (76)$$

where the conditions (75) are taken into account.

The following can be stated concerning the possible values of the parameters a and b. In the first place, for small positive values of η the function $\beta_+(\eta)$ should grow, and the maximum possible rate of growth is $\eta/2$, so that the condition $0 < a < 1$ is imposed on the parameter a. In the second place, the function $\beta_+(\eta)$ will be nondecreasing in the interval (–1, 1), if for not very small $\alpha_\perp$ the condition $-1 < b_- < b < b_+ < 1$ ($b_- < 0$, $b_+ > 0$) is satisfied and the quantities $b_\pm$ depend on the values of the parameter a and the parameter $\alpha_\perp$ (the explicit form of these dependences is extremely complicated). Thus, the parameter b can assume positive and negative values.

Using the expression (76), Eq. (74) becomes

$$\eta = \frac{\eta(1 - a + a\eta^2)}{4\alpha_\perp + (1-4\alpha_\perp)\eta^2(1 - b + b\eta^2)}. \quad (77)$$

It is easy to verify that $\eta_1 = 0$ and $\eta_{2,3} = \pm 1$ are solutions of Eq. (77) for arbitrary values of the parameters a, b, and $\alpha_\perp$. If

$$0 < \frac{a + 4\alpha_\perp - 1}{(1-4\alpha_\perp)b} < 1,$$

then the biquadratic equation

$$(1-4\alpha_\perp)b\eta^4 + [(1-4\alpha_\perp)(1-b) - a]\eta^2 + 4\alpha_\perp - 1 + a = 0,$$

equivalent to Eq. (77), possesses two more solutions in the physical range ($-1 \le \eta \le 1$):

$$\eta_\pm = \pm\sqrt{\frac{a - (1-4\alpha_\perp)}{(1-4\alpha_\perp)b}}. \quad (78)$$

So, we have found that depending on the values of the parameters a, b, and $\alpha_\perp$ Eq. (77) can possess from three to five solutions lying in the physical range: the solution $\eta_1 = 0$, corresponding to the isotropic state of the system; the solutions $\eta_{2,3} = \pm 1$, corresponding to completely orientationally ordered states; and, the solutions $\eta_\pm$, corresponding to partially ordered states. We note that partially ordered states $\eta_\pm$ exist for positive and negative values of the parameter b: for $b > 0$ if the inequalities $1 - 4\alpha_\perp < a < (1 - 4\alpha_\perp)(1 + b)$ are satisfied and for $b < 0$ if $(1 - 4\alpha_\perp)(1 + b) < a < 1 - 4\alpha_\perp$.

The stability of the states obtained can be determined according to the value of the derivative of the right-hand side of Eq. (77) with respect to the variable η: if the value of this derivative in front of η, equal to a solution of Eq. (77), is greater than 1, then this solution corresponds to an unstable state; if it is less than 1, then it corresponds to a stable state.

A stability analysis of the solutions of Eq. (77) showed that for $b > 0$ the states $\eta_\pm$ are unstable, the state $\eta_1 = 0$ is stable for $a > 1 - 4\alpha_\perp$, and the states $\eta_{2,3} = \pm 1$ are stable for $a < (1 - 4\alpha_\perp)(1 + b)$. This means that for $b > 0$ and $1 - 4\alpha_\perp < a < (1 - 4\alpha_\perp)(1 + b)$ the isotropic and completely ordered states are stable, and the unstable states $\eta_\pm$ separate them. In this range of values of the parameters the transitions from an ordered into an isotropic state and vice versa can occur only after a threshold is overcome; outside this interval the state of the system changes abruptly, and for a cyclic change of the parameters hysteresis phenomena will be observed in the system, i.e., the typical bistability situation occurs (see Fig. 5a).

For $b < 0$ the picture is somewhat different: the states $\eta_\pm$ are stable, the state $\eta_1 = 0$ is stable for $a > 1 - 4\alpha_\perp$, and the states $\eta_{2,3} = \pm 1$ are stable for $a < (1 - 4\alpha_\perp)(1 + b)$. The regions where the ordered and disordered states are stable do not overlap, and the degree of ordering of the system in the transitional region varies smoothly from minimum to maximum or, vice versa, from maximum to minimum. This means that for $b < 0$ a transition from an ordered into a disordered state (and vice versa) occurs smoothly with no jumps in a certain finite range of values of the parameters, i.e., softly (Fig. 5b).

The parameters $\alpha_\perp$ and a can depend on the external conditions, specifically, the temperature of the system.

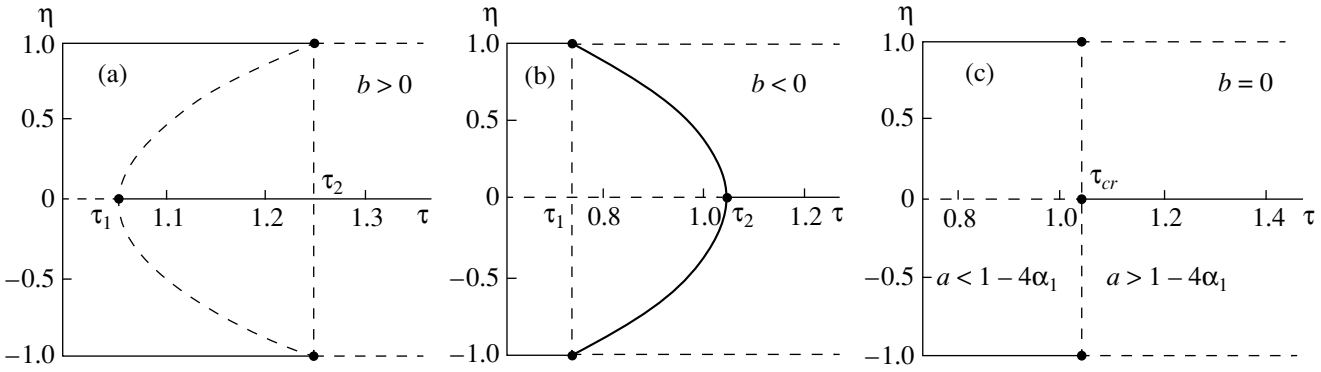


Fig. 5. Bifurcation diagram for the order parameter, characterizing the orientational ordering of the system, for $b = 0.3$ (a), –0.3 (b), and 0 (c). The order parameter is the dimensionless temperature $\tau = kT/E_r$. The stable states are marked by solid lines and the unstable states by dashed lines. Bistability is observed in the temperature range from τ_1 to τ_2.

Indeed, the effective stiffness of a chain (the parameter $\alpha_\perp$) and the ordering action of the environment (the parameter a) are functions of temperature. It is natural to assume that $\alpha_\perp(T)$ varies from zero at $T = 0$ up to 1/4 in the limit $T \longrightarrow \infty$ according to the Arrhenius law $4\alpha_\perp(T) = \exp(-E_r/kT)$, and $a(T)$ follows the same law but with a different activation energy: $a(T) = \exp(-E_{or}/kT)$, where $1 - a(T)$ decreases from 1 to 0 as temperature increases (E_r is the bending energy of a chain, $E_{or} = E_r/2$). The corresponding bifurcation diagrams, on which the dimensionless temperature $\tau = kT/E_r$ is chosen as the ordering parameter and the parameter $b = \pm 0.3$, are presented in Figs. 5a and 5b. If the width of the transitional interval decreases to zero, the state of the system changes abruptly (Fig. 5c).

Thus, we have been able to describe on the basis of the model of orientationally correlated walks the process of temperature orientational ordering of a polymeric melt: a system which is isotropic at high temperatures passes, as temperature decreases, into a highly oriented state. The transition described, in contrast to a standard transition, possesses a characteristic feature. Ordinarily, as temperature decreases, a system consisting of anisotropic components makes a transition into an anisotropic state on account of the fact that the tendency toward ordering resulting from steric constraints and (or) orienting interaction of the anisotropic components begins to predominate over the disordering entropy factors, whereas in our case, as temperature decreases, an additional ordering factor arises: the anisotropy of the components of the system increases. Indeed, the decrease of the parameter $\alpha_\perp(T)$ ($4\alpha_\perp(T) = \exp(-E_r/kT)$) with decreasing temperature, which corresponds to an increase in the effective stiffness of the chain, results in an increase of the persistence length (Kuhn segment) of the chain, i.e., an increase in the length of the rectilinear sections of the chain and, ultimately, rectification of the polymeric coil. It is the presence of such rectilinear sections that results in the appearance of orientational order in the system (first local and then global). Thus, given a system which is completely anisotropic at high temperatures (including locally), a liquid crystal system can be obtained by lowering the temperature of the system.

A different mechanism for the appearance of orientational order is also possible: the orienting role of the steric constraints increases when a system consisting of anisotropic components is compressed. On the basis of the formalism employed this corresponds to an increase of $1 - a$ (decrease of the parameter a) with all other parameters remaining fixed. The corresponding bifurcation diagrams for positive, negative, and zero values of the parameter b have the same form as the temperature bifurcation diagrams presented in Fig. 5. This means that three types of transitions from isotropic to orientationally ordered states are possible when a system undergoes compression as well as when the temperature decreases: hard (for $b > 0$), soft (for $b < 0$), and abrupt (for $b = 0$).

Unfortunately, it is impossible to determine on the basis of the phenomenological model examined here which of the possible transitions occur in a particular system. For this it is necessary to determine the sign of the parameter b on the basis of a microscopic analysis of the interchain interaction; this falls outside the scope of the problem considered here.

6. CONCLUSIONS

The analysis performed shows that the statistics of the conformational states of chains possessing a finite interlink flexural stiffness can differ strongly from Gaussian statistics. If the length of a chain is comparable to the length of a Kuhn segment, then the molecule is strongly anisotropic (almost rectilinear), i.e., it is a slightly curved rod. As its length increases, a molecule starts to bend, its anisotropy decreases, and ultimately the molecule coils up. However, this ball is not truly Gaussian. This is because the ball contains almost rectilinear chain segments. At the same time, the dependence of the average size of such a coil is the same as

the size of the Gaussian coil: $\langle \mathbf{R}^2 \rangle \sim N$. This apparent contradiction can be easily explained. Indeed, on the smallest scales the spatial distribution function of the mass of a molecule is different from a Gaussian distribution. However, if this distribution is averaged over a scale of the order of the length of "rectilinear" segments of the chain and it is assumed that the "elementary" link in the chain is a fragment with this scale, then the spatial distribution of such elementary links will be Gaussian. Therefore the length of a Kuhn segment ($n_K \approx \langle n \rangle$) corresponds to the length of a "rectilinear" chain section, whose size can be quite easily estimated on the basis of the model of directed self-avoiding walks.

Indeed, for directed self-avoiding walks the average number of steps up to the first rotation (i.e., the length of a Kuhn segment) is

$$\begin{aligned}\langle n \rangle &= 1(\alpha_+ + \alpha_-) + 2(\alpha_+ + \alpha_-)^2 + 3(\alpha_+ + \alpha_-)^3 + \ldots \\ &= (\alpha_+ + \alpha_-)\frac{d}{d\xi}\sum_n \xi^n\bigg|_{\xi = \alpha_+ + \alpha_-} = \frac{\alpha_+ + \alpha_-}{(1 - \alpha_+ - \alpha_-)^2}. \end{aligned} \quad (79)$$

Here we have summed a geometric progression.

Now, using the formula (79) it is easy to estimate the number of chain links comprising a Kuhn segment for fixed values of the parameters α_+ and α_- and to compare this number to the Kuhn length obtained from other considerations. Setting $\alpha_+ + \alpha_- = 0.9$ we obtain that a Kuhn segment consists of 90 links. Assuming that $\alpha_+ \gg \alpha_-$, we can set $\theta = \alpha_+ - \alpha_- \sim \alpha_+ + \alpha_- = 0.9$. Turning to Fig. 4 we see that the dependence of $\langle \mathbf{R}^2 \rangle$ on N becomes linear approximately for this value of N, while the estimate using the formula $n_K \approx -1/\ln\theta$ gives an order of magnitude smaller value (see remarks in the text for Fig. 4 at the end of Section 3). It seems to us that the proposed criterion for estimating the Kuhn length (giving a value close to that obtained from the formula $l_0 n_K = \langle \mathbf{R}^2 \rangle / N$ [2]) is more suitable for the physical meaning of this concept than the ordinarily employed criterion, which is associated with the exponential smallness of the correlation function.

A nongaussian spatial distribution of the links of a polymer chain can arise in two cases: if the resolution of the instrument (characteristic scale of the interaction of a probe with a polymer molecule) used to estimate the state of a polymeric molecule is less than the Kuhn length, introduced using the method proposed above, and if the length of the chain is comparable or not much greater than this scale. The latter case is more interesting, because it is for such relatively short molecules (called oligomers) that in our opinion the effects of a nongaussian spatial distribution function of the chain links should be strongest.

But, even for sufficiently long chains the nongaussian spatial distribution of the chain links at small scales should result in macroscopic effects. We are talking here about the orientational ordering, mentioned in the introduction, of a rigid-chain polymeric melt. The analysis performed in the preceding section for a two-dimensional rigid-chain concentrated system showed that a highly ordered state (of the liquid-crystal type) does indeed arise in such a system, and transitions between ordered and disordered states can follow two fundamentally different scenarios. A possible variant is one where in some range of values of the control parameter (this could be, for example, the temperature or external pressure) the system can be in ordered and disordered states, i.e., a typical bistability situation occurs (see Fig. 5a). Hysteresis should be observed for such systems. This is confirmed experimentally in a number of investigations of oligomeric systems (see, for example, [18]). According to a different possible variant, the system passes smoothly from one state into another (see Fig. 5b). Which of the above described scenarios is realized in a particular polymeric system seems to depend on the form of the interchain interactions in the system and is determined by the characteristic features of the chemical structure of the monomeric link of a polymer.

In summary, the proposed mathematical apparatus makes it possible not only to describe the statistics of conformations of polymers with finite stiffness but also to investigate their macroscopic physical properties, which are determined by the small-scale structure of rigid macromolecules.

ACKNOWLEDGMENTS

I thank V.I. Irzhak, L.I. Manevich, and S.M. Mezhikovskiĭ for many helpful discussions of this problem.

APPENDIX

Small-Scale Structure of a Rigid-Chain Polymeric Melt

It is well known that a single coiled polymer chain is a loose formation in which the volume density of matter is all the lower, the longer the chain. Indeed, the volume occupied by a freely-jointed chain consisting of N links of size a is of the order of $V_{ch} \sim a^3 N^{3\nu}$ ($1/3 < \nu < 3/5$), while the chain volume itself is $a^3 N$. If a chain possesses some stiffness, then a persistence length (Kuhn segment) is introduced, and the chain is treated as freely-jointed but consisting of Kuhn segments instead of elementary links. The volume occupied by such a chain will be even larger:

$$\begin{aligned} V_{\mathrm{ch}} &\sim (l_K)^3\left(\frac{N}{n_K}\right)^{3\nu} \\ &= (an_K)^3\left(\frac{N}{n_K}\right)^{3\nu} = a^3 n_K^{3(1-\nu)} N^{3\nu}. \end{aligned} \quad (A.1)$$

Here N is the number of links in the chain; $l_K = an_K$ is the length of a Kuhn segment containing n_K links; a is

the length of a single link; for a globule $\nu = 1/3$, for a Gaussian ball $\nu = 1/2$, and for a swelled ball $\nu = 3/5$.

The density of matter in a single ball (here and below the density is expressed in units of the excluded volume) decreases as the length of a Kuhn segment and the chain length increase:

$$\rho_{ch} \sim \frac{a^3 N}{V_K} \sim \frac{1}{n_K^{3(1-\nu)} N^{3\nu-1}}, \quad 1 < 3\nu < 3. \qquad (A.2)$$

Since the density of matter in the melt (we are studying a polymeric melt) is of the order of $\Theta \sim 1$ and is essentially independent of the length of a Kuhn segment and the chain length, not one but M chains, whose number is determined by the relation

$$M \sim \Theta n_K^{3(1-\nu)} N^{3\nu-1}, \quad 1 < 3\nu < 3 \qquad (A.3)$$

are present in the volume V_{ch} occupied by the trial chain.

A volume per Kuhn segment can be identified in the volume occupied by the chain under study:

$$V_K = \frac{V_{ch}}{N/n_K} \sim (an_K)^3 \left(\frac{N}{n_K}\right)^{3\nu-1} = a^3 n_K^{3(1-\nu)+1} N^{3\nu-1}. \qquad (A.4)$$

This volume per Kuhn segment of the trial chain (ellipsoid with axes $L_1 \sim an_K$, $L_{2,3} = L_K \propto an_K (N/n_K)^{(3\nu-1)/2} = an_K^{3(1-\nu)/2} N^{(3\nu-1)/2}$ is filled with M rods of length $l_K = an_K$ and diameter a—the Kuhn segments of other chains. The number density of such rods is

$$C_r \propto M/V_K \propto \Theta/a^3 n_K. \qquad (A.5)$$

Since the Kuhn segments of different chains are statistically independent, the situation under study is equivalent to the well-known model of rigid rods. Taking account of the excluded volume even in the simplest Onsager model [19] leads to the fact that for such a system there exists a critical concentration C_{cr} at which orientational ordering appears in the system. The value of this critical concentration is determined by the length and diameter of the rods:

$$C_{cr} \propto 1/al_K^2 \propto 1/a^3 n_K^2, \quad \Theta_{cr} \propto 1/n_K. \qquad (A.6)$$

Since we are studying rigid polymeric chains, whose Kuhn segments are long, the density of our system is greater than the critical density, and the orientation of the rods (Kuhn segments of different chains) in the ellipsoid which we have singled out are correlated.

Let us draw a plane perpendicular to the direction of a Kuhn segment in the trial chain. The section of an ellipsoid corresponding to a Kuhn segment of the trial chain by this plane forms a circle, and the intersection of the rods (Kuhn segments of other chains) located in the volume under study with this circle generates a system of points whose concentration is of the order of $C_{point} \propto \Theta/a^2$, and separated by an average distance $\Delta \sim a/\sqrt{\Theta}$.

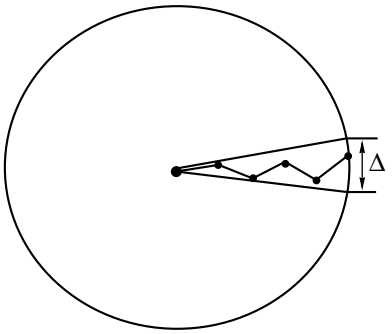


Fig. 6. Section of an ellipsoid by a plane perpendicular to a Kuhn segment of the trial chain. The points show the intersection of this plane with other Kuhn segments, located in a sector resting on the arc of a circle of length Δ (Δ is the average distance between the points of intersection of the Kuhn segments and the intersecting plane).

Let us single out in the circle a sector resting on an arc of a circle of length Δ, and let us construct a chain connecting the nearest points in this segment (Fig. 6). If each point is associated to a unit vector, whose direction is the same as the direction of the rod to which the point belongs, then an auxiliary chain is formed, whose properties are similar to the properties of a rigid polymeric chain, since the orientations of the two neighboring rods and therefore links of the auxiliary chain are correlated. The form of the correlation function of the orientation of the links along such a chain is well known (see, for example, [20, Section 127] or the expression (26) of the present paper):

$$g(l) \propto \exp\left(-\frac{l}{L_{cor}}\right) = \exp\left(-\frac{(l/\Delta)\Delta}{L_{cor}}\right) = \exp\left(-\frac{n\Delta}{L_{cor}}\right), \qquad (A.7)$$

where n is the number of links in the chain or the number of rods between two rods singled out in a sector of radius L_K and the arc Δ, and L_{cor} is the distance from the Kuhn segment of the trial chain in which the orientation of the Kuhn segments of other chains is essentially independent of the orientation of the Kuhn segment of the trial chain. In what follows we shall be interested precisely in this quantity.

For two neighboring "links" (rods) we have

$$g(\Delta) = g(1) \sim \exp\left(-\frac{\Delta}{L_{cor}}\right) = \exp\left(-\frac{a}{\sqrt{\Theta} L_{cor}}\right) \approx 1 - \frac{a}{\sqrt{\Theta} L_{cor}}. \qquad (A.8)$$

On the other hand this same correlation function can be expressed in terms of the relative orientation angle between two neighboring links [2, Section 2]:

$$g(\Delta) = g(1) = \langle \cos\theta \rangle \approx 1 - \frac{1}{2}\langle \theta^2 \rangle. \qquad (A.9)$$

The correlation of the orientation of two neighboring rods is completely determined by their packing density Θ. The required dependence $\langle \cos\theta(\Theta) \rangle$ can be found by using the Onsager model mentioned above, modified for the case of a high density of rods [21]:

$$\langle \cos\theta(\Theta) \rangle \approx 1 - [\ln(1-\Theta) n_K]^{-2}. \qquad (A.10)$$

Using Eqs. (A.8), (A.9), and (A.10), we find the desired length showing the distance from the chosen Kuhn segment at which the orientational effects are no longer manifested:

$$L_{\text{cor}} \sim \frac{a}{\sqrt{\Theta}} [\ln(1-\Theta)]^2 n_K^2. \qquad (A.11)$$

The value obtained for the correlation length L_{cor} must be compared with the characteristic scales in the system under study: the characteristic size of the region corresponding to a single Kuhn segment, $L_K \propto a n_K (N/n_K)^{(3\nu-1)/2} = a n_K^{3(1-\nu)/2} N^{(3\nu-1)/2}$, and the characteristic size of the entire chain, $R_{\text{ch}} \propto a n_K (N/n_K)^{\nu} = a n_K^{1-\nu} N^{\nu}$.

If $L_{\text{cor}} < L_K$, i.e., the size of an ordered region does not exceed the characteristic size of a region corresponding to a single Kuhn segment, then on averaging over a scale greater than this size local ordering effects do not appear. This means that the local ordering of the Kuhn segments of different chains, which, as shown above, is characteristic of rigid-chain polymeric systems, has virtually no effect on the macroscopic properties of the entire polymeric system as a whole. In this case the introduction of a persistence length (Kuhn segment) makes it possible to treat a polymeric chain as a freely-jointed chain. Such a situation is possible if the length of a Kuhn segment is quite small compared with the length of the entire chain:

$$n_K \leq |\ln(1-\Theta)|^{-4/(3\nu+1)} N^{1-2/(3\nu+1)}. \qquad (A.12)$$

If $L_K < L_{\text{cor}} < R_{\text{ch}}$, i.e., the size of the ordered region is greater than the characteristic size of the region corresponding to a single Kuhn segment, but does not exceed the characteristic size of the entire chain, then the local ordering of Kuhn segments of different chains affects the conformational structure of the polymeric chain on scales greater than the Kuhn segment, i.e., it influences the macroscopic properties of the entire polymeric system. This influence first appears when the Kuhn length is not small but still remains appreciably less than the length of the entire chain:

$$|\ln(1-\Theta)|^{-4/(3\nu+1)} N^{1-2(3\nu+1)} \leq n_K \leq |\ln(1-\Theta)|^{-2/(\nu+1)} N^{1-1/(\nu+1)}. \qquad (A.13)$$

If $L_{\text{cor}} > R_{\text{ch}}$, i.e., the size of the ordered region is greater than the size of the polymeric coil, then the local ordering of the Kuhn segments of different chains results in a radical rearrangement of the structure of the entire polymeric system. In the first place, orientational ordering of the entire chain occurs: the effective stiffness of the chain changes sharply and as a result the chain rectifies. In the second place, since as a result of rectification of chains a polymeric melt consists of virtually rectilinear rods, whose density is higher than the critical value, liquid-crystal ordering of the entire polymeric system as a whole occurs. These effects appear if the Kuhn length is sufficiently large:

$$n_K \geq |\ln(1-\Theta)|^{-2/(\nu+1)} N^{1-1/(\nu+1)}. \qquad (A.14)$$

This analysis shows that the small-scale structure of a linear polymeric molecule with finite stiffness to some degree influences the properties of a polymeric melt, and under certain conditions this influence could be decisive.

REFERENCES

1. P. G. de Gennes, *Scaling Concepts in the Physics of Polymers* (Cornell Univ. Press, Ithaca, 1979; Mir, Moscow, 1982).
2. A. Yu. Grosberg and A. R. Khokhlov, *Statistical Physics of Macromolecules* (Nauka, Moscow, 1989).
3. J. M. Ziman, *Models of Disorder: the Theoretical Physics of Homogeneously Disordered Systems* (Cambridge Univ. Press, Cambridge, 1979; Mir, Moscow, 1982).
4. A. E. Arinstein and S. M. Mezikovskii, Polym. Eng. Sci. **37**, 1339 (1997).
5. A. N. Semenov and P. R. Khokhlov, Usp. Fiz. Nauk **156**, 417 (1988).
6. W. J. Orr, Trans. Faraday Soc. **43**, 12 (1947).
7. F. Spitzer, *Principles of Random Walk* (Van Nostrand, Princeton, 1964; Mir, Moscow, 1968).
8. Yu. A. Rozanov, *Probability Theory, Stochastic Processes, and Mathematical Statistics* (Nauka, Moscow, 1989).
9. R. Balescu, *Equilibrium and Nonequilibrium Statistical Mechanics* (Wiley, New York, 1975; Mir, Moscow, 1978), Vol. 2.
10. V. Privman and N. Svrakič, *Directed Models of Polymers, Interfaces and Clusters: Scaling and Finite-Size Properties* (Springer-Verlag, Berlin, 1989).

11. P. Devillard and E. Stanley, Phys. Rev. A **41**, 2942 (1990).

12. B. Derrida, Physica A (Amsterdam) **20**, 491 (1990).

13. R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965; Mir, Moscow, 1968).

14. A. L. Kholodenko, Ann. Phys. (N.Y.) **202**, 186 (1990).

15. A. L. Kholodenko, J. Stat. Phys. **65**, 291 (1991).

16. A. E. Arinshtein, Dokl. Akad. Nauk **358**, 350 (1998).

17. R. P. Feynman, *Statistical Mechanics: a Set of Lectures* (Benjamin, Reading, Mass., 1972; Mir, Moscow, 1975).

18. M. P. Berezin and G. V. Korolev, in *Proceedings of 6th International Conference on Chemistry and Physicochemistry of Oligomers, Kazan, 1997* (Chernogolovka, 1997), Vol. I, p. 69.

19. L. Onsager, Ann. N. Y. Acad. Sci. **51**, 627 (1949).

20. L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Nauka, Moscow, 1995; Pergamon, Oxford, 1980), Part 1.

21. A. R. Khokhlov and A. N. Semenov, J. Stat. Phys. **38**, 161 (1985).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 1, 2000, pp. 31–47.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 1, 2000, pp. 36–55.

ATOMS, SPECTRA, RADIATION

Exchange Symmetry in a System of Nonrelativistic Spin-1/2 Fermions in the Feynman Quantum Statistics Representation

S. V. Shevkunov

St. Petersburg State Technical University, St. Petersburg, 195251 Russia
e-mail: root@shevk.hop.stu.neva.ru

Received November 29, 1999

Abstract—A compact representation is obtained for the quantum statistical sum of indistinguishable nonrelativistic spin-1/2 fermions in the form of Feynman path integrals which can be used as the basis to develop a fundamentally exact method of computer modeling for systems of strongly interacting electrons at nonzero temperature. A basis of symmetrized wave functions is constructed using Young symmetry operators. An exact permutation symmetrization procedure leads to an avalanche-like multiplication in the number of diagrams of linked Feynman integrals of the order of $N!$. The partition function can be simplified without introducing any approximations and this is performed numerically by computer by direct sorting of diagrams. The control tables obtained, containing combinatorial weights of diagrams, direct the Markov random walk process in virtual trajectory space which is achieved numerically by computer. The equilibrium characteristics of the quantum system are calculated by averaging. This approach is an expansion of the Monte Carlo–Metropolis method to systems of quantum indistinguishable particles with spin. Demonstration numerical calculations using this method were made for the simplest exchange systems, for a hydrogen molecule, a Be^+ ion, and a Li atom. The ground state of the hydrogen molecule is reproduced with a statistical error of 0.2%. Exchange-correlation effects lead to nontrivial structural changes in the thermally excited electron shells of ions in a state of strong plasma compression. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Numerical stochastic methods based on Markov processes are successfully used for systems obeying classical statistics [1–29]. The range of application of these methods extends from the modeling of ion plasmas [5], interaction of the gas phase with ions and ionic crystals [6–8, 11, 26], vapor condensation [13, 24], and calculations of the equilibrium properties of the liquid phase [17, 18, 21, 23, 29] to studies of metal clusters [27, 28]. Using the same methods in quantum statistical mechanics necessitates optimizing the choice of representation in which the problem is solved. On the one hand, the basis functions of the representation must be easy to calculate numerically and on the other, there should be a relatively simple formula for calculating the diagonal elements of the density matrix in this representation. The solution of the problem involves extending the ideas of the method of significant sampling to calculate these matrix elements so that summation over quantum states and calculations of the matrix elements of these states can be included in a single Markov process. Along these lines, all the difficulties of working in a specific representation can be transferred to the calculation of specific matrix elements and the basis functions can be selected as simply as possible, for example, the spectrum of eigenfunctions of the coordinate operator. The matrix elements in the coordinate representation may be written as Feynman path integrals [30] and these can then be approximated by multidimensional broken-line integrals. These ideas are implemented in the Monte Carlo method using path integrals, which is a method for computer modeling of systems of quantum particles under finite temperature conditions. The first study of this type for two ^4He atoms was made by Fosdick and Jordan [31, 32]. In addition to the method aimed at achieving a statistical description of systems of quantum particles at finite temperature, the eighties saw the development of stochastic methods of computer modeling of pure quantum states corresponding to zero temperature ($T = 0$). The most well-known are the variational Monte Carlo method [33–39] based on searching for the minimum of the energy functional and the diffusion Monte Carlo method based on the Green's function of the evolution operator [40–55]. The diffusion method is obtained by writing the Schrödinger equation in imaginary time. In imaginary time the Schrödinger equation has the form of a diffusion equation with sources and sinks and the time dependence of the steady states is converted into an exponentially damped one. At fairly long times, the state with the slowest damping survives, i.e., the ground quantum state. The Schrödinger equation is not solved in the ordinary meaning of this word. In addition, the diffusion random walk of the mapping point in multidimensional configuration space is simulated numerically. Assuming that the coordinate component of the wave function is antisymmetric, the permutation symmetry of the particles is represented as a system of

hypersurfaces ("nodes") on which the wave function changes sign and vanishes. The position of the nodes is specially defined by a selected reference function.

The eighties also the parallel development of an independent direction of quantum system modeling using an occupation number representation, i.e. "node" models [56–75] which are various modifications of the Hubbard model [59, 60, 66, 69, 70]. In this method the *d*-dimensional quantum problem is transformed into a $d + 1$ dimensional classical problem. The additional measurement is the imaginary time. The process of random walking over nodes of a $d + 1$ dimensional lattice with transition probabilities determined by the matrix elements of the evolution operator is simulated numerically. In [75] an original procedure is developed which can avoid the error caused by the discrete imaginary time scale.

The exact description of the permutation symmetry presents major difficulties along the way to developing the method of path integrals. Despite considerable efforts made over the last few years to develop this promising approach, so far specific results have only been achieved for systems of spin-zero particles [76–112].

Although the Hamilton operator of nonrelativistic fermions does not act on the spin variables, the statistical behavior of the system depends fundamentally on the particle spin since the type of permutation symmetry of the coordinate component of the wave function is the twin of the permutation symmetry of the spin component of the wave function, provided that the complete wave function is antisymmetric with respect to permutations of the coordinate and spin variables simultaneously. The fermion system in various spin states corresponds to various types of permutation systems of the coordinate component which determines the values of the canonical averages. Neglecting the spin variable [76] is an unsatisfactory approximation for the description of real fermions since no spin-zero fermions exist in nature.

The quantum state of an isolated spin 1/2 fermion is described by the function $\psi(\sigma, \eta)$ of the spin variable $\{\sigma_i\} = \pm 1/2$ and the variable η which depends on the specific quantum-mechanical representation. The quantum state of a system of *N* fermions is written in the form of a multiparticle wave function

$$\psi(\sigma_1, \sigma_2, \ldots, \sigma_N; \eta_1, \eta_2, \ldots, \eta_N) = \psi(\{\sigma_i\}, \{\eta_i\}), \quad (1)$$

which from the point of view of the theory of end group representations is a $2N$ rank tensor [113–115]. The principle of fermion indistinguishability establishes that only states described by the antisymmetric tensors (1) exist. All tensors of the type (1) are a substratum of the π' permutation group which is the direct product of the π_N group of all the permutations of *N* distinguishable elements with itself: $\pi' = \pi_N \otimes \pi_N$. Each permutation of the arguments in (1) corresponds to a certain transformation in the tensor space (1) and a linear transformation in the corresponding vector space V_N [116]. All the matrices of these linear transformations form a π_N group representation [113] which is then broken down into irreducible representations, i.e., sets of transformations which do not derive the element of the vector space V_N from the corresponding subspace $V_N^{(k)}$. The subspaces $V_N^{(k)}$ correspond to the subspaces $\Psi_N^{(k)}$ in *N* rank tensor space. The basis elements of these subspaces are obtained by the action of Young operators on an arbitrary nonsymmetrized tensor [117, 118]. A set of Young operators with different permutations of the arguments in the cells of its diagram generates a tensor subspace $\Psi_N^{(k)}$ invariant relative to pair permutations. It can be shown [116] that the space of these $2N$ rank antisymmetric tensors (1) is exhausted by the direct products $\Psi_N^{(i)} \otimes \Psi_N^{(i,\mathrm{tr})}$ of all the subspaces generated by Young operators with mutually transposed (dual) diagrams. The antisymmetric tensor space thus constructed is complete. The completeness follows from calculations of the space dimensions and the linear independence of the basis functions [116].

The problem of constructing the basis for a system of indistinguishable particles can be solved, in principle, by using Young operators although numerical calculations along these lines using the traditional Schrödinger formulation of quantum statistics encounter colossal computational difficulties. The linear combinations generated by the Young operators contain of the order of *N*! terms which must be calculated at each step of the quantum-system modeling procedure which makes these calculations inconceivable even for systems with small numbers of particles. The present study addresses the problem of formulating a procedure for making exact allowance for exchange symmetry in the path integral method and of obtaining a representation of the partition function and the equilibrium averages in a form suitable for numerical calculations of systems of indistinguishable nonrelativistic spin 1/2 fermions. Computer modeling of thermally excited quantum states using the Feynman path integral method was performed for two- and three-electron systems with exact allowance for exchange and spin.

2. REDUCTION TO DUMMY GRAPHS IN MATRIX FUNCTIONALS

Of particular interest for the calculation of quantum-mechanical observables are the matrix functionals determined in the space of the permutation operators $\hat{P}_k$ having the form

$$G(\hat{P}_k) = \int d^N r \int d^N x \times f^*(\{\mathbf{r}_i\}; \{\mathbf{x}_i\}) \hat{F} f(\{\mathbf{r}_i\}; \hat{P}_k\{\mathbf{x}_i\}), \quad (2)$$

where $\hat{F}$ is a symmetric operator with respect to permutations in the sequence of spatial variables $\{\mathbf{x}_i\}$, $f(\{\mathbf{r}_i\}; \{\mathbf{x}_i\})$ is the eigenfunction of the coordinate operator having the eigenvalues $\{\mathbf{r}_i\}$, $d^N r = d\mathbf{r}_1 d\mathbf{r}_2 \ldots d\mathbf{r}_N$, and the asterisk denotes complex conjugation. The result of calculating the matrix functional does not depend on the number of the vertices of the $\hat{P}_k$ graph (see Appendix), i.e., one for all elements of the same class:

$$\int d^N r \int d^N x f^*(\{\mathbf{r}_i\}; \{\mathbf{x}_i\}) \hat{F} f(\{\mathbf{r}_i\}; \hat{P}_k \hat{P}_n \hat{P}_k^{-1}\{\mathbf{x}_i\})$$
$$= \int d^N r' \int d^N x' f^*(\{\mathbf{r}_i'\}; \{\mathbf{x}_r'\}) \hat{F} f(\{\mathbf{r}_i'\}; \hat{P}_n\{\mathbf{x}_r'\}). \quad (3)$$

The reduction to dummy graphs considerably simplifies the problem of calculating functionals of the type (3) as is indicated by the data in Table 1.

The complete wave function has the property that the spin variable permutations are complementary

$$\psi(S, \hat{P}_l\{m_i\}, \{\mathbf{r}_i\}; \{\sigma_i\}, \{\mathbf{x}_i\})$$
$$= \psi^l(S, \{m_i\}, \{\mathbf{r}_i\}; \hat{P}_l^{-1}\{\sigma_i\}, \{\mathbf{x}_i\}), \quad (4)$$

and similarly for the coordinate variables. Here the subscript l indicates that positional permutation of the arguments of $\hat{P}_l$ is performed in the cells of the Young diagram of the spin functions:

$$\chi_j^l(S, \{m_i\}; \hat{P}_l^{-1}\{\sigma_i\})$$
$$= \sum_k (-1)^{c(k)} \chi(\{m_i\}; (\hat{P}_l^{-1}\hat{P}_k\hat{P}_l)\hat{P}_l^{-1}\{\sigma_i\})$$
$$= \sum_k (-1)^{c(k)} \chi(\hat{P}_l\{m_i\}; \hat{P}_k\{\sigma_i\}) \quad (5)$$
$$= \chi_J(S, \hat{P}_l\{m_i\}; \{\sigma_i\}).$$

Using the expansion (A.4) in the Appendix, we determine $N!$ spin functions:

$$\tilde{\chi}_k(\{\sigma_i\}) = \sum_n c_{nk} \chi(\hat{P}_n\{\sigma_i\}),$$
$$\tilde{\chi}(\{\sigma_i\}) \equiv \tilde{\chi}_1(\{\sigma_i\}). \quad (6)$$

It is easily shown that the complete wave function may be written in the diagonal form:

$$\psi(\{\sigma_i\}, \{\mathbf{x}_i\}) = \sum_k (-1)^{g(k)} \tilde{\chi}(\hat{P}_k\{\sigma_i\}) f(\hat{P}_k\{\mathbf{x}_i\}), \quad (7)$$

where $g(k)$ is the permutation parity of $\hat{P}_k$. For real spin and coordinate functions the diagonal elements of the operator $\hat{F}$, which is symmetric with respect to permutations of the coordinates, are written in terms of (6) in the form

$$\left\langle S, \{m_i\}, \{\mathbf{r}_i\} \middle| \hat{F} \middle| S, \{m_i\}, \{\mathbf{r}_i\} \right\rangle$$
$$= \sum_k \sum_{\{\sigma_i = \pm 1/2\}} \int d^N x \tilde{\chi}(S, \{m_i\}; \{\sigma_i\}) f(\{\mathbf{r}_i\}; \{\mathbf{x}_i\}) \quad (8)$$
$$\times \hat{F} \psi(S, \{m_i\}, \{\mathbf{r}_i\}; \{\sigma_i\}, \{\mathbf{x}_i\}).$$

All terms in the sum over k in the last expression (8) do not depend on k and may be combined. Using (7), we rewrite (8) in the representation of nonsymmetrized functions:

$$\left\langle S, \{m_i\}, \{\mathbf{r}_i\} \middle| \hat{F} \middle| S, \{m_i\}, \{\mathbf{r}_i\} \right\rangle$$
$$= N! \sum_n W_S(\hat{P}_n^{-1}) \left\langle \{\mathbf{r}_i\} \middle| \hat{F} \middle| \hat{P}_n^{-1}\{\mathbf{r}_i\} \right\rangle, \quad (9)$$

where we have introduced generalized weights (positive and negative) corresponding to the permutations of $\hat{P}_n$,

$$W_S(\hat{P}_n^{-1}) \equiv (-1)^{g(n)} \sum_{\{\sigma_i = \pm 1/2\}} \tilde{\chi}(S, \{m_i\}; \{\sigma_i\})$$
$$\times \tilde{\chi}(S, \{m_i\}; \hat{P}_n\{\sigma_i\}) = W_S(\hat{P}_n). \quad (10)$$

Quite clearly, $W_S(\hat{P}_n)$ does not depend on $\{m_i = \pm 1/2\}$. Assuming in (9) $\hat{F} = 1$, we obtain

$$1 = N! W_S(\hat{P}_1^{-1}) \langle \{\mathbf{r}_i\} | \{\mathbf{r}_i\} \rangle = N! W_S(\hat{P}_1^{-1}), \quad (11)$$

Table 1. Result of reduction to dummy graphs in matrix functionals

N	Number of numbered graphs	Number of dummy graphs
1	1	1
2	2	2
3	6	3
4	24	5
5	120	7
6	720	11
7	5040	15
8	40320	22
9	362880	30
10	3628800	42
11	39916800	56
12	479001600	77

in particular, for the identity permutation of $\hat{P}_1$

$$W_S(\hat{P}_1) = 1/N!. \tag{12}$$

The set of all possible functions of the type (A.4) (see Appendix) with various $\{m_i\}$ and $\{\mathbf{r}_i\}$, being complete in the space of permutationally antisymmetric pairs $(m_i, \mathbf{r}_i)$ with a given total spin S is not orthogonal, the orthogonality being impaired by the symmetrization procedure. In fact, we write the scalar product of two functions of the type (A.4):

$$\sum_{\{\sigma_i\}}\int d^N x \sum_{n,k,l,m} c^*_{nk}(S)c_{lm}(S)\chi^*(\{m_i\};\hat{P}_n\{\sigma_i\})$$
$$\times f^*(\{\mathbf{r}_i\};\hat{P}_k\{\mathbf{x}_i\})\chi(\{m'_i\};\hat{P}_l\{\sigma_i\}) \tag{13}$$
$$\times f(\{\mathbf{r}'_i\};\hat{P}_m\{\mathbf{x}_i\}).$$

For the pair of functions with $\{m'_i\} = \hat{P}_r\{m_i\}$ and $\{\mathbf{r}'_i\} = \hat{P}_t\{\mathbf{r}_i\}$ expression (13) has the form

$$\sum_{\{\sigma_i\}}\int d^N x \sum_{n,k,l,m} c^*_{nk}(S)c_{lm}(S)\chi^*(\hat{P}_l^{-1}\{m_i\};\{\sigma_i\})$$
$$\times f^*(\hat{P}_k^{-1}\{\mathbf{r}_i\};\{\mathbf{x}_i\})\chi(\hat{P}_l^{-1}\hat{P}_r\{m'_i\};\{\sigma_i\}) \tag{14}$$
$$\times f(\hat{P}_m^{-1}\hat{P}_t\{\mathbf{r}'_i\};\{\mathbf{x}_i\}).$$

From the group properties of the permutation operations it is clear that for any pair of indices n, k from the first bilinear combination (14) we can find a pair of indices l, m from the second bilinear combination (14) such that $\hat{P}_l^{-1}\hat{P}_r = \hat{P}_n^{-1}$ and $\hat{P}_m^{-1}\hat{P}_t = \hat{P}_k^{-1}$, i.e., the corresponding terms in the bilinear combinations are the same apart from the factor.

We isolate the linearly independent functions from the overfilled set (A.4). The spin operator $\hat{\mathbf{S}} = \mathbf{i}\hat{S}_x + \mathbf{j}\hat{S}_y + \mathbf{k}\hat{S}_z$, where $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are unit vectors and $2\hat{S}_x$, $2\hat{S}_y$, and $2\hat{S}_z$ are Pauli matrices, are uniquely related to the operator of an infinitely small spatial rotation $1 + (d\boldsymbol{\varphi}, \hat{S})$ through the angle $\delta\boldsymbol{\varphi}$ [118]. The spin wave function $\chi(\{\sigma_i\})$ can be conveniently represented as an Nth rank spinor with a corresponding transformation law for rotations of the coordinate system in spin space [117]. It follows from the linearity of the spin operator $\hat{S}$ that the transformation law should be linear:

$$\begin{pmatrix} \chi'(\sigma_i = +1/2) \\ \chi'(\sigma_i = -1/2) \end{pmatrix} \equiv \begin{pmatrix} \chi'^1 \\ \chi'^2 \end{pmatrix}$$
$$= \begin{vmatrix} a & b \\ c & d \end{vmatrix} \begin{pmatrix} \chi^1 \\ \chi^2 \end{pmatrix}. \tag{15}$$

If the ith and jth indices-arguments of the spinor are in the same column of the Young diagram, the spinor is antisymmetric with respect to these indices and the dependence of the spinor on these indices is expressed by the determinant

$$\chi'^1_i\chi'^2_k - \chi'^2_i\chi'^1_k = (ad - bc)(\chi^1_i\chi^2_k - \chi^2_i\chi^1_k), \tag{16}$$

where

$$\chi^1_i = \chi\left(\sigma_1, \ldots, \sigma_{i-1}, +\frac{1}{2}, \ldots, \sigma_N\right),$$

$$\chi^2_i = \chi\left(\sigma_1, \ldots, \sigma_{i-1}, -\frac{1}{2}, \ldots, \sigma_N\right).$$

The determinant (16) is an invariant scalar with respect to rotations and $ad - bc = 1$. The spin operator $\hat{S}$ does not in fact act on these spinor components. The spin state of the system is defined by a linear combination of symmetric $N_1 - N_2$ rank spinors retained after discarding all columns of the Young functions containing more than one cell. As a result of (4), these spinors are also symmetric with respect to m_i, i.e., permutations of values in the $\{m_i\}$ sequence do not result in new spin states—the spin state is uniquely determined by the sum $\sum_i m_i$.

The antisymmetry of the function (A.4) implies that if $\hat{P}_n = \hat{P}_m\hat{P}_j$, $\hat{P}_k = \hat{P}_l\hat{P}_j$ or $\hat{P}_m = \hat{P}_n\hat{P}_j$, $\hat{P}_l = \hat{P}_k\hat{P}_j$, then $c_{nk}(S) = \pm c_{ml}(S)$ depending on the parity of $\hat{P}_j$. Using the property (A.8), the bilinear combination (A.4) may be rewritten in the form

$$\psi(S, \{m_i\}, \{\mathbf{r}_i\}; \{\sigma_i\}, \{\mathbf{x}_i\})$$
$$= \sum_{n,k} c_{nk}(S)\chi(\hat{P}_n^{-1}\{m_i\};\{\sigma_i\})f(\hat{P}_k^{-1}\{\mathbf{r}_i\};\{\mathbf{x}_i\}). \tag{17}$$

The antisymmetry of (16) with respect to permutations of the pairs $(m_i, \mathbf{r}_i)$ will occur if $\hat{P}_m^{-1} = \hat{P}_n^{-1}\hat{P}_j$, $\hat{P}_l^{-1} = \hat{P}_k^{-1}\hat{P}_j$ or $\hat{P}_n^{-1} = \hat{P}_m^{-1}\hat{P}_j$, $\hat{P}_k^{-1} = \hat{P}_l^{-1}\hat{P}_j$ gives rise to $c_{nk}(S) = \pm c_{ml}(S)$ which, apart from replacing $\hat{P}_i$ by $\hat{P}_i^{-1}$, whose parities are the same, is equivalent to the anti-

symmetry conditions (A.4) for $(\sigma_i, \mathbf{x}_i)$. Since simultaneous permutations in $\{m_i\}$ and $\{\mathbf{r}_i\}$ do not lead to new quantum states and this applies separately to $\{m_i\}$, permutations only in $\{\mathbf{r}_i\}$ also do not lead to new quantum states. Thus, the reduced set obtained from (A.4) by selecting only functions with $\{m_i\}$ and $\{\mathbf{r}_i\}$ not interrelated by independent permutations, remains complete in function space with given spin S. It is easy to establish directly that this set will be orthogonal [see (16)]. Combining mutually orthogonal reduced sets with various $S = (1/2)(0), \ldots, N/2$, i.e., all irreducible representations of the group of permutations, we obtain a complete orthonormalized set in the space of antisymmetric functions with respect to simultaneous permutations in $\{\sigma_i\}$ and $\{\mathbf{x}_i\}$.

3. PARTITION FUNCTION

The Hamiltonian of a system of nonrelativistic particles does not act on the spin variables although the partition function has a component which owes its origin to the permutation symmetry. In a degenerate quantum system the exchange component of the partition function represents dominant contributions.

Summation (9) over the eigenvalues $\{m_i\}$, $\{\mathbf{r}_i\}$, $S(S+1)$ not related by permutations in $\{m_i\}$ and $\{\mathbf{r}_i\}$ for the density matrix $\hat{F} = \exp(-\beta\hat{H})$ where $\beta = 1/kT$ is the reciprocal temperature, yields the following partition function of the system:

$$Z = \sum_S (2S+1) \sum_n W_S(\hat{P}_n) \times \int d^N x \left\langle \{\mathbf{r}_i\} \middle| \exp(-\beta\hat{H}) \middle| \hat{P}_n \{\mathbf{r}_i\} \right\rangle. \tag{18}$$

The factor $1/N!$ on transition from (9) to (18) compensates for the $N!$ fold recurrent contributions associated with the permutations in $\{\mathbf{r}_i\}$. Summation over $\{m_i\}$ not related by permutations is replaced in (18) by summation over $m = \sum_i m_i$. Since m is the same as the eigenvalue of the complete spin projection operator of the system, its value for fixed S passes through $2S+1$ values in the range $-S \le n \le +S$ which give the same contributions to the sum (18) as a result of the independence of $W_S(\hat{P}_n)$ on m.

Using (12), we isolate the diagonal component in (18):

$$Z = \frac{1}{N!} \sum_S (2S+1) \int d^N r \left\langle \{\mathbf{r}_i\} \middle| \exp(-\beta\hat{H}) \middle| \{\mathbf{r}_i\} \right\rangle + \frac{1}{N!} \sum_S (2S+1) \sum_{n=2}^{N!} \frac{W_S(\hat{P}_n)}{W_S(\hat{P}_1)} \times \int d^N r \left\langle \{\mathbf{r}_i| \exp(-\beta\hat{H}) |\hat{P}_n \{\mathbf{r}_i\}\} \right\rangle. \tag{19}$$

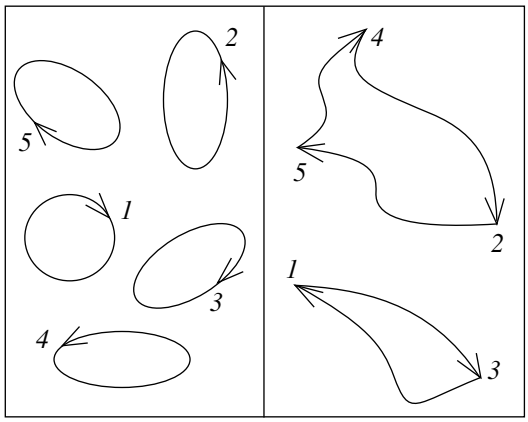


Fig. 1. Topological structure of Feynman paths comprising diagonal (left) and exchange (right) components of the partition function of indistinguishable quantum particles.

The exchange is represented by the off-diagonal component of the partition function.

Following Feynman [30], we write the matrix elements in (19) in the form of continuous path $\tilde{R}(t)$ in $3N$-dimensional space connecting the points $R^{(1)} = \{\mathbf{r}_i\}$ and $R^{(n)} = \hat{P}_n \{\mathbf{r}_i\}$:

$$Z = \sum_S (2S+1) \int dR \sum_n W_S(\hat{P}_n) \times \int_{R^{(1)}}^{R^{(n)}} \exp\left\{ \frac{i}{\hbar} \Phi([\tilde{R}(t)]) \right\} D\tilde{R}(t), \tag{20}$$

where $\Phi([\tilde{R}(t)])$ is the functional of the action on the path $\tilde{R}(t)$ apart from the formal replacement of the time t with the imaginary reciprocal temperature: $t \longrightarrow -i\hbar\beta$. Each $3N$-dimensional trajectory $\tilde{R}(t)$ uniquely corresponds to N three-dimensional paths $\mathbf{r}_i(t)$, $i = 1, 2, \ldots, N$. Since R and R_n are related by the permutation of $\hat{P}_n$, the end of any three-dimensional path $\mathbf{r}_i(t)$ is the beginning of another path, forming a ring structure topologically similar to the permutation graph of $\hat{P}_n$ (see Appendix). The diagonal component of the partition function [first term in (19)] only includes the identity permutation of $\hat{P}_1$, i.e., all the path close onto themselves (Fig. 1). The exchange component is obtained by integrating the set of topologically different ring structures of linked paths. Each topological structure has its own weight $W_S(\hat{P}_n)$.

We shall use a definition of the continuous integral in terms of the limiting multidimensional integral [119]

$$\int dR \int_{R^{(1)}}^{R^{(n)}} \exp\left\{ \frac{i}{\hbar} \Phi([\tilde{R}(t)]) \right\} D\tilde{R}(t)$$

$$= \lim_{M\to\infty}\left(\frac{Mm_0}{2\pi\beta\hbar^2}\right)^{3NM/2}\int dR(1)dR(2)\ldots dR(M) \quad (21)$$

$$\times \exp\left\{\frac{i}{\hbar}\Phi(R(1), \ldots, R(M))\right\},$$

where $R(k) \equiv \{\mathbf{r}_i(k)\} = \tilde{R}(kt/M)$, $k = 1, 2, 3, \ldots, M$, and the action $\Phi(R(1), R(2), \ldots, R(M))$ is calculated on the broken line connecting the points $R(1), R(2), \ldots, R(M)$, $\hbar$ is Planck's constant, and m_0 is the particle mass.

The continuous integrals can be calculated numerically using the Monte Carlo method [120, 121]. A computer is used to implement a Markov random walk process in Feynman path space and all methods of linking these. A random walk in path space with a fixed topological structure linking $\hat{P}_n$ is implemented with the weighting function

$$\sim\prod_{j=1}^{M}\left\langle \{\mathbf{r}_i(j)\}\left|\exp\left(-\frac{\beta\hat{H}}{M}\right)\right|\{\mathbf{r}_i(j+1)\}\right\rangle, \quad (22)$$

$$\{\mathbf{r}_i(M+1)\} = \hat{P}_n\{\mathbf{r}_i(1)\}.$$

A parallel random walk should be made in the topological structure space with the weighting function $(2S+1)W_S(\hat{P}_n)$. It is meaningful to first sum (18) over the graphs within classes, retaining summation over classes instead of summation over n. According to (10) and (6), the values of $W_S(\hat{P}_n)$ then depend on the method of constructing the bilinear combination (A.4) and require a knowledge of all $c_{nk}(S)$. Summation within classes eliminates this dependence.

In (8) we rewrite the complete wave function $\psi(S, \{m_i\}, \{\mathbf{r}_i\}; \{\sigma_i\}, \{\mathbf{x}_i\})$ on the right of the operator $\hat{F}$ in the form of a bilinear combination of spin $\chi_J^n(S, \{m_i\}; \{\sigma_i\})$ and coordinate $f_J^k(S, \{\mathbf{r}_i\}; \{\mathbf{x}_i\})$ wave functions symmetrized by Young symmetrization operators with mutually transposed dummy Young diagrams and various n and k permutations of the arguments in the cells of the Young diagrams. The integral (8) over $\{\mathbf{r}_i\}$ is a linear combination of matrix functionals with Young symmetrization operators $\hat{J}^k(S)$ having the same dummy Young diagrams and different k-permutations of the arguments:

$$\int d^N r\left\langle S, \{m_i\}, \{\mathbf{r}_i\}\left|\hat{F}\right|S, \{m_i\}, \{\mathbf{r}_i\}\right\rangle$$

$$= N!\sum_{\{\sigma_i=\pm1/2\}}\tilde{\chi}(S, \{m_i\}, \{\sigma_i\})$$

$$\times\sum_{nk} b_{nk}\chi_J^n(S, \{m_i\}; \{\sigma_i\}) \quad (23)$$

$$\times\int d^N r\int d^N x f(\{\mathbf{r}_i\}; \{\mathbf{x}_i\})\hat{F} f_J^k(S, \{\mathbf{r}_i\}; \{\mathbf{x}_i\}).$$

As a result of the degeneracy of the matrix functional with respect to permutations of the arguments in the cells of the Young diagram [see (3)], the index k can be removed from f_J^k in (23) and the matrix functional $G(\hat{J}(S))$ can be removed from the summation sign over k. Summing (23) over unrelated permutations of $\{m_i\}$ and also over S and dividing the result by $N!$ as in (18), we obtain the trace of the operator $\hat{F} = \exp(-\beta\hat{H})$, i.e., the partition function of the system:

$$Z = \sum_S (2S+1)\left(\sum_{\{\sigma_i=\pm1/2\}}\tilde{\chi}(S, \{m_i\}; \{\sigma_i\})\right.$$

$$\left.\times\sum_{nk} b_{nk}\chi_J^n(S, \{m_i\}; \{\sigma_i\})\right)G(\hat{J}(S)) \quad (24)$$

$$= \sum_S (2S+1)\tilde{W}_S G(\hat{J}(S)),$$

where

$$\tilde{W}_S = \sum_{\{\sigma_i=\pm1/2\}}\tilde{\chi}(S, \{m_i\}; \{\sigma_i\})$$

$$\times\sum_{nk} b_{nk}\chi_J^n(S, \{m_i\}; \{\sigma_i\}). \quad (25)$$

In (25) $\tilde{W}_S$ does not depend on $\{m_i\}$ for the same reasons as $W_S(\hat{P}_n)$ in (10).

In order to obtain an explicit expression for $\tilde{W}_S$ we could calculate (25) allowing for the orthonormalization of the spin functions. A simpler and faster method involves comparing the high-temperature limits of formula (24) and its equivalent notation (19). For $\beta \longrightarrow 0$ only the contributions from the diagonal matrix elements survive in both expressions and in particular (24) has the high-temperature asymptotic form

$$G(\hat{J}(S)) = \sum_n (\pm1)G(\hat{P}_n) \longrightarrow G(\hat{P}_1)$$

$$= \int d^N r\left\langle \{\mathbf{r}_i\}\left|\exp(-\beta\hat{H})\right|\{\mathbf{r}_i\}\right\rangle. \quad (26)$$

Equating (24) and (19) in this limit, we obtain

$$\sum_S (2S+1)\tilde{W}_S\int d^N r\left\langle \{\mathbf{r}_i\}\left|\exp(-\beta\hat{H})\right|\{\mathbf{r}_i\}\right\rangle$$

$$= \frac{1}{N!}\sum_S (2S+1)\int d^N r\left\langle \{\mathbf{r}_i\}\left|\exp(-\beta\hat{H})\right|\{\mathbf{r}_i\}\right\rangle, \quad (27)$$

whence

$$\tilde{W}_S = \frac{1}{N!} = W_S(\hat{P}_1). \quad (28)$$

Taking into account (28), the partition function has the form

$$Z = \frac{1}{N!}\sum_S (2S+1)\sum_n^S (-1)^{c(n)} \times \int d^N r \left\langle \{\mathbf{r}_i\}\middle|\exp(-\beta\hat{H})\middle|\hat{P}_n\{\mathbf{r}_i\}\right\rangle, \quad (29)$$

where the index S at the summation sign indicates that summation is performed over all permutations of $\hat{P}_n$ contained in the Young symmetrization operator corresponding to the spin S, for which the permutation of the arguments in the cells of the Young diagram is negligible. Thus, in order to obtain all topological structures of linked trajectories in (29), it is sufficient to construct all graphs of $\hat{P}_n$ obtained after removing the brackets in the Young symmetrization operator

$$\hat{J}(S) = (1 \pm \hat{n}_w)(1 \pm \hat{n}_v)\ldots(1 \pm \hat{n}_c)(1 \pm \hat{n}_b)(1 \pm \hat{n}_a), \quad (30)$$

and to reorient these: $\hat{N}_n \longrightarrow \hat{N}_n^{-1}$. The same result is achieved by reversing the order of the pair symmetrization operators in (30),

$$\hat{J}(S) = (1 \pm \hat{n}_a)(1 \pm \hat{n}_b)\ldots(1 \pm \hat{n}_s)(1 \pm \hat{n}_v)(1 \pm \hat{n}_w), \quad (31)$$

or by replacing the number pair permutations by position pair permutations,

$$\hat{J}(S) = (1 \pm \hat{p}_w)(1 \pm \hat{p}_v)\ldots(1 \pm \hat{p}_c)(1 \pm \hat{p}_b)(1 \pm \hat{p}_a). \quad (32)$$

Since the total number of graphs generated by the Young symmetrization operators in (29) is of the order of $N!$, numerical calculations of the partition function and the equilibrium averages can only be made after reducing these to dummy graphs. This reduction is achieved numerically by computer: graphs from one class with a representative graph of this class $\hat{\Pi}_k$ are grouped with a common factor $\omega_S(k)$:

$$Z = \frac{1}{N!}\sum_S (2S+1)\sum_k \omega_S(k) \times \int d^N r \left\langle \{\mathbf{r}_i\}\middle|\exp(-\beta\hat{H})\middle|\hat{\Pi}_k\{\mathbf{r}_i\}\right\rangle = \frac{1}{N!}\sum_k \omega(k)\int d^N r \left\langle \{\mathbf{r}_i\}\middle|\exp(-\beta\hat{H})\middle|\hat{\Pi}_k\{\mathbf{r}_i\}\right\rangle, \quad (33)$$

where

$$\omega(k) = \sum_S (2S+1)\omega_S(k)$$

is the sum over all spin states. The sum over k in (33) is taken over classes of the group of permutations and contains far fewer terms than (29) (see Table 1): for $N \sim 10$ it is of the order of a few tens. Each class is represented by a dummy graph which is then uniquely defined by dividing into dummy cycles: ν_i is the number of cycles from i vertices. Taking this factor into account, the notation (33) becomes clearer if k is replaced by the multidimensional index $\{\nu_i\}$:

$$Z = \frac{1}{N!}\sum_{\{\nu_i\}} \omega(\{\nu_i\}) \times \int d^N r \left\langle \{\mathbf{r}_i\}\middle|\exp(-\beta\hat{H})\middle|\hat{\Pi}_{\{\nu_i\}}\{\mathbf{r}_i\}\right\rangle. \quad (34)$$

Summation in (34) is performed over all divisions $\{\nu_i\}$ satisfying the condition

$$\sum_i i\nu_i = N.$$

Calculation of $\omega(\{\nu_i\})$ involves constructing all permissible Young dummy diagrams of the coordinate function and reducing the graphs of these diagrams to dummy graphs. If it is necessary to distinguish between states with different S for each $\{\nu_i\}$ we construct the distribution normalized to unity

$$\Gamma_{\{\nu_i\}}(S) = (2S+1)\frac{\omega_S(\{\nu_i\})}{\omega(\{\nu_i\})},$$

which reflects the contribution of the spin state S to this topological structure of $\{\nu_i\}$. The control tables $\Gamma_{\{\nu_i\}}(S)$ and $\omega(\{\nu_i\})$ are stored in the computer memory. An example of these data for $N = 5$ is given in Table 2. It is only advisable to make direct calculations of the control tables for a comparatively small number of fermions in the system and we made such calculations by computer for $N \le 10$. As N increases, the number of arithmetic operations required increases as $N!$ and direct calculations of the control tables become almost impracticable. For this case we developed a special fundamentally exact method aimed at calculations for large numbers of particles. This method is not described in the present study.

On the basis of (34), the spin state distribution function has the form

$$\rho(S) = \frac{2S+1}{N!}\sum_{\{\nu_i\}} \omega_S(\{\nu_i\}) \times \int \frac{d^N r}{Z}\left\langle \{\mathbf{r}_i\}\middle|\exp(-\beta\hat{H})\middle|\hat{\Pi}_{\{\nu_i\}}\{\mathbf{r}_i\}\right\rangle. \quad (35)$$

4. NUMERICAL CALCULATIONS OF PATH INTEGRALS FOR SYSTEMS OF SPIN PARTICLES WITH EXCHANGE

In this section we present results of applying the Monte Carlo path integral method to the modeling of a hydrogen atom, a hydrogen molecule, a Be^+ ion, and an Li atom. The principal aim of the calculations was to assess the possibilities of the method for application to more complex systems. It is fundamentally important to study those effects in the behavior of a multielectron system at room temperature which are caused by the

Table 2. Example of a control table for $N = 5$ spin 1/2 fermions containing relative combinatorial weights $\omega'(\{\nu_i\})$ of all methods of dividing unnumbered trajectories into cycles $\{\nu_i\}$

$\{\nu_i\}$	$\Gamma\{\nu_i\}(S)$			$\omega'(\{\nu_i\})$
	$S = 1/2$	$S = 3/2$	$S = 5/2$	
(0, 0, 0, 0, 1)	0	−0.2000000	1.2000000	10
(1, 0, 0, 1, 0)	−0.0714285	0	1.0714285	–14
(0, 1, 1, 0, 0)	0.0740741	−0.1851852	1.1111111	–9
(2, 0, 1, 0, 0)	−0.0606061	0.1515152	0.9090909	11
(1, 2, 0, 0, 0)	0.0625000	0	0.9375000	8
(3, 1, 0, 0, 0)	0.0476190	0.2380952	0.7142857	–7
(5, 0, 0, 0, 0)	0.1666666	0.3333333	0.5000000	1

Note: $\Gamma\{\nu_i\}(S)$ is the relative contribution of spin state S to $\omega'(\{\nu_i\})$. The weights $\omega'(\{\nu_i\}) \sim \omega(\{\nu_i\})$ are normalized using the condition $\omega'(N, 0, \ldots, 0) = 1$.

permutation symmetry. The Monte Carlo path integral method can combine an explicit description of permutation symmetry in spin systems with fundamentally exact calculations of the mixed quantum state at nonzero temperature. This level of detail is unique and merits detailed testing.

4.1. Single-Electron System

In [122] the path-integral Monte Carlo method was used to make test calculations of the ground quantum state of a hydrogen atom in a spherical cavity. The presence of a confining cavity is required because of the fundamental instability of a free atom in vacuum at any nonzero temperature. The calculations were made using a weakly singular interaction functional [123, 124]. In this study an analytic solution of the Schrödinger equation is obtained for this system when the cavity radius is comparable with the Bohr radius (strong compression). Excellent agreement is observed between the results of the modeling and the analytic calculations. In the Monte Carlo calculations the ground quantum state of the atom is achieved by cooling the atom with simultaneous pushing up of excited states by compressing the confining cavity.

The analytic calculations (solution of the Schrödinger equation) for the ground state of a hydrogen atom in a spherical cavity of radius 2 Å give –13.05 eV whereas Monte Carlo calculations using path integrals with a principal kinetic energy estimator [91, 125] give –13.02 ± 0.20 eV. Using a virial kinetic energy estimator [92, 126, 127] strictly speaking requires the inclusion of additional terms to reflect the contributions from interaction of the paths with the shell walls, which were neglected in the virial estimator in [122]. However, even calculations using an incomplete virial estimator yielded realistic values of –12.7 ± 0.5 eV which indicates that the neglected terms make relatively small contributions. On the basis of these results it was concluded that there were no systematic errors in the quantum statistics representation in the form of prelimiting Feynman paths, even in systems with a singular potential. It should be noted that as yet no general confirmation of this statement exists [119]. Our subsequent calculations using more powerful computers and larger amounts of collected statistics giving results with a smaller statistical error confirmed this conclusion.

4.2. Two-Electron System

The nonsymmetrized two-particle eigenfunctions of the coordinate and spin projection operators have the form

$$\begin{aligned} f(\mathbf{r}_1, \mathbf{r}_2; \mathbf{x}_1, \mathbf{x}_2) &= \delta(\mathbf{x}_1 - \mathbf{r}_1)\delta(\mathbf{x}_2 - \mathbf{r}_2), \\ \chi(m_1, m_2; \sigma_1, \sigma_2) &= \chi(m_1, \sigma_1)\chi(m_2, \sigma_2). \end{aligned} \tag{36}$$

Two irreducible representations of the group of permutations exist, corresponding to the singlet $S = 0$ and triplet $S = 1$ spin states. Accordingly, the complete two-particle wave function may be antisymmetrized by two methods:

$$\begin{aligned} &\psi\left(S = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, m_1, m_2, \mathbf{r}_1, \mathbf{r}_2; \sigma_1, \sigma_2, \mathbf{x}_1, \mathbf{x}_2\right) \\ &= \frac{1}{2}[\chi(m_1, m_2; \sigma_1, \sigma_2) \mp \chi(m_1, m_2; \sigma_2, \sigma_1)] \\ &\times [f(\mathbf{r}_1, \mathbf{r}_2; \mathbf{x}_1, \mathbf{x}_2) \pm f(\mathbf{r}_1, \mathbf{r}_2; \mathbf{x}_2, \mathbf{x}_1)] \\ &= \frac{1}{2}[\chi(m_1, m_2; \sigma_1, \sigma_2) \mp \chi(m_2, m_1; \sigma_1, \sigma_2)] \\ &\times [f(\mathbf{r}_1, \mathbf{r}_2; \mathbf{x}_1, \mathbf{x}_2) \pm f(\mathbf{r}_2, \mathbf{r}_1; \mathbf{x}_1, \mathbf{x}_2)]. \end{aligned} \tag{37}$$

The diagonal matrix elements have the following structure:

$$\left\langle S = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, M, \mathbf{r}_1, \mathbf{r}_2 \middle| \exp(-\beta\hat{H}) \middle| S = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, M, \mathbf{r}_1, \mathbf{r}_2 \right\rangle$$

$$= \frac{1}{2}(\langle \mathbf{r}_1, \mathbf{r}_2||\mathbf{r}_1, \mathbf{r}_2\rangle \pm \langle \mathbf{r}_1, \mathbf{r}_2||\mathbf{r}_2, \mathbf{r}_1\rangle \quad (38)$$

$$\pm \langle \mathbf{r}_2, \mathbf{r}_1||\mathbf{r}_1, \mathbf{r}_2\rangle + \langle \mathbf{r}_2, \mathbf{r}_1||\mathbf{r}_2, \mathbf{r}_1\rangle).$$

The partition function has the form

$$Z = \frac{3}{2!}\int d\mathbf{r}_1 d\mathbf{r}_2(\langle \mathbf{r}_1, \mathbf{r}_2||\mathbf{r}_1, \mathbf{r}_2\rangle - \langle \mathbf{r}_1, \mathbf{r}_2||\mathbf{r}_2, \mathbf{r}_1\rangle)$$

$$+ \frac{1}{2!}\int d\mathbf{r}_1 d\mathbf{r}_2(\langle \mathbf{r}_1, \mathbf{r}_2||\mathbf{r}_1, \mathbf{r}_2\rangle + \langle \mathbf{r}_1, \mathbf{r}_2||\mathbf{r}_2, \mathbf{r}_1\rangle) \quad (39)$$

$$= \int d\mathbf{r}_1 d\mathbf{r}_2(2\langle \mathbf{r}_1, \mathbf{r}_2||\mathbf{r}_1, \mathbf{r}_2\rangle - \langle \mathbf{r}_1, \mathbf{r}_2||\mathbf{r}_2, \mathbf{r}_1\rangle).$$

The matrix elements are written in the form of Feynman continuous integrals:

$$\langle \mathbf{r}_1, \mathbf{r}_2||\mathbf{r}_2, \mathbf{r}_1\rangle = \langle \mathbf{r}_1, \mathbf{r}_2|\exp(-\beta\hat{H})|\mathbf{r}_2, \mathbf{r}_1\rangle$$
$$= \int\int \exp\{-\beta\Phi([\mathbf{r}_1(t)], [\mathbf{r}_2(t)])\} D\mathbf{r}_1(t) D\mathbf{r}_2(t), \quad (40)$$

where

$$\Phi([\mathbf{r}_1(t)], [\mathbf{r}_2(t)]) = \lim_{M\to\infty}\left\{\sum_{j=1}^{M}\left[\frac{Mm_0}{2\beta^2\hbar^2}\right.\right.$$

$$\times[(\mathbf{r}_1(j+1) - \mathbf{r}_1(j))^2 + (\mathbf{r}_2(j+1) - \mathbf{r}_2(j))^2] \quad (41)$$

$$\left.\left. + \frac{1}{M}\left(\frac{e^2}{|\mathbf{r}_1(j) - \mathbf{r}_2(j)|} + U(\mathbf{r}_1(j)) + U(\mathbf{r}_2(j))\right)\right]\right\},$$

$U(\mathbf{r})$ is the operator of the electron interaction with an external potential field, $\mathbf{r}(j)$ is the broken curve connecting the points on the path $\mathbf{r}(t)$ with fixed time interval t.

The measure in the functional integral (40) is determined from the condition that for a free particle the density matrix in the coordinate representation

$$\rho(\mathbf{r}, \mathbf{r}') = \langle \mathbf{r}|\exp(-\beta\hat{H})|\mathbf{r}'\rangle$$

for $\beta \longrightarrow 0$ becomes a δ function, $\rho(\mathbf{r}, \mathbf{r}') \longrightarrow \delta(\mathbf{r} - \mathbf{r}')$ [30]:

$$D\mathbf{r}(t) = \lim_{M\to\infty}\left(\frac{2\pi\hbar^2\beta}{m_0 M}\right)^{-3M/2} d\mathbf{r}(1)\ldots d\mathbf{r}(M). \quad (42)$$

The two-particle partition function has the following structure:

$$Z = \int d\mathbf{r}_1 d\mathbf{r}_2(2(\mathbf{r}_1 \bigcirc\!\!\!\!\!\!\bigcirc\, \mathbf{r}_2 \bigcirc\!\!\!\!\!\!\bigcirc) - \mathbf{r}_1 \bigcirc\!\!\!\!\!\!\bigcirc\, \mathbf{r}_2). \quad (43)$$

The first, diagonal, term corresponds to the normalized distribution over spin states $\Gamma_{(2,0)}(0) = 1/4$, $\Gamma_{(2,0)}(1) = 3/4$, while the second, off-diagonal, term corresponds to $\Gamma_{(0,1)}(0) = -1/2$, $\Gamma_{(0,1)}(1) = 3/2$.

It can be seen from (43) that the linked trajectories should be taken into account with negative weight which makes it difficult to construct a standard Markov random walk which assumes the existence of a positive-definite measure of microstates. The problem is solved by redefining the measure with a corresponding compensating redefinition of the estimators. Let us assume that in the microstate space x

$$\langle F\rangle = \int dx F(x)\omega(x) \quad (44)$$

and $\omega(x)$ changes sign. We determine the normalized function $\Omega(x) \geq 0$:

$$\Omega(x) = \frac{|\omega(x)|}{\int dx|\omega(x)|}. \quad (45)$$

Then in terms of $\Omega(x)$ (44) is rewritten in the form

$$\langle F\rangle = \frac{\int dx\,\mathrm{sgn}(\omega(x))F(x)\Omega(x)}{\int dx\,\mathrm{sgn}(\omega(x))\Omega(x)}$$
$$= \frac{\langle \mathrm{sgn}(\omega(x))F(x)\rangle_+}{\langle \mathrm{sgn}(\omega(x))\rangle_+}, \quad (46)$$

where $\langle\ldots\rangle_+$ denotes averaging with a positive-definite weighting function $\Omega(x)$. The relative variance of the mean (46) calculated in a finite sample of Markov microstates, is made up of the relative variance of the numerator and denominator (46) where for $\langle \mathrm{sgn}(\omega(x))\rangle_+ \longrightarrow 0$ the latter diverges so that the presence of negative contributions in the partition function (39) makes it difficult to calculate the canonical averages. As the temperature decreases the variance, and with it the statistical error of the numerical value (46) calculated in finite samples, increases which necessitates generating longer Markov processes by computer and thus requires more computer time. The presence of negative contributions in the partition function is required to compensate for the contributions which do not satisfy the conditions of permutation symmetry. The expenditure of computer resources on the calculation of these mutually compensating contributions in fact does not increase the accumulated statistics and in this sense this is lost effort. As the temperature decreases, the fraction of this lost effort increases. The problem only becomes really important for systems with more than two electrons when $T \longrightarrow 0$. For two electrons the sum of the combinatorial weights in the partition functions and the denominator in (46) are nonzero even in the limiting case $T = 0$. It should be noted that the problem for the Monte Carlo method with path integrals is not to calculate the ground quantum state corresponding to $T = 0$. The method was developed to solve temperature problems belonging to the statisticomechanical category. The power of the method specifically lies in the possibility of making direct calculations of the statistical distribution over steady quantum states, bypassing their individual calculations. However, it is always interesting to broaden the range of efficacy of the method, in this case to lower temperatures. Our experience of using the method shows that as the temperature decreases, there

comes a time when the denominator in (46) decreases rapidly and calculations using the standard scheme become almost impossible. Physically this implies a transition to a highly degenerate electron component regime. The specific temperature at which this occurs depends strongly on the density of the system and the configuration of the external field. For atomic systems the critical boundary lies in the range 10^4–10^5 K. Experience shows that a value of the order of $\langle \operatorname{sgn}(\omega(x))\rangle_+ \approx 0.01$ should be taken as the permissible limit above which calculations can still be made without using special measures. However, in cases of quantum degeneracy we can propose special modifications of the calculation algorithm which allow the temperature to be reduced at least severalfold without introducing systematic error. The numerical results presented here were obtained without using these special measures since they were not necessary. The temperature of the calculations for a three-electron system is fairly high and no problems occurred with the divergence (46): the value of $\langle \operatorname{sgn}(\omega(x))\rangle_+$ did not fall below 0.1. The special case of strong degeneracy is outside the scope of this publication. A detailed analysis is planned for a future study where a method of solving the problem of negative sign will be developed and numerical results obtained with and without using special algorithms to solve the problem will be analyzed systematically.

An estimate of the statistical errors attributable to the finiteness of the significant samples obtained by the Monte Carlo method may be made using the same scheme as for problems without exchange symmetry and negative contributions since, from the formal point of view, as a result of redefining the measure using formula (46) and transferring the negative signs from the measure to the averageable quantities, the process of calculating the averages here does not differ in any respect from the calculations of averages in systems obeying classical statistics. In principle, the procedure remains valid regardless of the sign of the quantity being averaged. As a result of the statistical correlation of the configurations obtained in the Markov random walk process, the variance of the corresponding estimator cannot be used to estimate the error. The standard scheme [120] for calculating the errors of the numerator and the denominator (46) involves calculating the partial averages $\bar{F}_k$ along the Markov path and then, considering these as statistically independent quantities, estimating the absolute statistical error of the calculated average $\bar{F}$ using the formula

$$\Delta\bar{F} = \left\{\frac{1}{n-1}\left[\frac{1}{n}\sum_{k=1}^{n}\bar{F}_k^2 - \left(\frac{1}{n}\sum_{k=1}^{n}\bar{F}_k\right)^2\right]\right\}^{1/2}. \qquad (47)$$

Formula (47) becomes absolutely exact in the absence of correlations between the partial averages $\bar{F}_k$. This condition is better satisfied, the larger the length of the section of the Markov path L/n expended in calculating each $\bar{F}_k$. The theoretically optimum value of the parameter n depends on the total length of the Markov trajectory L although, if L is much greater than the characteristic autocorrelation length along the trajectory (a necessary condition for accuracy of the Monte Carlo calculations) an increase in L/n causes $\Delta\bar{F}$ (47) to reach a saturation value which depends weakly on n. This value should be considered to be an estimate of the statistical error. The characteristic autocorrelation length is estimated by the control map method [120] using the length of the nonsteady-state section of the Markov chain. Generally, the calculations are considered to be correct if this length does not exceed ~0.1% of the total length of the Markov process and this value with a factor of ten margin is used in the estimates. Since the error is estimated to one decimal place, it is meaningless to determine the optimum value of n with precision. Estimates are usually made for $n \approx 10$. This value was used in our calculations. The error was checked independently by reproducing the calculations for the same values of the parameters but from different initial configurations. The initial section of the chain, 10–20% of the length, no less than an order of magnitude shorter than the autocorrelation length, was considered to be nonsteady-state and was not used to calculate the averages. Some very effective measures to shorten the nonsteady-state section were used in the algorithm, but a detailed discussion of these is outside the scope of the present publication.

Numerical test calculations were made for a hydrogen molecule in which the exchange "interaction" is observed most clearly for the minimum number of electrons. In (40) as for the case of a hydrogen atom we used the weakly singular functional $\Phi(R(1), R(2), \ldots, R(M))$ [122] which contains no nonintegrable singularities at sites of atomic nuclei. In order to avoid effects caused by the motion of nuclei, these were fixed at an equilibrium distance $a = 0.7416$ Å. The calculations were made for $T = 2500$ K and the number of vertices in each trajectory was $M = 640$. The molecule was located in an impermeable spherical cavity of radius 3 Å in order to push the excited energy levels of the electron shell upward and reduce their weight in the partition function [128]. The energy and electron density distribution pattern were calculated. Calculations using the virial estimator [92, 126, 127] yielded the following results: kinetic energy of electron subsystem $\langle K\rangle = 28.5 \pm 0.1$ eV, potential energy of electron subsystem $\langle U\rangle = -79.8 \pm 0.1$ eV, total energy of electron subsystem $\langle C\rangle = -51.3 \pm 0.1$ eV, total molecular energy without kinetic energy of nuclei $\langle E\rangle = -31.9 \pm 0.1$ eV. This last value is the total ground-state energy of the molecule minus the energy of the zero-point vibrations of the nuclei $h\nu_{\text{nucl}}/2$ and may be determined experimentally [129]:

$$E^{\text{exp}} = 2E_{\text{H}} - E_{\text{dis}}^{\text{exp}} - 0.5h\nu_{\text{nucl}}^{\text{exp}}$$

$$= (-2 \times 13.598 - 4.478 - 0.5 \times 4.1357 \times 10^{-15}$$

$$\times 1.3181 \times 10^{14})\ \text{eV} = -31.946\ \text{eV},$$

where E_H is the ground-state energy of the hydrogen atom and E_{dis}^{exp} is the experimental dissociation energy of the hydrogen molecule. Our calculations reproduce to within 0.2% the experimentally determined value and yield the dissociation work of the hydrogen molecules $E_{dis} = \langle E \rangle - 2E_H = 4.4 \pm 0.1$ eV which agrees with the experimental value $E_{dis}^{exp} = 4.478$ eV.

The calculations showed that cooling the system causes an abrupt decrease in the variance of the principal energy estimator [91, 125] obtained by direct differentiation of the free energy with respect to the reciprocal temperature so that at $T = 2500$ K it also becomes suitable for calculations in addition to the virial estimator. The values obtained by averaging the principal estimator were:

$$\langle K \rangle = 28.5 \pm 0.5 \text{ eV}, \quad \langle C \rangle = -51.3 \pm 0.5 \text{ eV},$$
$$\langle E \rangle = -31.9 \pm 1.0 \text{ eV}.$$

The partition function of the system includes two spin states: $S = 0$ and $S = 1$. In our calculations the molecule spent 98% of the effective Markov time in the $S = 0$ state. An analysis of the fluctuation behavior suggests that an increase in the volume of the random sample leads to a more accurate and also lower than $\langle S \rangle = 0.02$ value of the average spin number.

4.3 Three-Electron System

We express the complete wave function of a system of three nonrelativistic electrons in the form of a symmetrized bilinear combination of spin and coordinate functions having $3! \times 3! = 36$ terms:

$$\psi(S, \{m_i\}, \{\mathbf{r}_i\}; \{\sigma_i\}, \{\mathbf{x}_i\}) = \sum_{n,k} c_{nk}(S) \times \chi(\{m_i\}; \hat{P}_n\{\sigma_i\}) f(\{\mathbf{r}_i\}; \hat{P}_k\{\mathbf{x}_i\}), \tag{48}$$

where $\chi(\{m_i\}; \{\sigma_i\}) = \chi(m_1; \sigma_1)\chi(m_2; \sigma_2)\chi(m_3; \sigma_3)$ are the eigenfunctions of the single-particle spin projection operators, and $f(\{\mathbf{r}_i\}; \{\mathbf{x}_i\}) = \delta(\mathbf{x}_1 - \mathbf{r}_1)\delta(\mathbf{x}_2 - \mathbf{r}_2)\delta(\mathbf{x}_3 - \mathbf{r}_3)$ are the eigenfunctions of the coordinate operator. At the same time the complete wave function of a state with spin S is a bilinear combination of spin $\chi_J(\{m_i\}; \{\sigma_i\})$ and coordinate $f_J(\{\mathbf{r}_i\}; \{\mathbf{x}_i\})$ functions symmetrized in accordance with mutually transposed Young diagrams from three cells with arguments as in Fig. 2. For example, for $S = 1/2$ in Fig. 2 we have

$$\begin{aligned} f_J(\{\mathbf{r}_i\}; \{\mathbf{x}_i\}) = {} & \delta(\mathbf{x}_1 - \mathbf{r}_1)\delta(\mathbf{x}_2 - \mathbf{r}_2)\delta(\mathbf{x}_3 - \mathbf{r}_3) \\ & + \delta(\mathbf{x}_2 - \mathbf{r}_1)\delta(\mathbf{x}_1 - \mathbf{r}_2)\delta(\mathbf{x}_3 - \mathbf{r}_3) \\ & - \delta(\mathbf{x}_3 - \mathbf{r}_1)\delta(\mathbf{x}_2 - \mathbf{r}_2)\delta(\mathbf{x}_1 - \mathbf{r}_3) \\ & - \delta(\mathbf{x}_2 - \mathbf{r}_1)\delta(\mathbf{x}_3 - \mathbf{r}_2)\delta(\mathbf{x}_1 - \mathbf{r}_3). \end{aligned} \tag{49}$$

The linearly independent functions obtained by various permutations of the arguments of the cells of the Young diagram form the basis of an irreducible representation of the permutation group corresponding to spin S.

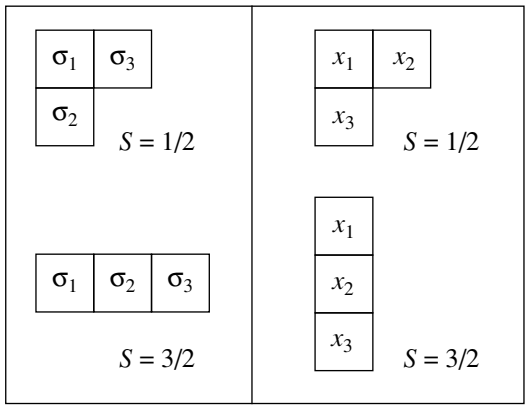


Fig. 2. Young diagrams of the spin and coordinate components of the wave function of a system of three electrons in two possible spin states which are eigenfunctions of the spin square operator of the system.

The diagonal matrix elements of the symmetrized coordinate-spin representation are reduced to a linear combination of off-diagonal elements of the nonsymmetrized coordinate representation:

$$\begin{aligned} &\langle S, m_1, m_2, m_3, \mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3 | e^{-\beta\hat{H}} | S, m_1, m_2, m_3, \mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3 \rangle \\ &\quad = \sum_{n=1}^{6} a_n' \langle \mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3 | \exp(-\beta\hat{H}) | \hat{P}_n(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) \rangle, \end{aligned} \tag{50}$$

where the set of operators $\hat{P}_n$ appearing in (50) is a Young symmetrization operator corresponding to spin $S = 1/2$ or $S = 3/2$ with filled cells. Each of the six operators $\hat{P}_n$ in (50) corresponds to its numbered graph. Since the matrix elements in (50) do not depend on the numbering of the graph vertices, the summation over six numbered graphs $\hat{P}_n$ in Fig. 3 reduces to summation over three graphs with unnumbered vertices $\hat{\Pi}_n$ in Fig. 4 with a corresponding redistribution of weights:

$$\begin{aligned} &\langle S, m_1, m_2, m_3, \mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3 | e^{-\beta\hat{H}} | S, m_1, m_2, m_3, \mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3 \rangle \\ &\quad = \sum_{n=1}^{3} a_n \langle \mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3 | \exp(-\beta\hat{H}) | \hat{\Pi}_n(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) \rangle. \end{aligned} \tag{51}$$

In accordance with Fig. 4 for $S = 1/2$ we have $a_1 = 1$, $a_2 = 1 - 1 = 0$, $a_3 = -1$, and for $S = 3/2$ we obtain $a_1 = 1$, $a_2 = -3$, $a_3 = 2$.

Since the complete set of Young symmetrization operators generates a complete set of antisymmetric wave functions and the functions unrelated by permutations in m_1, m_2, m_3, or in $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$ are orthogonal, integration (51) over $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$ yields a partition function

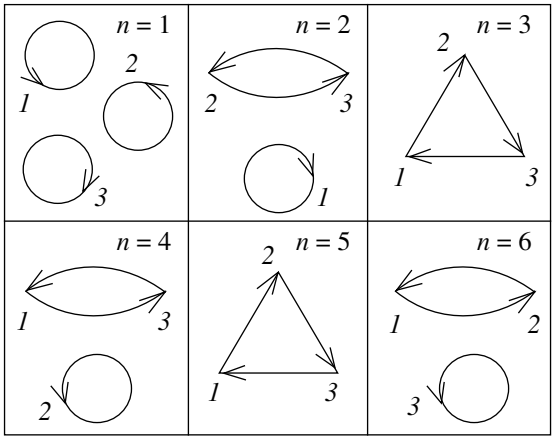


Fig. 3. Graphs representing all possible permutations contained in the Young symmetry operators of three electrons.

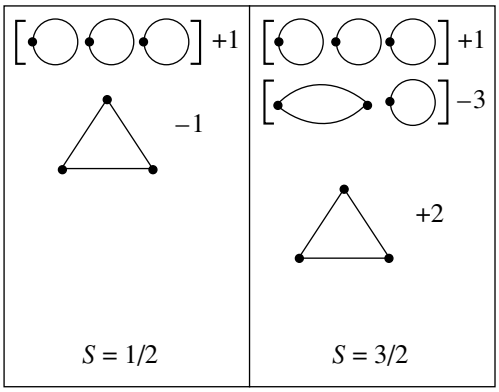


Fig. 4. Result of summation within classes of permutations in the Young symmetry operators of the coordinate part of the wave function in two possible spin states of a system of three electrons. The numbers denote the relative weights of the corresponding dummy graphs within this spin state. The negative weights indicate that the corresponding continuous integrals appear in the partition function with a negative sign.

where the factor $2S + 1$ replaces summation over m_1, m_2, m_3, and the divisor $N! = 6$ allows integration to be extended to all values of $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$ including those related by permutations:

$$Z = \frac{1}{6}\sum_{S=1/2,\,3/2}(2S+1)\sum_{n=1}^{3}a_n(S)\int d\mathbf{r}_1 d\mathbf{r}_2 d\mathbf{r}_3 \times\langle \mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3|\exp(-\beta\hat{H})|\hat{\Pi}_n(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)\rangle = \sum_{n=1}^{3}A_n\int d\mathbf{r}_1 d\mathbf{r}_2 d\mathbf{r}_3 \times\langle \mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3|\exp(-\beta\hat{H})|\hat{\Pi}_n(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)\rangle, \quad (52)$$

where

$$A_n = \frac{1}{6}\sum_{S=1/2,\,3/2}(2S+1)a_n(S),$$

$$A_1 = 1, \quad A_2 = -2, \quad A_3 = 1.$$

The relationship

$$A_1 + A_2 + A_3 = 0 \quad (53)$$

is an expression of the Pauli principle that it is impossible to place more than two electrons in a single quantum state. A simple superposition of single-electron orbitals neglecting the mutual correlations in the electron motion inside the orbitals does not completely reflect the pattern of exchange effects: the largest contributions to the partition function are made by those relative positions of the virtual trajectories for which the repelling "simultaneous" points of the trajectories are at the maximum distances. The spatial separation of these points influences the dependence of the exponential weights on the method of linking the trajectories. The multielectron wave function of the mixed state of the system is a result of a balance of these factors. A change in the external field configuration influences the weight distribution between the various methods of linking, preference being given to those inscribed in the external field geometry with the lowest values of the action calculated along them. Redistribution of the weights then gives rise to a change in the spin state of the system. Thus, the external field influences the resultant spin of the system.

Numerical calculations of equilibrium averages in the form of continuous integrals were made using a unified program for Markov random walk of virtual electron trajectories in the electric field of a beryllium nucleus. Each trajectory is represented by a broken line with $M = 40$ links and a weakly singular potential [122]. An increase in temperature makes the most representative trajectories shorter and reduces the constraint on the degree of discretization of the trajectories M. Calculations with smaller M are less cumbersome and the statistical fluctuations of the principal energy estimator are smaller. A temperature of 10–30 eV (the second ionization potential of beryllium is 18.2 eV) corresponds to strong thermal excitation of the electron shell. Although the system was placed in a solid-walled spherical cavity of radius $L = 3.78$ Å centered on the nucleus to ensure equilibrium stability at this temperature, in the thermally bound state the electron trajectories did not have time to move more than 2 Å away from the nucleus over the random walk time (10^6 steps). The existence of walls only became important after the detachment of an electron. In the first case, the system was in a locally stable metastable bound Be^+ state separated from the secondary ionization states by a high surface barrier. After electron detachment the state should be interpreted as Be^{++} against the background of a dense electron component. Figure 5 shows the evolu-

tion of the radial distribution of the electron density when the system is heated from $T = 120 \times 10^3$ K to $T = 144 \times 10^3$ K. At $T = 120 \times 10^3$ K an electron density maximum is observed at a distance of two Bohr radii which, from the point of view of the single-electron approximation, may be interpreted as the contribution of 2*s* and 2*p* orbitals. As a result of comparatively weak heating to $T = 144 \times 10^3$ K this maximum is doubled and a deep dip forms between the first high-density region of the electron cloud located near the nucleus and the second at a distance of two Bohr radii. The areas below the maxima of the electron density are in the ratio 1 : 2. The first maximum has the form of a hydrogen-like orbital with the nuclear charge number $z = 4$. It is interesting that the position of the second maximum does not change with temperature. Quite clearly, the second maximum forms mainly as a result of the motion of two electrons simultaneously on the outer 2*s*, 2*p* orbital. The $n = 2$ orbital is entropically more favorable that the $n = 1$ orbital because of the larger number of subshells. The position of the second maximum is determined by the balance of two opposing tendencies. First, as the temperature increases, higher quantum states should be included which would shift the maximum to larger distances from the nucleus. Second, the transition of a second electron from the $n = 1$ shell to $n = 2$ reduces the screening of the nuclear field for the third, outer, electron since the second and third electrons now participate in joint correlated motion in the 2*s*, 2*p* subshells, which reduces the effective radius of the second electron shell. As a result, the second maximum remains where it is and the transition of the second electron to a higher orbital intensifies the coupling between the outer electron and the Be^+ ion as a result of exchange correlation effects.

Overcoming the metastable barrier, i.e., electron detachment, occurred at $T = 200 \times 10^3$ K. In this case the second electron returned to the $n = 1$ inner orbital and the mutual screening of the nuclear field by the first and second electrons increased the radius of the inner electron orbital to half the Bohr radius followed by continuous smearing of the radial distribution of the electron density with increasing temperature.

It can be seen from these data that exchange correlation mechanisms lead to substantial layering of the electron shell in the pre-ionization state. Emptying of the inner orbital reduces the screening of the nuclear field for the inner electron, i.e., the radius of the inner orbital decreases and it becomes separated from the outer one. Conditions may be established when the outer electron orbitals are filled by emptying of inner ones followed by an electron transition from the outer orbital back to the inner one with increasing temperature. In principle, the radius of the outer electron shell may decrease during heating as a result of reduced screening of the nuclear field when inner electrons are transferred to higher orbitals. This behavior should be predicted primarily in systems with compensated charge where the role of interelectron correlations at outer orbitals is particularly important.

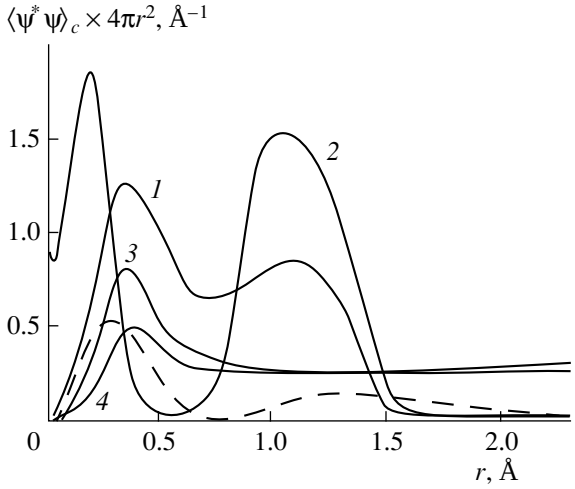


Fig. 5. Electron density distributions normalized to unity in the mixed Be^+ state at $T = 120 \times 10^3$ K (*1*), $T = 144 \times 10^3$ K (*2*) and Be^{++} at $T = 248 \times 10^3$ K (*3*), $T = 299 \times 10^3$ K (*4*) calculated by the Monte Carlo method using path integrals. The dashed curve gives the 1/3 normalized outer single-electron orbital of Be^+ in the quantum ground state using results of quantum chemical calculations [109].

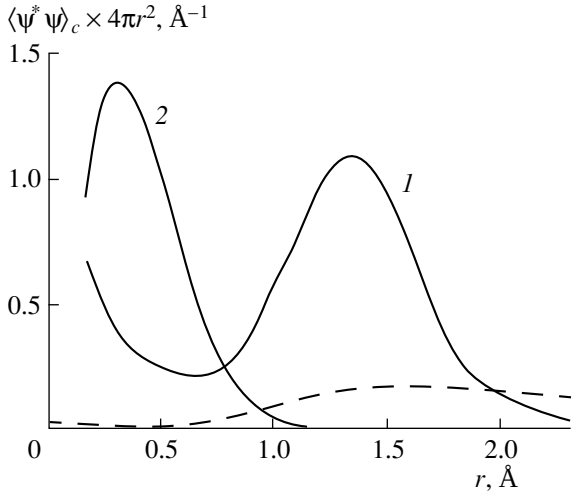


Fig. 6. Electron density distributions normalized to unity for Li at $T = 95 \times 10^3$ K (*1*) and Li at $T = 104 \times 10^3$ K (*2*), calculated by the Monte Carlo method using path integrals. The dashed curve gives the 1/3 normalized outer single-electron orbital of Li in the quantum ground state using results of quantum chemical calculations [109].

We repeated our calculations for the electrically neutral Li atom (Fig. 6). At $T = 95 \times 10^3$ K a two-electron outer shell is clearly formed as a result of the transition of one of the inner electrons to an outer orbital whose radius decreases to 1.3 Å instead of 1.65 Å for the 2*s* orbital in the field of a nucleus screened by two 1*s* electrons in the Li ground state [129]. Figuratively

speaking the atom is "compressed" before the imminent ionization jump. In the Li^+ ionized state at $T = 104 \times 10^3$ K (Fig. 6) the second electron is returned to the $n = 1$ inner orbital, partially screening it from the nuclear field. Thus, the observed irregularity in the electron shell behavior under strong thermal excitation is not a characteristic merely of Be^+ and Li.

APPENDIX

Topological Representation of Symmetry Operators

Since the Hamiltonian $\hat{H}$ of a system of N nonrelativistic fermions contains no spin variables $\{\sigma_i\} = (\sigma_1, \ldots, \sigma_N)$ and is symmetric with respect to permutations of the coordinate variables $\{\mathbf{x}_i\} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$, the complete wave function may be expressed as a symmetrized bilinear combination of spin χ and coordinate f functions:

$$\psi(\{\sigma_i\}, \{\mathbf{x}_i\}) = \sum_{n,k} c_{nk} \chi(\hat{P}_n\{\sigma_i\}) f(\hat{P}_k\{\mathbf{x}_i\}), \quad \text{(A.1)}$$

where $\hat{P}_n$ is the position permutation operator in the sequence of arguments. The permutation of the arguments may be interpreted as the transition to another function $f'(\{\mathbf{x}_i\}) = f(\hat{P}_k\{\mathbf{x}_i\})$ with the transition operator $f'(\{\mathbf{x}_i\}) = \hat{\pi}_k f(\{\mathbf{x}_i\})$. The composition law for the $\hat{\pi}_k$ operators is clearly

$$\hat{\pi}_k\hat{\pi}_n f(\{\mathbf{x}_i\}) = f(\hat{P}_n\hat{P}_k\{\mathbf{x}_i\}). \quad \text{(A.2)}$$

In addition to the position pair permutations $\hat{p}_a = \hat{p}_{ij}$, number pair permutations $\hat{n}_a = \hat{n}_{ij}$ also exist, involving the mutual permutation of two arguments having numbers i and j irrespective of their initial positions. It is easily established directly that any pair permutation operator $\hat{n}$ commutes with any pair permutation operator $\hat{p}$ and thus any number permutation operator $\hat{N}$ commutes with any position operator $\hat{P}$ whence

$$\hat{N}_n\hat{P}_k\{\mathbf{x}_i\} = \hat{P}_k\hat{N}_n\{\mathbf{x}_i\} = \hat{P}_k\hat{P}_n\{\mathbf{x}_i\} = (\hat{P}_k\hat{P}_n\hat{P}_k^{-1})\hat{P}_k\{\mathbf{x}_i\}. \quad \text{(A.3)}$$

It can be seen from (A.3) that the action of the operator $\hat{N}_n$ on the kth permutation in the sequence $\{\mathbf{x}_i\}$ reduces to the position permutations $\hat{P}_k\hat{P}_n\hat{P}_k^{-1}$ which implies that the operators $\hat{P}_n$ and $\hat{N}_n$ belong to the same class [113, 114].

We shall construct a bilinear combination (A.1) of the eigenfunctions $\chi(\{m_i\}, \{\sigma_i\})$ of the single-particle spin projection operators acting on the variables σ_1, σ_2, …, σ_N with the eigenvalues m_1, m_2, …, m_3, respectively, and the eigenfunctions $f(\{\mathbf{r}_i\}, \{\mathbf{x}_i\})$ of the coordinate operators acting on the variables $\mathbf{x}_1$, $\mathbf{x}_2$, …, $\mathbf{x}_N$ with the eigenvalues $\mathbf{r}_1$, $\mathbf{r}_2$, …, $\mathbf{r}_N$. Since the spin square operator commutes with the spin projection operator [117], the bilinear combination (A.1) may also be the eigenfunction of the spin square operator:

$$\psi(S, \{m_i\}, \{\mathbf{r}_i\}; \{\sigma_i\}, \{\mathbf{x}_i\}) = \sum_{n,k} c_{nk}(S) \times \chi(\{m_i\}; \hat{P}_n\{\sigma_i\}) f(\{\mathbf{r}_i\}; \hat{P}_k\{\mathbf{x}_i\}). \quad \text{(A.4)}$$

The complete wave function of the state having the total spin S is a bilinear combination of spin $\chi_J(\{m_i\}; \{\sigma_i\})$ and coordinate $f_J(\{\mathbf{r}_i\}; \{\mathbf{x}_i\})$ functions symmetrized in accordance with the mutually transposed Young diagrams [117, 118]. The spin function is initially symmetrized along the rows and then antisymmetrized along the columns. All linearly independent functions obtained by various permutations of the arguments in the cells of the Young diagram form the basis of the irreducible representation of the permutation group corresponding to the total spin S [116]. The Young symmetrization operator is a linear combination of the number permutation operators of the arguments. The coefficients $c_{nk}(S)$ are the result of opening the brackets in the complete bilinear combination

$$\chi_J(S, \{m_i\}; \{\sigma_i\}) = \sum_n^S (-1)^{c(n)} \chi(\{m_i\}; \hat{N}_n\{\sigma_i\}) = \sum_n^S (-1)^{c(n)} \chi(\{m_i\}; \hat{P}_n\{\sigma_i\}). \quad \text{(A.5)}$$

The index S after the sum implies that summation is performed over all permutations contained in the Young symmetrization operator corresponding to spin S, $c(n)$ is the number of pair permutations of the arguments along the columns of the Young diagram contained in the series

$$\hat{N}_n = \hat{n}_w\hat{n}_v\ldots\hat{n}_c\hat{n}_b\hat{n}_a. \quad \text{(A.6)}$$

The pair permutation operators of the arguments along the rows of the Young diagram are positioned at the end of the sequence (A.6): …, $\hat{n}_b$, $\hat{n}_a$ whereas those along the columns are positioned at the beginning: $\hat{n}_w$, $\hat{n}_v$, $\hat{n}_s$, …, and for the $\hat{p}$ operators the converse holds. The functions $\chi_J(\{m_i\}; \hat{N}_n\{\sigma_i\})$ in (A.1) being products of single-particle functions are symmetric with respect to the pair permutations (m_i, σ_i') where $\{\sigma_i'\} = \hat{N}_n\{\sigma_i\}$:

$$\chi(\{m_i\}; \hat{N}_n\{\sigma_i\}) = \chi(\hat{P}_k\{m_i\}; \hat{P}_k\hat{N}_n\{\sigma_i\}), \quad \text{(A.7)}$$

and in particular

$$\chi(\{m_i\}; \hat{N}_n\{\sigma_i\}) = \chi(\{m_i\}; \hat{P}_n\{\sigma_i\}) = \chi(\hat{P}_n^{-1}\{m_i\}; \{\sigma_i\}). \quad \text{(A.8)}$$

It can be seen from (A.7) and (A.8) that the $\{m_i\}$ symmetry of the wave function is described by a group of inverse permutations with respect to the $\{\sigma_i\}$ symmetry group. The same applies to the coordinate function. Any permutation $\hat{P}_k$ uniquely corresponds to an oriented graph constructed of N numbered vertices. The directional couplings (arrows) in the graph connect the old and new numbers of the positions of the permuted arguments. An inverse permutation corresponds to a reoriented graph (where the directions of all the couplings are changed). In the sense of the permutation operation, each vertex is both the beginning of one and the end of another or the same arrow which yields the ring structure of the graph. The set of rings in the graph corresponds to an equivalent set of commuting cyclic permutations. Each graph uniquely corresponds to a spin or coordinate function with permutated arguments. The structure of the function symmetrized in accordance with the Young diagram is represented by a set of graphs with coefficients (generalized weights) ±1. Permutation of the arguments in the cells of the Young diagram leads to renumbering of the vertices of all the graphs so that any permutations of the arguments in the cells of the Young diagram correspond to the same set of graphs constructed using unnumbered vertices, i.e., dummy graphs. If the symmetry of the coordinate function with respect to the arguments is described by a certain set of graphs, the symmetry of the same function in terms of the eigenvalues $\{\mathbf{r}_i\}$ is described by a set of reoriented graphs. Let us assume that

$$f_J(S, \{\mathbf{r}_i\}; \{\mathbf{x}_i\}) = \sum_n (-1)^{c(n)} f(\{\mathbf{r}_i\}; \hat{N}_n\{\mathbf{x}_i\}) = \sum_n (-1)^{c(n)} f(\{\mathbf{r}_i\}; \hat{P}_n\{\mathbf{x}_i\}) \quad \text{(A.9)}$$

is a coordinate function symmetrized by the Young symmetrization operator with initial permutation of the arguments in the cells of the Young diagram. Going over to a different permutation of the arguments in the cells using the permutation $\hat{P}_m$ implies renumbering the vertices of all the graphs appearing in the linear combination (A.9), i.e., going over to different graphs from the same classes:

$$f_J^m(S, \{\mathbf{r}_i\}; \{\mathbf{x}_i\}) = \sum_n (-1)^{c(n)} f(\{\mathbf{r}_i\}; \hat{P}_m\hat{P}_n\hat{P}_m^{-1}\{\mathbf{x}_i\}). \quad \text{(A.10)}$$

The specific Young diagram with empty cells is characterized by a set of dummy graphs.

REFERENCES

1. *Physics of Simple Liquids*, Ed. by H. N. V. Temperley, J. S. Rowlinson, and G. S. Rushbrooke (North-Holland, Amsterdam, 1968; Mir, Moscow, 1973).
2. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, and H. A. Teller, J. Chem. Phys. **21**, 1087 (1953).
3. R. D. Etters and J. Kaelberer, Phys. Rev. A **11**, 1068 (1975).
4. Ch. Margheritis and C. Sinistri, Z. Naturforsch. A **30**, 83 (1975).
5. P. N. Vorontsov-Vel'yaminov and V. P. Chasovskikh, Teplofiz. Vys. Temp. **13**, 1153 (1975).
6. S. V. Shevkunov, Zh. Éksp. Teor. Fiz. **104**, 3032 (1993) [JETP **77**, 413 (1993)].
7. S. V. Shevkunov, Zh. Éksp. Teor. Fiz. **105**, 1258 (1994) [JETP **78**, 677 (1994)].
8. S. V. Shevkunov, Zh. Éksp. Teor. Fiz. **108**, 1373 (1995) [JETP **81**, 753 (1995)].
9. A. A. Martsinovski, S. V. Shevkunov, and P. N. Vorontsov-Velyaminov, Mol. Simul. **6**, 143 (1991).
10. S. V. Shevkunov, P. N. Vorontsov-Velyaminov, and A. A. Martsinovski, Mol. Simul. **5**, 119 (1990).
11. S. V. Shevkunov and A. Vegiri, J. Chem. Phys. **111**, 9303 (1999).
12. K. Kiyohara, K. E. Gubbins, and A. Z. Panagiotopoulos, Mol. Phys. **94**, 803 (1998).
13. K. J. Oh and X. C. Zeng, J. Chem. Phys. **110**, 4471 (1999).
14. D. Colognesi, A. de Santis, and D. Rocca, Mol. Phys. **88**, 465 (1996).
15. P. Jedlovszky, I. Bako, G. Palinkas, *et al.*, J. Chem. Phys. **105**, 245 (1996).
16. P. Jedlovszky, J. P. Brodholt, F. Bruni, *et al.*, J. Chem. Phys. **108**, 8528 (1998).
17. P. Jedlovszky and J. Richardi, J. Chem. Phys. **110**, 8019 (1999).
18. P. Jedlovszky, J. Chem. Phys. **111**, 5975 (1999).
19. E. Allahyarov, I. D'Amico, and H. L. Lowen, Phys. Rev. Lett. **81**, 1334 (1998).
20. J. Wu, D. Bratko, and J. M. Prausnitz, Proc. Natl. Acad. Sci. USA **95**, 15 169 (1998).
21. J. Z. Wu, D. Bratko, H. W. Blanch, and J. M. Praushitz, J. Chem. Phys. **111**, 7084 (1999).
22. N. Gronbech-Jensen, K. M. Beardmore, and P. Pincus, Physica A (Amsterdam) **261**, 74 (1998).
23. L. Degreve and F. L. B. da Silva, J. Chem. Phys. **110**, 3070 (1999).
24. A. D. Mackie, J. Hernandez-Cobos, and L. F. Vega, J. Chem. Phys. **111**, 2103 (1999).
25. E. V. Akhmatskaya, M. D. Cooper, N. A. Burton, *et al.*, Chem. Phys. Lett. **267**, 105 (1997).
26. S. J. Vaughn, E. V. Akhmatskaya, M. A. Vincent, *et al.*, J. Chem. Phys. **110**, 4338 (1999).
27. F. Calvo, Phys. Rev. B **60**, 15 601 (1999).
28. U. Hansen, P. Vogl, and V. Fiorentini, Phys. Rev. B **60**, 5055 (1999).
29. N. Matubayasi, Ch. Wakai, and M. Nakahara, J. Chem. Phys. **110**, 8000 (1999).

30. R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965; Mir, Moscow, 1968).
31. L. D. Fosdick and H. F. Jordan, Phys. Rev. **143**, 58 (1966).
32. H. F. Jordan and L. D. Fosdick, Phys. Rev. **171**, 128 (1968).
33. V. R. Panharipande, S. C. Pieper, and R. B. Wiringa, Phys. Rev. B **34**, 4571 (1986).
34. D. S. Lewart, V. R. Pandharipande, and S. C. Pieper, Phys. Rev. B **37**, 4950 (1988).
35. P. A. Whitlock, G. V. Chester, and M. H. Kalos, Phys. Rev. B **38**, 2418 (1988).
36. B. Tanatar and D. M. Ceperley, Phys. Rev. B **39**, 5005 (1989).
37. H. Yokoyama and H. Shiba, J. Phys. Soc. Jpn. **56**, 3570 (1987); **59**, 3669 (1990).
38. Y. Liu, J. Dong, Chang-de Gong, and T. Chen, Phys. Rev. B **48**, 1308 (1993).
39. D. Bressanini, G. Fabbri, M. Mella, and G. Morosi, J. Chem. Phys. **111**, 6230 (1999).
40. J. B. Anderson, J. Chem. Phys. **63**, 1499 (1975); **65**, 4121 (1976); **73**, 3897 (1980).
41. R. D. Amos, Adv. Chem. Phys. **67**, 99 (1987).
42. P. Sandler, V. Buch, and D. C. Clary, J. Chem. Phys. **101**, 6353 (1994).
43. M. H. Muser and J. Ankerhold, Europhys. Lett. **44**, 216 (1998).
44. R. N. Barnett, P. J. Reynolds, and W. A. Lester, J. Chem. Phys. **84**, 4992 (1986).
45. M. Caffarel and P. Claverie, J. Chem. Phys. **88**, 1088 (1988).
46. D. F. Coker and R. O. Watts, J. Chem. Phys. **86**, 5703 (1987).
47. D. M. Ceperley and H. Partridge, J. Chem. Phys. **84**, 820 (1986).
48. R. M. Grimes, B. L. Hammond, P. J. Reynolds, and W. A. Lester, J. Chem. Phys. **85**, 4779 (1986).
49. M. Caffarel and P. Claverie, J. Chem. Phys. **88**, 1100 (1988).
50. G. J. Martyna and B. J. Berne, J. Chem. Phys. **88**, 4516 (1988); **90**, 3744 (1989).
51. P. A. Christensen, J. Chem. Phys. **88**, 4867 (1988).
52. T. Yoshida and K. Iguchi, J. Chem. Phys. **88**, 1032 (1988).
53. M. Caffarel, P. Claverie, C. Mijoule, *et al.*, J. Chem. Phys. **90**, 990 (1989).
54. B. Tanatar and D. M. Ceperley, Phys. Rev. B **39**, 5005 (1989).
55. J. Carlson, J. W. Moskowitz, and K. E. Schmidt, J. Chem. Phys. **90**, 1003 (1989).
56. J. E. Hirsch, D. J. Scalapino, and R. L. Sugar, Phys. Rev. Lett. **47**, 1628 (1981).
57. J. E. Hirsch, R. L. Sugar, D. J. Scalapino, and R. Blankenbecler, Phys. Rev. B **26**, 5033 (1982).
58. D. J. Scalapino and R. L. Sugar, Phys. Rev. B **24**, 4295 (1981).
59. J. E. Hirsch and D. J. Scalapino, Phys. Rev. B **27**, 7169 (1983).
60. J. E. Hirsh, Phys. Rev. B **31**, 4403 (1985).
61. Ad. Lagendijk and H. DeRaedt, Phys. Rev. Lett. **49**, 602 (1982).
62. S. R. White, D. J. Scalapino, R. L. Sugar, and N. E. Bickers, Phys. Rev. Lett. **63**, 1523 (1989).
63. M. Imada, J. Phys. Soc. Jpn. **57**, 3128 (1988).
64. M. Imada and Y. Hatsugai, J. Phys. Soc. Jpn. **58**, 3752 (1989).
65. H. Otsika, J. Phys. Soc. Jpn. **59**, 2916 (1990).
66. W. von der Linden, I. Morgenstern, and H. DeRaedt, Phys. Rev. B **41**, 4669 (1990).
67. H. Aoki and K. Kuroki, Phys. Rev. B **42**, 2125 (1990).
68. Y. Takada and T. Kita, J. Phys. Soc. Jpn. **60**, 25 (1991).
69. R. T. Scalettar, D. J. Scalapino, R. L. Sugar, and S. R. White, Phys. Rev. B **44**, 770 (1991).
70. L. Tan, Q. Li, and J. Callaway, Phys. Rev. B **44**, 341 (1991).
71. A. Moreo, D. Scalapino, and E. Dagatto, Phys. Rev. B **43**, 11 442 (1991).
72. G. G. Batrouni and R. T. Scalettar, Phys. Rev. B **46**, 9051 (1992).
73. R. Valenti and C. Gros, Phys. Rev. Lett. **68**, 2402 (1992).
74. A. F. Elesin and V. A. Kashurnikov, Zh. Éksp. Teor. Fiz. **106**, 1773 (1994) [JETP **79**, 961 (1994)].
75. N. V. Prokof'ev, B. V. Svistunov, and I. S. Tupitsyn, Pis'ma Zh. Éksp. Teor. Fiz. **64**, 853 (1996) [JETP Lett. **64**, 911 (1996)].
76. D. Chandler and P. G. Wolynes, J. Chem. Phys. **74**, 4078 (1981).
77. V. S. Filinov, Phys. Lett. A **54**, 259 (1975).
78. V. S. Filinov and G. E. Norman, Phys. Lett. A **55**, 219 (1975).
79. R. M. Fye, Phys. Rev. B **33**, 6271 (1986).
80. H. DeRaedt and B. DeRaedt, Phys. Rev. A **28**, 3575 (1983).
81. V. S. Filinov, Nucl. Phys. B **271**, 717 (1986).
82. N. Makri and W. H. Miller, Chem. Phys. Lett. **139**, 10 (1987).
83. J. D. Doll, D. L. Freeman, and M. J. Gillan, Chem. Phys. Lett. **143**, 277 (1988).
84. S. V. Shevkunov, O. M. Roschinenko, and P. N. Vorontsov-Velyaminov, Mol. Simul. **7**, 205 (1991).
85. M. Imada, J. Phys. Soc. Jpn. **53**, 2861 (1984).
86. D. M. Ceperley and E. L. Pollock, Phys. Rev. Lett. **56**, 351 (1986).
87. E. L. Pollock and D. M. Ceperley, Phys. Rev. B **36**, 8343 (1987).
88. D. M. Ceperley and E. L. Pollock, Phys. Rev. B **39**, 2084 (1989).
89. S. R. White and J. W. Wilkins, Phys. Rev. B **37**, 5024 (1988).
90. H. Q. Lin and J. E. Hirsch, Phys. Rev. B **34**, 1964 (1986).
91. M. Takahashi and M. Imada, J. Phys. Soc. Jpn. **53**, 963 (1983).
92. C. L. Cleveland, U. Landman, and R. N. Barnett, Phys. Rev. B **39**, 117 (1989).

93. E. L. Pollock and D. M. Ceperley, Phys. Rev. B **30**, 2555 (1984).
94. B. A. Mason and K. Hess, Phys. Rev. B **39**, 5051 (1989).
95. B. J. Berne and D. Thirumalai, Annu. Rev. Phys. Chem. **37**, 401 (1986).
96. K. Mackeown, Am. J. Phys. **53**, 880 (1985).
97. D. Marx and M. Parrinello, Z. Phys. B **95**, 143 (1994); J. Chem. Phys. **104**, 4077 (1996).
98. M. E. Tuckerman, D. Marx, M. E. Klein, and M. Parrinello, J. Chem. Phys. **104**, 5579 (1996).
99. Ph. Sindzingre, D. M. Ceperley, and M. L. Klein, Phys. Rev. Lett. **67**, 1871 (1991).
100. D. Scharf, G. J. Martyna, and M. L. Klein, J. Chem. Phys. **99**, 8997 (1993).
101. D. Marx, S. Sengupta, and P. Nielaba, J. Chem. Phys. **99**, 6031 (1993).
102. Ch. Chakravarty, J. Chem. Phys. **99**, 8038 (1993).
103. Ch. Chakravarty, J. Chem. Phys. **102**, 956 (1995).
104. L. M. Sese and R. Ledesma, J. Chem. Phys. **102**, 3776 (1995).
105. D. M. Ceperley, Rev. Mod. Phys. **67**, 279 (1995).
106. M. Skorobogatiy and J. D. Joannopoulos, Phys. Rev. B **60**, 1433 (1999).
107. M. C. Gordillo, Phys. Rev. B **60**, 6790 (1999).
108. B. Hetenyi, E. Rabani, and B. J. Berne, J. Chem. Phys. **110**, 6143 (1999).
109. S. Jang and G. A. Voth, J. Chem. Phys. **111**, 2357 (1999); **111**, 2371 (1999).
110. P.-N. Roy and G. A. Voth, J. Chem. Phys. **110**, 3647 (1999).
111. R. Rousseau and D. Marx, J. Chem. Phys. **111**, 5091 (1999).
112. P.-N. Roy, S. Jang, and G. A. Voth, J. Chem. Phys. **111**, 5303 (1999).
113. J. P. Serre, *Représentations linéaires des groupes finis* (Herman, Paris, 1967; Mir, Moscow, 1970).
114. A. A. Kirillov, *Elements of the Theory of Representations* (Nauka, Moscow, 1978; Springer-Verlag, Berlin, 1976).
115. V. D. Lyakhovskiĭ and A. A. Bolokhov, *Symmetry Groups and Elementary Particles* (Leningrad. Gos. Univ., Leningrad, 1983).
116. H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover, New York, 1950; Mir, Moscow, 1983).
117. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory* (Nauka, Moscow, 1974; Pergamon, New York, 1987).
118. A. S. Davydov, *Quantum Mechanics* (Fizmatgiz, Moscow, 1963).
119. F. A. Berezin, *Method of Second Quantization* (Nauka, Moscow, 1986; Academic, New York, 1986).
120. V. M. Zamalin, G. É. Norman, and V. S. Filinov, *Monte Carlo Method in Statistical Thermodynamics* (Mir, Moscow, 1977).
121. S. M. Ermakov, *Monte Carlo Method and Related Problems* (Nauka, Moscow, 1975).
122. S. V. Shevkunov, Teplofiz. Vys. Temp. **28**, 1 (1990).
123. N. Marki and W. H. Miller, J. Chem. Phys. **90**, 904 (1989).
124. R. P. Feynman, *Statistical Mechanics: A Set of Lectures* (Benjamin, Reading, Mass., 1972; Mir, Moscow, 1978).
125. M. F. Herman, E. Bruskin, and B. J. Berne, J. Chem. Phys. **76**, 5150 (1982).
126. M. Parrinello and A. Rahman, J. Chem. Phys. **80**, 860 (1984).
127. R. N. Barnett, U. Landman, C. L. Cleveland, and J. Jortner, J. Chem. Phys. **88**, 4429 (1988).
128. T. L. Hill, *Statistical Mechanics: Principles and Selected Applications* (McGraw-Hill, New York, 1956; Inostrannaya Literatura, Moscow, 1960).
129. K. S. Krasnov, V. S. Timoshinin, T. G. Danilova, and S. V. Khandozhko, *Molecular Constants of Inorganic Compounds* (Khimiya, Leningrad, 1968).
130. A. A. Radtsig and B. M. Smirnov, *Reference Data on Atoms, Molecules, and Ions* (Atomizdat, Moscow, 1980; Springer-Verlag, Berlin, 1985).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 1, 2000, pp. 48–66.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 1, 2000, pp. 56–76.

ATOMS, SPECTRA, RADIATION

Tunneling and Above-Barrier Ionization of Atoms in a Laser Radiation Field

V. S. Popov

Institute of Theoretical and Experimental Physics, Moscow, 117218 Russia

Received December 16, 1999

Abstract—Calculations are made of the energy and angular distributions of photoelectrons during tunneling ionization of an atom or an ion under the action of high-power laser radiation (for all values of the Keldysh parameter γ). Cases of linear, circular, and elliptic polarizations of the electromagnetic wave are considered. The probability of above-barrier ionization of hydrogen atoms in a low-frequency ($\gamma \ll 1$) laser field is calculated. Formulas are given for the momentum spectrum of the electrons when an atomic level is ionized by a general type of alternating electric field (for the case of linear polarization). An analysis is made of tunneling interference in the energy spectrum of the photoelectrons. Analytic approximations are discussed for the asymptotic coefficient C_κ of the atomic wave function at infinity (for *s*-wave electrons). © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Numerous studies have been devoted to the ionization of atoms as a result of interaction between high-intensity laser radiation and matter. The theory of these processes was first considered by Keldysh [1] who showed that the tunneling effect and multiphoton ionization are two extreme cases of a single nonlinear photoionization process whose character depends on the parameter γ. In [2–4] the authors developed a systematic quantum-mechanical theory of level ionization associated with short-range forces for all values of the optical frequency ω and electric field strength F satisfying the semiclassical conditions of validity:

$$\begin{aligned} &\omega \ll \kappa^2, \quad F \ll \kappa^3, \\ &0 \le \gamma < \infty, \quad (\hbar = m = e = 1), \end{aligned} \tag{1.1}$$

where $\kappa = \sqrt{2E_I}$, E_I is the ionization potential of the level, and $\gamma = \omega\kappa/F$ is the Keldysh parameter [1]. Cases of linear and circular [2, 3] and elliptic [4] polarizations of the electromagnetic wave have been considered.[1] Allowance for Coulomb interaction between the outgoing electron and the atomic core was considered in [5, 6]. The first numerical calculations of the multiphoton ionization probability in high orders of perturbation theory with respect to the field F were presented in [7] and references to subsequent studies along these lines can be found in [8–11]. Results of further studies in this field are presented in [10, 11].[2]

Now, as a result of developments in laser technology it has become possible to study not only the total ionization probability but also finer details, in particular the energy and angular distributions of the outgoing photoelectrons. Research has been developed in the low-frequency range and this has resulted in the appearance of numerous theoretical studies devoted to the special case $\gamma \ll 1$ [see [11]). The aim of the present study is to investigate the energy and angular distributions of the photoelectrons for tunneling ionization over all values of the Keldysh parameter γ and also to calculate the probability of above-barrier ionization in the adiabatic ($\gamma \ll 1$) range.

We shall briefly describe the contents of this study. In Sections 2–5 we systematically analyze the cases of linear, circular, and elliptic polarizations of the incident radiation (in the tunneling ionization regime). We pay particular attention to changes in the momentum, energy, and angular distributions of the photoelectrons with increasing γ. In Section 6 we calculate the probability of above-barrier ionization of hydrogen atoms in a low-frequency field and compare these with results of other authors. The formulas from Section 2 for the momentum spectrum of the photoelectrons are generalized to the case of an arbitrary field $F(t)$ (for linear polarization see Section 7). The effect of tunneling interference in the energy spectrum of the outgoing photoelectrons is briefly discussed in Section 8. The main conclusions and some observations are put forward in Section 9. In the Appendix we give details of the calculations and asymptotic forms, we discuss the relationship between the static and adiabatic ionization probabilities, and put forward additional formulas.

[1] We take this opportunity to note that the statement made in [3] that for linearly polarized radiation the formulas obtained in [2] and [3] differ twofold is based on a misunderstanding and arises because the normalization condition adopted in [2] yields $\int \psi^2 d^3r = 1/2$ in the nonrelativistic limit. In fact the results of [2, 3] are exactly the same. I am grateful to A.I. Nikishov and V.I. Ritus for discussing this point.

[2] We must note that in [10, 11], as in other works by the authors of these reviews, the history of the problem is distorted and the contribution of [2–6] is described extremely subjectively (for further details on this see [12]).

This article uses atomic units, $F_a = m^2e^5/\hbar^4 = 5.14 \times 10^9$ V/cm is the atomic unit of the electric field strength, $\epsilon = F/\kappa^3F_a$ is the "reduced" field (for the ground level of a hydrogen atom $\kappa = 1$ and $\epsilon \equiv F$), and $K_0 = E_I/\omega$ is the multiquantization parameter of the process. It is subsequently assumed everywhere that $\epsilon \ll 1$ and $K_0 \gg 1$. By c_i we denote the dimensionless coefficients (which generally depend on γ). Some of the results presented below were announced in [13, 14].

2. LINEAR POLARIZATION

The momentum spectrum of the photoelectrons has the form[3]

$$w(\mathbf{p}) = \text{const}\exp\left\{-\frac{2}{3\epsilon}g(\gamma) - \frac{1}{\omega}[c_1(\gamma)p_\parallel^2 + c_2(\gamma)p_\perp^2]\right\}, \tag{2.1}$$

where

$$g(\gamma) = \frac{3}{2\gamma}\left[\left(1+\frac{1}{2\gamma^2}\right)\text{arcsinh}\gamma - \frac{\sqrt{1+\gamma^2}}{2\gamma}\right], \tag{2.2}$$

$c_1(\gamma) = \text{arcsinh}\gamma - \gamma(1+\gamma^2)^{-1/2}$, $c_2(\gamma) = \text{arcsinh}\gamma$, $\gamma = \omega/\omega_t = \omega\kappa/F$, ω and F are the laser radiation frequency and the electric field strength, $F(t) = F\cos\omega t$, $\omega_t = F/\sqrt{2E_I}$ is the tunneling frequency, and $E_I = \kappa^2/2$. Expression (2.1) yields the angular momentum distribution of the electrons

$$w(\theta) = w(0)\exp\left\{-\frac{\kappa p^2}{F\sqrt{1+\gamma^2}}\sin^2\theta\right\}, \tag{2.3}$$

where θ is the angle between the direction of electron emission and the axis of linear polarization of the radiation. Formulas (2.1) and (2.3) barely require comment. For $\gamma \ll 1$ (adiabatic case, i.e., low frequency ω, high field strength F) the angular distribution has an abrupt maximum in the direction of the field. In this case $p_\perp \sim \sqrt{F/\kappa} \ll \kappa$ and the longitudinal momentum $p_\parallel \sim \gamma^{-1}p_\perp$ may exceed κ (the characteristic bound-state momentum) since an electron may be accelerated along the slowly varying field $F(t)$. In the opposite case $\gamma \gg 1$ (rapidly varying field $\omega \gg \omega_t$ we have $p_\parallel \sim p_\perp \sim \sqrt{\omega/(\ln 2\gamma)} \ll \kappa$. From (2.1) and (2.3) for the average values of the transverse and longitudinal (with respect to **F**) electron momenta we obtain:

$$\frac{\langle p_\perp^2\rangle^{1/2}}{\langle p_\parallel^2\rangle^{1/2}} = \left(1 - \frac{\gamma}{\sqrt{1+\gamma^2}\,\text{arcsinh}\gamma}\right)^{1/2} = \begin{cases} 0.577\gamma, & \gamma \ll 1, \\ 1-(2\ln\gamma)^{-1}, & \gamma \gg 1, \end{cases} \tag{2.4}$$

and also

$$\langle \sin^2\theta\rangle = 1 - \frac{1}{2\beta}\left(\frac{\sqrt{\beta}}{w(\sqrt{\beta})} - 1\right) = \begin{cases} \frac{2}{3} - \frac{4}{45}\beta + \ldots, & \beta \ll 1, \\ 1/2\beta, & \beta \gg 1. \end{cases} \tag{2.5}$$

Here we have

$$\beta = \frac{(c_2 - c_1)p_n^2}{\omega} = \frac{2\gamma}{\sqrt{1+\gamma^2}}(n-\nu),$$

$$p_n = \sqrt{2\omega(n-\nu)},$$

n is the number of absorbed quanta,

$$\nu = \frac{\kappa^2}{2\omega}\left(1 + \frac{1}{2\gamma^2}\right)$$

is the photoionization threshold for the linearly polarized wave, and

$$w(z) = e^{-z^2}\int_0^z e^{y^2}dy$$

is a known special function [15]. If $\gamma \ll 1$ we find $n - \nu \lesssim \gamma^{-3}$ [see (2.7) below], $\beta \sim \gamma^{-2}$, and $\sqrt{\langle\theta^2\rangle} \sim \gamma$, i.e., the angular distribution of the photoelectrons is highly elongated along **F**. If $\gamma \gg 1$, then $\beta \approx 2(n-\nu) \lesssim 1$ and the distribution approaches isotropic.

The total probability (or rate of ionization) of the level is

$$W(F,\omega) = \sum_{n>\nu} w_n, \tag{2.6}$$

$$w_n = \int w(\mathbf{p})\delta\left(\frac{p^2 - p_n^2}{2}\right)d^3p = p_n\int w(p_n\mathbf{n})d\Omega_\mathbf{n}$$

($\mathbf{n} = \mathbf{p}/p$). Integration (2.1) over angles is performed exactly [2, 3]. The probability of ionization of the lm state is proportional to $\epsilon^{|m|}$, i.e., decreases rapidly [2, 3] with increasing $|m|$ (here m is the projection of the orbital angular momentum l on the direction of the electric field). Finally, we obtain

$$w_n = W_l\frac{4\gamma^2}{\sqrt{3\pi}}\exp\left\{-\frac{2}{3}\gamma^3(n-\nu)\right\}w(\sqrt{2\gamma(n-\nu)}),$$

$$\gamma \ll 1,$$

$$W_l(F,\omega) = (2l+1)\kappa^2C_{\kappa l}^2\sqrt{\frac{3\epsilon^3}{\pi}}\left(\frac{2}{\epsilon}\right)^{2\eta} \times \exp\left\{-\frac{2}{3\epsilon}\left(1 - \frac{1}{10}\gamma^2\right)\right\}, \tag{2.7}$$

where w_n is the probability of n-photon ionization, and W_l is the total probability of ionization of level l, $\eta = Z/\kappa$, $w(z)$ is the same function as in (2.5), and $C_{\kappa l}$ is

[3] See formula (53) in [3]. We note that the function $g(\gamma)$ was first calculated by Keldysh [1].

the asymptotic ($r \longrightarrow \infty$) coefficient of the atomic wave function:[4]

$$\psi_{klm}(\mathbf{r}) = \frac{\chi_{\kappa l}(r)}{r} Y_{lm}\left(\frac{\mathbf{r}}{r}\right), \quad \int \psi^2_{\kappa lm} d^3 r = 1, \qquad (2.8)$$
$$\chi_{\kappa l}(r) \approx 2\sqrt{\kappa} C_{\kappa l} e^{-\kappa r} (\kappa r)^{\eta}, \quad r \gg \kappa^{-1}.$$

The distribution of w_n (2.7) has a smooth maximum at $n = n_{\max}$ where $n_{\max} - \nu = 0.43\gamma^{-1} \gg 1$ and $(n_{\max} - \nu)/\nu = 0.855\gamma/K_0 \ll 1$. For $n - \nu \sim \gamma^{-3}$ the probabilities w_n begin to decrease exponentially. Summing in (2.6) and (2.7) over n, it is easy to confirm that there is a correct limit for $\omega \longrightarrow 0$, i.e., a transition to well-known [16, 21–23] formulas for a static field F allowing for adiabatic [1–3] and Coulomb [5] corrections.

If $\gamma \gg 1$, to within logarithmic accuracy we have $c_1 \approx c_2$ and

$$w_n \propto e^{-\Lambda(n-\nu)} w(\sqrt{2\Lambda(n-\nu)}), \qquad (2.9)$$
$$\Lambda = \ln 2\gamma \gg 1,$$

so that as n increases, the probabilities w_n decrease rapidly, beginning from the threshold $n = \nu \approx K_0$ [1].

We shall make another observation on the photoelectron momentum $\mathbf{p}$. After overcoming the barrier, an electron moves along a classical trajectory along which $\mathbf{p}_\perp$ = const (integral of motion) and

$$p_\parallel(t) = p_\parallel(0) + F\int_0^t \cos\omega t dt.$$

This last integral has a unique meaning after the field has been adiabatically switched off at $t \longrightarrow +\infty$: $\Delta p_\parallel = p_\parallel(\infty) - p_\parallel(0) = 0$, since

$$J = \lim_{\alpha\to 0}\int_0^\infty e^{-\alpha t}\cos\omega t dt = \lim_{\alpha\to 0}\frac{\alpha}{\omega^2 + \alpha^2} = 0. \qquad (2.10)$$

Thus $\mathbf{p} = (p_\parallel, \mathbf{p}_\perp)$ in (2.1) is not only the electron momentum at the instant of escaping from below the barrier ($t = 0$) but is also the time-averaged (drift) momentum in the alternating field $F(t)$ which does not vary when the time envelope of the field is switched off adiabatically slowly and is recorded at infinity.[5]

[4] These coefficients are frequently encountered in quantum mechanics, atomic and nuclear physics [16–20]. Note that we normalized twice compared with earlier studies [3–5]: for the ground state of a hydrogen atom we now have $C_\kappa = 1$, and for the three-dimensional δ-potential $C_\kappa = 1/\sqrt{2}$ (subsequently for conciseness $C_\kappa \equiv C_{\kappa_0}$, $l = 0$).

[5] This conclusion holds for short laser pulses when $\delta = \alpha/\omega \ll 1$ (we shall give an estimate: $\delta \sim vT/L \sim (v/c)(\lambda/L)$, where v is the electron velocity, λ is the wavelength, $T = 2\pi/\omega$, and $L \gtrsim \lambda$ is the size of the laser radiation focusing region, $L \gtrsim \lambda$. For long pulses changes in the electron drift momentum under the action of the gradient force are substantial [24–27]. In this case, in order to calculate the electron distribution over finite kinetic energies we need to analyze their motion in a spatially inhomogeneous field near the laser focus and take into account ponderomotive acceleration. In simple models this can be performed analytically [26], but for realistic field profiles numerical analyses are required [11, 27]. The author would like to thank S.P. Goreslavskiĭ for clarifying this point.

3. CIRCULAR POLARIZATION

In this case, the probability of n-photon ionization of the s-level is [3]

$$w_n = \kappa^2 C_\kappa^2 R(t,\gamma)\exp\{-4\nu_c\varphi(t,\gamma)\}, \quad n > \nu_c, \qquad (3.1)$$

where

$$\varphi(t,\gamma) = \frac{\operatorname{arctanh} u - u}{1+t}, \quad u = \sqrt{\frac{t^2+\gamma^2}{1+\gamma^2}}, \qquad (3.2)$$

n is replaced by the auxiliary variable $t = (2\nu_c/n) - 1$ ($-1 < t < 1$ at the threshold $t = 1$), $\nu_c = K_0(1 + \gamma^{-2})$ is the photoionization threshold for circularly polarized radiation, $C_\kappa \equiv C_{\kappa 0}$ is the asymptotic coefficient for $l = 0$, and the preexponential factor R was also calculated in [3] but is not required here.

The function $\varphi(t, \gamma)$ has a minimum at the point $t = t_0(\gamma)$ which corresponds to the maximum ionization probability and is determined from the equation $t = 1 - u/\operatorname{arctanh} u$, where $n = n_0(\gamma) = 2\nu_c(1 + t_0)$. Near the maximum the distribution over n, i.e., the electron energy spectrum, has a Gaussian profile:

$$w_n = w_{\max}\exp\left\{-c_3\frac{(n-n_0)^2}{n_0}\right\} = w_{\max}\exp\left\{-\frac{(E_n - E_{\max})^2}{2\Delta^2}\right\}, \qquad (3.3)$$

where (see also Fig. 1)

$$c_3(\gamma) = \frac{t_0^2 + u_0^2}{(1-t_0)^2 u_0} = \begin{cases}\gamma + \frac{1}{3}\gamma^3 + \ldots, & \gamma \ll 1,\\ 2(\ln\gamma)^2[1 + \ln\ln\gamma/\ln\gamma], & \gamma \gg 1,\end{cases} \qquad (3.4)$$

$$\Delta = [\omega E_I(1+\gamma^{-2})/(1+t_0)c_3(\gamma)]^{1/2}, \qquad (3.4')$$

$\Delta = F^{3/2}/\omega\sqrt{2\kappa} \sim \gamma^{-3/2}\sqrt{\omega E_I}$ for $\gamma \ll 1$, $\Delta = \sqrt{\omega E_I}/2\ln\gamma$ for $\gamma \gg 1$. The most probable energy and momentum of the photoelectrons are given by

$$E_{\max} = (n_0 - \nu_c)\omega = \frac{F^2}{2\omega^2}(1+\gamma^2)\frac{1-t_0}{1+t_0} = \begin{cases}\gamma^{-2}E_I, & \gamma \ll 1,\\ E_I/2\ln\gamma, & \gamma \gg 1,\end{cases} \qquad (3.5)$$

where for $\gamma \ll 1$

$$p_{\max} = \frac{F}{\omega}\left(1 + \frac{1}{6}\gamma^2\right), \quad n_0 = \frac{F^2}{\omega^3}\left(1 + \frac{2}{3}\gamma^2\right),$$

and for $\gamma \gg 1$ we have $p_{\max} = \kappa/\sqrt{2\ln\gamma} \ll \kappa$,

$$w_n = w_{\max}\exp\left\{-\frac{2(\ln\gamma)^2}{K_0}(n-n_0)^2\right\}, \qquad n_0 = \left(1+\frac{1}{2\ln\gamma}\right)K_0. \tag{3.6}$$

The most probable number of absorbed photons n_0 for $\gamma \ll 1$ is twice the photoionization threshold ν_c and for $\gamma \gg 1$ approaches this albeit logarithmically slowly (see Fig. 1 in [13]). The coefficient c_3 increases monotonically with γ (curve *2* in Fig. 1).

Since

$$\frac{\Delta n}{\nu_c} \sim \begin{cases} w/\sqrt{\kappa F} \sim \sqrt{\gamma/K_0}, & \gamma \ll 1, \\ (\sqrt{K_0}\ln\gamma)^{-1}, & \gamma \gg 1, \end{cases} \tag{3.7}$$

the distribution over n has a comparatively narrow peak [2, 3] (in units of ν_c or n_0). It can be seen from Fig. 1 that for $\gamma \ll 1$ the distribution of w_n is considerably broader than a Poisson distribution: $\Delta n \approx \sqrt{n_0/2\gamma} \gg \sqrt{n_0}$ whereas for $\gamma \gg 1$ it is narrower ($\Delta n \approx \sqrt{n_0}$ for $\gamma \approx 0.468$).

In cases of circular polarization the distribution of w_n covers many values of n (unlike linear polarization when this only holds for $\gamma \ll 1$ [1–3]). Integrating (3.3) over n, we obtain

$$w_{\max} = \frac{c_4 W_c}{\sqrt{K_0}}, \quad c_4(\gamma) = \gamma\sqrt{\frac{(1+t_0)c_3}{2\pi(1+\gamma^2)}}, \tag{3.8}$$

where $w_{\max}$ corresponds to $n = n_0$, and W_c is the total probability of ionization of the s-wave level:

$$W_c(F,w) = \kappa^2 C_\kappa^2 P(\gamma)\epsilon\exp\left\{-\frac{2}{3\epsilon}g_c(\gamma)\right\}, \tag{3.9}$$

$$g_c(\gamma) = \frac{3t_0u_0}{\gamma^3(1-u_0^2)}, \qquad P(\gamma) = \frac{\gamma^2(1-t_0)}{u_0}\sqrt{\frac{1-u_0^2}{t_0^2+u_0^2}}, \tag{3.9'}$$

$u_0 = \sqrt{(t_0^2+\gamma^2)/(1+\gamma^2)}$, and $t_0(\gamma)$ is determined above. It then follows that

$$w_{\max} \equiv w_{n_0} = \frac{\kappa^2 C_\kappa^2}{\sqrt{8\pi}K_0^{3/2}}c_5(\gamma)\exp\left\{-\frac{2}{3\epsilon}g_c(\gamma)\right\}, \tag{3.10}$$

where for $\gamma \gtrsim 1$ the coefficient c_5 is almost independent of γ and is numerically close to unity (Fig. 2). Expansions of these functions are given in Appendix A, see also (A.5).

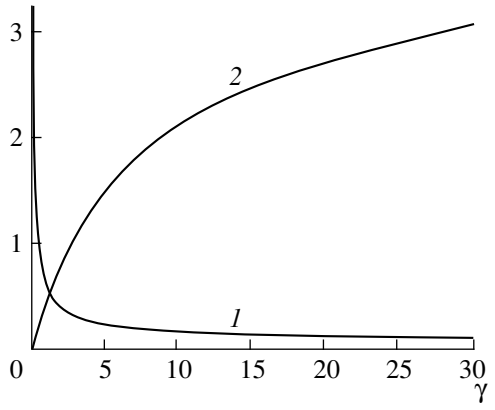


Fig. 1. Circularly polarized radiation. The ratio $\Delta n/\sqrt{n_0}$ (curve *1*) and the coefficient c_3 (curve *2*) are plotted as a function of γ.

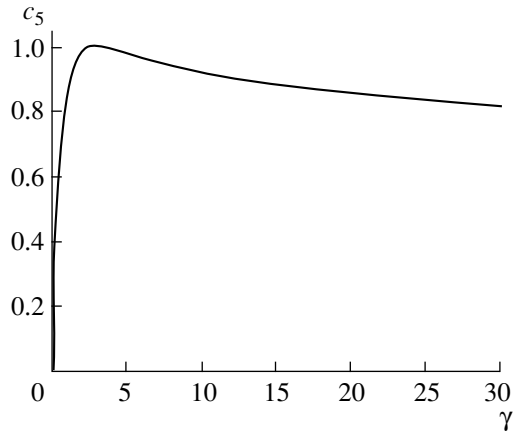


Fig. 2. The coefficient $c_5(\gamma)$ from (3.10).

The dependence of the probabilities W_c and $w_{\max}$ on F and ω is primarily determined by the function $g_c(\gamma)$ in Fig. 3 and the preexponential factor $P(\gamma)$ is plotted in Fig. 4. In the adiabatic limit we have

$$W_c = N\kappa^2\epsilon^{1-2\eta}\exp\left\{-\frac{2}{3\epsilon}\left(1-\frac{1}{15}\gamma^2\right)\right\}, \quad \gamma \ll 1, \tag{3.11}$$

where $N = 2^{2\eta}C_\kappa^2$, $\eta = Z/\kappa$ is the effective principal quantum number[6] and allowance is also made for the

[6] Frequently also denoted [17–19] by n^* (here Z is the charge of the atomic core: $Z = 1$, 2, and 0 for ionization of neutral atoms and singly charged positive and negative ions.) For the hydrogen atom $\eta = n = 1, 2, \ldots$ is an integer.

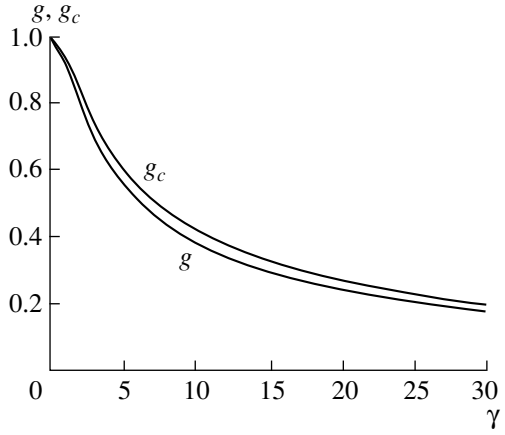


Fig. 3. The functions $g(\gamma)$ and $g_c(\gamma)$ in the exponential function for linear and circular polarizations.

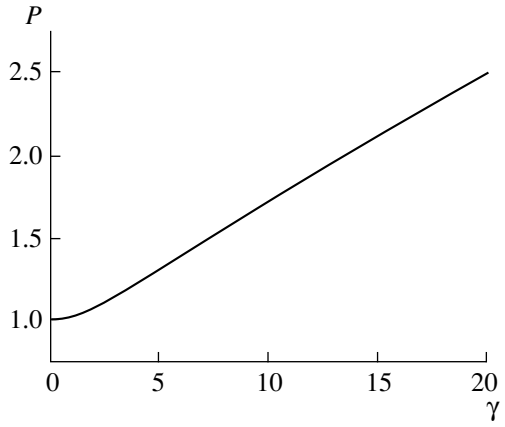


Fig. 4. Circular polarization. Preexponential factor P versus γ.

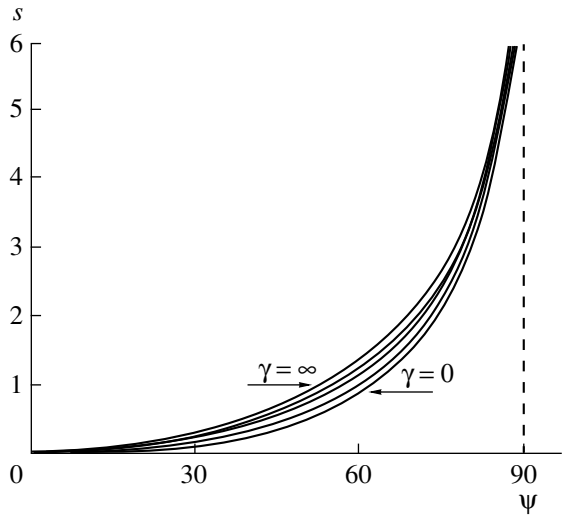


Fig. 5. The function $s(\psi, \gamma)$ determining the angular distribution of the photoelectrons. The curves correspond to the following values: $\gamma = 0$, 0.5, 1.0, 1.5, and ∞.

Coulomb correction in accordance with [5]. In this case, we have

$$W_c(F, \omega) = N\frac{\omega}{(\ln\gamma)^{3/2}} \times \exp\{-2K_0[\ln((\gamma\sqrt{\ln\gamma}) - c_6)]\}, \quad \gamma \gg 1, \tag{3.12}$$

$c_6 = (1 - \ln 2)/2 \approx 0.153$. Unlike (3.11) and (2.7), formulas (3.10) and (3.12) refer to the case $\eta = 0$, i.e., to ionization of negative ions such as H^-, Na^-, and so on.

4. PHOTOELECTRON DISTRIBUTIONS

It is also interesting to study the photoelectron distribution over the angle ψ between their direction of emission and the plane of polarization of the laser radiation. This is determined by formula (64) from [3] in which we need to substitute $n = n_0$ and the momentum $p_0 = \sqrt{2\omega(n_0 - \nu_c)}$ corresponding to the maximum ionization probability:

$$w(\psi) \propto [J_{n_0}(n_0\zeta)]^2, \quad \zeta = \sqrt{\frac{1 - t_0^2}{1 + \gamma^2}}\cos\psi, \tag{4.1}$$

where $J_{n_0}(z)$ is a Bessel function. Using the Langer [28, 29] asymptotic form for this for $n_0 \gg 1$, we find

$$w(\psi) = u_0\eta^{-1}\exp\{-n_0 s(\psi, \gamma)\}w(0), \tag{4.2}$$

where

$$s = 2[\operatorname{arctanh}\eta - \eta - (\operatorname{arctanh}u_0 - u_0)], \quad \eta = \sqrt{1 - \zeta^2}, \tag{4.2'}$$

and the width of the angular distribution is

$$\Delta\psi \sim \begin{cases} \omega/\sqrt{F\kappa} = \gamma\sqrt{\epsilon}, & \gamma \ll 1, \\ 1/\sqrt{K_0}, & \gamma \gg 1. \end{cases} \tag{4.3}$$

Figure 5 shows that the function $s(\psi, \gamma)$varies fairly sluggishly as the parameter γ increases. Thus, the change in $w(\psi)$ on transition from the adiabatic region to the case $\gamma \gtrsim 1$ is determined mainly by the change in $n_0(\gamma)$. In particular this yields the scaling relationship (9.1).

It can be seen from formula (A.7) in Appendix A that $\eta = \sqrt{\gamma^2 + \psi^2}$ in the range $\psi, \gamma \ll 1$ and formula (4.2) is simplified:

$$w(\psi) = (1 + \chi^2)^{-1/2}\exp\left\{-\frac{2\kappa^3}{3F}[(1 + \chi^2)^{3/2} - 1]\right\}, \tag{4.4}$$

where $\chi = \psi/\gamma$. For very small angles we obtain the Gaussian distribution [30–33]:

$$w(\psi) \approx \exp\left(-\frac{F\kappa}{\omega^2}\psi^2\right), \quad \psi \ll \gamma \ll 1, \tag{4.5}$$

and then the number of electrons decreases more rapidly:

$$w(\psi) \approx \gamma\psi^{-1}\exp\left(-\frac{2F^2}{3\omega^3}\psi^3\right), \quad \gamma \lesssim \psi \ll 1, \qquad (4.6)$$

where most of the electrons are emitted near the plane of polarization of the light:

$$\Delta\psi = \sqrt{\langle\psi^2\rangle} = \sqrt{\frac{\omega^2}{2F\kappa}}\left\{1-\frac{3F}{8\kappa^3}\right\} = 0.71\gamma\sqrt{\epsilon}\left(1-\frac{3}{8}\epsilon\right) \qquad (4.7)$$

[compare with (2.5) for linear polarization]. Using (A.8), we find

$$w(\psi) = w(0)(\cos\psi)^{2n_0}\left(1+\frac{n_0}{\gamma^2\ln\gamma}\sin^2\psi\right), \quad \gamma \gg 1. \qquad (4.8)$$

For $\gamma \longrightarrow \infty$ we have $w(\psi) \propto (\cos\psi)^{2K_0}$ although reaching the limiting distribution is a fairly slow process.

For small angles and any value of γ we have

$$w(\psi) = \text{const}\exp\left\{-c_7(\gamma)\frac{F\kappa}{\omega^2}\psi^2\right\}, \qquad (4.9)$$

where $c_7 = \sqrt{(1+\gamma^{-2})(t_0^2+\gamma^2)}/(1+t_0)$. If $\gamma \ll 1$, then $c_7 = 1 + (2/9)\gamma^2$ and formula (4.9) yields (4.5). However, for $\gamma \gg 1$ the coefficient is $c_7 = \gamma/2$ and $w(\psi) \approx \exp(-K_0\psi^2)$. As γ increases, the angular distribution becomes considerably broader but even for $\gamma \gg 1$ it is still fairly narrow: $\Delta\psi \sim 1/\sqrt{K_0} \ll 1$.

Expressions (3.3) and (4.9) directly yield the photoelectron distribution over the momentum components $p_\perp$ (in the plane of polarization) and p_z(in the direction of propagation of the electromagnetic wave):

$$w(p_\perp, p_z) \propto \exp\left\{-\frac{\kappa^3}{F}\left[\frac{2}{3}g_c(\gamma) + c_8\frac{p_z^2}{\kappa^2} + c_9\left(\frac{p_\perp - p_{\max}}{\kappa}\right)^2\right]\right\}, \qquad (4.10)$$

$$c_8(\gamma) = \frac{u_0}{\gamma(1-t_0)},$$

$$c_9(\gamma) = \frac{t_0^2+u_0^2}{\gamma u_0(1-t_0)} = \frac{1-t_0}{\gamma}c_3(\gamma).$$

In the two limiting cases we obtain

$$w(p_\perp, p_z) \propto \exp\left\{-\left[\frac{2}{3F}(\kappa^2+p_z^2)^{3/2} + \frac{\kappa}{F}\left(p_\perp - \frac{F}{\omega}\right)^2\right]\right\}, \quad \gamma \ll 1 \qquad (4.11)$$

(see also [11, 33]) and for $\gamma \longrightarrow \infty$ to within logarithmic accuracy we have

$$w(p_\parallel, p_\perp) \propto \exp\left\{-\frac{\Lambda}{\omega}[\kappa^2 + p_z^2 + 2(p_\perp - p_{\max})^2]\right\}, \qquad (4.12)$$

where $\Lambda = \ln(\gamma\sqrt{\ln\gamma}) \gg 1$ and the momentum $p_{\max} = \sqrt{2E_{\max}}$ was determined in (3.5).

We note that the dependence on p_z appears here in the combination $\kappa^2 + p_z^2$ (see [4]). The average photoelectron kinetic energy for $\gamma \ll 1$ is the same as its vibrational energy $F^2/2\omega^2$ in the field of a circularly polarized wave but for $\gamma \gg 1$ is $\gamma^2/2\ln\gamma$ times higher than this.

5. ELLIPTIC POLARIZATION

In the field of an electromagnetic wave

$$\mathbf{F}(t) = \{F\cos\omega t, \xi F\sin\omega t, 0\}, \quad -1 \le \xi \le 1 \qquad (5.1)$$

(ξ is the ellipticity of the light, which was denoted as ε in [4]) the photoionization threshold is given by

$$\nu = \left(1+\frac{1+\xi^2}{2\gamma^2}\right)K_0 = \frac{F^2}{4\omega^3}(1+\xi^2+2\gamma^2), \qquad (5.2)$$

and the momentum spectrum of the outgoing electrons was calculated in [4] and has the form of the sum of two anisotropic Gaussian distributions having centers at the points $\pm\mathbf{p}_{\max}$. The most probable momentum $\mathbf{p}_{\max}$ is directed along the y-axis [the minor axis of the field ellipse $\mathbf{F}(t)$] where

$$p_{\max} = |\xi|\frac{F}{\omega}\frac{\sinh\tau_0}{\tau_0} = |\xi|\kappa\frac{\sinh\tau_0}{\gamma\tau_0}. \qquad (5.3)$$

Here $\tau_0(\gamma, \xi)$ is determined from

$$\sinh^2\tau_0 - \xi^2\left(\cosh\tau_0 - \frac{\sinh\tau_0}{\tau_0}\right)^2 = \gamma^2 \qquad (5.4)$$

and has the simple physical meaning: $\tau_0 = -i\omega t_0$, where t_0 is the initial (purely imaginary) time or the complete "time" of subbarrier electron motion. Note that for $\gamma \lesssim 10$ the momentum $\mathbf{p}_{\max}$ is almost proportional to the ellipticity of the light (see Fig. 2 in [4]). In the two limiting cases the formulas are simplified considerably.

(a) If $\gamma \ll 1$ (adiabatic region), the n-photon ionization probability is given by

$$w_n = \int \delta\left(\frac{p^2}{2} - (n-\nu)\omega\right) dw(\mathbf{p}), \quad n > \nu, \tag{5.5}$$

$$dw(\mathbf{p}) = A\exp\left\{-\left[(1-\xi^2)\frac{\omega^2\kappa^3}{3F^3}p_x^2 + \frac{\kappa}{F}p_z^2\right]\right\} \times\left\{\exp\left[-\frac{\kappa}{F}\left(p_y - \xi\frac{F}{\omega}\right)^2\right] + \exp\left[-\frac{\kappa}{F}\left(p_y + \xi\frac{F}{\omega}\right)^2\right]\right\} d^3p, \tag{5.6}$$

$$A = \frac{C_\kappa^2}{4\pi^3}\frac{\omega^2}{F}\left(\frac{2}{\epsilon}\right)^{2\eta} \times \exp\left\{-\frac{2}{3\epsilon}\left[1 - \frac{1}{10}\left(1 - \frac{\xi^2}{3}\right)\gamma^2\right]\right\}. \tag{5.6'}$$

Here we allowed for a correction of the order of γ^2 in the exponential function [4] and also a Coulomb correction in accordance with [5]; $C_\kappa \equiv C_{\kappa_0}$ is the asymptotic coefficient [see (2.8) for $l = 0$], $\eta = Z/\kappa \equiv n^*$ [17, 18]. A simple and fairly accurate approximation for these coefficients was proposed by Hartree [34], see also Section 9.

It can be seen from (5.6) that $\Delta p_z \lesssim \sqrt{F/\kappa} \ll \kappa$ so that electrons are emitted at a small angle ψ to the plane of polarization: $\psi \sim \Delta p_z/p_{\max} \sim \xi_1/\xi \ll 1$ if $\xi \gg \xi_1 = \omega/\sqrt{F\kappa} = \gamma\sqrt{\epsilon}$. Then, we find $\Delta p_x/p_{\max} \sim \sqrt{\epsilon/\xi^2(1-\xi^2)}$ and $\Delta p_y/\Delta p_x \sim \gamma$. Thus, under the condition

$$\xi^2(1-\xi^2) \gg \epsilon \tag{5.7}$$

we have $p_{\max} \gg \Delta p_x \gg \Delta p_y = \Delta p_z$. Assuming in (5.6) $p_x = p_\perp \sin\varphi$, $p_y = p_\perp \cos\varphi$, and bearing in mind that $\varphi^2 \ll 1$, we rewrite (5.6) in the form

$$\frac{dw(\mathbf{p})}{d^3p} = A \times \exp\left\{-\frac{\kappa}{F}\left[p_z^2 + \left(p_\perp - |\xi|\frac{F}{w}\right)^2\right] - \frac{\xi^2(1-\xi^2)}{3\epsilon}\sin^2\varphi\right\}, \tag{5.8}$$

which is asymptotically ($\epsilon \longrightarrow 0$) equivalent to (5.6) but is more convenient for the transition to circular polarization. Here $p_\perp$ is the component of the electron momentum in the plane of polarization and φ is the azimuthal angle in this plane measured from the minor axis of the field ellipse $\mathbf{F}(t)$. The anisotropy of the angular distribution in the tunneling regime depends strongly on the ellipticity [35]. If inequality (5.7) is satisfied, the distribution (5.8) is concentrated near the angles $\varphi = 0$ and π. For $1 - \xi^2 \sim \epsilon$ this distribution begins to spread over the angle φ and for $\xi \longrightarrow \pm 1$ it is converted to (4.11) and all the directions of electron emission in the plane of polarization become equally justified [4] (this situation is similar to the appearance of zero modes in quantum field theory).

The angular distribution of the electrons also becomes broader for low ellipticity values $|\xi| < \xi_2 = \sqrt{\epsilon}$ and for even smaller $|\xi| < \xi_1$ the electrons are emitted mainly along the x axis. Here, however, formula (5.6) is completely valid and for $\xi = p_{\max} = 0$, i.e., for linear polarization, it is converted into (2.1) if $\gamma \ll 1$ is assumed in this last formula. Note that the condition (5.7) is satisfied for $\xi_2 \lesssim \xi \lesssim \xi_3 = 1 - \epsilon$, i.e., in the main region of ellipticity values (since $\epsilon = F/\kappa^3 \ll 1$, then $\xi_1 \ll \xi_2 \ll 1$, $1 - \xi_3 \ll 1$).

The spectrum of values $p_n = \sqrt{2\omega(n-\nu)}$ is almost continuous in the region $\gamma \ll 1$. Replacing summation over n by integration in (2.6), we obtain the total probability w_a (or the rate of ionization) of the s-level in a low-frequency laser field. If the polarization of the light is not too close to circular (and specifically $1 - \xi \gg \epsilon$), it then follows that [4]

$$w_a = \frac{me^4\kappa^2}{\hbar^2} N \sqrt{\frac{3}{\pi(1-\xi^2)}}\left(\frac{\hbar^4 F}{m^2e^5\kappa^3}\right)^{1.5-2\eta} \times \exp\left\{-\frac{2m^2e^5\kappa^3}{3\hbar^4 F}\left[1 - \frac{1}{10}\left(1 - \frac{\xi^2}{3}\right)\gamma^2\right]\right\}, \tag{5.9}$$

where $\kappa = \sqrt{2\hbar^2 E_I/me^4}$, $\eta = Z\sqrt{E_H/E_I}$, $E_H = 13.6$ eV, N is the same coefficient as in (3.11), corrections $\sim\gamma^2$ in the preexponential function are neglected, and the dimensions of the quantities appearing here are reconstructed.

It can be shown (see [3] and Appendix B) that for any ellipticity

$$w_a(F, \kappa, \xi) = \frac{C_\kappa^2\kappa^2}{|\xi|} a\left(\frac{1-\xi^2}{6\epsilon\xi^2}\right) A, \quad \gamma \ll 1, \tag{5.10}$$

where the function $a(x)$ is determined in (B.5) and we again converted to atomic units. If $1 - \xi^2 \gg \epsilon$, this formula is the same as (5.9). In the narrow transition region $1 - \xi^2 \sim \epsilon \ll 1$, the dependence on the field F in the preexponential function is not a simple power dependence. Formula (B.6) gives the transition from a static to an adiabatic level ionization probability in general form.

A detailed study (including numerical calculations) of the momentum, angular, and energy distributions of photoelectrons in an elliptically polarized strong low-frequency laser field was made by Goreslavskiĭ and Popruzhenko [35]. Unfortunately, this study contained some inaccuracies.

(1) In [35] it is stated (page 1201) that the "formulas for the tunneling regime from [4] … do not describe the

transition to circular polarization." However, this is not the case: converting in (5.6) from the Cartesian components of the electron momentum to the projections $p_\perp$ and p_z we obtain formula (5.8) which for $\xi \longrightarrow \pm 1$ is the same as expression (4.11) for circular polarization. For the total ionization probability this transition is described by formula (5.10) which appears in [3].

(2) Equations (29) and (34) in [35] contain the function $a(z)$ introduced in [3] which describes the transition region bordering on circular polarization. However, its argument $z = \xi^2(1 - \xi^2)/3\kappa^2\epsilon$ in [35] differs from (5.10) by the factor $2\xi^4/\kappa^2$. Consequently, when equation (36) is derived from (34) a superfluous factor $\kappa/\sqrt{2}$ appears, which may differ substantially from unity particularly for states having low binding energy. In addition, no passage to the limit occurs from (29), (34) to the case of linear ($\xi = 0$) polarization since the expressions for W and dW/dp from [35] for $\xi \longrightarrow 0$ go to zero and are only valid for $\xi \gg \sqrt{\epsilon}$. We also note that equation (36) from [35] also appears in [3] [see also formulas (4) and (11) in which we need to set $l = \lambda = 0$ and $C_{\kappa l}^2 = 2$, which corresponds to the δ-potential] and a correction of the order of γ^2 in the exponential function was also calculated in [4].

Note that the calculations in [3,4] and [35] were made using different electromagnetic field gauges.

(b) In the opposite case $\gamma \gg 1$ we shall confine ourselves to the logarithmic approximation $\Lambda = \ln(2\gamma/\sqrt{1-\xi^2})$ when [4]

$$g(\gamma, \xi) \approx \frac{3\Lambda}{2\gamma}, \quad c_x \approx c_y \approx c_z \approx \Lambda, \qquad p_{\max} = \frac{|\xi|\kappa}{\Lambda\sqrt{1-\xi^2}}, \tag{5.11}$$

$$\frac{dw(\mathbf{p})}{d^3p} \propto \exp\left\{-\frac{\Lambda}{\omega}(\mathbf{p} - \mathbf{p}_{\max})^2\right\} + \exp\left\{-\frac{\Lambda}{\omega}(\mathbf{p} + \mathbf{p}_{\max})^2\right\}, \tag{5.12}$$

where $\nu = K_0(1 + \delta)$, $n_0 = K_0(1 + 2\delta)$, where $\delta = \xi^2/(1 - \xi^2)\Lambda^2 \ll 1$. In addition $\Delta p/\kappa \sim \sqrt{\omega/\kappa^2\Lambda} \ll 1$ and the ratio $\Delta p/p_{\max} \sim \sqrt{\Lambda(\xi^{-2} - 1)/K_0}$ is also small if the ellipticity is not too close to 0 or ±1. Integrating over angles allowing for $\delta((p^2 - p_n^2)/2)$, which expresses the energy conservation law, we find

$$w_n = w_{\max}\exp\left\{-c_{10}\frac{(n - n_0)^2}{2n_0}\right\}, \qquad c_{10} = \frac{1}{2}\Lambda^3(\xi^{-2} - 1)(1 + 2\delta) \tag{5.13}$$

($c_{10} \gg 1$ if $1 - \xi^2 \gg \Lambda^{-3}$). The width of this distribution is $\Delta n \sim \sqrt{K_0/\Lambda^3}$ and $\Delta n/(n_0 - \nu) \sim \Delta p/p_{\max} \sim \sqrt{\Lambda/K_0} \ll 1$ if $\gamma \ll \exp(K_0)$ which is satisfied almost everywhere ($K_0 \gg 1$ is the multiquantization parameter).

Numerical calculations show that the function $g(\gamma, \xi)$ for $0 < \xi^2 < 1$ has qualitatively the same form as for $\xi = 0$ and 1 and is "compressed" between the two curves in Fig. 3. As the ellipticity $|\xi|$ increases, the value of g increases monotonically and the ionization probability decreases. The decrease in $g(\gamma, \xi)$ with increasing γ leads to an abrupt increase in the ionization probability on transition from the low-frequency region to the region $\gamma \gg 1$ (for the same value of the maximum field F).

6. ABOVE-BARRIER IONIZATION

This process takes place in strong fields $\mathcal{F} \equiv n^4F \gtrsim \mathcal{F}_{cr}$. Here $\mathcal{F}_{cr} = n^4F_{cr}$ is the "critical" value of $\mathcal{F}$ at which the level energy touches the top of the potential barrier (in the direction of electron emission). For atomic hydrogen states the values of $\mathcal{F}_{cr}$ vary between 0.130 and 0.383 depending on the parabolic quantum numbers (n_1n_2m), see [36–38].

We calculate the ionization probability (per unit time) in a low-frequency field in the adiabatic approximation:

$$w_a = \frac{1}{2\pi}\int_0^{2\pi} w_{st}(F(\phi))d\phi, \quad \gamma \ll 1, \tag{6.1}$$

where $\phi = \omega t$ is the phase of the field, and $w_{st}(F)$ is the ionization probability (i.e., the Stark resonance width) for a static electric field F. For the case of an elliptically polarized wave (5.1) we have

$$F(\phi) = F\sqrt{\cos^2\phi + \xi^2\sin^2\phi}.$$

Substituting this into (6.1) we can easily obtain $w_a(F, \xi)$ numerically. For circular polarization we have $F(\phi) = F = \text{const}$ so that $w_a = w_{st}(F)$.

We used values of the static probability $w_{st}(F)$ for a hydrogen atom obtained [39–42] using summation of higher orders of perturbation theory with respect to the field F and a $1/n$ expansion (in the range $F \lesssim 0.2$ these calculations agree with results obtained by other authors [43–50] and were recently confirmed by independent calculations [49]). Results of calculations

Table 1. Parameters of intermediate asymptotic form (6.2) in the above-barrier region for a hydrogen atom

ξ	k	F_0	σ, %	F_1	F_2
			Ground state		
1	1.473	0.122	0.12	0.3	1.0
1	1.482	0.124	0.13	0.25	0.5
0.9	1.40	0.128	0.12	0.3	1.0
0.8	1.33	0.135	0.10	0.3	1.0
0.7	1.26	0.143	0.08	0.3	1.0
0.5	1.13	0.157	0.12	0.3	1.0
0.5	1.10	0.148	0.21	0.25	0.5
0.3	1.02	0.166	0.21	0.3	1.0
0	0.913	0.163	0.20	0.3	1.0
			Rydberg states		
(0, 0, $n-1$)	0.810	0.260	0.42	0.3	2.0
	0.807	0.249	0.59	0.25	2.0

Note: The values of the parameters k and F_0 were determined by least squares fitting and refer to the case of a low-frequency field where for $\xi = 1$ the parameters are the same as those for a static field. Also given are the parameters for the (0, 0, $n-1$) states with $n \gg 1$ in a static field F [54].

using formula (6.1) are plotted in Fig. 6. As $|\xi|$ increases, the probability $w_a(F, \xi)$ increases since the period-averaged field acting on the electron increases (for fixed F).

Figure 6 shows the field strength F_{BSI} at which the level energy is the same as the maximum of the effective potential and the atomic state can undergo above-barrier decay according to classical mechanics (in [38] this field was called "critical"). For Rydberg states (0, 0, $n-1$) with $n \gg 1$ corresponding to circular electron orbits, we have $\mathcal{F}_{cr} = 0.208$ which is significantly higher than the value of $\mathcal{F}_{cr} = 1/16$ for the one-dimensional model of the Stark effect $V(x) = -x^{-1} - Fx$. Above-barrier ionization takes place for $F > F_{cr}$.

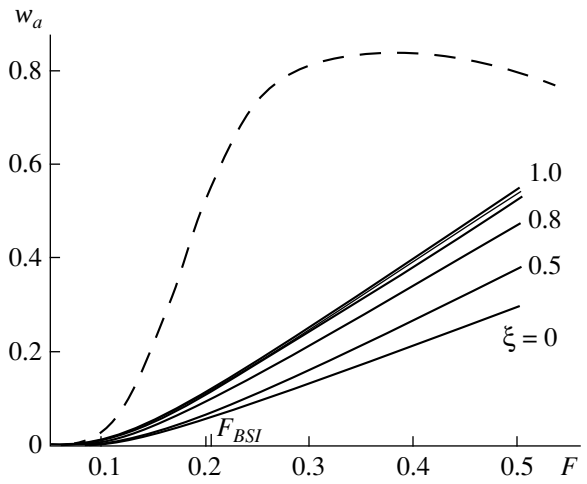


Fig. 6. Probability of above-barrier ionization of a hydrogen atom ($n = 1$, ground state) in a low-frequency laser field. The solid curves correspond to (from bottom to top) ellipticity values $\xi = 0$, 0.5, 0.8, 0.9, 0.95, and 1.0. The dashed curve gives the results of [58] for circular polarization. The values of w_a and F are given in atomic units.

It can be seen from Fig. 6 that in the above-barrier region the dependence of the Stark resonance width on the field F is surprisingly close to linear:[7]

$$w_a(F, \xi) \approx k(F - F_0), \quad F > F_0 \sim F_{cr}, \tag{6.2}$$

where k and F_0 are parameters which depend on the ellipticity ξ and on the quantum numbers ($n_1 n_2 m$) and are determined numerically. Their values for the ground and Rydberg ($n \gg 1$) states of the hydrogen atom are given in Table 1 which also gives the fitting interval $F_1 < F < F_2$ and the mean-square deviation σ (in percent). The parameters k and F_0 depend weakly on the choice of fitting interval and on the number of fitting points in it (see the values for $\xi = 1$ and 0.5 in Table 2). In this case the linear dependence (6.2) is not satisfied for weak fields $F \ll F_{cr}$ where $w \propto \exp(-2/3F)$ or in the far asymptotic form $F \longrightarrow \infty$, where $w(F) \propto (F \ln F)^{2/3}$ [51]. At the same time, as can be seen from Fig. 6, this dependence is well satisfied for $F_{cr} < F \lesssim 1$, i.e., is an example of the so-called intermediate asymptotic form.[8] An explanation of this (approximate) linearity in

[7] This fact was first noted in [40] for a static field F. Recent numerical calculations [50] for a helium atom confirm the presence of a linear regime for $w_{st}(F)$ in the above-barrier region $F > F_{BSI}$.

[8] An interesting discussion of the problem of the intermediate asymptotic form in hydrodynamics and mathematical physics can be found in [52, 53]. Concepts of dimensionality and similarity play a decisive role in finding these asymptotic forms as in the case of the $1/n$ expansion in quantum dynamics (for example, the scaling $E \longrightarrow n^2 E$, $F \longrightarrow n^4 F$ in the problem of the Stark effect [40, 54]).

the theory of the Stark effect in a strong field was obtained [54] using a $1/n$ expansion which in quantum mechanics problems not only works well at high quantum numbers but is fairly frequently "pulled" as far as $n \sim 1$ (in this context see [55]). In the low-frequency range for $F \longrightarrow 0$

$$w_a(F, \xi) = \frac{1}{|\xi|} a\left(\frac{(1-\xi^2)\kappa^3}{6\xi^2 F}\right) w_{\rm cl}(F) \qquad (6.3)$$

(see [3] and Appendix B) where $a(x)$ was determined in (B.5), and $w_{\rm cl}(F)$ is the probability of tunneling ionization calculated using semiclassical formulas for a static field: for example, $w_{\rm cl}(F) = 4F^{-1}\exp(-2/3F)$ [16] for the ground state of the hydrogen atom, $w_{\rm cl}(F) = 1.255F^{-1} \times \exp(-1/12F)$ for the (100) state with $n = 2$, and for an arbitrary state formulas for $w_{\rm cl}(n_1 n_2 m)$ can be found, for example, in [23].

We shall now analyze the ratio of the ionization probability $\bar{w}_a$ calculated using formula (6.1) using exact [39–41] values of $w_{\rm st}(F)$ to the tunneling ionization probability (6.3):

$$q(F, \xi) = \bar{w}_a(F, \xi)/w_a(F, \xi), \qquad (6.4)$$

where $q(F, \pm 1) = w_{\rm st}(F)/w_{\rm cl}(F)$ for circular polarization. Figure 7 gives the dependence of q_0 (ground state) on F. It can be seen that the semiclassical formula $w_{\rm cl}(F)$ being asymptotically exact in the weak field limit ($q \longrightarrow 0$ for $F \longrightarrow 0$) ceases to hold at a comparatively early stage $F \sim 0.01$ [56, 57]. For $F = 0.2$–0.4, i.e., for fields of the order of 10^9 V/cm, $w_{\rm cl}(F)$ is several times higher than the level ionization probability. Thus, the tunneling formulas such as (6.3) do not continue into the above-barrier region. It is interesting to note that $q_0(F, \xi)$ is almost independent of the ellipticity of the radiation (Fig. 7).

The dashed curves in Fig. 7 give the results of [33, 58]. In the range of fields between $F \sim 0.05$ and $F \sim 0.4$ these approximations sharply contradict our calculations and also the known [59] expansions of $q_0(F)$ in powers of F (see Appendix C) and cannot be taken to be satisfactory.

For $F \gtrsim 0.2$, the probability $w_a(F)$ reaches values such that a hydrogen atom is ionized within one or two electron revolutions about the nucleus. It is then necessary to allow for a saturation effect: in the simplest case of a field applied at time T, the probability of ionization of an atom is not $w_a T \gtrsim 1$ but $1 - \exp(-w_a T)$ or (in general) $1 - \exp\{-\int w_a(F(t))dt\}$. Problems associated with the experimental observation of above-barrier ionization are discussed in [11, 27, 32].

7. GENERALIZATION FOR THE LINEAR POLARIZATION

For linearly polarized radiation, formula (2.1) can be generalized to the case of an arbitrary time dependence of the electric field $F(t)$. It is then assumed that

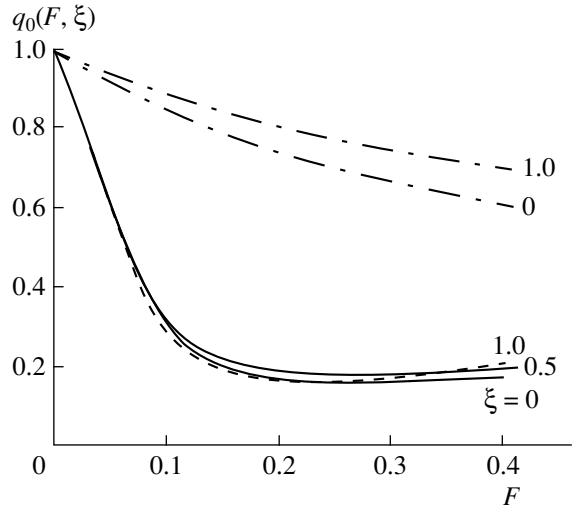


Fig. 7. The functions $q_0(F)$ from (6.4) for the ground state of a hydrogen atom. Solid curves give our calculations for $\xi = 0$ and 0.5, the dashed curve gives these calculations for $\xi = 1$. The dash–dotted curves give the results of [33, 58]. The values of the ellipticity ξ are indicated on the curves.

$F(t)$ is an analytic function of t which can be continued into the complex plane which is required to apply an "imaginary time" method [4]. Omitting details of the calculations, we give the final formulas.

Let us assume that $F(t)$ is an alternating electric field directed along the x axis:

$$F(t) = F\varphi(\omega t), \quad -\infty < t < \infty, \qquad (7.1)$$

F is its peak value where $\varphi(-t) = \varphi(t)$ and $|\varphi(t)| \le \varphi(0) = 1$ (it is convenient to assume that $t = 0$ is the time of maximum field when the particle escapes from below the barrier). The extreme subbarrier trajectory[9] is obtained from the equation of motion with the boundary conditions:

$$\ddot{\xi} = -\gamma^{-1}\varphi(i\tau), \quad \xi(\tau_0) = \dot{\xi}(0) = 0, \quad \dot{\xi}(\tau_0) = -1, \qquad (7.2)$$

where $\xi = \omega x/\kappa = \gamma\epsilon\kappa x$, $\dot{\xi} = d\xi/d\tau$, $\tau = -i\omega t$ is the imaginary "time" (in subbarrier motion) $0 < \tau < \tau_0$) and $\gamma = \omega\kappa/F$ is the Keldysh parameter for the field (7.1).

It follows from (7.2) that the initial time of subbarrier motion for the extreme trajectory is $\tau_0 = \tau(\gamma)$, where τ is a function determined by the profile of the external field:

$$\tau(u) = h^{-1}(u), \quad h(u) = \int_0^u \varphi(it)dt, \qquad (7.3)$$

[9] I.e. the "classical" trajectory which minimizes the imaginary part of the action function and determines the most probable path for particle tunneling (for further details see [4, 60]).

where $h^{-1}(u)$ here implies the function which is the reciprocal of $h(u)$.

We shall consider a pencil of classical subbarrier trajectories (i.e., trajectories satisfying the classical equations of motion but with an imaginary time) close to the extreme trajectory. These differ in respect of the transverse momentum $\mathbf{p}_\perp$ which is an integral of motion. By calculating the imaginary part of the action function along these trajectories, we find the electron momentum spectrum:

$$w(\mathbf{p}) \propto \exp\left\{-\left[2K_0 f(\gamma) + \frac{1}{\omega}(c_1 p_\parallel^2 + c_2 p_\perp^2)\right]\right\}, \quad (7.4)$$

where $K_0 = \kappa^2/2\omega \gg 1$,

$$f(\gamma) = 2\int_0^1 \tau(\gamma u)u\,du = \tau_0(\gamma) - \int_0^1 \tau_1(\gamma u)u\,du, \quad (7.5)$$
$$c_1 = \tau_0(\gamma) - \tau_1(\gamma), \quad c_2 = \tau_0(\gamma)$$

and we have introduced the notation $\tau_1(u) = u\tau'(u)$. These expressions can be conveniently used at fixed laser frequency. In the adiabatic region and also for the passage to the limit $\omega \longrightarrow 0$ it is more convenient to rewrite (7.4) in a slightly different form:

$$w(\mathbf{p}) \propto \exp\left\{-\frac{1}{\epsilon}\left[\frac{2}{3}g(\gamma) + b_1(\gamma)\frac{p_\parallel^2}{\kappa^2} + b_2(\gamma)\frac{p_\perp^2}{\kappa^2}\right]\right\}, \quad (7.6)$$

where $g(\gamma) = 3f(\gamma)/2\gamma$ and $b_i(\gamma) = \gamma^{-1}c_i(\gamma)$ where $g(0) = b_2(0) = 1$ and $b_1 \sim \gamma^2$ for $\gamma \longrightarrow 0$.

Thus, all quantities are expressed in terms of a single function $\tau(\gamma)$ which determines the position of the saddle point in the complex plane. The closed analytic expressions (7.3)–(7.6) reduce the calculation of the ionization probability to quadratures. We shall confine ourselves to a few examples.

(a) For a monochromatic laser field $\varphi = \cos\omega t$ we have $\tau_0 = \operatorname{arcsinh}\gamma$, $\tau_1 = \gamma/\sqrt{1+\gamma^2}$, and equations (7.5) rapidly yield (2.2).

(b) The momentum field $\varphi = 1/\cosh^2\omega t$ is similar to the exactly solvable potential in quantum mechanics [16]. In this case, $\tau_0 = \arctan\gamma$, $\tau_1 = \gamma/(1+\gamma^2)$,

$$f(\gamma) = (1+\gamma^{-2})\arctan\gamma - \gamma^{-1},$$
$$c_1(\gamma) = \arctan\gamma - \gamma(1+\gamma^2)^{-1}, \quad (7.7)$$
$$c_2(\gamma) = \arctan\gamma.$$

From this it follows that for $\gamma \gg 1$

$$f(\gamma) = \frac{\pi}{2}\left(1 - \frac{4}{\pi\gamma} + \frac{1}{\gamma^2} + \ldots\right),$$
$$c_1 = \frac{\pi}{2} - 2\gamma^{-1} + O(\gamma^{-3}), \quad c_2 = \frac{\pi}{2} - \gamma^{-1}, \quad (7.8)$$
$$w(\mathbf{p}) \propto \exp\left\{-\frac{\pi}{2\omega}(\kappa^2 + p^2)[1 + O(\gamma^{-1})]\right\},$$

which differs substantially from the behavior of the probability of ionization by monochromatic radiation:

$$w \propto \exp\{-(\kappa^2 + p^2)\omega^{-1}\ln 2\gamma\} \propto \gamma^{-2K_0}, \quad \gamma \gg 1. \quad (7.9)$$

This difference has a simple explanation: the momentum spectrum $\varphi(\omega t)$ contains not only the fundamental frequency ω but also higher harmonics whose contribution, although decreasing exponentially, still "clogs up" the multiphoton ionization as a result of the frequency ω. For example, for $\varphi = 1/\cosh^2\omega t$ the amplitude of the harmonic at frequency Ω is

$$|\varphi\Omega|^2 = \left\{\frac{\Omega}{2\omega^2\sinh(\pi\Omega/2\omega)}\right\}^2 \propto \exp\left(-\frac{\pi\Omega}{\omega}\right), \quad (7.10)$$
$$\Omega \gg \omega,$$

and for $\Omega > E_I$ the level can still be ionized by this harmonic in the first order of perturbation theory. However, for laser light containing no higher harmonics, only the multiphoton ionization mechanism remains for $\gamma \gg 1$.

(c) If $\gamma \ll 1$, ionization takes place at times near the electric field maximum. Assuming that at $t \longrightarrow 0$

$$\varphi(t) = 1 - \frac{a_2}{2!}t^2 + \frac{a_4}{4!}t^4 - \ldots, \quad a_2 > 0 \quad (7.11)$$

and using (7.3)–(7.6) we obtain the expansions

$$g(\gamma) = 1 - \frac{a_2}{10}\gamma^2 - \frac{1}{280}(a_4 - 10a_2^2)\gamma^4$$
$$- \frac{1}{15120}(a_6 - 56a_4a_2 + 280a_2^3)\gamma^6 + \ldots,$$
$$c_1 = \frac{1}{3}a_2\gamma^3 + \frac{1}{30}(a_4 - 10a_2^2)\gamma^5 + \ldots, \quad (7.12)$$
$$c_2 = \gamma - \frac{1}{6}a_2\gamma^3 + \ldots,$$

which determine the adiabatic corrections to the ionization probability and the momentum spectrum of the electrons. We note the numerical smallness of the coefficients in (7.12) as a result of which the adiabatic approximation is not abruptly truncated at $\gamma \sim 1$ but is "pulled" toward slightly higher values of the parameter

γ (it is assumed that the coefficients a_2, a_4 … are of the order of unity).

8. TUNNELING INTERFERENCE

For linearly polarized laser radiation an interference effect occurs in the energy spectrum of the photoelectrons, which was noted in [3]. The initial condition for subbarrier motion $\mathbf{p}^2(t) = -\kappa^2$ determines the saddle points in the complex plane t. For a periodic field $F(t) = F\cos\omega t$ their position is given by

$$\sin\omega t_k = \gamma[i(-1)^k\sqrt{1+q_\perp^2} - q_\parallel], \quad \mathbf{q} = \mathbf{p}/\kappa \qquad (8.1)$$

($k = 0, \pm1, \pm2, \ldots$) whence for $p \ll \kappa$ we have

$$\omega t_k = k\pi + i\left(\tau_0 + \frac{1}{2}\tau_1 q_\perp^2 + \tau_2 q_\parallel^2\right) + (-1)^k q_\parallel(\tau_1 - \tau_2 q_\perp^2 - \tau_3 q_\parallel^2) + O(q^4), \qquad (8.2)$$

where

$$\tau_0 = \operatorname{arcsinh}\gamma, \quad \tau_1 = \gamma/\sqrt{1+\gamma^2},$$

$$\tau_2 = \frac{1}{2}\tau_1^3, \quad \tau_3 = \frac{2\gamma^2 - 1}{3(1+\gamma^2)}\tau_2$$

and so on. The amplitude A_k of an electron transition from the bound state ($E_0 = -E_I = -\kappa^2/2$) to states of the continuous spectrum is determined by the action function calculated along the path from the point t_k to the real time axis. Summation of the amplitudes corresponding to the two saddle points positioned within the same period $T = 2\pi/\omega$ (for example, A_0 and A_1) gives an interference factor in the photoelectron momentum spectrum. For example, for ionization of the s-level

$$w(\mathbf{p}_n) \longrightarrow \frac{1}{2}[1 + (-1)^n\cos\phi_n]w(\mathbf{p}_n), \qquad (8.3)$$

where $p_n^2 = 2\omega(n - \nu)$ is given by formula (2.1) and the phase of the oscillations is[10]

$$\phi_n = \frac{\kappa^2 p_\parallel}{F}\left[\frac{2}{1+\sqrt{1+\gamma^2}} + \frac{1-\gamma^2}{3(1+\gamma^2)^{3/2}}\frac{p_\parallel^2}{\kappa^2}\right]\Bigg|_{p=p_n}. \qquad (8.4)$$

If $\gamma \gg 1$, then $p^2/\kappa^2 \sim 1/\ln 2\gamma$ so that the second term can be neglected in (8.4). However, for $\gamma \ll 1$ we have

$$p_\parallel \sim \sqrt{\frac{F^3}{\omega^2\kappa^3}} = \frac{\sqrt{\epsilon}F}{\omega},$$

and for values of $\gamma \lesssim \sqrt{\epsilon} \ll 1$ the contribution of the second term is comparable with the first.

Oscillations in the momentum spectrum are observed for $p_\parallel \neq 0$. This is because a real phase shift ϕ occurs between the amplitudes A_k and A_{k+1}, originating from a displacement of the points t_k and t_{k+1} along the real time axis.

In the semiclassical case (i.e., when the conditions $\epsilon \ll 1$, $\omega \ll \kappa^2$ are satisfied) the phase is $\phi_n \gg 1$and averaging over angles [in accordance with (2.6)] leads to an appreciable decrease in the amplitude of the oscillations in the probabilities w_n, i.e., in the energy spectrum of the photoelectrons. This agrees qualitatively with [61, 62] in which manifestations of quantum interference were observed experimentally in an elliptically polarized field: the change in the momentum distribution of the photoelectrons with the ellipticity of the radiation was investigated, see Figs. 1 and 2 in [61].

Near the photoionization threshold the phase is $\phi_n \propto p_n = \sqrt{2\omega(n-\nu)} \longrightarrow 0$, so that the threshold behavior of the ionization probabilities noted by Nikishov and Ritus [2] follows directly from (2.6) and (8.3): $w_n \propto (n-\nu)^{1/2} \propto \sqrt{E_n}$ for even n and $w_n \propto (n-\nu)^{3/2}$ for odd n. As for the sign factor $(-1)^n$ in (8.3) this difference occurs because for $p_\parallel = 0$ the transition amplitudes for neighboring (every half-period) saddle points have different signs.

Formulas (8.3) and (8.4) can be generalized to the case of an arbitrary field of the type (7.1) for which $F(t) = -F(t + T/2) = F(t + T)$.

9. ASYMPTOTIC COEFFICIENTS C_κ

The probability of tunneling ionization is proportional to the square of the asymptotic coefficient [see (2.8)] of the wave function of the atomic level. The values of C_κ for atoms and ions are determined numerically by the Hartree–Fock method [17]. Analytic approximations were also obtained by Hartree [34] (see also [18, 19]) and in [6, 20]. The error of the Hartree formula

$$C_{\kappa l} = \frac{2^{\eta-1}}{\sqrt{\eta\Gamma(\eta+l+1)\Gamma(\eta-l)}}, \quad \eta > l \qquad (9.1)$$

for the first s- and p-wave terms in the Rb atom is 2.5% and 2%, respectively and is even lower for higher terms [34]. The refinement (9.1) was obtained [20] using the effective range expansion (ERE). For the factor N from

[10] We take the opportunity to correct a misprint in [3]: in formula (53) in this study in the interference phase ϕ, $\sqrt{1+\gamma^2}$ should be replaced by $\sqrt{1+\gamma^2} - 1$ as a result of which this expression is the same as the first term in (8.4).

Table 2. Comparison of various approximations for the coefficient N

Atom or ion	E_I, eV	η	N			δ
			H	ERE	HF	
He (1*s*)	24.59	0.744	2.336	2.374	2.766	A
H (1*s*)	13.60	1.000	4.000	4.000	4.000	A
Au (6*s*)	9.226	1.215	5.88	5.90	6.55	A
Cu (4*s*)	7.726	1.326	7.03	7.06	7.33	A
U (7*s*)	6.194	1.482	8.83	8.87	9.15	A
Ca (4*s*)	6.113	1.492	8.96	8.99	8.81	C
Sr (5*s*)	5.695	1.546	9.63	9.67	9.37	C
Li (2*s*)	5.392	1.589	10.19	10.23	10.5	B
Na (3*s*)	5.139	1.627	10.69	10.72	10.3	B
K (4*s*)	4.341	1.770	12.7	12.7	10.5	C
Rb (5*s*)	4.177	1.805	13.2	13.2	10.7	C
Cs (6*s*)	3.894	1.869	14.1	14.1	11.4	C
Fr (7*s*)	4.0	1.84	13.7	13.7	–	–
Li^+	75.64	0.848	2.94	2.96	3.38	C
Be^+	18.21	1.729	12.08	12.09	10.23	A
Mg^+	15.04	1.903	14.57	14.58	14.69	A
Ca^+	11.87	2.141	18.09	18.12	18.3	B
Sr^+	11.03	2.221	19.27	19.29	18.6	B

Note: N is the coefficient in formulas (3.11) and (5.9), $\eta \equiv n^*$ is the effective principal quantum number; $\eta = (2E_I)^{-1/2}$ for neutral atoms, $\eta = (E_I/2)^{-1/2}$ for singly charged positive ions. Notation used: H is the Hartree approximation [34], ERE is the effective range expansion for $r_{cs} = 0$ [20], HF are numerical calculations using the Hartree–Fock method [17] whose classes of accuracy (A, B, C, see text) are given in the last column.

(3.11) in the commonly encountered case of *s*-states in the Hartree (H) and ERE approximations we obtain

$$N_{\mathrm{H}} = \left[\frac{2^{2\eta-1}}{\Gamma(\eta+1)}\right]^2, \qquad (9.2)$$

$$N_{\mathrm{ERE}} = N_{\mathrm{H}}\left\{1-\left(\frac{\sin\pi\eta}{\pi\eta}\right)^2\left[\eta^2\psi'(\eta)-\left(\eta+\frac{1}{2}\right)\right]\right\}^{-1},$$

where $\psi(\eta)$ is the logarithmic derivative of the gamma function (tables of functions $\Gamma(x)$ and $\psi'(x)$ are readily available, for example [29]). In this case we have $N_{\mathrm{ERE}} \geq N_{\mathrm{H}}$.

Results of calculations for neutral atoms (including atoms of all alkali metals for which a single valence electron in the n_s state is situated outside the filled shells) and various positive ions are presented in Table 2 where the atoms are given in order of increasing parameter η. It can be seen that these approximations are very similar and generally agree with the results of the Hartree–Fock calculations whose error δ (for the coefficients C_κ) is [17] $\delta < 1\%$ (accuracy class A), $\delta = 1$–3% (B), and $\delta = 3$–10% (C). The approximations (9.2) are even more useful since, in many cases, the error of calculations using the Hartree–Fock method is 30% [17].

For excited *s*-wave states both approximations are almost the same:

$$N_{\mathrm{ERA}} = N_{\mathrm{H}}\left(1+\frac{\sin^2\pi\eta}{6\pi^2\eta^3}\right), \quad \eta \gg 1. \qquad (9.3)$$

However, for negative ions the values of N_{H} and N_{ERE} differ by a factor of two:

$$N_{\mathrm{H}} = \frac{1}{4} + k_1\eta + \ldots,$$
$$N_{\mathrm{ERE}} = \frac{1}{2} + k_2\eta + \ldots, \quad \eta \longrightarrow 0 \qquad (9.4)$$

($k_1 = \ln 2 + (1/2)C = 0.982$, $k_2 = 2k_1 - 1 = 0.964$, and $C = 0.5772\ldots$ is the Euler constant) and only N_{ERE} agrees with the limit of zero range of the forces in the expansion

$$N = C_\kappa^2 = \frac{1}{2}[1-\kappa r_s + O((\kappa r_s)^3)]^{-1}, \qquad (9.5)$$
$$\kappa r_s \ll 1,$$

which is valid for shallow *s*-wave levels in the short-range ($\eta = 0$) potential; here r_s is the effective interaction range [16].

In the presence of a Coulomb potential, r_s should be replaced by the so-called nuclear-Coulomb effective radius r_{cs} for which an accurate definition can be found, for example, in [63–65]. The expansion (9.5) has the form [20, 66]

$$N = N_H \times\left\{1-\left(\frac{\sin\pi\eta}{\pi\eta}\right)^2\left[\lambda(\eta)+\frac{1}{2}\kappa r_{cs}+O((\kappa r_0)^3)\right]\right\}^{-1}, \quad (9.6)$$

where

$$\lambda(x) = x^2\psi'(x)-x-\frac{1}{2} = \begin{cases}\frac{1}{2}-x+\frac{\pi^2}{6}x^2+\ldots, & x\longrightarrow 0,\\ \frac{1}{6x}-\frac{1}{30x^3}+\ldots, & x\longrightarrow\infty.\end{cases}$$

From this it can be seen that the value of N_{ERE} (9.2) corresponds to the limit $r_{cs}\longrightarrow 0$ (note that this approximation has already been considered in calculations of the ionization of systems coupled by short-range and Coulomb forces [6] and also in the theory of the lightest hadron atoms [20]).

For a pure Coulomb field $V(r) = -Z/r$ the parameter is $\eta = n = 1, 2, \ldots$ and $C_H = C_{ERE} = 2^{n-1}/n!$ yields the well-known formula [see (36.15) in [16]] and $N = (2^{2n-1}/n!)^2$.

We stress that the expansions (9.5) and (9.6) contain no quadratic terms with respect to the effective radius [66] which extends their range of validity. However, this occurs when the effective range corresponding to the exact level energy $E = -\kappa^2/2$ is taken as r_s (respectively, r_{cs}). We shall illustrate this for the example of a square well (of radius R, depth $U_0 = g^2/2R^2$, g is the dimensionless coupling constant), when [3, 64]

$$C_\kappa^2 = \left[1-\frac{(\kappa R)^2}{g^2}\right]\frac{e^{2\kappa R}}{2(1+\kappa R)}, \quad l = 0, \quad (9.7)$$

$$r_s = R\left(1-\frac{1}{g^2\xi}-\frac{1}{3\xi^2}\right), \quad \xi = 1-\frac{\tan g}{g}. \quad (9.8)$$

The bound state ns exists for $g > g_n = (n-1/2)\pi$, $n = 1, 2, \ldots$. For a shallow ($h = g - g_n \longrightarrow 0$) level we find

$$\kappa R = \frac{1}{2}h-\frac{1}{8}(1-g_n^{-2})h^2+\ldots,$$
$$C_\kappa^{-2} = 2\left[1-\kappa r_s+\left(\frac{1}{3}-\frac{1}{2g_n^2}\right)(\kappa r_s)^3+O((\kappa r_s)^4)\right], \quad (9.9)$$

where

$$r_s = R\left[1-\frac{\kappa R}{g_n^2}+\left(\frac{1}{3}-\frac{1}{2g_n^2}-\frac{2}{g_n^4}\right)(\kappa R)^2+\ldots\right].$$

If $\kappa\tilde{r}_s$ is taken as the expansion parameter where $\tilde{r}_s = R$ is the effective range at the level formation time, then we have

$$C_\kappa^{-2} = 2\left[1-\kappa R+g_n^{-2}(\kappa R)^2+\left(\frac{2}{3}-g_n^{-2}-2g_n^{-4}\right)(\kappa R)^3+\ldots\right], \quad (9.10)$$

i.e., a term proportional to $\tilde{r}_s^2$ is already present. Expansions similar to (9.9) and (9.10) for various other potentials are given in [66].

To conclude this section, we make two observations.

(1) For an arbitrary attractive potential we have $\tilde{r}_s^2 > 0$ (see [16, page 632]) so that it follows from (9.5) that the asymptotic coefficient C_κ increases as the level becomes deeper. Although this conclusion is obtained for $\kappa r_s \ll 1$, this tendency may be conserved [66] as far as $\kappa r_s \sim 1$ when the level can no longer be called "shallow."

(2) For a square well $\tilde{r}_s$ does not depend on the level number n. However, this is a specific sharp-edged finite potential. For the case of smooth potentials having an exponential (for $r\longrightarrow\infty$) "tail" the range $\tilde{r}_s$ increases logarithmically with n [66].

10. CONCLUSIONS

Thus, we have analyzed the form of the energy and angular distributions of the photoelectrons over the range of variation of the Keldysh parameter $0 < \gamma < \infty$ and we make some concluding remarks.

(1) Although the relative width of the energy distribution increases with increasing γ, it still remains fairly narrow: $\Delta/E_{max} \sim \sqrt{\gamma/K_0}$ and $1/\sqrt{K_0}$ for $\gamma \ll 1$ and $\gamma \gg 1$, respectively, see (3.3) and (3.5). Slightly unexpectedly the function $s(\psi, \gamma)$ in (4.2) is almost independent of γ. Thus, the angular distribution of the photoelectrons is mainly determined by the average number of absorbed photons $n_0(\gamma)$ which yields the scaling relation (for circular polarization)

$$w(\psi, \gamma_2) \approx [w(\psi, \gamma_1)]^\rho, \quad \rho = n_0(\gamma_2)/n_0(\gamma_1), \quad (10.1)$$

where the normalization $w(0, \gamma) = 1$ is assumed.

(2) The general case of elliptically polarized radiation has been considered for $\gamma \ll 1$ and $\gamma \gg 1$. For arbitrary γ the calculations can be made using formulas from [4] although the expressions obtained are fairly cumbersome and require separate discussion.

(3) The probability of above-barrier ionization of hydrogen atoms has been calculated, It has been shown that extrapolating the semiclassical (tunneling) formulas to the above-barrier region, in the same way as using the formulas from [35, 58] strongly exaggerates the ionization probability in this range.

(4) For linearly polarized radiation the formulas for the photoelectron momentum spectrum have been generalized for an arbitrary time dependence of the electric field.

(5) We have analyzed the properties of the asymptotic coefficients C_κ and we have shown that the Hartree approximation (9.1) is highly accurate for neutral atoms and positive ions.

Bates and Damgaard [18] note that the formula (9.1) was not strictly substantiated in [34]. This substantiation was obtained later using the quantum defect method [8] and also from the effective radius expansion [20, 66].

(6) The formulas obtained above refer to ionization of the nondegenerate *s*-wave level (the case frequently encountered in practice: in accordance with [17] for all neutral atoms, from hydrogen to uranium, in 61 cases the valence electron is located in the *s*-wave state and in 30 cases it is in the *p*-wave state and we only have $l = 2$ for the Pd atom). For $l \neq 0$ in calculations of the complex energy [which describes both the shift and the level width $\Gamma(F)$] at short distances where the atomic potential is spherically symmetric we need to allow for mixing of lM states with various projections of the orbital angular momentum l. This was performed explicitly in [67] for the case of a *p*-wave level bound by short-range forces in a circularly polarized wave field. In this case states having a specific quasienergy correspond to $M = 0$ and two superpositions of $|\pm\rangle$ states with $M = \pm 1$, where M is the projection of the orbital angular momentum on the direction of wave propagation. In the limit $\omega \longrightarrow 0$ the state $|+\rangle$ corresponds to the projection $m = 0$ on the direction of the electric field $\mathbf{F}$ and $|-\rangle$ corresponds to a superposition (with different weights) of states with $m = \pm 1$. In a low-frequency field (and specifically subject to the condition $\gamma \ll \epsilon\kappa r_c$ where r_c is the radius of the atomic core) the width is $\Gamma^{(+)}(F) \gg \Gamma^{(-)}(F)$ since [2, 3] $\Gamma_{lm} \propto \epsilon^{|m|+1}$ and $\epsilon = F/\kappa^3 \ll 1$. The highest ionization probability is obtained for states with $m = 0$ for which the formulas are the same as those for the *s*-wave level. For other states the mixing effect leads to some changes in the preexponential function [67] although the total ionization width of the l-wave level remains the same. For linearly polarized radiation the projection of m on the direction of the field remains the same so the formulas from [2, 3] retain the same form.

ACKNOWLEDGMENTS

The author thanks V.I. Kogan, A.I. Nikishov, L.B. Okun', A.Ya. Parshin, and V.I. Ritus for discussions of problems relating to this study and for moral support. The author also thanks S.P. Goreslavskiĭ, B.M. Karnakov, N.B. Narozhnyĭ, S.V. Popruzhenko, and especially V.D. Mur for extremely useful discussions at various stages of the work, I. L. Beĭgman for drawing attention to [18], V.D. Taranukhin and N. Shubin for information on [49], S. G. Pozdnyakov, V.M. Vaĭnberg, and V.A. Gani for numerical calculations, N.M. Markina for assistance with preparation of the manuscript, and finally A.A. Alekhina, T.S. Nosova, and I.S. Tsukerman by whose efforts the author was able to familiarize himself with [18, 28, 33, 34].

This work was partly supported by the Russian Foundation for Basic Research (project no. 98-02-17007).

APPENDIX A

We shall give expansions of the functions contained in the formulas for the ionization probability.

(a) For $\gamma \longrightarrow 0$ and linear polarization

$$g(\gamma) = \sum_{n=0}^{\infty} (-1)^n \frac{3[(2n-1)!!]}{(2n+1)(2n+3)2^n n!}\gamma^{2n} = 1 - \frac{1}{10}\gamma^2 + \frac{9}{280}\gamma^4 - \frac{5}{336}\gamma^6 + \ldots, \quad \text{(A.1)}$$

$$c_1 = \frac{1}{3}\gamma^3 - \frac{3}{10}\gamma^5 + \ldots, \quad c_2 = \gamma - \frac{1}{3}\gamma^3 + \frac{3}{40}\gamma^5 + \ldots,$$

and for circular polarization

$$g_c(\gamma) = 1 - \frac{1}{15}\gamma^2 + \frac{39}{2835}\gamma^4 + \ldots, \qquad P(\gamma) = 1 + \frac{11}{810}\gamma^4 + \ldots. \quad \text{(A.2)}$$

The numerical smallness of this last coefficient c explains the "plateau" of the preexponential factor $P(\gamma)$ for $\gamma \lesssim 1$ which is clearly visible in Fig. 4.

(b) In the opposite case ($\gamma \longrightarrow \infty$) we have [see (7.4)]

$$f(\gamma) = \frac{2\gamma}{3}g(\gamma) = \ln 2\gamma - \frac{1}{2} + \frac{\ln 2\gamma}{2\gamma^2} - \frac{3}{32\gamma^4} + \ldots, \quad \text{(A.3)}$$

$$c_1 = \ln 2\gamma - 1 + \ldots, \quad c_2 = \ln 2\gamma + O(\gamma^{-2}),$$

and in (3.9)

$$f_c(\gamma) = \frac{2\gamma}{3}g_c(\gamma) = \ln(\gamma\sqrt{\ln\gamma}) - \frac{1}{2}(1 - \ln 2) + \ldots, \quad \text{(A.4)}$$

$$P(\gamma) = \gamma(\ln\gamma)^{-3/2}\left[1 - \frac{\ln\ln\gamma}{2\ln\gamma} + \ldots\right].$$

(c) The asymptotic forms for the coefficients in formulas (3.8) and (3.10) are as follows:

$$c_4 = \frac{\gamma^{3/2}}{\sqrt{2\pi}}, \quad c_5 = \sqrt{\gamma} \text{ at } \gamma \longrightarrow 0$$

and

$$c_4 = \sqrt{\frac{2}{\pi}}\ln\gamma, \quad c_5 = 2/\sqrt{\ln\gamma} \text{ at } \gamma \longrightarrow \infty.$$

It follows from (3.9), (3.10), and the preceding formulas that

$$\frac{w_{\max}}{W_c} = \begin{cases} \sqrt{\dfrac{\omega^4\kappa}{\pi F^3}} = \dfrac{\gamma^{3/2}}{\sqrt{2\pi K_0}}, & \gamma \ll 1, \\ \sqrt{\dfrac{2}{\pi K_0}}\ln\gamma, & \gamma \gg 1. \end{cases} \tag{A.5}$$

Even in the case $\gamma \gg 1$ the value of $w_{\max}$ is small compared with the probability W_c. This corresponds to the fact that in cases of circular polarization the distribution of w_n always covers many values of n and thus summation over n in (2.6) can be replaced by integration.

(d) From (4.1) we have

$$\eta = \sqrt{1-\zeta^2} = \sqrt{\frac{\sin^2\psi + t_0^2\cos^2\psi + \gamma^2}{1+\gamma^2}} = \begin{cases} u_0, & \psi = 0, \\ 1, & \psi = \pi/2. \end{cases} \tag{A.6}$$

From this it follows that for $\gamma \longrightarrow 0$

$$\eta = \begin{cases} \sqrt{\sin^2\psi + \gamma^2\cos^2\psi} + O(\psi^4), & 0 < \psi < \pi/2 \\ \gamma - \dfrac{4}{9}\gamma^3 + O(\gamma^5), & \psi = 0, \end{cases} \tag{A.7}$$

and for $\gamma \longrightarrow \infty$

$$\eta = 1 - \frac{\cos^2\psi}{\gamma^2\ln\gamma}, \qquad s = -\ln\cos^2\psi - \frac{\sin^2\psi}{\gamma^2\ln\gamma} + \ldots. \tag{A.8}$$

In the range of small angles for any γ

$$\eta = u_0\left(1 + \frac{1-u_0^2}{2u_0^2}\psi^2\right), \quad s = u_0\psi^2 + O(\psi^4), \tag{A.9}$$

which gives (4.9). For $\psi = \pi/2$ the function $s(\gamma, \psi)$ has a logarithmic singularity and the ionization probability goes to zero,

(e) The expansions (A.1) follow directly from (7.5) if we assume that $\tau_1(u) = u(1+u^2)^{-1/2}$. In order to obtain the formulas (A.2) it is convenient to write equation (70) from [3] in the form

$$\gamma^2 = \left[u^2 - \left(1 - \frac{u}{\operatorname{arctanh} u}\right)^2\right](1-u^2)^{-1} = u^2 + \frac{8}{9}u^4 + \frac{112}{135}u^6 + \ldots.$$

Inverting this series we find

$$u^2 = \sum_{n=1}^{\infty} d_n\gamma^{2n}, \qquad t_0^2 = (1+\gamma^2)u^2 - \gamma^2 = \sum_{n=2}^{\infty}(d_n + d_{n-1})\gamma^{2n}, \tag{A.10}$$

where $d_1 = 1$, $d_2 = -8/9$, $d_3 = 304/405$, $d_4 = -78472/127575$, and so on, whence

$$t_0(\gamma) = \frac{1}{3}\gamma^2 - \frac{28}{135}\gamma^4 + \frac{236}{1701}\gamma^6 - \frac{5212}{54675}\gamma^8 + \ldots, \qquad u_0 = \sqrt{\frac{t_0^2+\gamma^2}{1+\gamma^2}} = \gamma - \frac{4}{9}\gamma^3 + \frac{112}{405}\gamma^5 + \ldots \tag{A.11}$$

Finally for $\gamma \longrightarrow \infty$ we have

$$t_0 = 1 - \frac{1}{\ln\gamma} + \frac{\ln\ln\gamma}{2(\ln\gamma)^2} + \ldots, \qquad u_0 = 1 - \frac{1}{\gamma^2\ln\gamma} + \ldots \tag{A.12}$$

(f) The variables used in Section 3 are related to τ_0 from (5.4) by

$$t_0 = \frac{\sqrt{\sinh^2\tau_0 - \gamma^2}}{\cosh\tau_0}, \quad u_0 = \tanh\tau_0, \tag{A.13}$$

where [see (3.9')]

$$g_c(\gamma) = \frac{3}{\gamma}\sqrt{v(1+v)}, \quad P(\gamma) = \frac{\gamma}{\tau_0\sqrt{1+2v}}, \tag{A.14}$$

where

$$v = \begin{cases} \dfrac{\sinh^2\tau_0 - \gamma^2}{\gamma^2} = \dfrac{1}{9}\gamma^2 - \dfrac{11}{405}\gamma^4 + \ldots, & \gamma \longrightarrow 0, \\ \dfrac{1}{2}\ln(\gamma\sqrt{2\ln\gamma}), & \gamma \longrightarrow \infty. \end{cases}$$

Using these formulas we can easily obtain expansions for g_c, P, c_3, c_4, and other quantities.

APPENDIX B

We shall now give the derivation of formulas (5.8)–(5.10).

Writing the probability of level ionization by a static electric field in the form $w_{st}(F) = w_0\epsilon^{-\alpha}\exp(-2/3\epsilon)$, where $\epsilon = F/\kappa^3$ is the reduced field ($\epsilon \ll 1$) and w_0 and α are various constants,[11] for the case of elliptic polarization (5.1) we have

$$F(t) = \frac{F}{f(\omega t)},$$
$$f(\tau) = \left\{\frac{1}{2}[(1+\xi^2)+(1-\xi^2)\cos 2\tau]\right\}^{-1/2}. \quad \text{(B.1)}$$

Using the adiabatic approximation (6.1) and converting from $\tau = \omega t$ to the integration variable $x = f^2(\tau)$, we obtain

$$w_a(F,\xi) = \frac{2}{T}\int_{-T/4}^{T/4} w_{st}(F(t))dt = \frac{w_{st}(F)}{\pi}$$
$$\times \int_1^{1/\xi^2} \exp\left\{-\frac{2}{3\epsilon}(\sqrt{x}-1)\right\}\frac{x^{(\alpha-2)/2}}{\sqrt{(x-1)(1-\xi^2 x)}}dx. \quad \text{(B.2)}$$

If the polarization is not too close to circular, the integral is taken near the lower limit which gives the relationship between the adiabatic and static ionization probabilities [3]:

$$w_a(F,\xi) = \sqrt{\frac{3\epsilon}{\pi(1-\xi^2)}}w_{st}(F), \quad 1-\xi^2 \gg \epsilon. \quad \text{(B.3)}$$

This formula is also easily obtained directly from (6.1) assuming that in this case ionization takes place at times when the field is close to its maximum

$$f(\tau) = 1 + \frac{1}{2}(1-\xi^2)\tau^2 + \frac{1}{24}(1-\xi^2)(5-9\xi^2)\tau^4 + \ldots,$$

we confine ourselves to the quadratic term in τ. If $1-\xi^2 \lesssim \epsilon \ll 1$ the quadratic approximation cannot be applied and $w_{st}(F/f(t))$ must be averaged over the entire period T. We then have $1/\xi^2 \longrightarrow 1$; assuming everywhere in the integrand (B.2) that $x \approx 1$ where this is permissible and taking into account the value of the integral

$$\int_0^q \frac{e^{-px}dx}{\sqrt{x(q-x)}} = \pi a\left(\frac{1}{2}pq\right), \quad \text{(B.4)}$$

$$a(x) = e^{-x}I_0(x) = \begin{cases} 1 - x + \frac{3}{4}x^2 - \ldots, & x \longrightarrow 0, \\ \frac{1}{\sqrt{2\pi x}}\left(1 + \frac{1}{8x} + \ldots\right), & x \longrightarrow \infty, \end{cases} \quad \text{(B.5)}$$

[11] For example, $\alpha = -1$ for the short-range potential, $\alpha = 1$ for the ground state ($n = 1$), and $\alpha = n - n_1 + n_2 = 2n_2 + |m| + 1$ for excited states (n_1n_2m) of the hydrogen atom.

where $p, q > 0$ and $I_0(x)$ is a modified Bessel function, we obtain

$$\frac{w_a(F,\xi)}{w_{st}(F)} = \frac{1}{\sqrt{\xi^2}}a\left(\frac{1-\xi^2}{6\epsilon\xi^2}\right), \quad \text{(B.6)}$$
$$\gamma \ll 1, \quad \epsilon \ll 1,$$

which is essentially the same as the formula obtained in [3]. If the polarization is not too close to circular, (B.6) yields the simpler formula (B.3).

APPENDIX C

The ionization probabilities (or the widths of the Stark resonances) for various states (n_1n_2m) of the hydrogen atom

$$w_{n_1n_2m}(F) = \Gamma^{(n,n_2,m)}(F) = -2\mathrm{Im}E^{(n_1n_2m)}(F),$$

were obtained as far as $n^4F \sim 1$ using numerical calculations.[12] We shall confine ourselves to the ground state $\kappa = 1$ and use the notation $w_0 \equiv w_{000}(F)$; n_1, n_2, and m are parabolic quantum numbers [16]). In [58] the authors used an analytic approximation (subsequently $\tilde{w}_0$—see formula (25) in [58], which can be conveniently written in the form

$$\tilde{w}_0(F) = 4F^{-1}\exp(-2/3F)\tilde{q}_0(F), \quad \text{(C.1)}$$

$$\tilde{q}_0(F) = 8\pi k[\mathrm{Ai}'^2(k) - k\mathrm{Ai}^2(k)]\exp\left(\frac{4}{3}k^{3/2}\right), \quad \text{(C.2)}$$
$$k = \kappa^2(2F)^{-2/3} = (2\epsilon)^{(-2/3)},$$

having isolated the explicitly semiclassical factor $w_{cl}(F) = 4F^{-1}\exp(-2/3F)$ which is asymptotically exact in the weak field limit [16]. This expression refers to circular polarization and contains no frequency ω so that it should be the same as the ionization probability $w_0 \equiv w_{st}(F)$ in a static electric field. However, we shall consider the expansion of $\tilde{q}_0(F)$ for $F \ll 1$. Using the well-known [29] asymptotic forms for the Airy functions $\mathrm{Ai}(k)$ and $\mathrm{Ai}'(k)$ for $k \longrightarrow \infty$, we obtain from (C.2)

$$\tilde{q}_0(F) = 1 - \frac{17}{12}F + \frac{1225}{288}F^2 - 19.2F^3 + \ldots, \quad \text{(C.3)}$$
$$F \longrightarrow 0,$$

[12] This topic is dealt with extensively in the literature. Various computational methods have been used (the Wehl method [43], numerical solution of the Schrödinger equation after separation of the variables [44], summation of diverging perturbation theory series using the Borel method [45, 47] and using the Pade–Hermite approximant [39, 41], the $1/n$ expansion [40], nonstandard perturbation theory [46], the method of complex rotations [50], and so on) whose results in the range $F < 0.2$ show good agreement. The values of $w_0(F)$ in the range between $F_{cr} \simeq 0.2$ and $F = 1.5$ were calculated in [39–41] and for $F = 1$ our result agrees with that obtained in [45].

whereas the exact expansion has the form [59]

$$q_0(F) = 1 - \frac{107}{12}F + \frac{7363}{288}F^2 - 158.7F^3 + \ldots . \quad \text{(C.4)}$$

Thus, the first coefficients in (C.3) are far from the true ones. It is therefore not surprising that the difference between $w_0(F)$ and $\tilde{w}_0(F)$ becomes appreciable for $F \sim 0.05$ (i.e., in the range of weak fields where the values of $w_0(F)$ are established highly accurately [39–49]) and in the range $F_{cr} \sim 0.2 < F \lesssim 1$ which corresponds to above-barrier ionization, the Krainov–Shokri approximation [58] differs qualitatively from $w_0(F)$ and yields an error of a factor of 3–5.

A similar observation may be made in relation to the formulas for $w_a(F)$ proposed in [11, 33] for the case of linear polarization (see the curves with $\xi = 0$ in Fig. 7).

The reason for this error is evidently the inadequate allowance for Coulomb interaction between the outgoing electron and the atomic nucleus. In this context we note that in the case of a static field F the approximation in which the exact wave function of the final state is replaced by the Volkov function, although giving the exponential factor $\exp(-2/3F)$ (first obtained by the Oppenheimer method [68] for a hydrogen atom) does not yield the correct preexponential factor [23].

REFERENCES

1. L. V. Keldysh, Zh. Éksp. Teor. Fiz. **47**, 1945 (1964) [Sov. Phys. JETP **20**, 1307 (1964)].
2. A. I. Nikishov and V. I. Ritus, Zh. Éksp. Teor. Fiz. **50**, 255 (1966) [Sov. Phys. JETP **23**, 168 (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)].
4. A. M. Perelomov, V. S. Popov, and M. V. Terent'ev, Zh. Éksp. Teor. Fiz. **51**, 309 (1966) [Sov. Phys. JETP **24**, 207 (1966)].
5. A. M. Perelomov and V. S. Popov, Zh. Éksp. Teor. Fiz. **52**, 514 (1967) [Sov. Phys. JETP **25**, 336 (1967)].
6. A. I. Nikishov and V. I. Ritus, Zh. Éksp. Teor. Fiz. **52**, 223 (1967) [Sov. Phys. JETP **25**, 145 (1967)].
7. A. Gold and H. B. Bebb, Phys. Rev. Lett. **14**, 60 (1965); Phys. Rev. **143**, 1 (1966).
8. L. P. Rapoport, B. A. Zon, and N. L. Manakov, *Theory of Multiphoton Processes in Atoms* (Atomizdat, Moscow, 1978).
9. A. Maquet, V. Veniard, and T. A. Marion, J. Phys. B **31**, 3743 (1998).
10. N. B. Delone and V. P. Krainov, *Multiphoton Processes in Atoms* (Springer-Verlag, Berlin, 1994).
11. N. B. Delone and V. P. Kraĭnov, Usp. Fiz. Nauk **168**, 531 (1998) [Phys. Usp. **41**, 469 (1998)].
12. V. S. Popov, Usp. Fiz. Nauk **19**, 819 (1999).
13. V. S. Popov, Pis'ma Zh. Éksp. Teor. Fiz. **70**, 493 (1999) [JETP Lett. **70**, 502 (1999)].
14. V. S. Popov, in *Scientific Session of Moscow Engeneering Physics Institute–2000* (Mosk. Inzh. Fiz. Inst., Moscow, 2000), Vol. 5, p. 155.
15. K. A. Karpov, *Tables of w(z) Function on the Complex Plane* (Vychisl. Tsentr Akad. Nauk SSSR, Moscow, 1954).
16. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory* (Fizmatgiz, Moscow, 1963; Pergamon, New York, 1987).
17. A. A. Radtsig and B. M. Smirnov, *Reference Data on Atoms, Molecules, and Ions* (Énergoatomizdat, Moscow, 1986; Springer-Verlag, Berlin, 1985).
18. D. R. Bates and A. Damgaard, Philos. Trans. R. Soc. London **242**, 101 (1949).
19. I. I. Sobelman, *Atomic Spectra and Radiative Transitions* (Nauka, Moscow, 1977; Springer-Verlag, Berlin, 1979).
20. V. D. Mur, B. M. Karnakov, and V. S. Popov, Zh. Éksp. Teor. Fiz. **115**, 521 (1999) [JETP **88**, 286 (1999)].
21. Yu. N. Demkov and G. F. Drukarev, Zh. Éksp. Teor. Fiz. **47**, 918 (1964) [Sov. Phys. JETP **20**, 614 (1964)].
22. B. M. Smirnov and M. I. Chibisov, Zh. Éksp. Teor. Fiz. **49**, 841 (1965) [Sov. Phys. JETP **22**, 585 (1965)].
23. T. Yamabe, A. Tachibana, and H. J. Silverstone, Phys. Rev. A **16**, 877 (1977).
24. A. V. Gaponov and M. A. Miller, Zh. Éksp. Teor. Fiz. **34**, 242 (1958) [Sov. Phys. JETP **7**, 168 (1958)].
25. H. G. Muller, A. Tip, and M. J. van der Wiel, J. Phys. B **16**, L679 (1983).
26. S. P. Goreslavskiĭ, N. B. Narozhnyĭ, and V. P. Yakovlev, Pis'ma Zh. Éksp. Teor. Fiz. **46**, 173 (1987) [JETP Lett. **46**, 219 (1987)]; S. P. Goreslavsky, N. B. Narozhny, and V. P. Yakovlev, J. Opt. Soc. Am. B **6**, 1752 (1989).
27. J. H. Eberly, J. Javanainen, and K. Rzazewsky, Phys. Rep. **204**, 331 (1991).
28. R. E. Langer, Trans. Am. Math. Soc. **33**, 23 (1931).
29. *Handbook of Mathematical Functions*, Ed. M. Abramowitz and I. A. Stegun (Dover, New York, 1971; Nauka, Moscow, 1979).
30. P. B. Corkum, N. H. Burnett, and F. Brunel, Phys. Rev. Lett. **62**, 1259 (1989).
31. P. B. Corkum, Phys. Rev. Lett. **71**, 1994 (1993).
32. P. B. Corkum and P. Dietrich, Comments At. Mol. Phys. **28**, 357 (1993).
33. V. P. Krainov, J. Opt. Soc. Am. B **14**, 425 (1997); J. Phys. B **32**, 1607 (1999).
34. D. R. Hartree, Proc. Cambridge Philos. Soc. **24**, 89 (1927); **24**, 426 (1927).
35. S. P. Goreslavskiĭ and S. Popruzhenko, Zh. Éksp. Teor. Fiz. **110**, 1200 (1996) [JETP **83**, 661 (1996)].
36. D. Banks and J. G. Leopold, J. Phys. B **11**, 37 (1978).
37. M. B. Kadomtsev and B. M. Smirnov, Zh. Éksp. Teor. Fiz. **80**, 1715 (1981) [Sov. Phys. JETP **53**, 885 (1981)].
38. V. S. Popov, V. D. Mur, and A. V. Sergeev, Phys. Lett. A **193**, 159 (1994); V. S. Popov, V. D. Mur, and A. V. Sergeev, Zh. Éksp. Teor. Fiz. **106**, 1001 (1994) [JETP **79**, 547 (1994)].
39. V. M. Vaĭnberg, V. D. Mur, V. S. Popov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **44**, 9 (1986) [JETP Lett. **44**, 9 (1986)];

Zh. Éksp. Teor. Fiz. **93**, 450 (1987) [Sov. Phys. JETP **66**, 258 (1987)].
40. V. S. Popov, V. D. Mur, A. V. Shcheblykin, *et al.*, Phys. Lett. A **124**, 77 (1987).
41. V. S. Popov, V. D. Mur, A. V. Sergeev, *et al.*, Phys. Lett. A **149**, 418 (1990).
42. V. D. Mur and V. S. Popov, Laser Phys. **3**, 462 (1993).
43. M. Hehenberger, H. V. Intosh, and E. Brändes, Phys. Rev. A **10**, 1494 (1974).
44. R. J. Damburg and V. V. Kolosov, J. Phys. B **9**, 3149 (1976); **11**, 1921 (1978).
45. L. Benassy and V. Grecchi, J. Phys. B **13**, 911 (1980).
46. A. D. Dolgov and A. V. Turbiner, Phys. Lett. A **77**, 15 (1980).
47. V. Franceschini, V. Grecchi, and H. J. Silverstone, Phys. Rev. A **32**, 1338 (1985).
48. F. M. Fernandes, Phys. Rev. A **54**, 1206 (1996).
49. R. V. Kulyagin and V. D. Taranukhin, Kvantovaya Élektron. (Moscow) **23**, 889 (1996); Laser Phys. **7**, 623 (1997).
50. A. Scrinzi, M. Geissler, and T. Brabec, Phys. Rev. Lett. **83**, 706 (1999).
51. L. Benassi, V. Grecchi, E. Harrell, and B. Simon, Phys. Rev. Lett. **42**, 704 (1979); **42**, 1430 (1979).
52. G. I. Barenblatt and Ya. B. Zel'dovich, Usp. Mat. Nauk **26**, 115 (1971).
53. G. I. Barenblatt, *Scaling, Selfsimilarity, and Intermediate Asymptotics* (Cambridge Univ. Press, Cambridge, 1996).
54. V. S. Popov, Phys. Lett. A **173**, 63 (1993).
55. V. S. Popov, in *New Methods in Quantum Theory* (Kluwer, Dordrecht, 1996), p. 149.
56. H. J. Kull, L. Dimon, and J. L. Sanz, Laser Phys. **9**, 48 (1999).
57. V. S. Popov, B. M. Karnakov, and V. D. Mur, Zh. Éksp. Teor. Fiz. **115**, 1642 (1999) [JETP **88**, 902 (1999)].
58. V. P. Kraĭnov and B. Shokri, Zh. Éksp. Teor. Fiz. **107**, 1180 (1995) [JETP **80**, 657 (1995)].
59. H. J. Silverstone, B. G. Adams, J. Cizek, *et al.*, Phys. Rev. Lett. **43**, 1498 (1979).
60. V. S. Popov, V. P. Kuznetsov, and A. M. Perelomov, Zh. Éksp. Teor. Fiz. **53**, 331 (1967) [Sov. Phys. JETP **26**, 222 (1967)].
61. G. G. Paulus, F. Zacher, H. Walther, *et al.*, Phys. Rev. Lett. **80**, 484 (1998).
62. W. Becker, M. Kleber, A. Lohr, *et al.*, Laser Phys. **8**, 56 (1998).
63. J. D. Jackson and J. M. Blatt, Rev. Mod. Phys. **22**, 77 (1950).
64. T.-Y. Wu and T. Ohmura, *Quantum Theory of Scattering* (Prentice Hall, New York, 1962).
65. G. E. Brown and A. D. Jackson, *The Nucleon-Nucleon Interaction* (North-Holland, Amsterdam, 1976).
66. V. D. Mur, B. M. Karnakov, and V. S. Popov, Dokl. Akad. Nauk **365**, 329 (1999) [Dokl. Phys. **44**, 156 (1999)]; Zh. Éksp. Teor. Fiz. **115**, 521 (1999) [JETP **88**, 286 (1999)].
67. S. P. Andreev, B. M. Karnakov, and V. D. Mur, Pis'ma Zh. Éksp. Teor. Fiz. **37**, 155 (1983) [JETP Lett. **37**, 187 (1983)]; S. P. Andreev, B. M. Karnakov, V. D. Mur, and V. A. Polunin, Zh. Éksp. Teor. Fiz. **86**, 866 (1984) [Sov. Phys. JETP **59**, 506 (1984)].
68. J. R. Oppenheimer, Phys. Rev. **31**, 66 (1928).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 1, 2000, pp. 67–75.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 1, 2000, pp. 77–86.

ATOMS, SPECTRA, RADIATION

Elliptic Dark States: Explicit Invariant Form

A. V. Taĭchenachev*, A. M. Tumaĭkin, and V. I. Yudin

Novosibirsk State University, Novosibirsk, 630090 Russia
**e-mail: llf@admin.nsu.ru*

Received February 8, 2000

Abstract—An analysis is made of coherent population trapping as a result of resonant interaction of elliptically polarized light with atoms whose energy levels are degenerate with respect to the projection of the angular momentum and are coupled by a dipole transition. Explicit invariant expressions for dark states are obtained in tensor form for all transitions where population trapping occurs. A correspondence is established between the vector of the elliptic polarization and the pair of associated spinors. It is shown that all dark states can be constructed from these spinors by means of a multiple tensor product. For integer values of the angular momenta of the transitions these constructions reduce to spherical functions of a complex variable. As applications analytic expressions are obtained for the dark magneto-optic and geometric potentials, and the change in their profile with increasing angular momenta is analyzed. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Following its discovery in 1976 [1], the effect of coherent population trapping (CPT) has become firmly established in the arsenal of modern atomic and laser physics. This effect has found important applications in ultrahigh-resolution spectroscopy [2], lasing without inversion [3], laser cooling [4], including dark optical and magneto-optic lattices [5–8], and in atomic optics and interferometry [9].

As we know, CPT is a nonlinear interference effect whereby, under certain conditions, there is a coherent superposition of atomic states which do not interact with the field (so-called dark states or CPT states). For example, if we have a bichromatic field which resonantly couples different energy levels of the ground state (these are usually components of the hyperfine structure) with a single common excited level (Λ-system), under conditions of two-photon resonance the dark state will be a coherent superposition of the wave functions of the two lower energy levels. Another important example is the presence of dark states as a result of the Zeeman degeneracy of the ground energy level. For example, it was shown in [10] that for transitions of the type $J_g = J \longrightarrow J_e = J - 1$ and $J_g = J' \longrightarrow J_e = J'$ (where J' is an integer, and J_g and J_e are the angular momenta of the ground and excited energy levels, respectively) under conditions of resonant interaction with elliptically polarized radiation there are always dark states comprising a coherent superposition of Zeeman wave functions of the ground level g. These states only depend on the elliptic polarization vector and do not depend on the detuning or the light intensity. Consequently the CPT effect for these transitions can be observed not only in a monochromatic field but also in fields having an arbitrary spectral composition (obviously when the resonance condition is satisfied). Recent experiments [11] directly confirmed the basic assumptions of [10], i.e. that dark states exist for arbitrary elliptic polarization and they do not depend on the intensity and frequency of the light.

However, it should be noted that for a detailed analysis of various phenomena it may be difficult to use the results of [10] directly. This is because the explicit form of the CPT states was determined in [10] for a particular choice of coordinates where the quantization axis is orthogonal to the plane of the polarization ellipse. Thus, in the general case of fields with a polarization gradient (for example, laser cooling by CPT) it is necessary to introduce a local coordinate system for which the direction of the quantization axis changes from one point to another depending on the orientation of the local polarization ellipse and only then can the form of the spatially nonuniform dark states be determined in the laboratory coordinates by rotating the basis (the angles of rotation are local). Quite clearly this procedure is laborious, particularly for transitions having a high angular momentum J_g.

In the present study we find the explicit invariant form of all the dark states which appear as a result of resonant interaction of elliptically polarized light with the transitions $J_g = J \longrightarrow J_e = J - 1$ and $J_g = J' \longrightarrow J_e = J'$, where J' is an integer (following [11], we shall frequently use the term elliptic dark states). In addition, we also demonstrate the existence of two associated spinors for any elliptic vector which contain information on the polarization of the light. In this case, all the dark states are expressed in terms of these spinors using a multiple tensor product. For the case of integer J_g these constructions are reduced to a form similar to the invariant form of spherical functions.

As applications of these results for all the transitions specified above we determine analytic expressions for

the dark magneto-optic potentials formed when a static magnetic field is applied in cases where the condition $\Omega \ll V$ is satisfied ($\hbar\Omega$ is the Zeeman splitting in the ground state and V is the Rabi frequency) [8]. This problem is important for the theory of laser cooling and trapping of atoms, in particular for the theory of optical gratings. Taking as an example the optical field configuration *lin* ⊥ *lin* we analyze the profile of the magneto-optic potential and observe some interesting characteristics as the angular momentum J_g increases. For example, for $J_g = J \longrightarrow J_e = J - 1$ transitions which have two potentials, the curvature of these potentials at the minimum increases with increasing J_g whereas for $J_g = J' \longrightarrow J_e = J'$ transitions (J' is an integer) the profile of the magneto-optic potential becomes close to stepped as J_g increases.

In addition, for $J_g = J' \longrightarrow J_e = J'$ transitions (J' is an integer) analytic expressions are obtained for the dark geometric potentials (the vector potential and the scalar potential [12–14]).

2. CONDITION FOR COHERENT POPULATION TRAPPING IN INVARIANT TENSOR FORM

We shall consider a $J_g \longrightarrow J_e$ optical transition in resonance with the field

$$\mathbf{E}(t) = \mathbf{e}f(t) + \text{c.c.}, \quad (\mathbf{e}\cdot\mathbf{e}^*) = 1,$$
$$\mathbf{e} = \sum_{q=0,\pm1} (-1)^{-q} e_{-q}\mathbf{e}_q, \tag{1}$$

where $\mathbf{e}$ is the unit complex vector of the elliptic polarization, and e_q are its covariant components in the spherical basis $\{\mathbf{e}_0 = \mathbf{e}_z, \mathbf{e}_{\pm1} = \mp(\mathbf{e}_x \pm i\mathbf{e}_y)/\sqrt{2}\}$; $f(t)$ is a certain function of time (for example, $\exp\{-i\omega t\}$ which determines the resonant nature of the interaction.

As we know from [10], the elliptic dark states $|\Psi^{(NC)}\rangle$ are obtained from the solution of the operator equation:

$$(\hat{\mathbf{d}}\cdot\mathbf{e})|\Psi^{(NC)}\rangle = 0. \tag{2}$$

Since any dark state lies entirely on the lower energy level g, it can be considered to be a tensor of rank J_g which we shall subsequently denote by $\Psi^{(NC)}_{J_g}$. The covariant components $\Psi^{(NC)}_{J_g\mu_g}$ determine the expansion of $|\Psi^{(NC)}\rangle$ in terms of the Zeeman wave functions $|J_g, \mu_g\rangle$ of the ground state:

$$|\Psi^{(NC)}\rangle = \sum_{\mu_g} (-1)^{-\mu_g}\Psi^{(NC)}_{J_g-\mu_g}|J_g, \mu_g\rangle. \tag{3}$$

Then, using the Wigner–Eckart theorem, we write the expression for the matrix element:

$$\langle J_e, \mu_e|(\hat{\mathbf{d}}\cdot\mathbf{e})|\Psi^{(NC)}\rangle = \frac{\langle J_e\|\hat{d}\|J_g\rangle}{\sqrt{2J_e+1}}$$
$$\times \sum_{q,\mu_g} C^{J_e\mu_e}_{J_g\mu_g;\,1q}(-1)^{-q-\mu_g} e_{-q}\Psi^{(NC)}_{J_g-\mu_g} \tag{4}$$
$$= \frac{\langle J_e\|\hat{d}\|J_g\rangle}{\sqrt{2J_e+1}}(-1)^{-\mu_e}\{\mathbf{e}\otimes\Psi^{(NC)}_{J_g}\}_{J_e-\mu_e},$$

where $|J_e, \mu_e\rangle$ is the wave function of the magnetic sublevel of the excited state. Here we used the following properties of the Clebsch–Gordan coefficients: $\mu_e = q + \mu_g$ and $C^{J_e\mu_e}_{J_g\mu_g;\,1q} = C^{J_e-\mu_e}_{1-q;\,J_g-\mu_g}$, along with the standard definition (see, for example, [15]) of the irreducible tensor product $\{\ldots\otimes\ldots\}$; $\langle J_e\|\hat{d}\|J_g\rangle$ is the reduced matrix element of the dipole moment.

On the basis of (4), equation (2) can now be rewritten in invariant form as a tensor product of rank J_e,

$$\{\mathbf{e}\otimes\Psi^{(NC)}_{J_g}\}_{J_g} = 0, \tag{5}$$

with the normalization condition

$$(\Psi^{(NC)*}_{J_g}\cdot\Psi^{(NC)}_{J_g}) = 1.$$

Equation (5) is the basis for the following analysis. Here and subsequently we shall frequently omit the component indices for conciseness.

3. $J_g = J \longrightarrow J_e = J$ TRANSITIONS (INTEGER J)

For these transitions equation (5) has the form

$$\{\mathbf{e}\otimes\Psi^{(NC)}_{J}\}_{J} = 0. \tag{6}$$

In order to obtain the explicit invariant form $\Psi^{(NC)}_J$ we need to introduce the following tensor constructions constructed only from the vector $\mathbf{e}$ [16]:

$$\{\mathbf{e}\}_L = \{\ldots\{\{\mathbf{e}\otimes\mathbf{e}\}_2\otimes\mathbf{e}\}_3\ldots\otimes\mathbf{e}\}_L. \tag{7}$$

For the case of real $\mathbf{e}$ these constructions determine the invariant form of the notation for the spherical functions $Y_{Lq}(\mathbf{e})$ (see [15]):

$$Y_{Lq}(\mathbf{e}) = \sqrt{\frac{(2L+1)!!}{4\pi L!(\mathbf{e}\cdot\mathbf{e})^L}}\{\mathbf{e}\}_{Lq}. \tag{8}$$

However, it can be shown [16–18] that the main group properties of spherical functions will also be satisfied if $\mathbf{e}$ in expression (8) is an arbitrary complex vector. Then, for example, from the well-known Clebsch–Gordan expansion for the product of the spherical functions we can obtain $\{\{\mathbf{e}\}_m\otimes\{\mathbf{e}\}_n\}_l \sim C^{l0}_{m0;\,n0}\{\mathbf{e}\}_l$. It follows from the symmetry properties of the Clebsch–Gordan coefficients that $C^{l0}_{m0;\,n0} = 0$ if $m + n - l$ is an odd number. Then, setting $m = 1$, $n = l = J$ (i.e., $m + n - l = 1$) and bearing in mind that $\{\mathbf{e}\}_1 = \mathbf{e}$, we have

$$\{\mathbf{e}\otimes\{\mathbf{e}\}_J\}_J = 0. \tag{9}$$

Comparing this relationship with equation (6), we can easily find

$$\Psi_J^{(NC)} = A\{\mathbf{e}\}_J, \tag{10}$$

where the normalization constant A can be determined from the equality [16]

$$(\{\mathbf{a}\}_J \cdot \{\mathbf{b}\}_J) = \frac{J!}{(2J-1)!!} \times\left(\sqrt{(\mathbf{a}\cdot\mathbf{a})(\mathbf{b}\cdot\mathbf{b})}\right)^J P_J\left(\frac{(\mathbf{a}\cdot\mathbf{b})}{\sqrt{(\mathbf{a}\cdot\mathbf{a})(\mathbf{b}\cdot\mathbf{b})}}\right), \tag{11}$$

which is derived from the summation theorem for spherical harmonics for arbitrary **a** and **b** [17]. Here $P_L(x)$ is a Legendre polynomial. Assuming in (11) **a** = **e**, **b** = **e***, we obtain the expression for A:

$$A = \left[\frac{J!}{(2J-1)!!}|\mathbf{e}\cdot\mathbf{e}|^J P_J\left(\frac{1}{|\mathbf{e}\cdot\mathbf{e}|}\right)\right]^{-1/2}. \tag{12}$$

As was shown in [10], the solution (10) is unique.

It can be seen that the explicit invariant form (10) of the elliptic dark states for $J_g = J \longrightarrow J_e = J$ (J is an integer) transitions is simple and convenient to analyze, especially as the main symmetry and algebraic properties are similar to the properties of the spherical functions $Y_{Lq}(\mathbf{e})$. We also note that for the $J_g = 1 \longrightarrow J_e = 1$ transition (10) automatically yields the well-known result [19, 20] $\Psi_1^{(NC)} = \mathbf{e}$ which corresponds to the expression

$$|\Psi^{(NC)}\rangle = \sum_{q=0,\pm1} (-1)^{-q} e_{-q}|J_g, \mu_g = q\rangle.$$

4. $J_g = J \longrightarrow J_e = J - 1$ TRANSITIONS

In this case, the CPT condition (5) has the form

$$\{\mathbf{e}\otimes\Psi_J^{(NC)}\}_{J-1} = 0. \tag{13}$$

It was shown in [10] that for these transitions the dark states form a two-dimensional linear subspace. In order to express their general invariant form, we need to introduce the coordinates defined in [5] which are closely related to the polarization ellipse. We know that any ellipse is a cross section of a certain cylinder on a plane so that each vector of the elliptic polarization **e** can be made to correspond to a cylinder (generally two) whose cross section is the given ellipse. Then, if the z-axis is directed along the axis of one of these cylinders (Fig. 1), it follows from [5] that the elliptic polarization **e** is a superposition of a linear and only one circular component. In particular, if the y axis is directed along the minor semiaxis of the polarization ellipse, the vector **e** can always be expressed as

$$\mathbf{e} = \pm\sqrt{\cos(2\varepsilon)}\mathbf{e}_0 - \sqrt{2}\sin\varepsilon\mathbf{e}_{+1}, \tag{14}$$

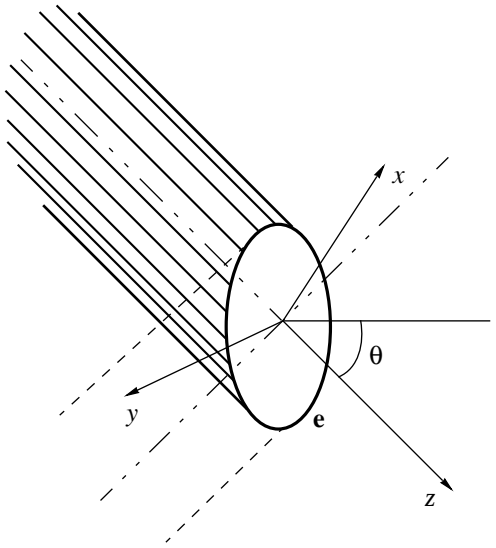


Fig. 1. System of coordinates related to the polarization ellipse ("natural" basis) in which the quantization axis z is directed along the axis of a cylinder constructed along the polarization ellipse (the dashed lines show the profile of the second cylinder).

where ε is the ellipticity parameter ($|\tan\varepsilon|$ is the ratio of the axes of the polarization ellipse) and the $\pm$ signs refer to the two different cylinders in Fig. 1. Here we impose no constraints on ε.

In coordinates related to one of the cylinders, the expansion (14) defines the selection rules for light-induced transitions in terms of the projection of the momentum $\mu_e - \mu_g = 0, 1$ which clearly reveals the existence of a dark state coinciding with the outer Zeeman sublevel $|J_g, \mu_g = J\rangle$. Another linearly independent dark state is determined similarly in coordinates related to the second cylinder. Quite clearly, in cases of integer J the invariant form of these CPT states can be written using the tensor construction (7) introduced above:

$$\Psi_J^{(m)} = \{\mathbf{c}^{(m)}\}_J, \quad m = 1, 2, \tag{15}$$

where each vector $\mathbf{c}^{(m)}$ apart from the phase factor is the same as the circular vector $\mathbf{e}_{+1}$ in the corresponding coordinates related to the mth cylinder. The explicit form of the dependence of $\mathbf{c}^{(m)}$ on **e** is determined from the invariant conditions:

$$(\mathbf{c}^{(m)}\cdot\mathbf{c}^{(m)}) = 0, \quad (\mathbf{c}^{(m)}\cdot\mathbf{e}) = 0, \quad (\mathbf{c}^{(m)*}\cdot\mathbf{c}^{(m)}) = 1. \tag{16}$$

This system has two solutions:

$$\mathbf{c}^{(1,2)} = \frac{\pm i\sqrt{(\mathbf{e}\cdot\mathbf{e})}[\mathbf{e}^*\times\mathbf{e}] + [\mathbf{e}\times[\mathbf{e}^*\times\mathbf{e}]]}{(1+|\mathbf{e}\cdot\mathbf{e}|)\sqrt{1-|\mathbf{e}\cdot\mathbf{e}|}}. \tag{17}$$

The dark states $\Psi_J^{(1)}$ and $\Psi_J^{(2)}$ are generally linearly independent but not orthogonal. From (11) we find the scalar products:

$$(\Psi_J^{(1)*} \cdot \Psi_J^{(2)}) = (\mathbf{c}^{(1)*} \cdot \mathbf{c}^{(2)})^J = \left(\frac{1 - |\mathbf{e} \cdot \mathbf{e}|}{1 + |\mathbf{e} \cdot \mathbf{e}|}\right)^J, \quad (\Psi_J^{(m)*} \cdot \Psi_J^{(m)}) = 1. \tag{18}$$

Since the product $(\Psi_J^{(1)*} \cdot \Psi_J^{(2)})$ is real, we have the following pair of dark orthonormalized tensors:

$$\Psi_J^{(\pm)} = \frac{\Psi_J^{(1)} \pm \Psi_J^{(2)}}{\sqrt{2(1 \pm (\Psi_J^{(1)*} \cdot \Psi_J^{(2)}))}} e^{i\beta_\pm}, \quad (\Psi_J^{(\pm)*} \cdot \Psi_J^{(\pm)}) = 1, \quad (\Psi_J^{(+)*} \cdot \Psi_J^{(-)}) = 0, \tag{19}$$

where $\exp\{i\beta_\pm\}$ are arbitrary phase factors. It should be noted that the dark states (19) are related to the results obtained earlier [10]. We recall that in [10] the dark states were determined in coordinates in which the quantization axis is orthogonal to the polarization ellipse. In this basis the system of light-induced $J_g = J \longrightarrow J_e = J - 1$ transitions has the form of two Λ-chains: the first begins with the extreme left sublevel $|J_g, \mu_g = -J\rangle$ and the second with the penultimate left sublevel $|J_g, \mu_g = -J + 1\rangle$. It was shown in [10] that each of these Λ-chains corresponds to a single dark state. The dark states $|\Psi^{(\pm)}\rangle$ determined using formulas (3) and (19) in the present study are the same as those from [10]. Here the $|\Psi^{(+)}\rangle$ state is associated with the first Λ-chain and the $|\Psi^{(-)}\rangle$ state is associated with the second Λ-chain.

We note that the states (19) are not the only possible choice of orthogonal basis since any unitary transformation in the two-dimensional subspace of the dark states converts $\Psi_J^{(\pm)}$ to another pair of orthonormalized tensors.

For half-integer J the invariant formula (15) for the dark states $\Psi_J^{(m)}$ cannot be applied because of the half-integer rank. However, there is a more general procedure for constructing invariant formulas for CPT states based on the correspondence between the circular vectors and the spinors. This correspondence is established as follows: for an arbitrary circular vector $\mathbf{c}$ (i.e., $(\mathbf{c} \cdot \mathbf{c}) = 0$) there always exists a spinor ξ (a tensor of rank 1/2) such that

$$\{\xi \otimes \xi\}_1 = \mathbf{c}.$$

Since this condition is quadratic, the spinor ξ is determined but not its sign. The inverse correspondence is single-valued: for an arbitrary spinor ξ the tensor product $\{\xi \otimes \xi\}_1$ is a circular vector, i.e.,

$$(\{\xi \otimes \xi\}_1 \cdot (\xi \otimes \xi)_1) = 0.$$

Thus, having determined the two spinors $\xi^{(1,2)}$ using the formula

$$\{\xi^{(m)} \otimes \xi^{(m)}\}_1 = \mathbf{c}^{(m)}, \quad m = 1, 2, \tag{20}$$

the dark states $\Psi_J^{(m)}$ for arbitrary (integer and half-integer) J may be expressed in the form

$$\Psi_J^{(m)} = \{\xi^{(m)}\}_J, \tag{21}$$

where the tensor $\{\xi\}_J$ is constructed of $2J$ spinors ξ similar to the construction (7):

$$\{\xi\}_J = \{\ldots\{\{\xi \otimes \xi\}_1 \otimes \xi\}_{3/2} \ldots \otimes \xi\}_J. \tag{22}$$

The general procedure for orthogonalizing the dark states is exactly the same as formulas (18) and (19), i.e., in these formulas J can have integer and half-integer values. The main difference is that for half-integer J the scalar product $(\Psi_J^{(1)*} \cdot \Psi_J^{(2)})$, although real, can have both positive and negative values depending on the choice of signs when determining $\xi^{(m)}$ from the quadratic expressions (20). However, the modulus of the scalar product will be determined by the right-hand side of (18) as before.

It should be noted that for half-integer J the $|\Psi^{(\pm)}\rangle$ states determined in accordance with (3) and (19), also correspond to two different Λ-chains from [10] when the quantization axis is orthogonal to the polarization ellipse. However, this relationship is nonunique and depends on the choice of positive direction of the z axis. This is because when the direction of the quantization axis is reversed for integer J these systems of coupled sublevels remain the same after conversion, i.e., the Λ-chain beginning from the left sublevel $|J_g, \mu_g = -J\rangle$ ($|J_g, \mu_g = -J + 1\rangle$) in the initial basis will also correspond to the Λ-chain beginning with $|J_g, \mu_g = -J\rangle$ ($|J_g, \mu_g = -J + 1\rangle$) in the new basis. However, for half-integer J when the positive direction of the quantization axis is changed, the Λ-chains convert one to the other, i.e., the Λ-chain which initially began with the extreme left sublevel $|J_g, \mu_g = -J\rangle$ will correspond to the Λ-chain beginning from the penultimate left sublevel $|J_g, \mu_g = -J + 1\rangle$ in the new basis, and conversely.

5. $J_g = J \longrightarrow J_e = J$ TRANSITIONS (HALF-INTEGER J)

To complete the picture it makes sense to consider this case. It was shown in [10] that for $J_g = J \longrightarrow J_e = J$ transitions (where J is a half-integer) a dark state only exists for a circularly polarized field, i.e., when $(\mathbf{e} \cdot \mathbf{e}) = 0$. For this polarization we have $\mathbf{c}_1 = \mathbf{c}_2 = -\mathbf{e}$ and consequently only one spinor $\xi = \xi^{(1)} = \xi^{(2)}$ which is defined as $\{\xi \otimes \xi\}_1 = -\mathbf{e}$. The dark state then has the invariant form: $\Psi_J^{(NC)} = \{\xi\}_J$. For dipole transitions of the type $J_g = J \longrightarrow J_e = J + 1$ no CPT effect occurs (see [10]).

6. ASSOCIATED SPINORS

It has been shown that by using the invariant formulas (17) and (20) each unit complex vector **e** can be associated with a pair of unit spinors $\xi^{(1,2)}$ which we shall call associated spinors. An important property is directly proven:

$$\{\xi^{(1)} \otimes \xi^{(2)}\}_1 = \pm\frac{\mathbf{e}}{\sqrt{1+|\mathbf{e}\cdot\mathbf{e}|}}, \tag{23}$$

i.e., the vector **e** is constructed of spinors $\xi^{(m)}$. The indeterminacy in the signs (23) occurs as a result of the quadratic nature of the formula (20) in terms of $\xi^{(m)}$. However, having fixed on one sign (for example, minus) in (23), we thereby achieve a rigid coupling between $\xi^{(1)}$ and $\xi^{(2)}$ when these are determined.

We can therefore talk of a correspondence between the complex vector space and the set of all possible pairs of spinors. This correspondence is achieved using formulas (17), (20), and (23) and has the following geometric meaning: each complex vector **e** corresponds to a pair of cylinders (Fig. 1); each cylinder has a circular vector (17) associated with it; according to formula (20) these circular vectors $\mathbf{c}^{(m)}$ (m = 1, 2) correspond to the associated spinors $\xi^{(m)}$, and the vector **e** is constructed of associated spinors according to formula (23).

Determining the explicit form of this correspondence is not only an important mathematical result but can also lead to new group-theoretical and geometric approaches to the analysis of polarization effects. For example, the CPT state (10) for $J_g = J \longrightarrow J_e = J$ transitions (where J is an integer) may be expressed in a different form:

$$\Psi_J^{(NC)} = \left[\frac{J!}{(2J-1)!!}\frac{|\mathbf{e}\cdot\mathbf{e}|^J}{(1+|\mathbf{e}\mathbf{e}|)^J}P_J\left(\frac{1}{|\mathbf{e}\cdot\mathbf{e}|}\right)\right]^{-1/2} \times\{\{\xi^{(1)}\}_{J/2}\otimes\{\xi^{(2)}\}_{J/2}\}_J. \tag{24}$$

This is easily demonstrated by systematically changing the coupling scheme (see [15]) of the spinors in expression (24) so that they are grouped in pairs $\{\xi^{(1)}\otimes\xi^{(2)}\}_1$. Then using (23) we reconstruct the invariant form (10) for $\Psi_J^{(NC)}$ apart from the phase factor. Hence, the elliptic dark states for all types of transitions can be explicitly expressed in terms of spinors $\xi^{(m)}$ using the constructions (22).

7. DARK GEOMETRIC AND MAGNETO-OPTIC POTENTIALS

It follows from formulas (10), (20), and (21) that in fields having spatially inhomogeneous polarization **e**(**r**) the dark states also acquire a coordinate dependence. As an application of these results, we consider the potentials acting on an atom being in these states $|\Psi^{(NC)}(\mathbf{r})\rangle$.

There are two types of potentials associated with the spatial inhomogeneity of the dark states. The first are geometric (or gauge) potentials [12–14] caused by the translational motion of the atoms. We recall that under certain conditions there is an overwhelming probability of an atom being in a CPT state. In this case, its dynamics on translational degrees of freedom are described by the effective Hamiltonian:

$$\hat{H}^{(\mathrm{eff})} = \frac{(\hat{\mathbf{p}}+\mathbf{A})^2}{2M} + W(\mathbf{r}). \tag{25}$$

Here **A** is the geometric vector potential whose Cartesian coordinates are defined as

$$A_j = -i\hbar\left\langle\Psi^{(NC)}(\mathbf{r})\middle|\frac{\partial}{\partial x_j}\Psi^{(NC)}(\mathbf{r})\right\rangle, \tag{26}$$

and $W(\mathbf{r})$ is the geometric scalar potential:

$$W(\mathbf{r}) = \frac{\hbar^2}{2M} \times\sum_j\left\langle\frac{\partial}{\partial x_j}\Psi^{(NC)}(\mathbf{r})\middle|\frac{\partial}{\partial x_j}\Psi^{(NC)}(\mathbf{r})\right\rangle - \frac{\mathbf{A}^2}{2M}. \tag{27}$$

Another type of potential occurs when static fields are applied and is caused by spatial modulation of the multipole moments of the atom in the dark state. For example, in a static magnetic field **B**, as a result of the inhomogeneity of the magnetic moment a dark magneto-optic potential is formed [8]:

$$U(\mathbf{r}) = -\langle\Psi^{(NC)}(\mathbf{r})|(\hat{\mu}\mathbf{B})|\Psi^{(NC)}(\mathbf{r})\rangle, \tag{28}$$

where $\hat{\mu}$ is the magnetic moment operator of the atom.

The results of the previous sections can be used to obtain explicit analytic expressions for the dark potentials (26), (27), and (28) for an arbitrary spatial dependence **e**(**r**). Without limiting the generality (see, for example, [21]) we shall subsequently assume everywhere that

$$\mathrm{Im}(\mathbf{e}(\mathbf{r})\cdot\mathbf{e}(\mathbf{r})) = 0. \tag{29}$$

For a spatially inhomogeneous monochromatic field having the general form

$$\mathbf{E}(\mathbf{r},t) = \mathbf{E}(\mathbf{r})e^{-i\omega t} + \text{c.c.} \tag{30}$$

the polarization unit vector **e**(**r**) satisfying (29) is invariantly determined as follows:

$$\mathbf{e}(\mathbf{r}) = \left(\frac{(\mathbf{E}(\mathbf{r})\cdot\mathbf{E}(\mathbf{r}))^*}{(\mathbf{E}(\mathbf{r})\cdot\mathbf{E}(\mathbf{r}))}\right)^{1/4}\frac{\mathbf{E}(\mathbf{r})}{|\mathbf{E}(\mathbf{r})|}. \tag{31}$$

We can then introduce the real ellipticity parameter $\varepsilon(\mathbf{r})$ which is an analytic function of the coordinates and satisfies the condition

$$\cos(2\varepsilon) = (\mathbf{e}(\mathbf{r})\cdot\mathbf{e}(\mathbf{r})). \tag{32}$$

7.1. Dark Magneto-Optic Potentials

From procedural concepts we begin by determining the magneto-optic potentials (28). Below we consider each type of transition separately.

7.1.1. $J_g = J \longrightarrow J_e = J$ transitions (integer J). For these transitions, substituting the expression for $|\Psi^{(NC)}(\mathbf{r})\rangle$ from (10) and (28) and using the Wigner–Eckart theorem, we have

$$U(\mathbf{r}) = A^2 \frac{\langle J_g \| \hat{\mu} \| J_g \rangle |\mathbf{B}|}{\sqrt{2J+1}} (\{\mathbf{e}^*\}_J \cdot \{\{\mathbf{e}\}_J \otimes \mathbf{n}\}_J). \quad (33)$$

Here $\langle J_g \| \hat{\mu} \| J_g \rangle$ is the reduced matrix element of the magnetic moment, $\mathbf{n} = \mathbf{B}/|\mathbf{B}|$ is the unit vector in the direction of the magnetic field. In accordance with [16], we can write

$$\{\{\mathbf{e}\}_J \otimes \mathbf{n}\}_J = -\frac{i}{\sqrt{J(J+1)}} (\mathbf{n} \cdot [\mathbf{e} \times \nabla_{\mathbf{e}}]) \{\mathbf{e}\}_J, \quad (34)$$

where $\nabla_{\mathbf{e}}$ denotes the gradient along the vector $\mathbf{e}$. Since the conjugate vector $\mathbf{e}^*$ is linearly independent in relation to $\mathbf{e}$, any function of $\mathbf{e}^*$ can be introduced after $\nabla_{\mathbf{e}}$. At this point we need to bear in mind that $\nabla_{\mathbf{e}}(\mathbf{e}^* \cdot \mathbf{e}) = \mathbf{e}^*$ and use the normalization condition $((\mathbf{e}^* \cdot \mathbf{e}) = 1$ after calculating the derivative. Then, applying formula (11) and expression (12) for the normalization constant A, we obtain from (33) and (34) the final expression for the potential:

$$U(\mathbf{r}) = \hbar\Omega \frac{i(\mathbf{n} \cdot [\mathbf{e} \times \mathbf{e}^*])}{\cos(2\varepsilon)} \frac{d}{dw} \ln P_J(w) \Big|_{w = 1/\cos(2\varepsilon)}. \quad (35)$$

We note that in formula (35) the direction of the magnetic field $\mathbf{n}$ and the Zeeman splitting $\hbar\Omega$ (i.e., the magnetic field $|\mathbf{B}|$) may also depend on the coordinates.

For the $J_g = 1 \longrightarrow J_e = 1$ transition the potential (35) reduces to the well-known result [22]:

$$U(\mathbf{r}) = \hbar\Omega i(\mathbf{n} \cdot [\mathbf{e} \times \mathbf{e}^*]). \quad (36)$$

7.1.2. $J_g = J \longrightarrow J_e = J - 1$ transitions. As we have already noted, for these transitions the dark states form a two-dimensional subspace. Consequently, the dark magneto-optic potentials are determined by the 2×2 matrix $\hat{U}$ whose matrix elements in the basis (19) have the form

$$U_{\pm} = -\langle \Psi^{(\pm)} | (\hat{\mu}\mathbf{B}) | \Psi^{(\pm)} \rangle,$$
$$U_{+-} = U_{-+}^* = -\langle \Psi^{(+)} | (\hat{\mu}\mathbf{B}) | \Psi^{(-)} \rangle. \quad (37)$$

Note that in the general case of spatially inhomogeneous polarization $\mathbf{e}(\mathbf{r})$ the dark states (19) are analytic functions of the coordinates with a suitable choice of spatial dependence of the phase factors $\exp\{i\beta_{\pm}(\mathbf{r})\}$. For these analytic states calculations yield the following result:

$$U_{\pm} = \hbar\Omega J \frac{i(\mathbf{n} \cdot [\mathbf{e} \times \mathbf{e}^*])}{2} \frac{(\cos\varepsilon)^{2J-2} \pm (\sin\varepsilon)^{2J-2}}{(\cos\varepsilon)^{2J} \pm (\sin\varepsilon)^{2J}}, \quad (38)$$

$$U_{+-} = U_{-+}^* = \frac{\hbar\Omega J}{2} [\{(\mathbf{n} \cdot \mathbf{e}) + (\mathbf{n} \cdot \mathbf{e}^*)\}(\cos\varepsilon)^{2J-2}$$
$$- \{(\mathbf{n} \cdot \mathbf{e}) - (\mathbf{n} \cdot \mathbf{e}^*)\}(\sin\varepsilon)^{2J-2}] \quad (39)$$
$$\times \left[\sum_{q=0}^{2J-1} (\sin^2\varepsilon)^q (\cos^2\varepsilon)^{2J-1-q} \right]^{-1/2}.$$

The eigenvalues of the matrix $\hat{U}$ determine the adiabatic potentials. In general, the motion of an atom is bipotential similar to the motion of a two-level atom in the field of a strong standing wave [23].

It follows from (39) that the off-diagonal elements U_{+-} and U_{-+} are zero if the direction of the magnetic field $\mathbf{n}$ is orthogonal to the polarization vector $\mathbf{e}$. In this case, comparing (35) and (38), we can see that the potentials for the $J \longrightarrow J$ (integer J) and $J_g = J \longrightarrow J_e = J - 1$ transitions have the same vector component $(\mathbf{n} \cdot [\mathbf{e} \times \mathbf{e}^*])$ but different scalar components which only depend on the local value of the parameter ε(see (32)).

7.1.3. Magneto-optic potentials in an optical field *lin* ⊥ *lin*. We shall analyze the change in the profile of the dark magneto-optic potential with increasing angular momentum J for the well-known *lin* ⊥ *lin* field configuration.

This configuration is a superposition of two counterpropagating (along the z axis) traveling waves of the same frequency and amplitude, having mutually orthogonal (along $\mathbf{e}_x$ and $\mathbf{e}_y$) linear polarizations. In this case we direct the magnetic field $\mathbf{B}$ along the z axis, i.e., $\mathbf{n} = \mathbf{e}_z$.

In this configuration the spatially inhomogeneous polarization vector (31) is defined as

$$\mathbf{e}(z) = \frac{\mathbf{e}_x e^{ikz} + \mathbf{e}_y e^{-ikz}}{\sqrt{2}}. \quad (40)$$

We can easily find

$$(\mathbf{e} \cdot \mathbf{e}) = \cos(2kz), \quad [\mathbf{e} \times \mathbf{e}^*] = i\sin(2kz)\mathbf{e}_z$$

from which we determine the analytic parameter in accordance with (32)

$$\varepsilon = kz.$$

For $J_g = J \longrightarrow J_e = J$ transitions (integer J) we then have from (35)

$$U = -\hbar\Omega \tan(2kz) \frac{d}{dw} \ln P_J(w) \Big|_{w = 1/\cos(2kz)}. \quad (41)$$

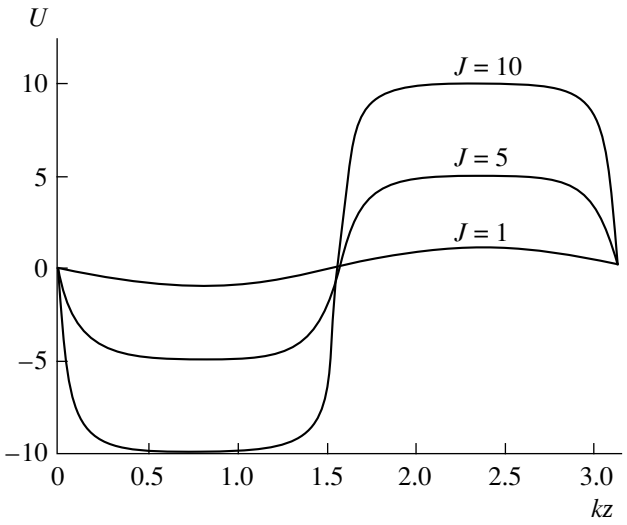


Fig. 2. Dark magneto-optic potentials (in units of $\hbar\Omega$) for $J_g = J \longrightarrow J_e = J$ transitions (J is an integer) for the *lin* ⊥ *lin* field configuration.

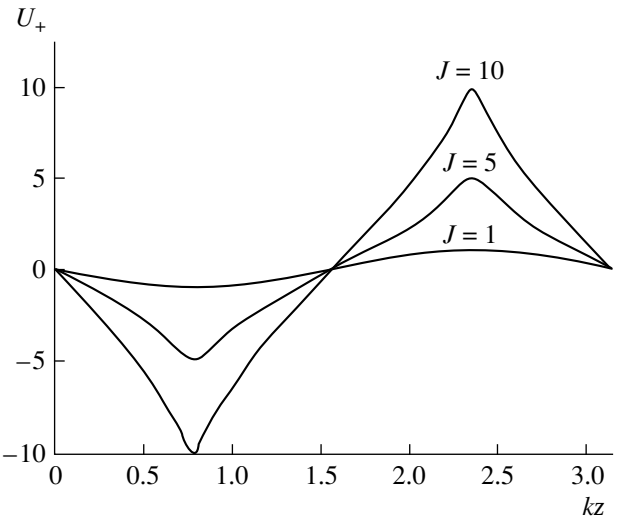


Fig. 3. Dark magneto-optic potentials U_+ (in units of $\hbar\Omega$) for $J_g = J \longrightarrow J_e = J - 1$ transitions (J is an integer) for the *lin* ⊥ *lin* field configuration.

It can be seen from Fig. 2 that as J increases, the potential profile becomes close to stepped, as was noted in [6] for the particular case $J = 4$.

For $J_g = J \longrightarrow J_e = J - 1$ transitions in this particular case of an *lin* ⊥ *lin* optical configuration and a magnetic field directed along the z axis, the off-diagonal elements U_{+-} and U_{-+} (39) are zero (since $\mathbf{e} \perp \mathbf{n}$) and the expressions for the diagonal elements (38) reduce to:

$$U_\pm = -\hbar\Omega J \frac{\sin(2kz)}{2} \frac{(\cos kz)^{2J-2} \pm (\sin kz)^{2J-2}}{(\cos kz)^{2J} \pm (\sin kz)^{2J}}. \quad (42)$$

Figure 3 shows the spatial dependence of the potential U_+ for several values of J (the behavior of U_- is basically similar). It can be seen that as J increases, the curvature of the potentials at the minimum increases. It is easy to show that the second derivative increases proportionately as $2J(J-1)$ for U_+ and $4J^2/3 - 2J + 2/3$ for U_-. Consequently, strong spatial localization of trapped atoms can be obtained for high values of J.

7.2. Dark Geometric Potentials

The method of calculating the geometric potentials for $J_g = J \longrightarrow J_e = J - 1$transitions (where J is an integer) is essentially similar to the method of determining the dark magneto-optic potentials and reduces to differentiating expressions (12) and (11) with respect to the spatial coordinates $\{x_j\}$ where we need to set $\mathbf{a} = \mathbf{e}^*$, $\mathbf{b} = \mathbf{e}$ and impose the normalization condition $(\mathbf{e} \cdot \mathbf{e}^*) = 1$ only after the differentiation operation. Omitting cumbersome intermediate procedures, we give the final results. For the Cartesian components A_j of the vector potential (26) we have the expression

$$A_j = -\hbar \frac{i\left(\mathbf{e}^* \cdot \frac{\partial}{\partial x_j}\mathbf{e}\right)}{\cos(2\varepsilon)} \frac{d}{dw} \ln P_J(w)\Big|_{w = 1/\cos(2\varepsilon)} \quad (43)$$
$$(j = 1, 2, 3).$$

The scalar potential (27) is determined as follows:

$$W = \frac{\hbar^2}{2M} \sum_{j=1,2,3} \left\{ \frac{\tan^2(2\varepsilon)\left(\frac{\partial}{\partial x_j}\varepsilon\right)^2 + \left|\frac{\partial}{\partial x_j}\mathbf{e}\right|^2}{\cos(2\varepsilon)} \frac{d}{dw} \right.$$
$$\left. + \frac{\tan^2(2\varepsilon)\left(\frac{\partial}{\partial x_j}\varepsilon\right)^2 + \left|\mathbf{e}^* \cdot \frac{\partial}{\partial x_j}\mathbf{e}\right|^2}{\cos^2(2\varepsilon)} \frac{d^2}{dw^2} \right\} \quad (44)$$
$$\times \ln P_J(w)\big|_{w = 1/\cos(2\varepsilon)}.$$

For the $J_g = 1 \longrightarrow J_e = 1$ transition we obtain the well-known result [12]:

$$A_j = -i\hbar\left(\mathbf{e}^* \cdot \frac{\partial}{\partial x_j}\mathbf{e}\right),$$
$$W = \frac{\hbar^2}{2M} \sum_{j=1,2,3} \left\{ \left|\frac{\partial}{\partial x_j}\mathbf{e}\right|^2 - \left|\mathbf{e}^* \cdot \frac{\partial}{\partial x_j}\mathbf{e}\right|^2 \right\}. \quad (45)$$

As an example we now consider the one-dimensional field configuration *lin* ⊥ *lin* (40). In this case no vector

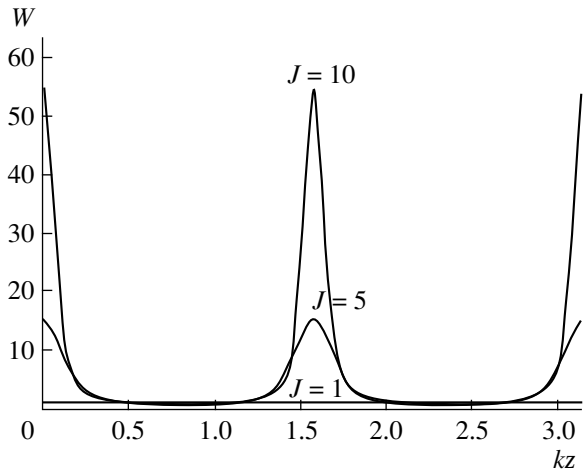


Fig. 4. Dark geometric scalar potentials W (in units of $\hbar^2k^2/(2M)$) for $J_g = J \longrightarrow J_e = J$ transitions (J is an integer) for the *lin* ⊥ *lin* field configuration.

potential is present ($A_j = 0$) and the scalar potential has the form

$$W = \frac{\hbar^2 k^2}{2M} \times \left\{ \frac{1}{\cos^3(2kz)} \frac{d}{dw} + \frac{\sin^2(2kz)}{\cos^4(2kz)} \frac{d^2}{dw^2} \right\} \times \ln P_J(w)\big|_{w = 1/\cos(2kz)} \quad (46)$$

and is given in Fig. 4 for several values of J.

The dark geometric potentials for $J_g = J \longrightarrow J_e = J - 1$ transitions can be determined similarly. At this point we do not give the explicit form of these potentials because of the cumbersome nature of the expressions. We merely note that for this type of transition because of the two-dimensional subspace of the dark states, both the Cartesian components $\hat{A}_j$ of the vector potential and the scalar potential $\hat{W}$ are square 2×2 matrices. These were determined in [14] for the particular *lin* ⊥ *lin* field configuration.

8. CONCLUSIONS

The present study is a continuation in a series of studies [10, 18, 24, 25] dealing with the fundamental problem of the steady state of atoms in an elliptically polarized field. Invariant analytic expressions were obtained for all elliptic dark states, allowing their tensor properties to be analyzed in detail. As well as being of general theoretical value, the results of the present study may also have important practical applications in all problems involving coherent population trapping in elliptically polarized fields, especially in those cases where the polarization vector varies in a complex fashion in space and (or) time. For example, we determined the explicit analytic form of the dark magneto-optic and geometric potentials in an inhomogeneous monochromatic field for all transitions where the CPT effect occurs.

Other possible applications may include the kinetics of atoms in nonuniformly polarized fields [26], adiabatic transfer of momentum and populations [9], and so on. In addition, the results of the present study may be of procedural interest as a demonstration of the high efficiency of the apparatus of irreducible tensors and spherical functions in analyses of the interaction of atoms with polarized radiation, and the geometric correspondence observed between the complex vector space and the set of all possible spinor pairs is, in our opinion, an interesting mathematical result.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project no. 98-02-17794).

REFERENCES

1. G. Alzetta, A. Cozzini, L. Moi, and G. Orriols, Nuovo Cimento B **36**, 5 (1976); E. Arimondo and G. Orriols, Lett. Nuovo Cimento **17**, 333 (1976); R. M. Whitley and C. R. Stroud, Phys. Rev. A **14**, 1498 (1976).
2. A. Akulshin, A. Celikov, and V. Velichansky, Opt. Comm. **84**, 139 (1991); A. M. Akulshin, S. Barreiro, and A. Lezama, Phys. Rev. A **57**, 2996 (1998).
3. M. O. Scully, Phys. Rep. **129**, 191 (1992); O. Kocharovskaya, Phys. Rep. **129**, 175 (1992); G. G. Padmabandu, G. Welch, I. N. Shubin, *et al.*, Phys. Rev. Lett. **76**, 2053 (1996).
4. A. Aspect, E. Arimondo, R. Kaiser, *et al.*, Phys. Rev. Lett. **61**, 826 (1988).
5. A. M. Tumaĭkin and V. I. Yudin, Zh. Éksp. Teor. Fiz. **98**, 81 (1990) [Sov. Phys. JETP **71**, 43 (1990)].
6. G. Grynberg and J.-Y. Courtois, Europhys. Lett. **27**, 41 (1994).
7. A. Hemmerich, M. Wiedemüller, T. Esslinger, *et al.*, Phys. Rev. Lett. **75**, 37 (1995).
8. N. P. Konopleva, A. V. Taichenachev, A. M. Tumaikin, and V. I. Yudin, Quantum Semiclassic. Opt. **8**, 837 (1996).
9. P. Marte, P. Zoller, and J. L. Hall, Phys. Rev. A **44**, 4118 (1991); C. Valentin and P. Pillet, J. Phys. II **4**, 1925 (1994); P. D. Featonby, G. S. Summy, C. L. Webb, *et al.*, Phys. Rev. Lett. **81**, 495 (1998).
10. V. S. Smirnov, A. M. Tumaĭkin, and V. I. Yudin, Zh. Éksp. Teor. Fiz. **96**, 1613 (1989) [Sov. Phys. JETP **69**, 913 (1989)].
11. V. Milner and Y. Prior, Phys. Rev. Lett. **80**, 940 (1998).
12. A. V. Taichenachev, A. M. Tumaikin, and V. I. Yudin, Laser Phys. **2**, 575 (1992).

13. R. Dum and M. Olshanii, Phys. Rev. Lett. **76**, 1788 (1996).

14. P. M. Visser and G. Nienhuis, Phys. Rev. A **57**, 4581 (1998).

15. D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskii, *Quantum Theory of Angular Momentum* (Nauka, Leningrad, 1975; World Scientific, Singapore, 1988).

16. N. L. Manakov and A. V. Merem'yanin, Zh. Éksp. Teor. Fiz. **111**, 1984 (1997) [JETP **84**, 1080 (1997)].

17. N. Ya. Vilenkin, *Special Functions and the Theory of Group Representations* (Nauka, Moscow, 1991; American Mathematical Society, Providence, 1968).

18. A. V. Taĭchenachev, A. M. Tumaĭkin, V. I. Yudin, and G. Nienhuis, Zh. Éksp. Teor. Fiz. **114**, 125 (1998) [JETP **87**, 70 (1998)].

19. A. V. Taĭchenachev, A. M. Tumaĭkin, M. A. Ol'shanyĭ, and V. I. Yudin, Pis'ma Zh. Éksp. Teor. Fiz. **53**, 336 (1991) [JETP Lett. **53**, 351 (1991)].

20. M. A. Ol'shanii and V. G. Minogin, Opt. Commun. **89**, 393 (1992).

21. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Nauka, Moscow, 1973; Pergamon, Oxford, 1975).

22. A. V. Taĭchenachev, A. M. Tumaĭkin, and V. I. Yudin, Zh. Éksp. Teor. Fiz. **113**, 2056 (1998) [JETP **86**, 1127 (1998)].

23. A. P. Kazantsev, G. I. Surdutovich, and V. P. Yakovlev, *Mechanical Action of Light on Atoms* (Nauka, Moscow, 1991; World Sci., Singapore, 1991).

24. A. V. Taĭchenachev, A. M. Tumaĭkin, V. I. Yudin, and G. Nienhuis, Zh. Éksp. Teor. Fiz. **108**, 415 (1995) [JETP **81**, 224 (1995)].

25. A. V. Taĭchenachev, A. M. Tumaĭkin, and V. I. Yudin, Zh. Éksp. Teor. Fiz. **110**, 1727 (1996) [JETP **83**, 949 (1996)].

26. J. Lawall, F. Bardou, B. Saubamea, *et al.*, Phys. Rev. Lett. **73**, 1915 (1994); J. Lawall, S. Kulin, B. Saubamea, *et al.*, Phys. Rev. Lett. **75**, 4194 (1995).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 1, 2000, pp. 76–83.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 1, 2000, pp. 87–95.

ATOMS, SPECTRA, RADIATION

Semiclassical Description of Shell Effects in Atomic Clusters

G. V. Shpatakovskaya

Institute of Mathematical Modeling, Russian Academy of Sciences, Moscow, 125047 Russia
e-mail: shpat@imamod.ru

Received February 23, 2000

Abstract—The density functional method for a system of electrons in an external field is applied to derive an expression for the correction to the free energy for a small change in electron density. This expression is used for a semiclassical description of shell effects in atomic clusters at both zero and finite temperatures. The spherical "jelly" model is used to obtain a formula for the oscillating component of the free electron energy in medium and large clusters in the form of an explicit sum of contributions from supershells having quantum numbers $2n_r + l$, $3n_r + l$, $4n_r + l$, … or corresponding trajectories (linear, triangular, square, …) of the classical motion of an electron whose energy is equal to the chemical potential in a self-consistent potential. The conditions for the appearance of a new supershell and its relative contribution are analyzed for various sizes of clusters, potential profile, and electron temperature. Specific calculations are made for sodium clusters with two types of self-consistent potential: a "square well" and the Woods–Saxon potential. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

One topic in mesoscopic physics which has seen rapid development over the last two decades is atomic clusters, in particular metallic clusters (see the reviews presented in [1, 2]). Studies of these complexes, whose size ranges from a few particles to several tens of thousands of atoms, provide a unique possibility for studying the evolution of the properties of matter with increasing scales from atomic to solid-state. A characteristic feature of the experimental mass spectra of clusters (dependences of the number of N-atomic clusters on N) is the existence of "magic" numbers: there are considerably more clusters with these numbers than neighboring ones. As the number N increases, these deviations decrease and then begin to increase again, i.e., oscillations with beats are observed. As the temperature increases, these effects are reduced appreciably.

Theoretical analyses of metal clusters frequently use the spherical "jelly" model in which the ions are considered to be uniformly distributed over the volume of a cluster of radius[1] $R = r_s N_e^{1/3}$ (where r_s is the Wigner–Seitz electron radius of the condensed phase of the element, $N_e = wN$ is the number of electrons in an N-atomic cluster, and w is the valence of the element). Self-consistent calculations are made to describe interacting valence electrons situated in the field of ions (see, for example, [3] for sodium clusters which uses the density functional method). One of the results of these calculations is the reproduction of the experimentally determined characteristic two types of periodicity in the oscillating component of the electron energy δE_{sh} as a function of the cluster radius: small-period oscillations are accompanied by beats whose period is an order of magnitude larger. Similar results were obtained in calculations [4] using the phenomenological Woods–Saxon potential and in [5] for various metals having potentials close to self-consistent.

For small clusters with $N < 100$, the numerical results can be interpreted quite easily: cusps of energy minima of the electron shell correspond to clusters in which the shell having the maximum orbital angular momentum $l = l_{\max}$ is completely filled. However, for $N > 100$ this dependence breaks down and the nature of the specific oscillations described has been discussed in the literature over the last ten years.

In a fundamental study [6] for a spherical cavity Balian and Bloch put forward a theory which explains the beats in the electron level density as the superposition of contributions from closed periodic classical electron trajectories and they show that it is sufficient to use triangular and square orbits to obtain an approximate description of these effects. This theory was successfully applied [7] in nuclear physics to describe the shell structure of atomic nuclei.

Detailed numerical calculations [5] of clusters having more complex potentials reproduce behavior very similar to oscillations with beats in a spherical cavity. It was noted that electron energy levels $\varepsilon(n_r, l)$ with high angular momenta are grouped into supershells: $\varepsilon(n_r, l) \simeq \varepsilon(n_r + 1, l - K)$ with pseudoquantum numbers $Kn_r + l$, where n_r is the radial quantum number, $K = 2, 3, 4, \ldots$.

It seems that a correspondence can be established between the quantum supershells and the classical trajectories. In the classical treatment (see, for example [8]) the integer K characterizing the supershell is equal to the ratio of the frequencies of the radial and angular motion for the corresponding closed orbit. For $K = 2$ the pseudoquantum number is the same as the principal

[1] Atomic units are used in this article.

quantum number and characterizes the distribution of the single-electron levels in small clusters. The corresponding classical electron motion takes place along a linear orbit passing through the point $r = 0$. As N increases, supershells with $K = 3$ appear, followed by $K = 4$, which correspond to triangular and square orbits.

Thus, a scrupulous analysis of the results of laborious calculations made in these studies suggests that first, the expansion in terms of classical trajectories is equivalent to the expansion in terms of supershells and second, the result obtained in [6] for a spherical cavity is also valid for a spherical cluster potential of a more general nature. This last assumption was partially confirmed in a recent study [9] for the Woods–Saxon potential by means of an expansion in terms of the parameter a/R (a is the surface width) near the well-known result [6].

It is interesting to obtain this expansion in terms of trajectories analytically and to study the origin of supershells and the laws governing their appearance for a general type of potential.

In the present study we show that this problem may be solved by a semiclassical method of isolating shell effects which was previously used successfully in atomic physics [10, 11] and in plasma physics [12, 13] on the basis of the Thomas–Fermi model.

Here the density functional method is used to derive an expression for the correction to the free energy for a small variation in electron density (Section 2). Then a semiclassical theory of shell effects for a spherically symmetric attractive potential is constructed using the generalized Thomas–Fermi model taken as the zeroth approximation (Section 3). These analytic expressions are used to calculate the shell oscillations of the free electron energy in sodium clusters for the "square well" potential and the Woods–Saxon potential (Section 4). The main results of this study are presented in the Conclusions (Section 5).

A brief description of the proposed method and some preliminary results are published in [14].

2. CORRECTION TO THE FREE ENERGY FOR A SMALL DENSITY VARIATION

We shall consider a system of N_e interacting electrons in an external field $U_{\text{ext}}(\mathbf{r})$ at temperature T in a given volume. In accordance with the Hohenberg–Kohn–Mermin theorem the electron density functional $n(\mathbf{r})$

$$F[n] = K[n] + \int d\mathbf{r} n(\mathbf{r}) U_{\text{ext}}(\mathbf{r}) + \frac{1}{2}\int d\mathbf{r} d\mathbf{r}' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + F_{xc}[n], \quad (1)$$

where $K[n]$ and $F_{xc}[n]$ are the functionals of the kinetic and exchange correlation energy, reaches a minimum equal to the free energy of the system F_e when its functional argument $n_e(\mathbf{r})$ is the same as the true equilibrium electron density $n_e(\mathbf{r})$ in a given external field $U_{\text{ext}}(\mathbf{r})$. The extremum condition of the functional $F[n]$ if the total number of particles is conserved yields the Euler–Lagrange equation:

$$\frac{\delta K}{\delta n} = \mu - \tilde{U}(\mathbf{r}),$$
$$\tilde{U}(\mathbf{r}) = U_{\text{ext}}(\mathbf{r}) + \int d\mathbf{r}' \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + \frac{\delta F_{xc}}{\delta n}. \quad (2)$$

Here the Lagrange factor μ is the chemical potential of the system.

Since the exact form of the functionals $K[n]$ and $F_{xc}[n]$ is unknown, various approximations are used for these. The following statement is useful in this respect.

Let us assume that the solution $n_0(\mathbf{r})$, μ_0, $\tilde{U}_0(\mathbf{r})$ of the problem (1), (2) with the approximate functional $F_0[n]$ is known and the small correction $\Delta n(\mathbf{r}, \mu, T)$ to the electron density as a result of various effects neglected in $F_0[n]$ is also known. Then, to within quadratic terms in Δn the corresponding correction ΔF to the free energy $F_0[n_0]$ has the form:[2]

$$\Delta F = -\int_{-\infty}^{\mu_0} d\mu \int d\mathbf{r} \Delta n(\mathbf{r}, \mu, T). \quad (3)$$

We shall now demonstrate this.

In accordance with equation (2) the functional $K[n]$ in (1) is the potential relative to $\eta \equiv \mu - \tilde{U}(\mathbf{r})$. Converting to the new variable η instead of the density n, we construct the functional

$$\Psi[\eta] = K - \int d\mathbf{r} n\eta = F - F_{xc} - \int d\mathbf{r} n\left(\mu - \frac{1}{2}\int d\mathbf{r}' \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} - \frac{\delta F_{xc}}{\delta n}\right), \quad (4)$$

which is the potential relative to n:

$$\delta\Psi[\eta]/\delta\eta = -n.$$

The density variation Δn leads to a variation in the potential Ψ by $\Delta\Psi$ where quite clearly

$$\delta\Delta\Psi[\eta]/\delta\eta = -\Delta n.$$

It follows from expression (2) that the density variation $n \longrightarrow n(\eta) + \Delta n$ leads to variations in the chemical and self-consistent potentials $n \longrightarrow \eta + \Delta\eta$ so that the total correction to the density has the form $\Delta n + (\partial n/\partial\eta)\Delta\eta$. Considering the total variation in Ψ taking this into account, we obtain from (4) the equal potential variations: $\Delta F = \Delta\Psi$ from which it follows that

$$\delta(\Delta F)/\delta\eta = -\Delta n.$$

[2] The derivation of expression (3) presented below generalizes the corresponding proof for the correction to the Thomas–Fermi model presented in [10].

Taking $\delta\eta$ to be the variation of the chemical potential and integrating over this, we obtain expression (3).

Isolating the explicit temperature dependence (see [10]), we can rewrite expression (3) in the following form:

$$\Delta F = -\int_{-\infty}^{\mu_0} d\mu \frac{\pi\hat{k}_\mu}{\sinh(\pi\hat{k}_\mu)}\Delta N(\mu, 0), \quad (5)$$

where we use the operator $\hat{k}_\mu = -iT\partial/\partial\mu$ and

$$\Delta N(\mu, 0) = \int d\mathbf{r}\Delta n(\mathbf{r}, \mu, 0)$$

is the correction to the number of states as a result of this effect without the explicit dependence on temperature. Note that an implicit temperature dependence is contained in (3) and (5) in terms of the values of μ_0 and $\tilde{U}_0(\mathbf{r})$ used to calculate the corrections $\Delta n(\mathbf{r})$ and ΔN.

Since the corresponding density variations are assumed to be small, when allowance is made for the various effects, there is no interference between them and each corresponds to a separate term in the correction to the density and the free energy.

We stress that expressions (3) and (5) are of a most general nature and can be used in the density functional method for any type of functional in the zeroth approximation, including in the single-particle Kohn–Sham treatment.

3. SEMICLASSICAL SHELL CORRECTION TO THE FREE ENERGY OF A SPHERICALLY SYMMETRIC CLUSTER

In order to describe a spherically symmetric system containing bound states, we shall use the extended Thomas–Fermi model as the zeroth approximation. This model allows for the local exchange correlation term $F_{xc}[n]$ and the term $K[n]$ includes second- and fourth-order quantum gradient corrections in addition to the kinetic energy of a homogenous gas. This model is the best (in the sense of allowing for the largest number of effects) known approximation for F_0 other than the Kohn–Sham model. In particular, the extended Thomas–Fermi model very accurately describes the average (smooth) component of the density and the electron binding energy as a function of the cluster size even for small clusters [15]. Here shell effects associated with the discreteness of the electron spectrum are neglected.

Below we propose a semiclassical method of describing shell effects for a spherically symmetric potential constructed using expression (5). This approach is the most systematic in this case since the extended statistical Thomas–Fermi model used as the zeroth approximation is a set of several first terms of the expansion in terms of the semiclassical behavior parameter.

We shall estimate the contribution to the free energy in the form (5) of the shell correction to the density of electron states Δn_{sh} or to the number of states having energies lower than or equal to μ:

$$\Delta N(\mu, 0) = \Delta N_{sh}(\mu) = \int d\mathbf{r}\Delta n_{sh}(r, \mu) = N(\mu) - N_0(\mu). \quad (6)$$

Here for filled shells we have

$$N(\mu) = 2\sum_{n_r, l}(2l+1)\theta(\mu - \varepsilon_{n_r, l}), \quad (7)$$

and the level energies $\varepsilon_{n_r, l}$ are determined from the quantization condition

$$S_{\varepsilon l} = \int dr\, p_{\varepsilon l}(r) = \pi\left(n_r + \frac{1}{2}\right), \quad (8)$$

where $S_{\varepsilon l} \equiv S_{\varepsilon\lambda}$ and

$$p_{\varepsilon l}(r) = \sqrt{2[\varepsilon - U(r)] - (l+1/2)^2/r^2} \equiv \sqrt{p_\varepsilon^2(r) - \lambda^2/r^2} \equiv p_{\varepsilon\lambda}$$

are the classical radial action and the momentum of an electron having energy ε and orbital angular momentum λ, and the region of integration in the action integral is limited by the turning points. By means of simple procedures using the Poisson formula to replace the sums over the quantum numbers n_r, l with integrals, we can rewrite expression (7) in the following form:

$$N(\mu) = \frac{2}{\pi}\sum_{k, s=-\infty}^{\infty}\frac{(-1)^{k+s}}{k} \times \int_0^{\lambda_\mu} d\lambda\lambda\sin[2\pi(k\nu_\mu(\lambda) + s\lambda)]. \quad (9)$$

Here we introduce the notation $\nu_\varepsilon(\lambda) = S_{\varepsilon\lambda}/\pi$ and λ_ε determines the boundary of the phase region of the classically allowed motion of an electron of energy ε: $\nu_\varepsilon(\lambda_\varepsilon) = 0$. Thus, the values of λ and ν are the orbital and radial quantum numbers shifted by 1/2 and considered as continuous parameters.

In formula (9) the term containing $k = s = 0$ corresponds to a smooth nonoscillating dependence of N_0 on μ. It has already been taken into account in the zeroth approximation of the model and satisfies the normalization condition. Thus, the sum (9) minus this term in accordance with formula (6) determines the required quantity $\Delta N_{sh}(\mu)$.

It follows from expression (9) that the shell oscillations are completely determined by the radial action function $S_{\mu\lambda} = \pi\nu_\mu(\lambda)$ for the motion of an electron having energy $\mu = \mu_0$ and orbital angular momentum λ in

the spherically symmetric potential of the zeroth approximation $U(r) = \tilde{U}_0(r)$.

In what follows, we shall confine our analysis to attractive potentials $U(r)$ finite at zero, which are typical of clusters.[3] For all potentials $U(r)$ of this type the slope of the curve $\nu_\mu(\lambda)$ at $\lambda = 0$ is the same [16]:

$$\left.\frac{\partial \nu_\mu(\lambda)}{\partial \lambda}\right|_0 = -\frac{1}{2},$$

and the derivative for $\lambda = \lambda_\mu$

$$\left.\frac{\partial \nu_{\mu\lambda}}{\partial \lambda}\right|_{\lambda_\mu} \equiv -\nu'_\mu \tag{10}$$

is determined by the potential profile. Semiclassical calculations of λ_μ and ν'_μ give (see Appendix)

$$\lambda_\mu = p_\mu(r_0) r_0,$$
$$\nu'_\mu = \left[3 + U''(r_0)\frac{r_0^2}{p_\mu^2(r_0)}\right]^{-1/2}, \tag{11}$$

where r_0 is the maximum point of the function $p_\mu(r)r$.

Figure 1 demonstrates the characteristic behavior of (a) the function $\nu_\mu(\lambda)$ and (b) its derivative for the case of sodium clusters with the Woods–Saxon potential for various numbers of electrons.

The main contribution to the integral of the oscillating function in (9) is made by the limits of integration and the steady-state phase points $\bar{\lambda}$. These are determined from

$$\left.\frac{\partial \nu_\mu(\lambda)}{\partial \lambda}\right|_{\bar{\lambda}} = -\frac{s}{k}, \quad 0 \le \bar{\lambda} \le \lambda_\mu, \tag{12}$$

which isolates the principal terms in the sum over k:[4]

$$k = (2 + j)s, \quad \left.\frac{\partial \nu_\mu(\lambda)}{\partial \lambda}\right|_{\bar{\lambda}} = -\frac{1}{2 + j},$$
$$j = 0, 1, \ldots, j_{\max}, \quad j_{\max} = \left[\frac{1}{\nu'_\mu} - 2\right]. \tag{13}$$

Figure 1b gives the horizontal lines $\partial\nu_\mu(\lambda)/\partial\lambda = -1/(2 + j)$, $j = 1, 2, \ldots$. The points of intersection of the derivative curve with these lines in accordance with (13) determine the steady-state phase points $\bar{\lambda}_j$.

In calculations of the integral in (9) terms with $k = 2s$ ($j = 0$) should be analyzed separately because in this case, the steady-state phase point is the same as the lower limit of integration. As a result we obtain

$$\Delta N_{\rm sh} = \frac{1}{2}\sum_{s=1}^{\infty}\frac{(-1)^s}{(\pi s)^2}\left\{\frac{\cos(2\pi s \times 2\nu_\mu(0))}{\delta_\mu^{(0)}} - \frac{\lambda_\mu \cos(2\pi s \lambda_\mu)}{0.5 - \nu'_\mu} + \sum_{j=1}^{j_{\max}}\frac{8\pi\sqrt{s}\bar{\lambda}_j(-1)^{js}}{\sqrt{\delta_\mu^{(j)}}(2 + j)^{3/2}} \times \cos\left[2\pi s[(2 + j)\nu_\mu(\bar{\lambda}_j) + \bar{\lambda}_j] - \frac{\pi}{4}\right]\right\}. \tag{14}$$

Here we have

$$\delta_\mu^{(j)} \equiv \left.\frac{\partial^2 \nu_\mu(\lambda)}{\partial \lambda^2}\right|_{\bar{\lambda}_j},$$

and the first two terms in braces are the contribution from the steady-state phase point with $j = 0$.

By substituting expressions (14) into (5), the result of the action of the "temperature" operator on $\Delta N_{\rm sh}(\mu')$ can be estimated by differentiating only the rapidly varying cosines. Then, integrating by parts and isolating the principal term with respect to the semiclassical

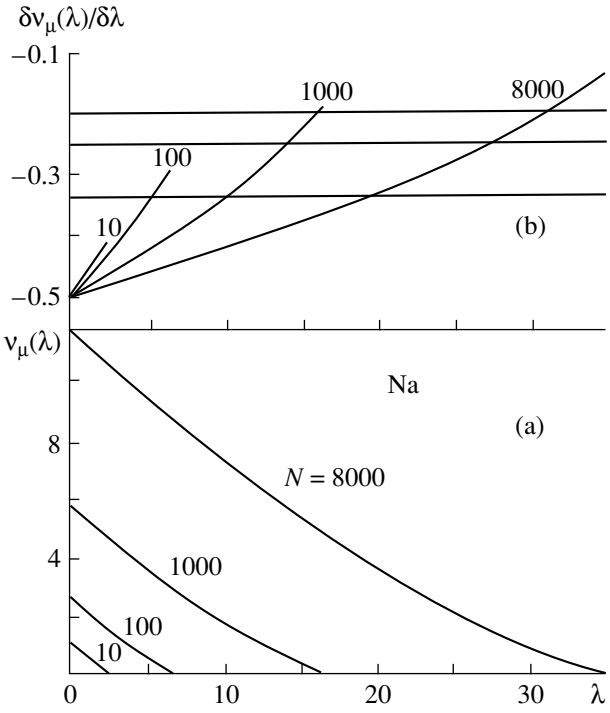


Fig. 1. Dependences of (a) the radial action $\nu_\mu(\lambda) = S_{\mu\lambda}/\pi$ and (b) its derivative $\partial\nu_\mu(\lambda)/\partial\lambda$ on the orbital angular momentum $\lambda = l + 1/2$ for sodium clusters with the Woods–Saxon potential (20) for various numbers of atoms N.

[3] Corresponding results for attractive potentials having a Coulomb singularity at zero are presented in [10–12].

[4] Here we neglect very soft potentials (for example, in aluminum clusters, see literature cited in [9]) for which terms with a different ratio of s and k are the principal ones in a certain range of electron numbers.

behavior parameter, we obtain the formula for the shell correction to the free energy of the cluster electrons:

$$\Delta F_{\rm sh} = \frac{1}{4}\sum_{s=1}^{\infty}\frac{(-1)^s}{(\pi s)^3}\Bigg\{C_\mu^{(s)}(T)\frac{\lambda_\mu\sin(2\pi s\lambda_\mu)}{(0.5-\nu_\mu')\partial\lambda_\mu/\partial\mu} - C_0^{(s)}(T)\frac{\sin(2\pi s\times 2\nu_\mu(0))}{\delta_\mu^{(0)}\partial 2\nu_\mu(0)/\partial\mu} - \sum_{j=1}^{j_{\max}} C_j^{(s)}(T)\frac{8\pi\sqrt{s\bar\lambda_j}(-1)^{js}}{\sqrt{\delta_\mu^{(j)}}(2+j)^{5/2}} \times\frac{\sin[2\pi s((2+j)\nu_\mu(\bar\lambda_j)+\bar\lambda_j)-\pi/4]}{\partial\nu_\mu(\lambda)/\partial\mu|_{\bar\lambda_j}}\Bigg\}, \tag{15}$$

where

$$C_\mu^{(s)}(T) = \frac{sT/T_\mu}{\sinh(sT/T_\mu)},\quad C_0^{(s)}(T) = \frac{sT/T_0}{\sinh(sT/T_0)},\quad C_j^{(s)}(T) = \frac{sT/T_j}{\sinh(sT/T_j)}. \tag{16}$$

In formula (16) we use the notation T_0, T_μ, and T_j for the characteristic temperatures which determine the damping of the corresponding oscillations as the electron temperature increases:

$$\frac{1}{T_0} = 2\pi^2\frac{\partial 2\nu_\mu(0)}{\partial\mu},\quad \frac{1}{T_\mu} = 2\pi^2\frac{\partial\lambda_\mu}{\partial\mu},\quad \frac{1}{T_j} = 2\pi^2(2+j)\frac{\partial\nu_\mu(\lambda)}{\partial\mu}\bigg|_{\bar\lambda_j}. \tag{17}$$

The terms of the sum over j in formulas (14) and (15) are the contributions of supershells having the quantum numbers $n_j = Kn_r + l$, $K = 2 + j$, where quantization is significant at the energy level equal to the chemical potential of the system μ.

In the classical treatment the index j in formulas (14) and (15) numbers the type of electron orbit: the index $j = 0$ ($K = 2$) corresponds to a linear orbit passing through the point $r = 0$ and terms with $j \geq 1$ ($K \geq 3$) correspond to plane regular polygons having the number of vertices K. The index s is equal to the number of periods which include the trajectory (j, s) so that for each j-orbit the sum over s is the expansion over the trajectory lengths.

At zero temperature the factors (16) are unity and (15) gives the expression for the shell correction to the electron energy. At temperatures higher than the characteristic values (17) the contribution of the corresponding shell becomes negligible.

4. SHELL OSCILLATIONS OF THE ELECTRON ENERGY

In this section we apply the general expressions obtained above to analyze shell effects in sodium clusters. We shall assume that the form of the self-consistent potential of the zeroth approximation is known and we shall approximate this by two types of model potentials: the limiting hard potential of a "square well" and the Woods–Saxon potential having parameters taken from the self-consistent calculations.

4.1. Square Well Potential

For a potential having the form

$$U(r) = \begin{cases}-2\varepsilon_F, & r \leq R\\ 0, & r > R,\end{cases}\qquad R = r_s N_e^{1/3},\quad \varepsilon_F = \frac{1}{2r_s^2}\left(\frac{9\pi}{4}\right)^{2/3},\quad \mu = -\varepsilon_F, \tag{18}$$

all the necessary quantities contained in expression (15) can be calculated analytically:

$$\nu_\mu(\lambda) = \frac{\lambda_\mu}{\pi}[\sqrt{1-x^2} - x\arccos x],\quad x = \frac{\lambda}{\lambda_\mu},$$

$$\lambda_\mu = p_\mu R,\quad \delta_\mu^{(0)} = \frac{1}{\pi\lambda_\mu},\quad \nu_\mu' = 0,\quad \nu_\mu(0) = \frac{\lambda_\mu}{\pi},$$

$$\frac{\partial 2\nu_\mu(0)}{\partial\mu} = \frac{2}{\pi}\frac{\partial\lambda_\mu}{\partial\mu},\quad \bar\lambda_j = \lambda_\mu\cos\left(\frac{\pi}{2+j}\right),$$

$$(2+j)\nu_\mu(\bar\lambda_j) + \bar\lambda_j = \frac{\lambda_\mu}{\pi}(2+j)\sin\left(\frac{\pi}{2+j}\right),$$

$$\delta_\mu^{(j)} = \left[\pi\lambda_\mu\sin\left(\frac{\pi}{2+j}\right)\right]^{-1}.$$

Since $\nu_\mu' = 0$ for this potential, in a cluster of any size there is an infinite number of steady-state phase points although calculations show that only the first few values with $j \leq 4$ make any realistic contribution. Figures 2 and 3 give results of calculations of the shell oscillations of the electron energy using formula (15) at zero temperature for up to 8000 atoms per cluster together with an analysis of the contribution of each jth supershell (Fig. 2) and its relative role in the beat formation (Fig. 3). It can be seen from this analysis that the $j = 0$ shell makes an insignificant contribution with the $j = 1$, 2 supershells playing the main role, while for larger clusters terms with $j = 3$, 4 must also be taken into account.

The dependence of the oscillation amplitude on the cluster size is easily estimated and is proportional to $\lambda_\mu^{3/2}/(\partial\lambda_\mu/\partial\mu) \sim N_e^{1/6}$.

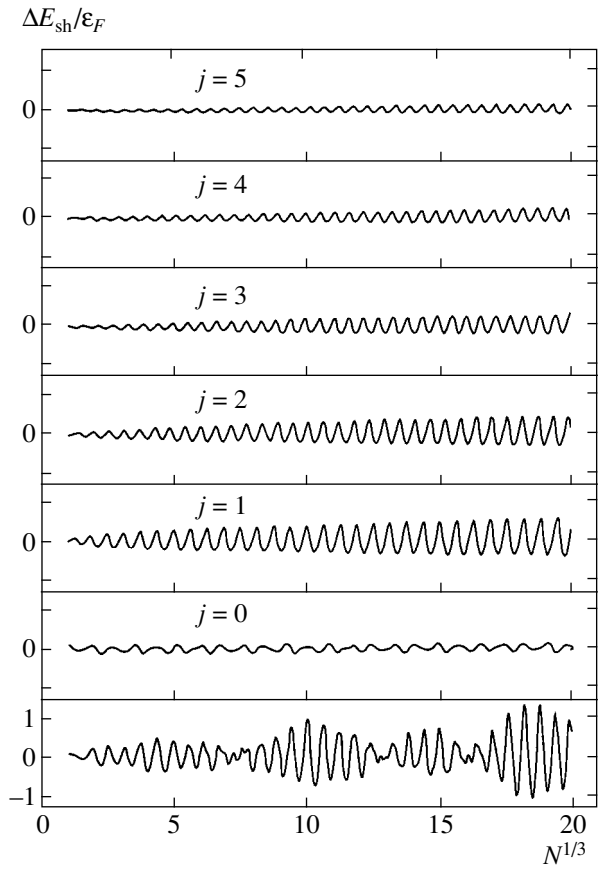


Fig. 2. Shell correction to the electron energy of a cluster having the spherical square-well potential (18) in units of the Fermi energy, calculated using formula (15) (lower curve) and various contributions to this value from supershells with various j (upper curves) at zero temperature.

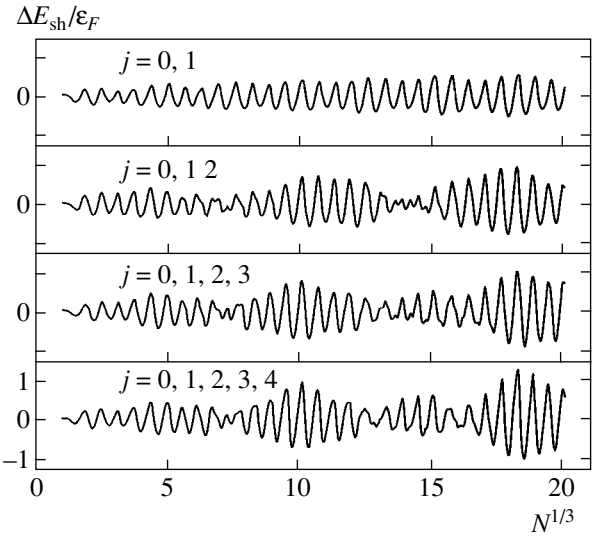


Fig. 3. Analysis of the relative role of supershells with various j in the formation of beats at zero temperature. The potential and the notation are the same as in Fig. 2.

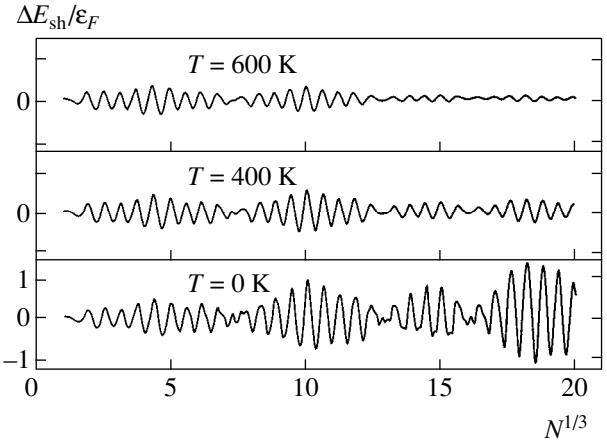


Fig. 4. Temperature damping of shell oscillations of the free energy of clusters. The potential and notation are the same as in Fig. 2.

A comparison with the numerical calculations [5] for the potential (18) shows that for $N_e > 50$ our analytic results show good agreement in terms of oscillation frequency, curve profile, and even fine-structure details. We note that in [5] the shell correction was isolated not from the total electron energy but from

$$E(\mu) = 2\sum_{n_r, l}(2l+1)\varepsilon_{n_r, l}\theta(\mu - \varepsilon_{n_r, l}). \qquad (19)$$

The role of the temperature in suppressing the oscillations of the free energy is demonstrated in Fig. 4 in which results of calculations using formula (15) for two finite temperatures and for zero temperature are compared. Since the ratio of the characteristic temperature T_j to the Fermi energy is small,

$$\frac{T_j}{|\mu|} = \left[\pi\lambda_\mu(2+j)\sin\left(\frac{\pi}{2+j}\right)\right]^{-1} \simeq \frac{1}{3\pi\lambda_\mu} \ll 1,$$

the degeneracy condition is known to be satisfied for the medium and large clusters considered here at temperatures $T < 3T_j$ and the potential calculated at zero temperature can be used.

4.2. Woods–Saxon Potential

We shall now consider the more realistic example of the potential in the zeroth approximation, the Woods–Saxon potential:

$$U(r) = -V_0\left[1 + e^{(r-R)/a}\right]^{-1}. \qquad (20)$$

For sodium clusters the parameters in (20) have the values

$$V_0 = 0.22, \quad a = 1.4, \quad r_s = 3.93,$$

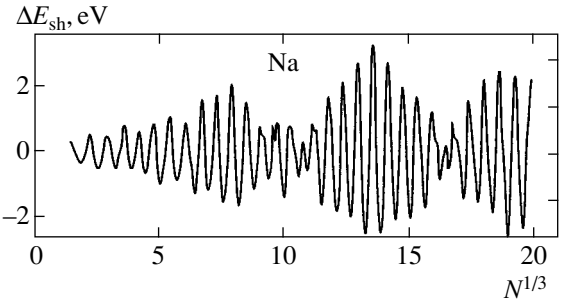


Fig. 5. Shell correction to the electron energy in sodium clusters having the potential (20) as a function of the cluster size at zero temperature.

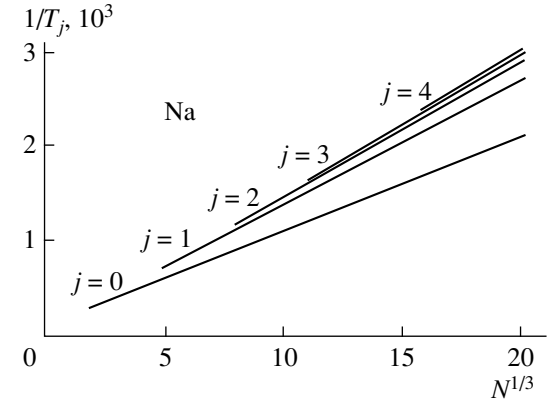


Fig. 6. Reciprocal characteristic temperatures (17) (in atomic units) for various supershells as a function of the size of a sodium cluster having the potential (20).

and the chemical potential is $\mu = -0.1015$.

Figure 5 gives the shell oscillations of the electron energy as a function of the cluster size calculated using formula (15) for sodium over a wide range of variation of the number of electrons up to 8000. For $N_e^{1/3} \geq 7$ our curve shows very good agreement with the results of the exact calculations [4] for the same potential and in particular, the magic numbers N_e corresponding to the position of the minima are calculated to the third decimal place and the beat period, curve profile, and fine-structure details are reproduced exactly. Note that in [4] as in [5] the shell correction was isolated not from the total electron energy but from (19).

When our results are compared with the self-consistent calculations [3] for sodium, the amplitude of the oscillations and the position of the local minima show good agreement but in [3] the beat pattern is shifted toward lower N_e. For $N_e = 1000$, it was shown in [5] that our approximate potential (20) differs from the true self-consistent value. The latter is characterized in particular by a larger effective cluster size which explains the observed shift. This reaffirms the conclusion that the form of the beats is sensitive to the potential profile near the cluster surface.

The damping of the oscillations of the free energy with increasing temperature for the potential (20) is qualitatively similar to that shown in Fig. 4. For a quantitative analysis more information can be obtained from Fig. 6 which gives the reciprocal characteristic temperatures (17) as a function of the cluster size for various supershells. To a high degree of accuracy, these are segments of straight lines with various slopes. Although the slowest temperature damping corresponds to the lowest $j = 0$ line, the contribution of the corresponding shell is negligible and damping of the $j = 1$ supershell for which the linear dependence has the form $1/T_1 = 62.4 + 132.4N^{1/3}$plays the dominant role. From this it follows that the characteristic exponential factor determining the dependences of the oscillation damping (in atomic units) on temperature and cluster size is $\exp(-132.4TN^{1/3})$. We note the difference between the characteristic temperature T_1 and the estimate T_0 used in the literature by analogy with a harmonic oscillator.

5. CONCLUSIONS

We have derived an expression for the first-order correction to the free energy of a system of electrons in an external field for a small variation in particle density.

This expression was used to make a semiclassical analysis of shell effects in atomic clusters at zero and finite temperatures. The spherical jelly model was used to obtain an analytic expression for the oscillating component of the free electron energy for medium and large clusters as an explicit sum of the contributions of closed trajectories (linear, triangular, square, …, K-angle) of the classical motion of an electron whose energy is equal to the chemical potential in the self-consistent potential. We showed that there is a correspondence between these K-angle trajectories and supershells having pseudoquantum numbers $Kn_r + l$ which were observed earlier in numerical calculations.

We analyzed the dependence of the conditions of supershell formation on the cluster size, potential profile, and electron temperature and showed the relative role of the various supershells in the formation of beating oscillations observed in the mass spectra of the clusters.

The free-energy oscillations of the cluster electrons described by formula (15) introduced by us agree exceptionally well with the results of laborious calculations using direct quantum-mechanical or self-consistent methods. Our proposed approach is particularly effective for large clusters, i.e., in the range where the self-consistent calculations are the most laborious and difficult to interpret.

ACKNOWLEDGMENTS

This work was partly supported by the Russian Foundation for Basic Research (project no. 00-01-00397).

APPENDIX

We shall show how λ_μ, $\partial\lambda_\mu/\partial\mu$, and ν'_μ can be expressed in terms of the potential $U(r)$. The value of λ_μ is obtained from the condition

$$\lambda_\mu = p_\mu(r_0)r_0,$$

where r_0 is the extremum point of the function $p_\mu^2(r)r^2$ which is determined from the equation

$$U'(r_0)r_0 = p_\mu^2(r_0).$$

From this it follows that

$$\frac{\partial\lambda_\mu}{\partial\mu} = \frac{r_0^2}{\lambda_\mu}, \quad \nu'_\mu = \frac{\sqrt{2}\lambda_\mu}{\omega r_0},$$

where

$$\omega^2 = -(p_\mu^2(r)r^2)''_{r_0} = 2U''(r_0)r_0^2 + 6p_\mu^2(r_0).$$

REFERENCES

1. W. A. de Heer, Rev. Mod. Phys. **65**, 612 (1993).
2. M. Brack, Rev. Mod. Phys. **65**, 677 (1993).
3. O. Genzken and M. Brack, Phys. Rev. Lett. **76**, 3286 (1991).
4. H. Nishioka, K. Hansen, and B. R. Mottelson, Phys. Rev. B **42**, 9377 (1990).
5. K. Clemenger, Phys. Rev. B **44**, 12991 (1991).
6. R. Balian and C. Bloch, Ann. Phys. (N. Y.) **69**, 76 (1971).
7. V. M. Strutinsky, A. G. Magner, S. R. Ofengenden, and T. Dössing, Z. Phys. A **283**, 269 (1977).
8. S. Bjørnholm, in *Nuclear Physics Concepts in Atomic Cluster Physics*, Ed. by R. Schmidt, H. O. Lutz, and R. Dreizler (Springer-Verlag, New York, 1992), p. 26.
9. E. Koch, Phys. Rev. B **58**, 2329 (1998); E-prints archive, cond-mat/9803309.
10. D. A. Kirzhnits, Yu. E. Lozovik, and G. V. Shpatakovskaya, Usp. Fiz. Nauk **111**, 3 (1975) [Sov. Phys. Usp. **16**, 587 (1975)].
11. B. G. Englert, *Semiclassical Theory of Atoms*, in *Lecture Notes in Physics* (Springer-Verlag, New York, 1988), Vol. 300.
12. G. V. Shpatakovskaya, Teplofiz. Vys. Temp. **23**, 42 (1985).
13. E. A. Kuzmenkov and G. V. Shpatakovskaya, Int. J. Thermophys. **13**, 315 (1992).
14. G. V. Shpatakovskaya, Pis'ma Zh. Éksp. Teor. Fiz. **70**, 333 (1999) [JETP Lett. **70**, 334 (1999)]; E-prints archive, cond-mat/0001116.
15. C. Yannouleas and Uzi Landman, Phys. Rev. B **48**, 8376 (1993).
16. D. A. Kirzhnits and G. V. Shpatakovskaya, Zh. Éksp. Teor. Fiz. **62**, 2082 (1972) [Sov. Phys. JETP **35**, 1088 (1972)].

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 1, 2000, pp. 84–92.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 1, 2000, pp. 96–104.

ATOMS, SPECTRA, RADIATION

Collective Generation of Higher Optical Harmonics by Dipole Molecules

V. A. Kovarskiĭ and O. B. Prepelitsa*

Institute of Applied Physics, Academy of Sciences of Moldova, Chisinau, 277028 Moldova
**e-mail: exciton@cc.acad.md, referent@prozelit.com*

Received February 24, 2000

Abstract—An analysis is made of the collective resonant generation of higher harmonics by a spatially extended system of two-level molecules possessing an intrinsic electron dipole moment in the excited state. Frequency and angular dependences of the scattered radiation are studied. It is shown that for moderately small numbers of harmonics their intensity depends comparatively weakly on the number (plateau) and the intensity of the emitted harmonics then falls sharply as the number increases. The angular distribution of the harmonics is strongly anisotropic. It is also shown that collective effects significantly change the time profile of the generated harmonic pulse. In addition, as a result of cooperative effects harmonics of a certain parity are suppressed in the low-frequency part of the radiation spectrum. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The generation of higher optical harmonics by quantum systems is the subject of intensive theoretical and experimental research [1–19]. In particular, fairly detailed studies have been made of processes of resonant generation of higher radiation harmonics by molecules having different intrinsic dipole moments in the ground and excited states [15–19]. In this case, as was predicted in [16] and confirmed experimentally in [17, 18], harmonics are effectively generated at relatively low intensities of the exciting electromagnetic field. In this context we note that very high-intensity electromagnetic fields, of the order of atomic, are required for the effective generation of rf radiation by atoms. From this point of view, the resonant generation of harmonics by dipolar molecules has certain advantages compared with the generation of higher optical harmonics by atoms although it is inferior in terms of the possible generation of rf radiation, such as soft X-rays. It should be noted that the mechanisms for the generation of harmonics by atoms and dipole molecules, and also the corresponding parameters determining the characteristics of the spectrum (the intensity of the emitted harmonics, cutoff, and so on) differ. In particular, the generation of harmonics by atoms is accurately described by the well-known Kulander–Corkum model [7, 8] whereas the resonant generation of harmonics by dipole molecules considered here is a resonant multiphoton scattering process in which a single rf quantum is emitted as a result of the absorption of several quanta of the applied electromagnetic field (Fig. 1). In the first case, the cutoff in the harmonic emission spectrum is determined by the ponderomotive potential, i.e., it depends quadratically on the intensity of the exciting electromagnetic field whereas in the second case, as will be shown subsequently, it depends linearly on the field intensity and on the intrinsic electron dipole moment of the excited state.

In the present study we investigate the collective generation of higher harmonics by an extended system consisting of a large number of dipolar molecules. (A similar problem for a concentrated system of dipole molecules, i.e., systems having linear dimensions much shorter than the wavelength, was considered in an earlier study [19].) It should be stressed that in a fairly dense medium where there are a large number of emitters at distances of the order of the wavelength, effects caused by interaction of the emitters via the field of reemitted photons have a substantial influence on the characteristics of the harmonic generation process. This is primarily because the radiation intensity is directly related to the lifetime of the system in the excited state. Collectivization of the emitters leads to a sharp drop in the decay time of the system which in turn leads to a substantial change in the time profile of the emitted radiation pulse. This mechanism forms the basis of the

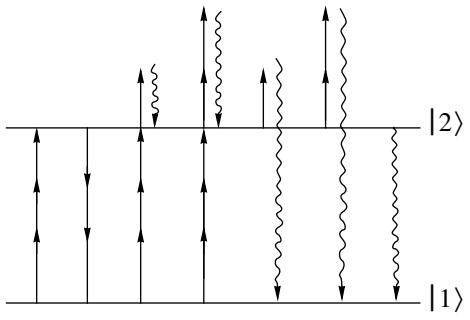


Fig. 1. Resonant multiphoton scattering process. The wavy lines indicate spontaneously generated photons, the straight lines denote photons of the applied electromagnetic field.

well-known Dicke superradiance process (see, for example, [20]). In processes involving the scattering of an electromagnetic field by molecules an important role is played by electronic–vibrational interactions which, although reducing the molecular lifetime in the excited state, also destroy the coherence of the system of emitters and are thus a destructive factor. As the object of the present investigation we consider macromolecules, for which a characteristic property is the existence of long-lived long-wavelength vibrations, for example, as a result of conformational degrees of freedom [21] which cause appreciable slowing of the vibrational relaxation. Macromolecules are also characterized by fairly high intrinsic electron dipole moments, for example, for the gamma-globulin protein molecule these values are of the order of several hundreds of Debye. This makes them interesting objects for nonlinear interaction with an external electromagnetic field. We know that the phase-matching conditions must be satisfied for effective generation of multifrequency radiation in spatially extended systems. In the model of two-level molecules considered below, phase-matching conditions can be satisfied under exact multiphoton resonance between the transition frequency of the system and the frequency of the external field. Since the refractive index tends to unity at the frequency of the external field, as a result of a large difference between the transition frequency of the system and the frequency of the higher harmonics, the system is optically transparent to rf radiation, i.e., the refractive index at these frequencies is also unity. In this case, the phase-matching conditions can easily be satisfied.

2. BASIC EQUATIONS FOR THE INTENSITY OF RADIATION HARMONICS

We shall consider an extended system of N two-level emitters (each having an intrinsic dipole moment in the excited state) in the field of a strong classical electromagnetic wave. The Hamiltonian of this system has the form

$$\begin{aligned}
\hat{H} &= \hat{H}_0 + \hat{V}(t) + \hat{W}(t) + \hat{H}_{\text{int}},\\
\hat{H}_0 &= \sum_{j=1}^{N} \hbar\omega_{21} R_{zj} + \sum_{\mathbf{k}} \hbar\omega_{\mathbf{k}} a_{\mathbf{k}}^{+} a_{\mathbf{k}},\\
\hat{V}(t) &= \sum_{j=1}^{N} (\mathbf{E}_0 \cdot \mathbf{d}_{22}) R_{zj} \cos(\omega_0 t - \mathbf{k}_0 \cdot \mathbf{r}_j),\\
\hat{W}(t) &= \sum_{j=1}^{N} (\mathbf{E}_0 \cdot \mathbf{d}_{21})(R_j^{+} + R_j^{-}) \cos(\omega_0 t - \mathbf{k}_0 \cdot \mathbf{r}_j),\\
\hat{H}_{int} &= i \sum_{\mathbf{k}} \sum_{j=1} (\mathbf{g}_{\mathbf{k}} \cdot \mathbf{d}_{21})(R_j^{+} + R_j^{-})\\
&\times (a_{\mathbf{k}}^{+} \exp(-i\mathbf{k} \cdot \mathbf{r}_j) - a_{\mathbf{k}} \exp(i\mathbf{k} \cdot \mathbf{r}_j)).
\end{aligned} \tag{1}$$

Here ω_{21} is the frequency of the transition between the ground and excited states of the emitter, $\mathbf{g}_{\mathbf{k}} = \sqrt{2\pi\hbar\omega_{\mathbf{k}}/V}\,\mathbf{e}_{\lambda}$, V is the quantization volume, $\mathbf{e}_{\lambda}$ is the polarization unit vector, $\omega_{\mathbf{k}}$ is the frequency of a photon having the wave vector $\mathbf{k}$, $R_j^{\pm}$, and R_{zj} are the quasispin operators of the jth emitter, $a_{\mathbf{k}}^{+}$ ($a_{\mathbf{k}}$) are the Bose operators of the electromagnetic field satisfying ordinary commutation rules:

$$[R_i^{+}, R_j^{-}] = 2\delta_{ij} R_{zj}, \quad [R_{zi}, R_j^{\pm}] = \pm\delta_{ij} R_j^{\pm},$$
$$[a_{\mathbf{k}_1}, a_{\mathbf{k}_2}^{+}] = \delta_{\mathbf{k}_1\mathbf{k}_2}, \quad [a_{\mathbf{k}_1}, a_{\mathbf{k}_2}] = 0,$$

$\mathbf{E}_0$ and ω_0 are the amplitude and frequency of the external electromagnetic field, respectively, $\mathbf{d}_{21}$ is the matrix element of the dipole moment of the transition from the ground to the excited state, and $\mathbf{d}_{22}$ is the intrinsic dipole moment of the excited state. It is then assumed that the matrix element $|\mathbf{d}_{22}|$ is much larger than $|\mathbf{d}_{21}|$; this situation is frequently obtained in dipolar molecules.

We shall assume that the transition frequency ω_{21} is a multiple of the frequency of the applied electromagnetic field ω_0, so that

$$\omega_{21} = n_0 \omega_0.$$

Multiphoton mixing of the ground and excited states described by the term $\hat{W}(t)$ in the Hamiltonian (1) will be analyzed in the resonance approximation. The corresponding constraint imposed on the field strength $\mathbf{E}_0$ and determining the limits of validity of the resonance approximation has the form

$$\frac{\mathbf{E}_0 \cdot \mathbf{d}_{21}}{\hbar\omega_0} \ll 1.$$

Since it is assumed that $|\mathbf{d}_{22}| \gg |\mathbf{d}_{21}|$, it is advisable to allow exactly for the interaction between the intrinsic dipole moment of the emitter excited states and the applied electromagnetic field, without having recourse to any approximations. For this purpose we perform a canonical transformation using a unitary operator having the following form:

$$U(t) = \exp\left(-\frac{i}{\hbar}\int_0^t d\tau \hat{V}(\tau)\right).$$

In this case, the statistical operator and also the dynamic variables of the system are transformed using well-known rules:

$$\rho(t) = U(t)\hat{\rho}(0)U^{+}(t), \quad Q(t) = U^{+}(t)\hat{Q}(t)U(t),$$
$$\hat{\rho}(0) = \hat{\rho}_a(0) \otimes \hat{\rho}_{ph}(0),$$

where $\hat{\rho}_a(0)$ and $\hat{\rho}_{ph}(0)$ are the statistical operators of the molecular and photon subsystems, respectively, at

time $t = 0$, and $\hat{Q}(t)$ is an arbitrary operator in the Heisenberg representation.

The transformed Hamiltonian has the form

$$H = H_0 + W(t) + H_{\text{int}}, \tag{2}$$

$$H_0 = \sum_{j=1}^{N} \hbar\omega_{21} R_{zj} + \sum_{\mathbf{k}} \hbar\omega_{\mathbf{k}} a_{\mathbf{k}}^{+} a_{\mathbf{k}}, \tag{3}$$

$$W(t) = \sum_{j=1}^{N} (\mathbf{E}_0 \cdot \mathbf{d}_{21})(R_j^{+} \exp(i\rho \sin(\omega_0 t - \mathbf{k}_0 \cdot \mathbf{r}_j)) + R_j^{-} \exp(-i\rho \sin(\omega_0 t - \mathbf{k}_0 \cdot \mathbf{r}_j)))\cos(\omega_0 t - \mathbf{k}_0 \cdot \mathbf{r}_j), \tag{4}$$

$$H_{\text{int}} = i\sum_{\mathbf{k}} \sum_{j=1}^{N} (\mathbf{g}_{\mathbf{k}} \cdot \mathbf{d}_{21})(R_j^{+} \exp(i\rho \sin(\omega_0 t - \mathbf{k}_0 \cdot \mathbf{r}_j)) + R_j^{-} \exp(-i\rho \sin(\omega_0 t - \mathbf{k}_0 \cdot \mathbf{r}_j))) \times (a_{\mathbf{k}}^{+} \exp(-i\mathbf{k} \cdot \mathbf{r}_j) - a_{\mathbf{k}} \exp(i\mathbf{k} \cdot \mathbf{r}_j)), \tag{5}$$

where

$$\rho = \frac{\mathbf{E}_0 \cdot \mathbf{d}_{22}}{\hbar\omega_0}.$$

It follows from expressions (4) and (5) that the interaction between the intrinsic dipole moment of the excited state and the electromagnetic field makes a contribution to the evolution of the system determined by the parameter ρ.

It can be shown that neglecting any delay, the radiation power scattered by the system is defined by

$$I = \sum_{\mathbf{k}} \hbar\omega_{\mathbf{k}} \frac{d}{dt} \langle a_{\mathbf{k}}^{+}(t) a_{\mathbf{k}}(t) \rangle,$$

where $\langle \ldots \rangle = \mathrm{Sp}(\rho(t)\ldots)$ denotes averaging with the total statistical operator of the system. We write the equation of motion for the operators of the electromagnetic field. In accordance with formulas (2)–(5), this has the form

$$\frac{da_{\mathbf{k}}}{dt} = -i\omega_{\mathbf{k}} a_{\mathbf{k}} + \sum_{j=1}^{N} \frac{\mathbf{g}_{\mathbf{k}} \cdot \mathbf{d}_{21}}{\hbar} \times (R_j^{+} \exp(i\rho \sin(\omega_0 t - \mathbf{k}_0 \cdot \mathbf{r}_j)) + R_j^{-} \exp(-i\rho \sin(\omega_0 t - \mathbf{k}_0 \cdot \mathbf{r}_j)))\exp(-i\mathbf{k} \cdot \mathbf{r}_j).$$

Using the Born–Markov approximation for quasispin operators:

$$R_j^{\pm}(t - t') \approx R_j^{\pm}(t) \exp(\mp i\omega_{21} t'),$$

and also the well-known expansion of the exponential functions in terms of Bessel functions of the real argument:

$$\exp(i\rho \sin\omega_0 t) = \sum_{m=-\infty}^{\infty} J_m(\rho) \exp(im\omega_0 t),$$

we can easily obtain the following solution of the equation of motion for the operators of the photon subsystem:

$$a_{\mathbf{k}}(t) = a_{\mathbf{k}}^{e}(t) + a_{\mathbf{k}}^{s}(t), \tag{6}$$

$$a_{\mathbf{k}}^{e}(t) = a_{\mathbf{k}}^{e}(0) \exp(-i\omega_{\mathbf{k}} t), \tag{7}$$

$$a_{\mathbf{k}}^{s}(t) = \pi \frac{\mathbf{g}_{\mathbf{k}} \cdot \mathbf{d}_{21}}{\hbar} \times \sum_{m=1}^{\infty} \sum_{j=1}^{N} (R_j^{+} \exp(-i(m + n_0)\omega_0 t) \times \exp(-i(\mathbf{k} - m_0\mathbf{k}_0) \cdot \mathbf{r}_j) J_{-(m+n_0)}(\rho) + R_j^{-} \exp(-i(m - n_0)\omega_0 t) \times \exp(-i(\mathbf{k} - m\mathbf{k}_0) \cdot \mathbf{r}_j) J_{m-n_0}(\rho))\delta(\omega_{\mathbf{k}} - m\omega_0). \tag{8}$$

At this point it should be noted that the validity of the Born–Markov approximation imposes a constraint on the photon residence time (lifetime) in the system. Specifically, the time taken for transit of a photon through the system of emitters should be much shorter than the characteristic times of the quantum transitions.

Using these last equations, we easily obtain an expression for the intensity of the radiation having the wave vector $\mathbf{k}$ scattered per unit solid angle:

$$\frac{dI(\mathbf{k})}{d\Omega_{\mathbf{k}}} = \frac{1}{4} \sum_{m=1}^{\infty} \hbar\omega_{\mathbf{k}} \left(\frac{\mathbf{g}_{\mathbf{k}} \cdot \mathbf{d}_{21}}{\hbar} \right)^2 \times \sum_{l,j=1}^{N} (J_{m-n_0}^2(\rho) \langle R_l^{+} R_j^{-} \rangle + J_{m+n_0}^2(\rho) \langle R_l^{-} R_j^{+} \rangle + J_{m+n_0}(\rho) J_{n_0-m}(\rho)(\langle R_l^{+} R_j^{+} \rangle + \langle R_l^{-} R_j^{-} \rangle)) \times \exp(i(\mathbf{k} - m\mathbf{k}_0) \cdot (\mathbf{r}_l - \mathbf{r}_j))\delta(\omega_{\mathbf{k}} - m\omega_0).$$

This last expression can be given in the form:

$$\frac{dI(\mathbf{k})}{d\Omega_{\mathbf{k}}} = \sum_{m=1}^{\infty} \frac{dI(\mathbf{k}_m)}{d\Omega_{\mathbf{k}_m}},$$

$$\frac{dI(\mathbf{k}_m)}{d\Omega_{\mathbf{k}_m}} = \frac{\hbar\omega_0}{8\pi\tau_{sp}} \sum_{l,j=1}^{N} m(J_{m-n_0}^2(\rho) \langle R_l^{+} R_j^{-} \rangle + J_{m+n_0}^2(\rho) \langle R_l^{-} R_j^{+} \rangle + J_{m+n_0}(\rho) J_{m-n_0}(\rho) \tag{9}$$

$$\times (\langle R_l^+ R_j^+ \rangle + \langle R_l^- R_j^- \rangle)) \exp(i(\mathbf{k} - m\mathbf{k}_0) \cdot (\mathbf{r}_l - \mathbf{r}_j)).$$

The quantity $dI(\mathbf{k}_m)/d\Omega_{\mathbf{k}_m}$ determines the intensity of the scattered electromagnetic wave having the wave vector $\mathbf{k}_m$ ($|\mathbf{k}_m| = m\omega_0/c$ where c is the velocity of light in vacuum) per unit solid angle. Here

$$\frac{1}{\tau_{sp}} = \frac{4}{3}\frac{\omega_{21}^3}{\hbar c^3}|\mathbf{d}_{21}|^2$$

is the probability of one-photon spontaneous decay of the emitter excited state.

In the derivation of these formulas we assumed that at time $t = 0$ the photon subsystem is in the vacuum state so that $\langle \ldots a_{\mathbf{k}}^e(t) \rangle = 0$ and thus only the source components $a_{\mathbf{k}}^e(t)$ of the electromagnetic field operators contribute to the radiation intensity. It is deduced from this formula that the frequency distribution of the radiation intensity is determined mainly by the functional dependence of the Bessel functions on the number and argument. Since the Bessel function $J_m(\rho)$ depends comparatively weakly on ρ in the range $\rho > m$, harmonics having numbers $m < \rho + n_0$ form a plateau in the radiation spectrum. From $\rho \sim m$ the value of $J_m(\rho)$ decreases sharply (exponentially) so that $m \sim \rho + n_0$ determines the cutoff in the radiation spectrum. The numerical values of the harmonic intensities are determined by the correlation functions $\langle R_i^\pm R_j^\mp \rangle$, $\langle R_i^\pm R_j^\pm \rangle$, $i, j = 1, 2, \ldots, N$. In order to find the time evolution of these correlation functions we need to derive and solve kinetic equations for the averages $\langle R_i^\pm R_j^\mp \rangle$ and $\langle R_i^\pm R_j^\pm \rangle$.

3. EQUATIONS OF MOTION FOR THE AVERAGES OF THE OPERATORS OF THE EMITTER SYSTEM

We write the equation of motion for the average of a certain operator of the system of emitters $Q(t)$:

$$\left\langle \frac{dQ}{dt} \right\rangle = \frac{i}{\hbar}\langle [H, Q(t)] \rangle.$$

Having eliminated the electromagnetic field operators using formulas (6)–(8), after the ordinary transformations (see, for example [20–22]) we can obtain the following kinetic equation for the averages:

$$\left\langle \frac{dQ}{dt} \right\rangle = i\Omega_{n_0}(\rho) \sum_{j=1}^{N} \langle [R_j^+ + R_j^-, Q] \rangle$$

$$+ \sum_{l,j=1}^{N} \sum_{m=1}^{\infty} (a_m (C_{jl}^m \langle R_j^+ [Q, R_l^-] \rangle + C_{lj}^m \langle [R_l^+, Q] R_j^- \rangle)$$

$$- b_m (C_{lj}^m \langle [Q, R_l^-] R_j^+ \rangle + C_{jl}^m \langle R_j^- [R_l^+, Q] \rangle) \qquad (10)$$

$$+ d_m (C_{jl}^m \langle R_j^+ [Q, R_l^+] \rangle + \langle R_j^- [Q, R_l^-] \rangle)$$

$$+ C_{lj}^m (\langle [R_l^+, Q] R_j^+ \rangle + \langle [R_l^-, Q] R_j^- \rangle)),$$

where the coefficients of the equation have the form

$$\Omega_{n_0}(\rho) = \omega_{21}\frac{\mathbf{E}_0 \cdot \mathbf{d}_{21}}{\mathbf{E}_0 \cdot \mathbf{d}_{22}} J_{n_0}(\rho),$$

$$a_m = \frac{1}{2\tau_{sp}}\left(\frac{m}{n_0}\right)^3 J_{m-n_0}^2(\rho),$$

$$b_m = \frac{1}{2\tau_{sp}}\left(\frac{m}{n_0}\right)^3 J_{m+n_0}^2(\rho),$$

$$d_m = \frac{1}{2\tau_{sp}}\left(\frac{m}{n_0}\right)^3 J_{m+n_0}(\rho) J_{n_0-m}(\rho).$$

Here

$$C_{lj}^m = \exp(i|\mathbf{k}_m||\mathbf{r}_j - \mathbf{r}_l|)$$
$$\times \exp(-im\mathbf{k}_0 \cdot (\mathbf{r}_j - \mathbf{r}_l))(i|\mathbf{k}_m||\mathbf{r}_j - \mathbf{r}_l|)^{-1}$$

are the diffraction factors which take into account the spatial extent of the system.

The coefficients of equation (10) describe the following processes: the Rabi frequency $\Omega_{n_0}(\rho)$ determines the n_0 quantum nutation of the system, the coefficients a_m, b_m and d_m determine the probabilities of dissipative processes involving m quanta of the external electromagnetic field (Fig. 1). For example, the coefficient a_m describes the rate of relaxation of the system and b_m describes its rate of excitation. We need to stress the differences between the induced excitation which is determined by the Rabi frequency $\Omega_{n_0}(\rho)$ and is a coherent process, and the excitation described by b_m which can be arbitrarily called spontaneously induced excitation since a single spontaneously generated photon and several photons of the applied electromagnetic field participate in this process (Fig. 1). Since the scattering of the electromagnetic field is essentially a dissipative process, the probability of harmonic generation is directly related to the values of a_m, b_m and d_m. It should be noted that processes caused by interaction of electrons with the phonon subsystem (for example, nonradiative transitions leading to an increase in the rates of longitudinal and transverse relaxation of the emitters) are not reflected in equation (10). Thus, strictly speaking, this equation describes the evolution of the system over times much shorter than the times for nonradiative transitions T_s in the molecules:

$$t \ll T_s. \qquad (11)$$

Using (10), we can obtain the following closed system of equations for the averages:

$$\frac{d\langle R_{zj}\rangle}{dt} = -i\Omega_{n_0}(\rho)(\langle R_j^+\rangle - \langle R_j^-\rangle)$$
$$-\sum_{n=1}^{N}\sum_{m=1}^{\infty}((C_{nj}^m a_m - C_{jn}^m b_m)\langle R_n^+ R_j^-\rangle + (C_{jn}^m a_m - C_{nj}^m b_m)\langle R_j^+ R_n^-\rangle), \quad (12)$$

$$\frac{d\langle R_l^+ R_j^-\rangle}{dt} = 2i\Omega_{n_0}(\rho)(\langle R_l^+\rangle\langle R_{zj}\rangle - \langle R_j^-\rangle\langle R_{zl}\rangle)$$
$$+2\sum_{n=1}^{N}\sum_{m=1}^{\infty}((C_{nl}^m a_m - C_{ln}^m b_m)\langle R_n^+ R_j^-\rangle\langle R_{zl}\rangle + (C_{jn}^m a_m - C_{nj}^m b_m)\langle R_l^+ R_n^-\rangle\langle R_{zj}\rangle), \quad l\neq j, \quad (13)$$

$$\frac{d\langle R_l^+ R_j^+\rangle}{dt} = -2i\Omega_{n_0}(\rho)(\langle R_l^+\rangle\langle R_{zj}\rangle + \langle R_l^+\rangle\langle R_{zj}\rangle)$$
$$+2\sum_{n=1}^{N}\sum_{m=1}^{\infty}(C_{nl}^m(a_m - b_m)\langle R_n^+ R_j^+\rangle\langle R_{zl}\rangle + C_{nj}^m(a_m - b_m)\langle R_n^+ R_l^+\rangle\langle R_{zj}\rangle), \quad l\neq j, \quad (14)$$

$$\frac{d\langle R_j^+\rangle}{dt} = -2i\Omega_{n_0}(\rho)\langle R_{zj}\rangle + 2\sum_{n=1}^{N}\sum_{m=1}^{\infty}C_{nj}^m(a_m - b_m)\langle R_n^+\rangle\langle R_{zj}\rangle. \quad (15)$$

For the derivation of equations (12)–(15) we used the semiclassical decoupling of the correlation functions [20]:

$$\langle R_l^+ R_{zn} R_j^-\rangle = \langle R_l^+ R_j^-\rangle\langle R_{zn}\rangle,$$
$$\langle R_l^\pm R_{zn} R_j^\pm\rangle = \langle R_l^\pm R_j^\pm\rangle\langle R_{zn}\rangle,$$
$$\langle R_l^+ R_{zj}\rangle = \langle R_l^+\rangle\langle R_{zj}\rangle,$$

and we also neglected the noncommutativity of the operators $R_j^\pm$ and R_{zj} below the averaging sign. As we know, this last factor leads to a loss of terms which allow for noncollective effects such as spontaneous emission, which only holds for short time intervals $t \ll a_m^{-1}, b_m^{-1}, d_m^{-1}$. In this particular case, this is quite justified because in molecules, processes caused by interaction with the phonon subsystem generally take place over times much shorter than the times of the radiative transitions, i.e.,

$$T_s \ll a_m^{-1}, b_m^{-1}, d_m^{-1}.$$

Thus, the condition for validity of equations (12)–(15) [see formula (11)] is stronger which demonstrates the validity of the approximation used. We also note that neglecting the longitudinal and transverse relaxation in the equations describing the collective evolution of the system of emitters in an external electromagnetic field imposes a lower constraint on its intensity. A correlation between the emitters is induced during migration of the system population. This is usually achieved as a result of spontaneous decay and stimulated transitions induced by the external electromagnetic field. In our case collectivization of the system only occurs as a result of stimulated transitions. In order for collectivization to occur, the probability of stimulated transitions must be much higher than the probability of destructive processes leading to a phase shift, in this case processes associated with the phonon subsystem. This condition is expressed by the inequality

$$\Omega_{n_0}(\rho) \ll T_s^{-1}.$$

We shall assume that the system of emitters occupies a cylindrical region of space. The axis of the cylinder is directed along the wave vector of the external electromagnetic field which is parallel to the z-axis. We also assume that the linear dimensions of the system are much greater than the wavelengths of the applied and scattered electromagnetic fields:

$$\lambda_m/L, \lambda_m/R_0 \ll 1, \quad \lambda_m = 2\pi c/m\omega_0, \quad m = 1, 2, 3\ldots$$

(L and R_0 are the length and radius of the cylinder, respectively) and the average distance between the emitters is much less than the wavelength λ_m. At the end of the previous section we noted that the probability of the generation of harmonics having numbers $m > \rho + n_0$ is negligible. It is thus sufficient for the conditions described above to be satisfied for harmonics having numbers $m \le \rho + n_0$.

We subsequently require the explicit form of the expression

$$\alpha(\mathbf{k}_m) = \frac{1}{4\pi N^2}\sum_{l,j=1}^{N}\exp(i(\mathbf{k}_m - m\mathbf{k}_0)\cdot(\mathbf{r}_l - \mathbf{r}_j)).$$

In spherical coordinates this can be written as

$$\alpha(\mathbf{k}_m) = \frac{1}{\pi}\times\left[\frac{\sin(1-\cos\theta)|\mathbf{k}_m|L/2}{(1-\cos\theta)|\mathbf{k}_m|L/2}\,\frac{J_1(|\mathbf{k}_m|R_0\sin\theta)}{|\mathbf{k}_m|R_0\sin\theta}\right]^2. \quad (16)$$

Using equations (12) and (13) it is easy to show that the system of equations (12)–(15) has two integrals of motion:

$$\frac{d}{dt}\sum_{l,j=1}^{N}\exp(i(\mathbf{k}_n - n\mathbf{k}_0)\cdot(\mathbf{r}_l - \mathbf{r}_j))\times(\langle R_l^- R_j^-\rangle + \langle R_l^+ R_j^+\rangle) = 0,$$

$$\frac{d}{dt}\sum_{l,j=1}^{N}\exp(i(\mathbf{k}_n - n\mathbf{k}_0)\cdot(\mathbf{r}_l - \mathbf{r}_j))$$
$$\times\left(\langle R_{zl}\rangle\langle R_{zj}\rangle + \frac{1}{2}(\langle R_l^+ R_j^-\rangle + \langle R_j^+ R_l^-\rangle)\right) = 0,$$

where $l \neq j$.

It follows from this last relationship that for times shorter than the times of the noncollective transitions (spontaneous transitions, transitions involving phonons, and so on) the square of the Bloch vector is conserved:

$$\frac{1}{4\pi}\sum_{l,j=1}^{N}\exp(i(\mathbf{k}_n - n\mathbf{k}_0)\cdot(\mathbf{r}_l - \mathbf{r}_j))$$
$$\times\left(\langle R_{zl}\rangle\langle R_{zj}\rangle + \frac{1}{2}(\langle R_l^+ R_j^-\rangle + \langle R_j^+ R_l^-\rangle)\right) = \alpha(\mathbf{k}_n)\frac{N^2}{4}.$$

We introduce the notation

$$R = \sum_{j=1}^{N}\langle R_{zj}\rangle, \quad D = \sum_{j=1}^{N}(\langle R_j^+\rangle - \langle R_j^-\rangle),$$

$$S(\mathbf{k}_m) = \frac{1}{4\pi}$$
$$\times\sum_{l,j=1}^{N}\exp(i(\mathbf{k}_m - m\mathbf{k}_0)\cdot(\mathbf{r}_l - \mathbf{r}_j))\langle R_l^+ R_j^-\rangle,$$

$$\Phi(\mathbf{k}_m) = \frac{1}{4\pi}\sum_{l,j=1}^{N}\exp(i(\mathbf{k}_m - m\mathbf{k}_0)\cdot(\mathbf{r}_l - \mathbf{r}_j))$$
$$\times(\langle R_l^+ R_j^+\rangle + \langle R_l^- R_j^-\rangle).$$

The integrals of motion, and also the weak dependence of the averages $\langle R_{zj}\rangle$ on the position of the atom, which allows us to use approximate rules for summation of expressions of the type [20]

$$\sum_{l=1}^{N}\langle R_{zl}\rangle\exp(i(\mathbf{k}_n - n\mathbf{k}_0)\cdot(\mathbf{r}_l - \mathbf{r}_j))C_{nl}^m$$
$$\approx\frac{R}{N}\sum_{l=1}^{N}\exp(i(\mathbf{k}_n - n\mathbf{k}_0)\cdot(\mathbf{r}_l - \mathbf{r}_j))C_{nl}^m,$$

means that the system (12)–(15) can be simplified substantially and reduced to the following form:

$$\frac{dR}{dt} = -i\Omega_{n_0}(\rho)D - 2\Gamma N\left(\frac{1}{4} - \frac{R^2}{N^2}\right), \tag{17}$$

$$\frac{dD}{dt} = -4i\Omega_{n_0}(\rho)R + 2\frac{\Gamma}{N}DR, \tag{18}$$

$$S(\mathbf{k}_m) = \alpha(\mathbf{k}_m)N^2\left(\frac{1}{4} - \frac{R^2}{N^2}\right), \tag{19}$$

$$\Phi(\mathbf{k}_m) = -2\alpha(\mathbf{k}_m)N^2\left(\frac{1}{4} - \frac{R^2}{N^2}\right), \tag{20}$$

where

$$\Gamma = N\sum_{m=1}^{\infty}\alpha_m(a_m - b_m), \quad \alpha_m = \int d\Omega_{\mathbf{k}_m}\alpha(\mathbf{k}_m).$$

For large and small Fresnel numbers $f_m = \pi R_0^2/\lambda_m L$ the expression α_m can be calculated analytically:

$$\alpha_m = \frac{\lambda_m}{4L}, \quad f_m \ll 1,$$

$$\alpha_m = \frac{\lambda_m^2}{4\pi^2 R_0^2}, \quad f_m \gg 1.$$

In the derivation of (19) and (20) we neglected small terms of the order $1/N$. Their role is only significant near the point $R = \pm N/2$, so that the formulas (19) and (20) can be applied everywhere except for the small vicinity of $R = \pm N/2$. Note that the value of $(2\Gamma)^{-1}$ determines the probability of collective spontaneous decay of an inverted system of emitters. Collective processes only take place in the system if, in addition to $\Omega_{n_0}(\rho) \gg T_s^{-1}$, the condition $\Gamma \gg T_s^{-1}$ is also satisfied. We shall assume that these conditions are satisfied.

Taking into account expressions (9), (19) and (20) we write the formula for the intensity of the emission at frequency $m\omega_0$ per unit solid angle in the compact form:

$$\frac{dI(\mathbf{k}_m)}{d\Omega_{\mathbf{k}_m}} = \frac{m^4\hbar\omega_0}{8\pi n_0^3\tau_{sp}}\alpha(\mathbf{k}_m)$$
$$\times N^2(J_{m-n_0}(\rho) - J_{m+n_0}(\rho))^2\left(\frac{1}{4} - \frac{R^2}{N^2}\right). \tag{21}$$

Thus, the integrated intensity of he scattered radiation at frequency $m\omega_0$ has the form

$$I(m\omega_0) = \frac{m^4\hbar\omega_0}{8\pi n_0^3\tau_{sp}}$$
$$\times\alpha_m N^2(J_{m-n_0}(\rho) - J_{m+n_0}(\rho))^2\left(\frac{1}{4} - \frac{R^2}{N^2}\right). \tag{22}$$

We shall analyze these expressions. The intensity of the collective emission increases as the correlations in the system increase and reaches a maximum when the system of emitters is completely correlated and the population difference is zero. The directionality of the harmonic emission is determined by the angular depen-

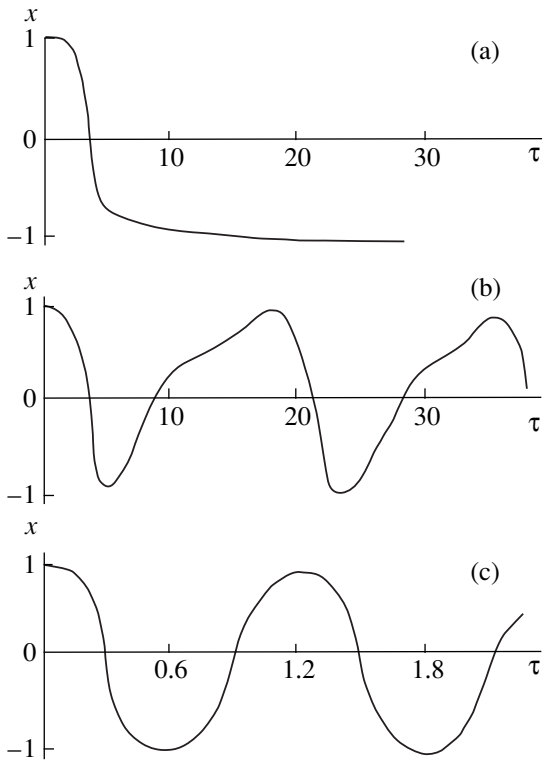


Fig. 2. The function $x(\tau)$ for various values of the parameter $\Theta = 0.1$ (a), 1.1 (b), and 5 (c).

dence of the function $\alpha(\mathbf{k}_m)$). In accordance with formula (16), the angular distribution of the radiation intensity has a narrow maximum at the point $\theta = 0$. The half-width of the distribution depends on the Fresnel number and is determined by the relationships $\Delta\theta \leq 2\pi/|\mathbf{k}_m|L$ for $f_m \ll 1$ and $\Delta\theta \leq (|\mathbf{k}_m|R_0)^{-2}$ for $f_m \gg 1$. A quadratic dependence of the radiation intensity on the number of emitters N is only obtained in the range $0 \leq \theta < \Delta\theta$. For $\theta \gg \Delta\theta$ we have

$$\frac{dI(\mathbf{k}_m)}{d\Omega_{\mathbf{k}_m}} \sim \frac{\hbar\omega_0 N^2}{2\pi^2 n_0^2 |\mathbf{k}_0|^4 \tau_{sp} L^2 R_0^2} \leq \frac{N\hbar\omega_0}{\tau_{sp}}.$$

In addition, it is easy to show that the integral intensity of the mth harmonic is proportional to the square of the density of emitters. It has been noted that the dependence of the harmonic intensity on the number has a characteristic form for this type of spectra [1–19, 23, 24]. For values of $m < \rho$ the harmonic intensity depends comparatively weakly on ρ (plateau), in the range $0 \leq \theta < \Delta\theta$ a slight increase in intensity is observed with increasing m as a result of the factor m^4 [see formula (21)], and then from values of $m \sim n_0 + \rho$ onward the intensity of the harmonic radiation drops rapidly. Note that for $\rho \gg n_0$ in the range $m < n_0$ there is also a dependence on the parity of the harmonic number m. Harmonics satisfying the conditions $n_0 - m = 2k$, $k = 1, 2, \ldots, m < n_0$ are suppressed to a certain extent. From the mathematical point of view, this is caused by the presence of the third term after the summation sign in formula (9). It is easy to see that if correlations between the emitters are neglected, this term is zero. Thus, the dependence of the harmonic intensity on the parity of the number is a purely collective effect.

In order to obtain a quantitative result we need to solve equations (17) and (18). It is convenient to introduce the new variables:

$$x = \frac{2R}{N}, \quad y = -2i\frac{\Omega_{n_0}(\rho)}{\Gamma N}D,$$

$$\tau = \Gamma t, \quad \Theta = \frac{2\Omega_{n_0}(\rho)}{\Gamma},$$

and then equations (17) and (18) have the form

$$\frac{dx}{d\tau} = x^2 - 1 + y, \tag{23}$$

$$\frac{dy}{d\tau} = x(y - \Theta^2). \tag{24}$$

The system of equations (23), (24) can be reduced to a single equation for the population difference:

$$\left(\frac{dx}{d\tau} - x^2 + 1\right)^2 = \Theta^2(1 - x^2). \tag{25}$$

The solutions of equation (25) with the initial conditions $x(0) = \pm 1$, $y(0) = 0$ corresponding to the initial and ground states of the emitter system at the instant of switching on an external electromagnetic field $\tau = 0$ have the form

$$x(\tau) = \pm\frac{\Theta^2 - \left(1 \pm \sqrt{1-\Theta^2}\coth\left(\sqrt{1-\Theta^2}(\tau \mp \tau_{01})/2\right)\right)^2}{\Theta^2 + \left(1 \pm \sqrt{1-\Theta^2}\coth\left(\sqrt{1-\Theta^2}(\tau \mp \tau_{01})/2\right)\right)^2}, \tag{26}$$

$$\Theta < 1,$$

$$x(\tau) = \pm\frac{\Theta^2 - \left(1 \mp \sqrt{\Theta^2-1}\tan\left(\sqrt{\Theta^2-1}(\tau \pm \tau_{02})/2\right)\right)^2}{\Theta^2 + \left(1 \mp \sqrt{\Theta^2-1}\tan\left(\sqrt{\Theta^2-1}(\tau \mp \tau_{02})/2\right)\right)^2}, \tag{27}$$

$$\Theta > 1,$$

$$\tau_{01} = \frac{2}{\sqrt{1-\Theta^2}}\operatorname{arctanh}\frac{1}{\sqrt{1-\Theta^2}},$$

$$\tau_{02} = \frac{2}{\sqrt{\Theta^2-1}}\arctan\frac{1}{\sqrt{\Theta^2-1}}.$$

For $\Theta < 1$ the motion of the level population is an aperiodic function and with time the system reaches a

steady state. This is because the probability of collective spontaneous decay is higher than the probability of stimulated excitation of the system which means in particular that the applied electromagnetic field cannot completely invert the system of emitters. For $\Theta > 1$ the probability of stimulated transitions is higher than the probability of collective decay of the system and the change in population is a periodic function so that in this time interval the system has no steady state. The value $\Theta = 1$, separating the two qualitatively different regimes of population migration and harmonic generation, aperiodic and periodic, can be considered to be the critical point of a nonequilibrium phase transition induced by the external electromagnetic field [25–28]. A comparison of (26), (27) with the expression describing population migration in two-level systems neglecting collectivization shows that collective effects slow the excitation of the emitter system and accelerate its relaxation. For a weak external electromagnetic field $\Theta \ll 1$ the case where the system is completely inverted at time $\tau = 0$, $x(0) = 1$, is of particular interest. At times $\tau < \Gamma/\Omega_{n_0}(\rho)$ stimulated processes play a comparatively minor role in the population migration and collective processes make the main contribution to the population evolution. In accordance with (26), (21), and (22) at time $\tau \sim \tau_{01}$ the population changes abruptly (Fig. 2a), accompanied by a multifrequency superradiance pulse. In this case, since the conditions $\Theta \ll 1$ and $\rho \gg n_0$ (which ensures effective generation of higher harmonics) are independent, the superradiance pulse also contains higher harmonics exceeding the transition frequency of the emitter $n_0\omega_0$. Thus, the external electromagnetic field leads to the generation of rf radiation and an appreciable reduction in the pulse duration is caused by collective processes in the system. For values of the parameter $\Theta \gtrsim 1$ slow excitation of the system ends in rapid relaxation (Fig. 2b) and the relatively slow change in the intensity of the generated harmonics is replaced by an abrupt increase. For the case of a very strong field $\Theta \gg 1$, collective effects barely influence the evolution of the system population (Fig. 2c) and consequently the time dependence of the intensity of the emitted harmonics. Thus, collective effects strongly influence the generation of higher harmonics for values of the parameter $\Theta \leq 1$.

4. CONCLUSIONS

This analysis of the generation of higher optical harmonics has shown that substantial collectivization of emitters occurs, even in strong electromagnetic fields for which the conventional resonance approximation cannot be used. In this case, the concept of a strong field is to a considerable extent determined not by the individual characteristics of an isolated emitter but by the collective parameters of the system as a whole, in this case by the ratio of the Rabi frequency to the probability of collective decay of the system. The collective generation of higher harmonics can be achieved experimentally by exposing a system of macromolecules such as gamma globulin molecules for which $\omega_{21} \approx 4.8 \times 10^{15}$ s^{-1}, $|\mathbf{d}_{21}| \approx 1$ D, $|\mathbf{d}_{22}| \approx 100$ D [29] to short electromagnetic field pulses of duration τ_f ($\Omega_{n_0}(\rho)$, $\Gamma \gg \tau_f^{-1} \gg T_s^{-1}$). At an external field frequency $\omega_0 \approx 1.6 \times 10^{15}$ s^{-1} and strength $|\mathbf{E}_0| \approx 1.5 \times 10^7$ V/cm we have $\rho \approx 7$ and harmonics having numbers up to $n_{\max} \approx 10$ are effectively generated.

ACKNOWLEDGMENTS

This work was supported in part by INTAS (project no. 99-01495).

REFERENCES

1. A. L'Huillier, K. J. Schafer, and K. C. Kulander, J. Phys. B **24**, 3315 (1991).
2. M. Ferray, A. L'Huillier, X. F. Li, *et al.*, J. Phys. B **21**, L31 (1988).
3. L.-A. Lompre, A. L'Huillier, M. Ferray, *et al.*, J. Opt. Soc. Am. B **7**, 754 (1990).
4. N. Sarakura, K. Hata, T. Adachi, *et al.*, Phys. Rev. A **43**, 1669 (1991).
5. J. J. Macklin, J. D. Kmetec, and C. L. Gordon, Phys. Rev. Lett. **70**, 766 (1993).
6. A. L'Huillier and Ph. Balcou, Phys. Rev. Lett. **70**, 774 (1993).
7. P. B. Corkum, Phys. Rev. Lett. **71**, 1994 (1993).
8. J. L. Krause, K. J. Schafer, and K. C. Kulander, Phys. Rev. Lett. **68**, 3535 (1992).
9. W. Becker, S. Long, and J. K. McIver, Phys. Rev. A **50**, 1540 (1994).
10. J. Peatross and D. D. Meyerhofer, Phys. Rev. A **51**, R906 (1995).
11. J. V. Watson, A. Sampera, and K. Burnett, Phys. Rev. A **51**, 1458 (1995).
12. D. F. Zaretskiĭ and É. A. Nersesov, Zh. Éksp. Teor. Fiz. **107**, 79 (1995) [JETP **80**, 41 (1995)].
13. D. F. Zaretskiĭ and É. A. Nersesov, Zh. Éksp. Teor. Fiz. **109**, 1994 (1996) [JETP **82**, 1073 (1996)].
14. V. T. Platonenko and V. V. Strelkov, Zh. Éksp. Teor. Fiz. **110**, 1641 (1996) [JETP **83**, 902 (1996)].
15. M. Lewenstein, Ph. Balcou, M. Yu. Ivanov, *et al.*, Phys. Rev. A **49**, 2117 (1994).
16. V. A. Kovarskiĭ and N. F. Perel'man, Zh. Éksp. Teor. Fiz. **60**, 509 (1971) [Sov. Phys. JETP **33**, 274 (1971)].
17. V. A. Kovarskiĭ and N. A. Ferdman, Pis'ma Zh. Éksp. Teor. Fiz. **15**, 483 (1972) [JETP Lett. **15**, 342 (1972)].
18. V. A. Kovarsky, B. S. Pillip, and E. V. Kovarsky, Phys. Lett. A **226**, 321 (1997).
19. V. A. Kovarskiĭ and O. B. Prepelitsa, Zh. Éksp. Teor. Fiz. **113**, 1206 (1998) [JETP **86**, 657 (1998)].

20. A. V. Andreev, V. I. Emel'yanov, and Yu. A. Il'inskiĭ, in *Super-Radiation* (Nauka, Moscow, 1988), p. 287.
21. M. V. Vol'kenshteĭn, *Molecular Biophysics* (Nauka, Moscow, 1975; Academic, New York, 1977).
22. N. N. Bogolyubov, Jr., B. I. Sadovnikov, and A. S. Shumovskiĭ, in *Mathematical Methods in Statistical Mechanics of Model Systems* (Nauka, Moscow, 1989), p. 295.
23. O. B. Prepelitsa, Opt. Spektrosk. **81**, 377 (1996) [Opt. Spectrosc. **81**, 340 (1996)].
24. O. B. Prepelitsa, Opt. Spektrosk. **85**, 825 (1998) [Opt. Spectrosc. **85**, 757 (1998)].
25. S. Ya. Kilin, Zh. Éksp. Teor. Fiz. **82**, 63 (1982) [Sov. Phys. JETP **55**, 38 (1982)].
26. S. V. Lavande, D. C. Khandekar, K. V. Bhagvat, and B. N. Jagtap, Physica A (Amsterdam) **134**, 598 (1986).
27. N. A. Enaki, Zh. Éksp. Teor. Fiz. **98**, 783 (1990) [Sov. Phys. JETP **71**, 435 (1990)].
28. N. A. Enaki and O. B. Prepelitsa, Opt. Spektrosk. **72**, 598 (1992) [Opt. Spectrosc. **72**, 322 (1992)].
29. V. A. Kovarskiĭ, Usp. Fiz. Nauk **169**, 889 (1999).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 1, 2000, pp. 93–105.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 1, 2000, pp. 105–118.

PLASMA, GASES

Nonlinear Theory of Resonant Beam–Plasma Interaction: Nonrelativistic Case

Yu. V. Bobylev, M. V. Kuzelev, and A. A. Rukhadze*

Institute of General Physics, Russian Academy of Sciences, Moscow, 117942 Russia
**e-mail: rukh@fpl.gpi.ru*

Received January 13, 2000

Abstract—Analytic and numerical methods are used to study the nonlinear dynamics of the resonant interaction between a dense nonrelativistic electron beam and a plasma in a spatially bounded system. Regimes such as collective (Raman) and single-particle (Thomson) Cherenkov effects are considered. It is shown that in the first case, the motion of both the beam and plasma electrons exhibits significant nonlinearities. However, because of the weak coupling between the beam and the plasma, the nonlinear dynamics of the instability can be studied analytically and it can be strictly shown that saturation of instability is caused by a nonlinear shift of the radiation frequency and loss of resonance. In the second case, the nonlinear instability dynamics can only be studied numerically. In this regime, at low beam densities significant nonlinearity is only observed in the motion of the beam electrons while the plasma remains linear and saturation of the instability is caused by trapping of beam electrons in the field of the beam-excited plasma wave. © *2000 MAIK "Nauka/Interperiodica".*

1. In some early studies on the nonlinear theory of resonant beam–plasma instabilities it was shown that these are stabilized as a result of the trapping of beam electrons by the beam-excited plasma wave [1, 2]. The trapping is accompanied by bunching of the electron beam and the formation of multiflux flows in these electron bunches. Under certain simplifying assumptions (see below) this process is described by a system of nonlinear integrodifferential equations containing no small parameter and permitting no analytic solution. An approximate solution of the problem of trapping of beam electrons by a plasma wave was obtained in [3] using conservation laws and methods of averages. We specify that the above reasoning only applies to resonant beam–plasma instability in the form of the single-particle stimulated Cherenkov effect [4]. These simplifications mainly involve assuming that the plasma is linear and the amplitude of the plasma wave is slow. For the single-particle Cherenkov effect these assumptions are justified when the electron beam density is low compared with the plasma density. At low beam and plasma densities it is meaningless to talk of the single-particle stimulated Cherenkov effect—in this case the instability is called two-beam (two-flux) [5] and is studied numerically [6].

In addition to the single-particle regime there is another resonant beam–plasma interaction regime, known as the collective stimulated Cherenkov effect [4]. Unlike the single-particle effect, the collective effect can only occur in transversely inhomogeneous beam–plasma systems at fairly high electron beam densities. The collective stimulated Cherenkov effect began to be discussed in the literature some time after the appearance of the first studies on beam instability in a plasma. Nonlinear stabilization of instabilities caused by the collective stimulated Cherenkov effect takes place as a result of a nonlinear frequency shift of the interacting beam and plasma waves and may be described analytically. As will become clear from the following, a similar analytic description is also suitable for beam–plasma systems with similar beam and plasma densities and thus does not assume plasma linearity. In the present study we consider various regimes of beam–plasma interaction as a result of the collective stimulated Cherenkov effect in nonlinear beams and plasma taking into account the excitation and interaction of harmonics of the beam and plasma waves. We use analytic methods or significantly simplified computer models based on a method of expanding the characteristics (trajectories) of the transport equation with a self-consistent field developed in an earlier study by the authors [7]. At the end we discuss some characteristics observed in the description of the single-particle stimulated Cherenkov effect under conditions of linear dispersion of the plasma waves.

2. We shall begin by formulating a fairly general mathematical model of a beam–plasma system exhibiting very substantial transverse instability. We shall consider a cylindrical metal waveguide having an arbitrary simply connected cross section containing an infinitely thin in cross section ("needle-shaped") nonrelativistic electron beam and plasma. The waveguide is placed in a longitudinal strong external magnetic field which impedes transverse motion of the beam and plasma electrons (the motion of heavy ions is completely neglected). The beam and plasma are cold.

1063-7761/00/9101-0093$20.00

The potential (electrostatic) perturbations in this system are described by the following equations:

$$\left(\Delta_\perp + \frac{\partial^2}{\partial z^2}\right)\varphi = -4\pi e\sum_\alpha S_\alpha \delta(\mathbf{r}_\perp - \mathbf{r}_\alpha) n_\alpha,$$
$$n_\alpha = \int f_\alpha(t, z, v)dv, \qquad (1)$$
$$\frac{\partial f_\alpha}{\partial t} + v\frac{\partial f_\alpha}{\partial z} + \frac{e}{m}\left(-\frac{\partial \varphi}{\partial z}\right)\frac{\partial f_\alpha}{\partial v} = 0.$$

Here φ is the scalar potential, α is the particle species, ($\alpha = p$ are plasma electrons, $\alpha = b$ are beam electrons), z is the coordinate in the longitudinal direction of the waveguide, $\mathbf{r}_\perp$ is the coordinate in its cross section, $\Delta_\perp$ is the transverse component of the Laplace operator, v is the velocity, e and m are the electron charge and mass, n_α is the density of particles of species α, f_α is the distribution function of particles of the corresponding species, S_α is the cross-sectional area of the system of species α particles, and $\mathbf{r}_\alpha$ is their average coordinate in the waveguide cross section. We shall assume that initially the distribution functions satisfy the conditions

$$f_\alpha(0, z, v) \equiv f_{0\alpha}(v) = n_{0\alpha}\delta(v - v_{0\alpha}). \qquad (2)$$

Here $n_{0\alpha}$ are the unperturbed densities of particles of species α,[1] and $v_{0\alpha}$ are their unperturbed velocities. In our case, for a cold quiescent plasma we have $v_{0p} = 0$ and for a moving monoenergetic beam we have $v_{0b} = u$. In (1) and (2) we use generalized Dirac δ functions.

We know (see, for example, [8, 9]) that in general the solution of the Vlasov equation for the distribution function f_α is expressed as the following integral over the initial data of the characteristic system of the Vlasov equation:

$$f_\alpha(t, z, v) = \iint dz_0 dv_0 f_{0\alpha}(v_0) \times \delta(z - z_\alpha(t, z_0, v_0))\delta(v - v_\alpha(t, z_0, v_0)). \qquad (3)$$

Here $z_\alpha(t, z_0, v_0)$ and $v_\alpha(t, z_0, v_0)$ are the solutions of the characteristic system

$$\frac{dz_\alpha}{dt} = v_\alpha, \quad \frac{dv_\alpha}{dt} = -\frac{e}{m}\frac{\partial \varphi}{\partial z}, \qquad (4)$$

supplemented with the initial conditions[2]

$$z_\alpha|_{t=0} = z_0, \quad v_\alpha|_{t=0} = v_0. \qquad (5)$$

The validity of the solution (3) is easily confirmed by directly substituting this into the Vlasov equation taking into account the properties of the δ-function, the Liouville theorem on the conservation of phase volume, and the equations (4). Then, substituting (3) into (1) and integrating over velocities, allowing for (2), we obtain the following equations:

$$\left(\Delta_\perp + \frac{\partial^2}{\partial z^2}\right)\varphi' = \sum_\alpha \omega_\alpha^2 S_\alpha \delta(\mathbf{r}_\perp - \mathbf{r}_\alpha) \times \int dz_0 \delta(z - z_\alpha(t, z_0, v_{0\alpha})),$$
$$\frac{dz_\alpha}{dt} = v_\alpha, \quad \frac{dv_\alpha}{dt} = \frac{\partial \varphi'}{\partial z}(t, z_\alpha, \mathbf{r}_\alpha), \qquad (6)$$
$$z_\alpha|_{t=0} = z_0, \quad v_\alpha|_{t=0} = v_{0\alpha} = \begin{cases} u, & \alpha = b \\ 0, & \alpha = p. \end{cases}$$

Here $\varphi' = -e\varphi/m$ and $\omega_\alpha = \sqrt{4\pi e^2 n_{0\alpha}/m}$ are the Langmuir frequencies of particles of species α. A further transformation of the system (6) reduces it to a more convenient formulation for solving the nonlinear initial problem.

We shall assume that the initial perturbation in this case has a characteristic longitudinal dimension (period) L. Then, all the perturbed quantities, including the potential φ', can be expressed in the form

$$\varphi' = \frac{1}{2}\sum_{n=1} [\varphi_n(\mathbf{r}_\perp, t)\exp(inkz) + \text{c.c.}], \qquad (7)$$

where $k = 2\pi/L$ is the dominant longitudinal wave number. Substituting (7) into (6) and using the orthogonality of the functions $\exp(inkz)$, we obtain the following equations for the expansion coefficients φ_n:

$$(\Delta_\perp - n^2k^2)\varphi_n = \sum_\alpha \omega_\alpha^2 S_\alpha \delta(\mathbf{r}_\perp - \mathbf{r}_\alpha)\rho_{\alpha n}, \qquad (8)$$

where $\rho_{\alpha n}$ are functions determined by the formulas

$$\rho_{\alpha n} = \frac{1}{\pi}\int_0^{2\pi} \exp(-iny_\alpha)dy_0, \quad y_\alpha = kz_\alpha. \qquad (9)$$

We shall now assume that the eigenfunctions and eigenvalues of the waveguide cross section are known, i.e., the solutions of the following problem are known:

$$\Delta_\perp \phi_m = -k_{\perp m}^2 \phi_m,$$
$$\phi_m|_{\sigma=0} = 0, \qquad (10)$$

where ϕ_m is an eigenfunction, $k_{\perp m}^2$ is the corresponding eigenvalue, $m = 1, 2, \ldots$, and $\sigma \equiv \sigma(\mathbf{r}_\perp) = 0$ is the equation for the metal surface of the waveguide. Then using the expansion

$$\varphi_n = \sum_{m=1}^{\infty} \tilde{\varphi}_{n,m}(t)\phi_m(\mathbf{r}_\perp), \qquad (11)$$

[1] They have the meaning of the product $S_\alpha n_{0\alpha}$ which are the instantaneous particle densities, i.e., the values measured experimentally.

[2] We write z_0 and not $z_{0\alpha}$ and similarly for the velocity since these are simply variables of integration.

we obtain from (8) the following relations for $\tilde{\varphi}_{n,m}$:

$$\tilde{\varphi}_{n,m} = -\frac{1}{\|\phi_m\|^2(k_{\perp m}^2 + n^2k^2)}\sum_\alpha \omega_\alpha^2 S_\alpha \phi_m(\mathbf{r}_\alpha)\rho_{\alpha n}. \quad (12)$$

We now need to collect together (7), (9), (11), and (12) and substitute (7) into the equations of motion for the plasma and beam electrons from the system (6). As a result, we arrive at the following equations:

$$\frac{d^2y_p}{dt^2} = -\frac{i}{2} \times \sum_n \frac{1}{n}[(g_{pn}\rho_{pn} + \tilde{\omega}_b^2 q_n \rho_{bn})\exp(iny_p) - \text{c.c.}],$$
$$\frac{d^2y_b}{dt^2} = -\frac{i}{2} \times \sum_n \frac{1}{n}[(\tilde{\omega}_p^2 q_n \rho_{pn} + g_{bn}\rho_{bn})\exp(iny_b) - \text{c.c.}]. \quad (13)$$

Here we use the notation:

$$g_{\alpha n} = \tilde{\omega}_\alpha^2 \sum_{m=1}^{\infty} \frac{n^2k^2\phi_m^2(\mathbf{r}_\alpha)}{\|\phi_m\|^2(k_{\perp m}^2 + n^2k^2)} = \tilde{\omega}_\alpha^2 \tilde{g}_{\alpha n},$$
$$q_n = \sum_{m=1}^{\infty} \frac{n^2k^2\phi_m(\mathbf{r}_p)\phi_m(\mathbf{r}_b)}{\|\phi_m\|^2(k_{\perp m}^2 + n^2k^2)} = \tilde{q}_n\sqrt{\tilde{g}_{pn}\tilde{g}_{bn}}, \quad (14)$$

$\tilde{\omega}_\alpha^2$ (cm^2 rad^2 s^{-2}) $= S_\alpha\omega_\alpha^2$, $\alpha = p, b$ are quantities proportional to the instantaneous densities of the beam and plasma electrons, and the geometric parameters $\tilde{g}_{\alpha n}$ (cm^{-2}) and dimensionless coefficient $\tilde{q}_n$ were introduced in (14) for convenience of the notation in some of the following formulas. The quantities $\sqrt{g_{\alpha n}}$ (rad/s) are the frequencies of the natural oscillations in the beam and the plasma at the wavelength $\lambda_n = \pi n/L$. They do not simply reduce to the corresponding plasma frequencies but also depend on the transverse geometry.

We obtain the first integrals of the equations (13). We integrate each of the equations (13) over y_0 between zero and 2π, multiply the first equation by $\tilde{\omega}_p^2$, the second equation by $\tilde{\omega}_b^2$, and add it to the first. As a result, we obtain the momentum conservation law:

$$\frac{\tilde{\omega}_p^2}{\pi}\int_0^{2\pi} \dot{y}_p dy_0 + \frac{\tilde{\omega}_b^2}{\pi}\int_0^{2\pi} \dot{y}_b dy_0 = \text{const}. \quad (15)$$

The first term in (15) is the momentum of the plasma electrons, the second term is the momentum of the beam electrons.

In order to obtain the energy conservation law, we multiply the first equation in (13) by $(2\tilde{\omega}_p^2/\pi)dy_p/dt$, the second equation by $(2\tilde{\omega}_b^2/\pi)dy_b/dt$, integrate with respect to y_0 between zero and 2π, and add these together. As a result we obtain

$$\tilde{\omega}_p^2\left\{\frac{1}{\pi}\int_0^{2\pi}\dot{y}_p^2 dy_0 + \sum_n \frac{g_{pn}}{n^2}|\rho_{pn}|^2\right\} + \tilde{\omega}_b^2\left\{\frac{1}{\pi}\int_0^{2\pi}\dot{y}_b^2 dy_0 + \sum_n \frac{g_{bn}}{n^2}|\rho_{bn}|^2\right\} + \tilde{\omega}_p^2\tilde{\omega}_b^2\sum_n q_n(\rho_{pn}\rho_{bn}^* + \rho_{bn}\rho_{pn}^*) = \text{const}. \quad (16)$$

Here the first term is the energy density of the perturbations in the plasma (kinetic plus electrostatic), the second term is the same in the beam, and the third term is the energy density of the electrostatic plasma–beam interaction.

The nonlinear equations (13) together with the formulas (9) and (14) form the basis for the following analysis.

3. We shall begin with a linear analysis of system (13). To this end, taking into account the initial values of the beam and plasma electron velocities, we express the solutions of the characteristic Vlasov equations in the form

$$y_p = y_0 + \tilde{y}_p,$$
$$y_b = y_0 + kut + \tilde{y}_b \quad (17)$$

and we linearize relations (13) and (9) with respect to the values of $\tilde{y}_{p,b}$. Since in the linear approximation perturbations (spatial harmonics) with different numbers n do not interact, in the linearized equations it is sufficient to allow only for a single harmonic with some arbitrary number n. As a result of linearizing, we arrive at the following system of equations:

$$\frac{d^2\rho_{pn}}{dt^2} + g_{pn}\rho_{pn} = -\tilde{\omega}_b^2 q_n \rho_{bn} e^{-inkut},$$
$$\frac{d^2\rho_{bn}}{dt^2} + g_{bn}\rho_{bn} = -\tilde{\omega}_p^2 q_n \rho_{pn} e^{inkut}. \quad (18)$$

Here the first equation describes the plasma oscillations excited by the electron beam and the second describes the beam oscillations "excited" by the plasma oscillations.

We seek a solution of system (18) in the form

$$\rho_{pn} = c_p\exp(-i\omega_n t),$$
$$\rho_{bn} = c_b\exp(-i\omega_n t + inkut), \quad (19)$$

where c_p and c_b are constants and ω_n is the frequency of the nth spatial harmonic. Substituting (19) into (18) and eliminating the constants, we obtain a dispersion equation to determine the complex frequency:

$$(\omega_n^2 - g_{pn})[(\omega_n - nku)^2 - g_{bn}] = g_{pn}g_{bn}\tilde{q}_n^2. \quad (20)$$

For $\tilde{q}_n = 0$ equation (20) breaks down into an equation for the frequencies of the plasma oscillations and an equation for the frequencies of the beam oscillations. We shall write the spectrum of the plasma oscillations in the form (we take the plus sign when extracting the root)

$$\omega_n = \sqrt{g_{pn}}. \quad (21)$$

For the beam oscillations we have one of the following two expressions:

$$\omega_n = nku + \sqrt{g_{bn}}, \quad \omega_n = nku - \sqrt{g_{bn}}. \quad (22)$$

The second expression (22) in accordance with (21) is the spectrum of the slow beam wave. Interaction between this wave and the wave (21) leads to the appearance of the resonant beam instabilities considered in this study.

Here we shall not study the spectra (21) and (22) in detail. Only the following properties of g_{pn}, g_{bn}, and $\tilde{q}_n$ which are easily deduced from the expressions (14) are important for the subsequent analysis:

$$g_{pn} \approx \begin{cases} n^2\tilde{\omega}_p^2, & n^2k^2 \ll k_\perp^2 \\ \tilde{\omega}_p^2, & n^2k^2 \gg k_\perp^2, \end{cases} \quad (23)$$

where $k_\perp$ is the reciprocal of the transverse dimension of the waveguide. The quantity g_{bn} has similar properties. The coefficients $\tilde{q}_n$ satisfy the inequalities

$$0 < \tilde{q}_n \le 1. \quad (24)$$

The exact equality in (24) is only achieved for $\mathbf{r}_b = \mathbf{r}_p$, i.e., when the position of the beam and the plasma coincide in the waveguide cross section. As $|\mathbf{r}_b - \mathbf{r}_p|$ increases the coefficients $\tilde{q}_n$ decrease rapidly, this decrease being stronger the higher the number n. For $\tilde{q}_n \longrightarrow 0$ the plasma and beam oscillations become independent which was used to obtain the spectra (21) and (22).

We shall now solve equation (20) for $\tilde{q}_n \neq 0$. We express the frequency ω_n in the form

$$\omega_n = \sqrt{g_{pn}} + \delta\omega_n = nku - \sqrt{g_{bn}} + \delta\omega_n, \quad (25)$$

where $\delta\omega_n$ is the growth rate of the resonant instability at the nth spatial harmonic. Substituting (25) into (20) and assuming that the following inequality is satisfied (see below)

$$|\delta\omega_n| \ll \sqrt{g_{pn}}, \quad (26)$$

we obtain:

$$2\sqrt{g_{pn}}\delta\omega_n[(\delta\omega_n)^2 - 2\sqrt{g_{bn}}\delta\omega_n] = g_{pn}g_{bn}\tilde{q}_n^2. \quad (27)$$

We shall consider two cases. We first assume that

$$|\delta\omega_n| \gg \sqrt{g_{bn}}. \quad (28)$$

It then follows from (27) that

$$\delta\omega_n = \frac{-1 + i\sqrt{3}}{2}\left(\frac{g_{bn}}{2g_{pn}}\tilde{q}_n^2\right)^{1/3}\sqrt{g_{pn}}. \quad (29)$$

Instability with this growth rate is caused by the single-particle stimulated Cherenkov effect. The condition for the existence of this effect is inequality (28) which reduces to:

$$g_{bn}/g_{pn} \ll \tilde{q}_n^4, \quad (30)$$

and inequality (26), on account of (24), is a consequence of (30).

Let us now assume that an inequality opposite to (28) is now satisfied:

$$|\delta\omega_n| \ll \sqrt{g_{bn}}. \quad (31)$$

From (27) we then obtain the following expression for the growth rate:

$$\delta\omega_n = \frac{i}{2}(\sqrt{g_{pn}g_{bn}}\tilde{q}_n^2)^{1/2} = \frac{i}{2}\left(\sqrt{\frac{g_{bn}}{g_{pn}}}\tilde{q}_n^2\right)^{1/2}\sqrt{g_{pn}}. \quad (32)$$

Instability with this growth rate is called the collective stimulated Cherenkov effect. Inequalities (26) and (31) reduce to the constraint:

$$\tilde{q}_n^4 \ll g_{bn}/g_{pn} \ll \tilde{q}_n^{-4}. \quad (33)$$

It follows from (33) that the collective Cherenkov effect can only occur in systems with a small parameter $\tilde{q}_n$. Moreover, for $\tilde{q}_n \ll 1$ the collective Cherenkov effect occurs even for g_{bn} of the order of g_{pn}. Note that for $\tilde{q}_n \approx 1$ the single-particle Cherenkov effect can only occur if $g_{bn} \ll g_{pn}$. For similar g_{bn} and g_{pn} two-beam instability develops which is difficult to assign to any resonant stimulated process of the Cherenkov type. Assuming that $\tilde{q}_n \ll 1$ we concentrate for now on the collective Cherenkov effect.

4. It follows from inequalities (26) and (31) that for the collective stimulated Cherenkov effect the growth rate is lower than the frequency of the plasma wave and the frequency of the beam wave in coordinates moving with the beam. Consequently, the collective effect is a resonant interaction between the plasma and beam waves. In the single-particle effect this is not the case since we cannot generally talk of beam waves because of inequality (28). For instabilities caused by resonant wave interaction a mechanism of instability stabilization such as a nonlinear frequency shift [10] may become the determining factor. This is the situation with the collective

Cherenkov effect which makes it possible to describe its nonlinear stage using a method of expanding the trajectories of the beam and plasma electrons [7, 9].

We express the coordinates of the plasma and beam electrons in the form [see (17)]

$$\begin{aligned} y_p &= y_0 + w_p(t) + x_p(y_0, t), \\ y_b &= y_0 + kut + w_b(t) + x_b(y_0, t), \end{aligned} \tag{34}$$

where

$$\begin{aligned} x_p(y_0, t) &= \frac{1}{2}\sum_{k=1}^{\infty}(a_{pk}(t)\exp(iky_0) + \text{c.c.}), \\ x_b(y_0, t) &= \frac{1}{2}\sum_{k=1}^{\infty}(a_{bk}(t)\exp(iky_0) + \text{c.c.}). \end{aligned} \tag{35}$$

Here the functions $w_{p,b}(t)$ describe the changes in the translational motion of the plasma electrons in laboratory coordinates and of the beam electrons in coordinates moving at the velocity u. We can say that the functions $w_{p,b}(t)$ are responsible for the nonlinear excitation of zeroth spatial harmonics of the perturbations in the plasma and the beam (nonlinear generation of the constant component). In turn, the expansion coefficients a_{pk} and a_{bk} from (35) describe the nonzero spatial harmonics of the perturbations in the plasma and the beam. They are responsible for the wave interaction and for the nonlinear excitation of high spatial harmonics of the initial perturbation. To be specific, we shall assume that in this particular system resonance of the plasma and slow beam waves occurs at the first spatial harmonic, i.e., the following equality is satisfied [see (21) and (22)]

$$\sqrt{g_{p1}} = ku - \sqrt{g_{b1}}. \tag{36}$$

Under the conditions (33) we can then assume that the coefficients a_{pk} and a_{bk} are quantities of the kth order of smallness. The concept "order of smallness" will be refined subsequently: it is found that this order is determined by the small parameter $\tilde{q}_1$.

Substituting (34) and (35) into (9), to within terms of the third order of smallness inclusive, we obtain

$$\begin{aligned} \rho_{p1} &= \left(-ia_{p1} - \frac{1}{2}a_{p1}^*a_{p2} + \frac{i}{8}|a_{p1}|^2a_{p1}\right)\exp(-iw_p), \\ \rho_{p2} &= (-2ia_{p2} - a_{p1}^2 + 2i|a_{p1}|^2a_{p2})\exp(-2iw_p), \\ \rho_{p3} &= \left(-3ia_{p3} - \frac{9}{2}a_{p1}a_{p2} + \frac{9}{8}ia_{p1}^3\right)\exp(-3iw_p). \end{aligned} \tag{37}$$

Similar formulas are used to determine ρ_{bk} but with the index "p" replaced by "b" and w_p replaced by $w_b + kut$. Then, substituting (34), (35), and (37) into equation (13), we obtain the following equations for the expansion coefficients a_{pk}, a_{bk}, $k = 1, 2, 3$, and the functions w_p, w_b, also given to within terms of the third order of smallness (for conciseness we only give the equations for a_{pk} and w_p):

$$\begin{aligned} \frac{d^2a_{p1}}{dt^2} &= -g_{p1}a_{p1} + i(g_{p1} - g_{p2})a_{p1}^*a_{p2} \\ &\quad + \frac{1}{2}(g_{p1} - g_{p2})|a_{p1}|^2a_{p1} \\ &\quad - \tilde{\omega}_b^2q_1\left(a_{b1} - \frac{1}{2}ia_{b1}^*a_{b2} - \frac{1}{8}|a_{b1}|^2a_{b1} - \frac{1}{4}|a_{p1}|^2a_{b1}\right) \\ &\quad \times \exp[i(w_p - w_b - kut)] \\ &\quad + \frac{\tilde{\omega}_b^2q_1}{2}\left(ia_{b1}^*a_{p2} + \frac{1}{4}a_{b1}^*a_{p1}^2\right)\exp[-i(w_p - w_b - kut)] \\ &\quad - \tilde{\omega}_b^2q_2\left(ia_{b1}^*a_{b2} + \frac{1}{2}a_{p1}^*a_{b1}^2\right)\exp[2i(w_p - w_b - kut)], \end{aligned} \tag{38a}$$

$$\begin{aligned} \frac{d^2a_{p2}}{dt^2} &= -g_{p2}a_{p2} + \frac{1}{2}i(g_{p2} - g_{p1})a_{p1}^2 \\ &\quad - \frac{1}{2}i\tilde{\omega}_b^2q_1a_{b1}a_{p1}\exp[i(w_p - w_b - kut)] \\ &\quad - \tilde{\omega}_b^2q_2\left(a_{b2} - \frac{1}{2}ia_{b1}^2\right)\exp[2i(w_p - w_b - kut)], \end{aligned} \tag{38b}$$

$$\begin{aligned} \frac{d^2a_{p3}}{dt^2} &= -g_{p3}a_{p3} + i\left(\frac{3}{2}g_{p3} - \frac{1}{2}g_{p1} - g_{p2}\right)a_{p1}a_{p2} \\ &\quad + \left(\frac{3}{8}g_{p3} + \frac{1}{8}g_{p1} - \frac{1}{2}g_{p2}\right)a_{p1}^3 \\ &\quad - \frac{1}{2}i\tilde{\omega}_b^2q_1\left(a_{b1}a_{p2} + \frac{1}{4}ia_{b1}a_{p1}^2\right)\exp[i(w_p - w_b - kut)] \\ &\quad - \frac{1}{2}i\tilde{\omega}_b^2q_2(2a_{b2}a_{p1} - ia_{b1}^2a_{p1})\exp[2i(w_p - w_b - kut)] \\ &\quad - \frac{1}{2}i\tilde{\omega}_b^2q_3\left(-2ia_{b3} - 3a_{b1}a_{b2} + \frac{3}{4}ia_{b1}^3\right) \\ &\quad \times \exp[3i(w_p - w_b - kut)], \end{aligned} \tag{38c}$$

$$\begin{aligned} \frac{d^2w_p}{dt^2} &= -\frac{1}{4}i\left\{\tilde{\omega}_b^2q_1\left(a_{b1}a_{p1}^* + \frac{1}{2}ia_{b1}a_{p1}a_{p2}^*\right.\right. \\ &\quad \left.\left. - \frac{1}{8}|a_{p1}|^2a_{b1}a_{p1}^* - \frac{1}{2}ia_{p1}^*a_{b1}^*a_{b2} - \frac{1}{8}|a_{b1}|^2a_{b1}a_{p1}^*\right) \right. \\ &\quad \times \exp[i(w_p - w_b - kut)] \end{aligned} \tag{38d}$$

$$+\frac{1}{2}\tilde{\omega}_b^2 q_2(4a_{b2}a_{p2}^* + 2ia_{b2}a_{p1}^{*2} - 2ia_{b1}^2 a_{p2}^* + a_{b1}^2 a_{p1}^{*2})$$

$$\times \exp[2i(w_p - w_b - kut)] - \text{c.c.}\Big\}.$$

The equations for a_{bk} and w_b are similar but with the index "p" replaced by "b" and $w_p - w_b - kut$ replaced by $-(w_p - w_b - kut)$.

Comparing the right-hand sides of equation (38d) and the similar equation for w_p, it is easy to see that these only differ in respect of the sign and the constant factor. We can then easily obtain the following:

$$\tilde{\omega}_p^2 \frac{dw_p}{dt} + \tilde{\omega}_p^2 \frac{dw_p}{dt} = 0. \tag{39}$$

The functions w_p and w_b satisfy the same relationship. However, equation (39) follows automatically from expressions (34), (35), and the integral (15), i.e., is a consequence of the conservation of momentum. Subsequent transformations of equations (38) are performed differently for various particular cases.

5. We shall begin with the case of nonresonant excitation of multiple harmonics of the plasma and beam density perturbations when $g_{p2} \neq 4g_{p1}$, $g_{p3} \neq 9g_{p1}$, $g_{b2} \neq 4g_{b1}$, and $g_{b3} \neq 9g_{b1}$. Returning to the property (23) of the coefficients g_{pn} and g_{bn}, nonresonant excitation of multiple (second, third, and so on) harmonics can occur in cases of nonlinear dispersion laws for the plasma and beam waves which is found, for example, when $k^2 > k_\perp^2$ (short-wavelength limit).

Taking into account inequalities (26) and (31), we seek a solution of equations (38) in the form [see also (19)]

$$\begin{aligned} a_{p1} &= \tilde{a}_{p1}\exp(-i\sqrt{g_{p1}}t), \\ a_{p2} &= \tilde{a}_{p2}\exp(-2i\sqrt{g_{p1}}t), \\ a_{b1} &= \tilde{a}_{b1}\exp(i\sqrt{g_{b1}}t), \\ a_{b2} &= \tilde{a}_{b2}\exp(2i\sqrt{g_{b1}}t), \end{aligned} \tag{40}$$

where $\tilde{a}_{p1}$, $\tilde{a}_{p2}$, $\tilde{a}_{b1}$, and $\tilde{a}_{b2}$ are slowly varying amplitudes. Using the resonance condition (36) and bearing in mind that the equations for a_{p3} and a_{b3} are not required to solve this problem with the given accuracy (we shall now seek a solution to within the third order of smallness), we obtain the following reduced equations from (38) and (40):

$$-2i\sqrt{g_{p1}}\frac{d\tilde{a}_{p1}}{dt} = i(g_{p1} - g_{p2})\tilde{a}_{p1}^*\tilde{a}_{p2}$$

$$+\frac{1}{2}(g_{p1} - g_{p2})|\tilde{a}_{p1}|^2\tilde{a}_{p1} - \tilde{\omega}_b^2 q_1 \tilde{a}_{b1}\exp[i(w_p - w_b)]$$

$$+ i\frac{1}{2}\tilde{\omega}_b^2 q_1 \tilde{a}_{b1}^*\tilde{a}_{p2}\exp[-i(w_p - w_b)]$$

$$+ i\frac{1}{2}\tilde{\omega}_b^2 q_1 \tilde{a}_{b1}^*\tilde{a}_{b2}\exp[i(w_p - w_b)],$$

$$-4i\sqrt{g_{p1}}\frac{d\tilde{a}_{p2}}{dt} = (4g_{p1} - g_{p2})\tilde{a}_{p2} + \frac{1}{2}(g_{p2} - g_{p1})\tilde{a}_{p1}^2$$

$$-\frac{1}{2}i\tilde{\omega}_b^2 q_1 \tilde{a}_{b1}\tilde{a}_{p1}\exp[i(w_p - w_b)],$$

$$\frac{d^2 w_p}{dt^2} = -\frac{1}{4}i\tilde{\omega}_b^2 q_1$$

$$\times\Big[\Big(\tilde{a}_{b1}\tilde{a}_{p1}^* + \frac{1}{2}i\tilde{a}_{b1}\tilde{a}_{p1}\tilde{a}_{p2}^* - \frac{1}{2}i\tilde{a}_{p1}^*\tilde{a}_{b1}^*\tilde{a}_{b2}\Big) \tag{41}$$

$$\times \exp[i(w_p - w_b)] - \text{c.c.}\Big],$$

$$2i\sqrt{g_{b1}}\frac{d\tilde{a}_{b1}}{dt} = i(g_{b1} - g_{b2})\tilde{a}_{b1}^*\tilde{a}_{b2} + \frac{1}{2}(g_{b1} - g_{b2})|\tilde{a}_{b1}|^2\tilde{a}_{b1}$$

$$-\tilde{\omega}_p^2 q_1 \tilde{a}_{p1}\exp[i(w_b - w_p)]$$

$$+ i\frac{1}{2}\tilde{\omega}_p^2 q_1 \tilde{a}_{p1}^*\tilde{a}_{b2}\exp[i(w_p - w_b)]$$

$$+ i\frac{1}{2}\tilde{\omega}_p^2 q_1 \tilde{a}_{p1}^*\tilde{a}_{p2}\exp[i(w_b - w_p)],$$

$$4i\sqrt{g_{b1}}\frac{d\tilde{a}_{b2}}{dt} = (4g_{b1} - g_{b2})\tilde{a}_{b2} + \frac{1}{2}i(g_{b2} - g_{b1})\tilde{a}_{b1}^2$$

$$-\frac{1}{2}i\omega_p^2 q_1 \tilde{a}_{p1}\tilde{a}_{b1}\exp[i(w_b - w_p)],$$

$$\frac{d^2 w_b}{dt^2} = -\frac{1}{4}i\tilde{\omega}_p^2 q_1$$

$$\times\Big[\Big(\tilde{a}_{p1}\tilde{a}_{b1}^* + \frac{1}{2}i\tilde{a}_{p1}\tilde{a}_{b1}\tilde{a}_{b2}^* - \frac{1}{2}i\tilde{a}_{b1}^*\tilde{a}_{p1}^*\tilde{a}_{p2}\Big)$$

$$\times \exp[i(w_b - w_p)] - \text{c.c.}\Big].$$

We supplement the system (41) with the initial conditions:

$$\begin{aligned} &a_{p1}(0) = 0, \quad a_{p2}(0) = 0, \quad a_{b1}(0) = a_0, \\ &a_{b2}(0) = 0, \quad \frac{dw_p}{d\tau}(0) = \frac{dw_b}{d\tau}(0) = 0, \end{aligned} \tag{42}$$

which correspond to the case of an electron beam modulated at the first harmonic in an unperturbed plasma. The initial modulation of the beam is proportional to a_0.

Bearing in mind the initial conditions (42), we find the following first integrals of equations (41):

$$\frac{dw_p}{dt} = \frac{1}{2}\sqrt{g_{p1}}(|\tilde{a}_{p1}|^2 + 4|\tilde{a}_{p2}|^2),$$
$$\frac{dw_b}{dt} = -\frac{1}{2}\sqrt{g_{b1}}(|\tilde{a}_{b1}|^2 + 4|\tilde{a}_{b2}|^2 - a_0^2), \quad (43)$$
$$\tilde{\omega}_p\frac{dw_p}{dt} + \tilde{\omega}_b\frac{dw_b}{dt} = 0,$$
$$\sqrt{g_{p1}}(|\tilde{a}_{p1}|^2 + 4|\tilde{a}_{p2}|^2) = \sqrt{g_{b1}}(|\tilde{a}_{b1}|^2 + 4|\tilde{a}_{b2}|^2 - a_0^2),$$

where the fourth integral is a consequence of the first three. It reflects the energy conservation law (in terms of slow amplitudes) and is a consequence of (16).

In cases of nonresonant excitation of charge-density harmonics of the plasma and beam waves when $g_{p2} \neq 4g_{p1}$ and $g_{b2} \neq 4g_{b1}$, derivatives of the slow amplitudes $\tilde{a}_{p2}$ and $\tilde{a}_{b2}$ can be neglected in equations (41). Making the additional change of variables

$$\tilde{a}_{p1} = b_{p1}e^{iw_p}, \quad \tilde{a}_{b1} = b_{b1}e^{iw_b}, \quad (44)$$

taking into account the integrals (43) and the initial conditions (42), we transform system (41) to give

$$-2i\sqrt{g_{p1}}\frac{db_{p1}}{dt} = -g_{p1}\tilde{a}_p|b_{p1}|^2b_{p1} - \tilde{\omega}_b^2q_1b_{b1},$$
$$2i\sqrt{g_{b1}}\frac{db_{b1}}{dt} = -g_{b1}\tilde{a}_b|b_{b1}|^2b_{b1} \quad (45)$$
$$-\tilde{\omega}_p^2q_1b_{p1} + g_{b1}a_0^2b_{b1},$$

where

$$\tilde{a}_p = 1 - \frac{3}{2}\frac{g_{p2} - g_{p1}}{g_{p2} - 4g_{p1}}, \quad \tilde{a}_b = 1 - \frac{3}{2}\frac{g_{b2} - g_{b1}}{g_{b2} - 4g_{b1}}. \quad (46)$$

The solutions of the system (45) are expressed in terms of elliptic functions and are extremely cumbersome. We only give the solutions for the case where the field was switched on adiabatically in the infinite past when $|b_{p1,\,b1}| \longrightarrow 0$ for $\tau \longrightarrow -\infty$:

$$|b_{b1}|^2 = \frac{4q_1\tilde{\omega}_p^2}{\tilde{a}_b\tilde{\omega}_p^2 + \tilde{a}_p\tilde{\omega}_b^2}\left(\frac{\tilde{\omega}_p^2\tilde{\omega}_b^2}{g_{b1}(g_{p1}g_{b1})^{1/2}}t\right)^{1/2}$$
$$\times \operatorname{arccosh}\left(\frac{q_1\tilde{\omega}_p\tilde{\omega}_b}{(g_{p1}g_{b1})^{1/4}}\right),$$
$$|b_{p1}|^2 = \frac{4q_1\tilde{\omega}_b^2}{\tilde{a}_b\tilde{\omega}_p^2 + \tilde{a}_p\tilde{\omega}_b^2}\left(\frac{\tilde{\omega}_p^2\tilde{\omega}_b^2}{g_{b1}(g_{p1}g_{b1})^{1/2}}\right)^{1/2} \quad (47)$$
$$\times \operatorname{arccosh}\left(\frac{q_1\tilde{\omega}_p\tilde{\omega}_b}{(g_{p1}g_{b1})^{1/4}}t\right).$$

From (47) we obtain expressions for the maximum amplitudes of the interacting first harmonics of the beam and plasma waves. For simplicity we only give these for the particular case $k \gg k_\perp$ [see (23)]:

$$|b_{b1}|_{\max} = 2\tilde{q}_1^{1/2}\left[\frac{\tilde{\omega}_p^2}{\tilde{\omega}_p^2 + \tilde{\omega}_b^2}\left(\frac{\tilde{\omega}_p}{\tilde{\omega}_b}\right)^{1/2}\right]^2,$$
$$|b_{p1}|_{\max} = 2\tilde{q}_1^{1/2}\left[\frac{\tilde{\omega}_b^2}{\tilde{\omega}_p^2 + \tilde{\omega}_b^2}\left(\frac{\tilde{\omega}_b}{\tilde{\omega}_p}\right)^{1/2}\right]^2. \quad (48)$$

The values (48) determine the percent density modulation of the beam and the plasma (in units corresponding to n_{0b} and n_{0p}). It can be seen that the modulation is deeper for the particle system for which $\tilde{\omega}_\alpha$ is lower, i.e., the instantaneous density. The condition for validity of this analysis is that the modulation of the beam and the plasma should be weak, which implies that the following inequality is satisfied

$$|b_{b1}|_{\max}, |b_{p1}|_{\max} \ll 1. \quad (49)$$

Satisfaction of both these inequalities simultaneously for any relationship between the instantaneous densities requires the following condition to be satisfied:

$$\tilde{q}_1^{1/2} \ll 1, \quad (50)$$

which agrees with the condition (33) for the existence of the collective stimulated Cherenkov effect.

As can be seen from formula (47), the time taken for the beam wave amplitude to increase, for example, from b_0 to $b_{b1\max}$ ($|b_0| \ll |b_{b1}|_{\max}$) is given by

$$t_0 = \frac{2(g_{p1}g_{b1})^{1/4}}{q_1\tilde{\omega}_p\tilde{\omega}_b}\ln 4\left|\frac{b_{b1\max}}{b_0}\right|. \quad (51)$$

The value (51) is the same as the reciprocal growth rate (32) in the linear theory of the collective Cherenkov effect. The mechanism for nonlinear saturation of this process is a nonlinear frequency shift of the beam and plasma waves described by cubic terms in the equations (45).

6. We shall now analyze the resonant excitation of second harmonics of the plasma and beam density perturbation. In this case we shall distinguish between the following three cases:

(1) $g_{b2} = 4g_{b1}$, but $g_{p2} \neq 4g_{p1}$;

(2) $g_{p2} = 4g_{p1}$, but $g_{b2} \neq 4g_{b1}$;

(3) $g_{p2} = 4g_{p1}$ and $g_{b2} = 4g_{b1}$.

In all these cases the system (41) does not have a sufficient number of first integrals and thus can only be solved by numerical methods. However, some idea of the behavior of the system may be obtained from the qualitative reasoning put forward below. For this we initially consider the case when no relationship exists between the beam and plasma components and consequently $q_1 = 0$. From (41) for the case of resonant exci-

tation of the second harmonic of the beam wave when $g_{b2} = 4g_{b1}$, we then have

$$\frac{d\tilde{a}_{b1}}{d\tau} = \frac{3}{4}i|\tilde{a}_{b1}|^2\tilde{a}_{b1} - \frac{3}{2}\tilde{a}_{b1}^*\tilde{a}_{b2},$$
$$\frac{d\tilde{a}_{b2}}{d\tau} = \frac{3}{8}\tilde{a}_{b1}^2. \tag{52}$$

Here $\tau = \sqrt{g_{b1}}\,t$ is the dimensionless time. This system of equations has the first integral:

$$\frac{1}{4}|\tilde{a}_{b1}|^2 + |\tilde{a}_{b2}|^2 = \frac{1}{4}a_0^2. \tag{53}$$

Taking this integral into account, the solutions of the system (52) may be written as follows:

$$|\tilde{a}_{b1}|^2 = a_0^2\mathrm{cn}^2(z, p) + \frac{1}{2}a_0^3\mathrm{sn}^2(z, p), \tag{54}$$

where $\mathrm{cn}(z, p)$ and $\mathrm{sn}(z, p)$ are elliptic functions whose argument and modulus are determined respectively by the expressions $z = (3/4)a_0\tau$ and $p = 1 - a_0$. From this it can be seen that the first harmonic of the beam wave decays, its amplitude decreases, and by the dimensionless time

$$\tau_1 = \frac{4}{3a_0}\ln\frac{4}{\sqrt{a_0}}, \tag{55}$$

it has a minimum given by

$$|\tilde{a}_{b1}|_{\min} = \frac{1}{2}a_0^{3/2}. \tag{56}$$

Since, if the initial modulation of the beam is weak, we find

$$\frac{|\tilde{a}_{b1}|_{\min}}{a_0} = \frac{1}{2}\sqrt{a_0} \ll 1,$$

the decay of the first harmonic is very significant. It should be noted that in cases of resonant excitation of the second harmonic of the plasma wave when $g_{p2} = 4g_{p1}$, the solutions will be determined by formulas of the type (54)–(56) in which a_0 should be taken to be the initial amplitude a_{p1}.

Thus, in the absence of beam–plasma interaction, energy is transferred [in accordance with formula (54)] from the first harmonic of the beam wave to the second and back. The same occurs in resonance between the harmonics of the plasma wave. When interaction takes place between the beam and the plasma but the harmonics are not in resonance, the first harmonics of the plasma and beam waves increase. Consequently, when both factors are present (beam–plasma interaction and harmonic resonance), both an increase in amplitude and transfer between harmonics should be observed. Assuming that the characteristic value in (55) is of the order of the first quantity (48) and equating the times (51) and (55), we find that until the amplitude of the first harmonic of the plasma wave saturates, multiple energy exchange takes place between the first and second harmonics of the beam wave. This is completely confirmed by results of numerical calculations some of which will now be analyzed. In order to reduce the number of free parameters in the calculations we assumed that the positions of the plasma and the beam in the waveguide are "symmetric" so that $\tilde{g}_{p1} = \tilde{g}_{b1}$. We shall analyze the cases of resonant excitation of the second harmonics of the beam and plasma densities listed above.

Let us assume that $g_{p2} = 4g_{p1}$ but $g_{b2} \neq 4g_{b1}$ (case 2). From system (41), after changing the variables (44) taking into account the integrals (43) and neglecting the derivative of $\tilde{a}_{b2}$, we obtain

$$\frac{db_{p1}}{d\tau} = -\frac{5}{4}i|b_{p1}|^2b_{p1}$$
$$+ \frac{3}{2}b_{p1}^*b_{p2} - \frac{1}{2}i\nu q b_{b1} - \frac{1}{4}\nu q b_{b1}^*b_{p2},$$
$$\frac{db_{p2}}{d\tau} = -\frac{3}{8}b_{p1}^2 + \frac{1}{8}\nu q b_{b1}b_{p1}, \tag{57}$$
$$\frac{db_{b1}}{d\tau} = \frac{1}{4}i\sqrt{\nu}\frac{5 - g_b}{g_b - 4}|b_{b1}|^2b_{b1}$$
$$-\frac{1}{2}i\sqrt{\nu}b_0^2b_{b1} + \frac{1}{2}i\frac{q}{\sqrt{\nu}}\left(b_{p1} - \frac{1}{2}ib_{p1}^*b_{p2}\right).$$

Here we introduce the dimensionless quantities:

$$\tau = \sqrt{g_{b1}}t, \quad \nu = \frac{\tilde{\omega}_b^2}{\tilde{\omega}_p^2},$$
$$q = \frac{q_1}{g_{p1}}\tilde{\omega}_p^2, \quad g_b = \frac{g_{b2}}{g_{b1}}, \tag{58}$$

τ is the dimensionless time, ν is a parameter equal to the ratio of the instantaneous beam and plasma densities, and g_b is a parameter characterizing the degree of "nonresonance" of the beam wave harmonics ($g_b \neq 4$).

Figure 1 gives results of a numerical solution of the system (57) for the following values of the parameters: $a_0 = 0.01$, $q = 0.03$, $g_b = 2$, $\nu = 1$ (a) and $\nu = 1/3$ (b). It can be seen from these values that Fig. 1a corresponds to the case when the instantaneous beam and plasma densities are the same and Fig. 1b corresponds to the case when the beam density is three times lower than the plasma density. The figures clearly show multiple interaction of the plasma wave harmonics over the time taken for the dominant beam instability to saturate.

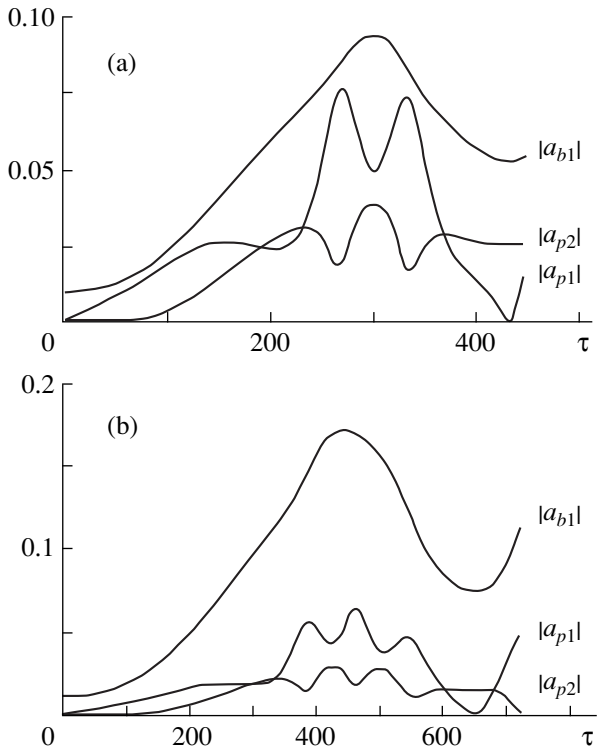


Fig. 1. Collective Cherenkov effect under conditions of resonance between the plasma wave harmonics: $a_0 = 0.01$, $q = 0.03$, $g_b = 2$, the beam and plasma densities are the same [(a) $\nu = 1$] and the beam density is lower than the plasma density [(b) $\nu = 1/3$].

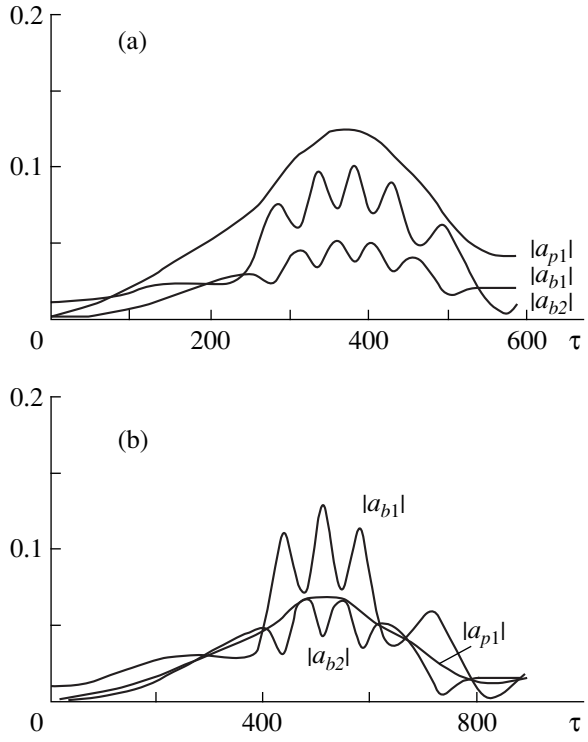


Fig. 2. Collective Cherenkov effect under conditions of resonance between the beam wave harmonics: $a_0 = 0.01$, $q = 0.03$, $g_p = 2$, the beam and plasma densities are the same [(a) $\nu = 1$] and the beam density is lower than the plasma density [(b) $\nu = 1/3$].

Let us now assume $g_{b2} = 4g_{b1}$ but $g_{p2} \neq 4g_{p1}$. Performing transformations similar to the previous case, we obtain from system (41)

$$\frac{db_{p1}}{d\tau} = \frac{1}{4}i\frac{g_p + 5}{g_p - 4}|b_{p1}|^2 b_{p1} - \frac{1}{2}i\nu q b_{b1} + \frac{1}{4}\nu q b_{b1}^* b_{b2},$$

$$\frac{db_{b1}}{d\tau} = \frac{1}{2}i\sqrt{\nu}(|b_{b1}^2| - b_0^2)a_{b1} - \frac{3}{2}\sqrt{\nu}b_{b1}^* b_{b2} + \frac{3}{4}i\sqrt{\nu}|b_{b1}^2|b_{b1} + \frac{1}{2}i\frac{q}{\sqrt{\nu}}b_{p1}^* b_{b2}, \quad (59)$$

$$\frac{db_{b2}}{d\tau} = \frac{3}{8}\sqrt{\nu}b_{b1}^2 - \frac{1}{8}\frac{q}{\sqrt{\nu}}b_{p1}b_{b1}.$$

Here the values of τ, ν, and q were determined in (58) and have the same meaning, while the parameter

$$g_p = \frac{g_{p2}}{g_{p1}} \quad (60)$$

characterizes the degree of nonresonance of the plasma wave harmonics ($g_p \neq 4$).

Figure 2 gives results of a numerical solution of the system (59) for the following values of the parameters: $a_0 = 0.01$, $q = 0.03$, $g_p = 2$, $\nu = 1$ (a) and $\nu = 1/3$ (b). The figures again clearly show the multiple interaction of the beam wave harmonics.

Finally let us assume $g_{p2} = 4g_{p1}$ and $g_{b2} = 4g_{b1}$. As in the previous cases we can obtain from system (41)

$$\frac{db_{p1}}{d\tau} = -\frac{5}{4}i|b_{p1}|^2 b_{p1} + \frac{3}{2}b_{p1}^* b_{p2} - \frac{1}{2}i\nu q b_{b1} - \frac{1}{4}\nu q b_{b1}^* b_{p2} - \frac{1}{4}\nu q b_{b1}^* b_{b2},$$

$$\frac{db_{p2}}{d\tau} = -\frac{3}{8}b_{p1}^2 + \frac{1}{8}\nu q b_{b1}b_{p1},$$

$$\frac{db_{b1}}{d\tau} = \frac{5}{4}i\sqrt{\nu}|b_{b1}^2|b_{b1} - \frac{1}{2}i\sqrt{\nu}b_0^2 b_{b1} - \frac{3}{2}\sqrt{\nu}b_{b1}^* b_{b2} + \frac{1}{2}i\frac{q}{\sqrt{\nu}}b_{p1} + \frac{1}{4}\frac{q}{\sqrt{\nu}}(b_{p1}^* b_{b2} + b_{p1}^* b_{p2}), \quad (61)$$

$$\frac{db_{b2}}{d\tau} = \frac{3}{8}\sqrt{\nu}b_{b1}^2 - \frac{1}{8}\frac{q}{\sqrt{\nu}}b_{p1}b_{b1}.$$

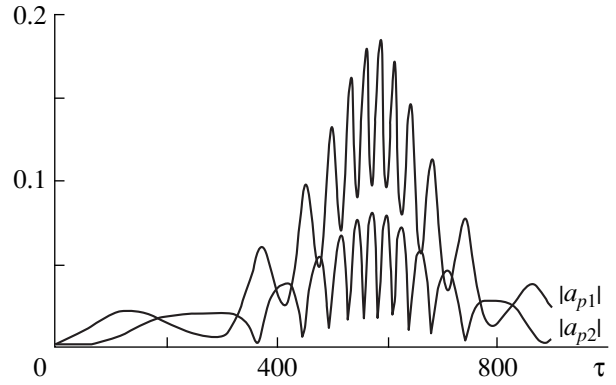


Fig. 3. Collective Cherenkov effect under conditions of resonance between the plasma and beam wave harmonics. The beam and plasma densities are the same ($a_0 = 0.01$, $q = 0.03$, $\nu = 1$).

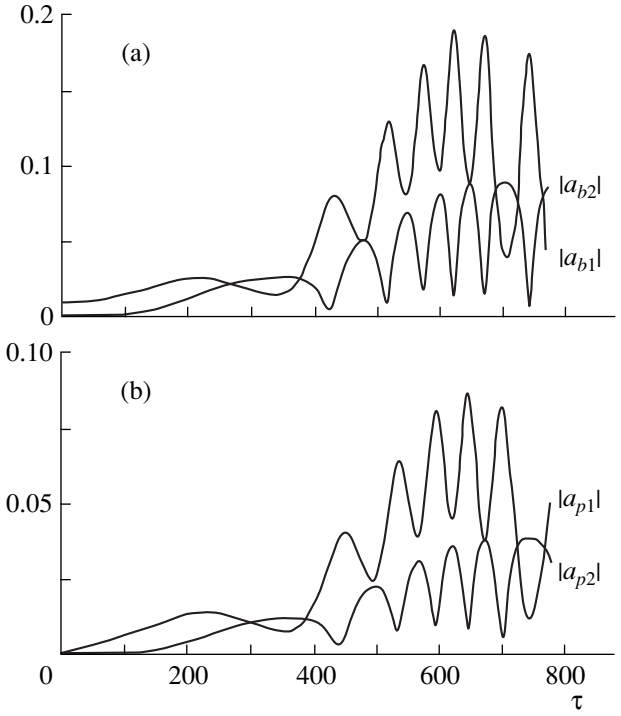


Fig. 4. Amplitudes of beam waves in the collective Cherenkov effect under conditions of resonance (a) between the plasma and beam wave harmonics and (b) between the beam wave harmonics. The beam density is lower than the plasma density ($a_0 = 0.01$, $q = 0.03$, $\nu = 1/3$).

The values of ν, τ, and q were determined in (58) and have the same meaning. Figures 3 and 4 give results of a numerical solution of the system (61) for the following values of the parameters: $a_0 = 0.01$, $q = 0.03$, $\nu = 1$ (Fig. 3), $a_0 = 0.01$, $q = 0.03$, $\nu = 1/3$ (Fig. 4). Figure 3 shows the dynamics of the first and second plasma wave harmonics for the case of equal plasma and beam densities. Since for $\nu = 1$ the equations in the system (61) are symmetric relative to the amplitudes of the beam and plasma wave harmonics, the dynamics of the first and second harmonics of the beam wave will be given by the same dependences as in Fig. 3 and are thus not given here. Figure 4 gives the dynamics of the harmonics of the beam (Fig. 4a) and plasma (Fig. 4b) oscillations for the case where the beam and plasma densities differ. Here we observe a difference between the harmonics of the plasma and beam waves, particularly at the maximum amplitude. These figures broadly show multiple transfer between the harmonics of the beam and plasma waves over the time taken for the dominant beam instability to evolve.

7. We shall now consider the single-particle Cherenkov effect, which can only occur when the electron beam density is low compared with the plasma density [see inequality (30)]. The growth rate of this instability is given by expression (29). We shall use this to calculate the phase velocity of the nth harmonic of the plasma wave excited in the system $V^{(\mathrm{ph})} = \mathrm{Re}\,\omega_n/nk$. Also taking into account (25) and inequality (26) we have

$$V^{(\mathrm{ph})} = u\left[1 - \frac{1}{2}\left(\tilde{q}_n^2 \frac{g_{bn}}{2g_{pn}}\right)^{1/3}\right]. \tag{62}$$

It can be seen that the phase velocity of the beam-excited plasma wave is slightly lower than the unperturbed beam velocity u, i.e., in the wave system the beam has an excess energy which serves as an energy source of instability in the single-particle Cherenkov effect.

It can be seen from (62) that the phase velocity of the plasma wave is close to the velocity of the beam electrons and far from that of the plasma electrons. We can therefore postulate that nonlinear effects are particularly important when describing the motion of the beam electrons whereas the plasma electrons can be described in the linear approximation. The criterion for plasma linearity can be expressed in the form

$$|\rho_{pn}| \ll 1, \tag{63}$$

where ρ_{pn} are the amplitudes of the harmonics of the plasma density perturbation (9). Assuming that inequality (63) is satisfied, we shall describe the plasma electrons in the linear approximation.

Linearizing in (13) the equations of motion of the plasma electrons [as in the derivation of (18)] and leaving the equations for the beam electrons unchanged, we obtain the following equations:

$$\frac{d^2\rho_{pn}}{dt^2} + g_{pn}\rho_{pn} = -\tilde{\omega}_b^2 q_n \rho_{bn} \exp(-inkut),$$

$$\frac{d^2 y_b}{dt^2} = -\frac{i}{2} \tag{64}$$

$$\times \sum_n \frac{1}{n}[(\tilde{\omega}_p^2 q_n \rho_{pn} + g_{bn}\rho_{bn})\exp(iny_b) - \text{c.c.}].$$

Note that in (64) the harmonics ρ_{pn} of the plasma wave are not directly interrelated but they interact with ρ_{bn}. The latter are associated with the beam nonlinearity.

For the following transformation of the equations (64) we make the substitutions

$$y_b = kut + y(y_0, t), \qquad \rho_{pn} = \rho_{pn}^M(t)\exp(-inkut). \tag{65}$$

The first of these is similar to the second expression in (17) but linearization with respect to $y(y_0, t)$ is not assumed. The second change in (65) is made to introduce the amplitudes of the plasma wave harmonics ρ_{pn}^M. As a result of inequality (26), these amplitudes are slow compared with the exponential factor exp(–*inkut*), i.e.,

$$o \equiv \frac{1}{nku}\left|\frac{1}{\rho_{pn}^M}\frac{d\rho_{pn}^M}{dt}\right| \ll 1. \tag{66}$$

Calculations to within $1 + o \approx 1$ and $(1 + o)^{-1} - 1 \approx -o$ have been called the slow amplitude method. By applying this method to transform the equations (64) and substituting (65) in these equations to within this accuracy, we obtain

$$2inku\frac{d\rho_{pn}^M}{dt} + [(nku)^2 - g_{pn}]\rho_{pn}^M = \tilde{\omega}_b^2 q_n \rho_{bn}^M,$$
$$\frac{d^2y}{dt^2} = -\frac{i}{2}\sum_n \frac{1}{n}[(\tilde{\omega}_p^2 q_n \rho_{pn}^M + g_{bn}\rho_{bn}^M)e^{iny} - \text{c.c.}]. \tag{67}$$

Here

$$\rho_{bn}^M = \frac{1}{\pi}\int_0^{2\pi} e^{-iny}dy_0 \tag{68}$$

are the slow amplitudes of the beam wave harmonics (as they were observed in the beam rest system).

We shall now assume that one of the plasma wave harmonics, to be specific, say the first, is in exact Cherenkov resonance with the electron beam. For the other harmonics we assume that no resonance occurs. This implies that $g_{p1} - k^2u^2 = 0$ but $g_{pn} - (nku)^2 \neq 0$ for $n = 2, 3, \ldots$. Then to within the method of slow amplitudes all ρ_{pn}^M with $n = 2, 3, \ldots$ can be expressed from the corresponding equations and substituted into the equation for y. As a result we obtain

$$2iku\frac{d\rho_{p1}^M}{dt} = \tilde{\omega}_b^2 q_1 \rho_{b1}^M,$$
$$\frac{d^2y}{dt^2} + \frac{1}{2}ig_{b1}\sum_n \frac{1}{n}[\hat{g}_{bn}\rho_{bn}^M e^{iny} - \text{c.c.}] = -\frac{1}{2}iq_1\tilde{\omega}_p^2(\rho_{p1}^M e^{iy} - \text{c.c.}), \tag{69}$$

where

$$\hat{g}_{bn} = \begin{cases} 1, & n = 1 \\ \dfrac{1}{g_{b1}}\left[g_{bn} + \dfrac{q_n^2\tilde{\omega}_b^2\tilde{\omega}_p^2}{(nku)^2 - g_{pn}}\right], & n = 2, 3, \ldots. \end{cases} \tag{70}$$

It can be seen from (70) that allowance for higher harmonics of the plasma wave merely yields numerical corrections to the coefficients at the amplitudes of the higher beam harmonics. In similar problems this situation is normal. We can now see that the higher harmonics are generally insignificant.

We can easily show that in the linear approximation (69) yields the growth rate (29). We shall use this growth rate to transform the equations (69). We introduce the dimensionless variables

$$\tau = |\delta\omega_1|t, \quad \varepsilon = -iq_1\tilde{\omega}_p^2|\delta\omega_1|^{-2}\rho_{p1}^M, \quad \rho_{b1}^M \equiv \rho, \tag{71}$$

where $\delta\omega_1$ is given by formula (29) for $n = 1$. In terms of the new variables the system of equations (69) is written in the form

$$\frac{d\varepsilon}{d\tau} = -\rho, \quad \rho = \frac{1}{\pi}\int_0^{2\pi} e^{-iy}dy_0,$$
$$\frac{d^2y}{d\tau^2} + \frac{i}{2}2^{2/3}\left(\frac{g_{b1}/g_{p1}}{\tilde{q}_1^4}\right)^{1/3}\mathcal{R} = \frac{1}{2}(\varepsilon e^{iy} + \text{c.c.}). \tag{72}$$

Here $\mathcal{R}$ denotes the sum contained in the second equation of the system (69). The term containing $\mathcal{R}$ describes part of the force from the electric field generated by the density modulation of the electron beam. In microwave electronics this force is called the rf space charge force of the beam. As a result of the inequality (30), this force appears in the equation with a small parameter. Thus, when describing the single-particle Cherenkov effect we can neglect the rf space charge of the beam so that (72) can be written in a universal form containing no individual parameters of the plasma–beam system:

$$\frac{d\varepsilon}{d\tau} = -\rho, \quad \rho = \frac{1}{\pi}\int_0^{2\pi} e^{-iy}dy_0,$$
$$\frac{d^2y}{d\tau^2} = \frac{1}{2}(\varepsilon e^{iy} + \text{c.c.}). \tag{73}$$

Equations of the type (73) were obtained in [1]. These should be supplemented by the initial conditions

$$\varepsilon(\tau = 0) = \varepsilon_0, \quad y(\tau = 0) = y_0 + b\sin(y_0 + \vartheta), \quad \frac{dy}{d\tau}(\tau = 0) = 0. \tag{74}$$

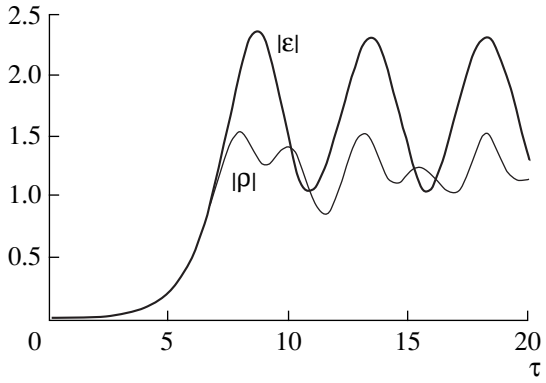


Fig. 5. Time dependences of the amplitudes of the plasma density (|ε|) and beam density (|ρ|) waves in the single-particle Cherenkov effect.

Here ε_0 is the initial (seed) amplitude of the plasma wave, b determines the percent initial seed density modulation of the beam, and ϑ is a certain constant. The last condition (74) follows from (65) and implies that at $t = 0$ all the beam electrons have the velocity u. For completeness we give the well-known numerical solutions of equations (73).

Figure 5 gives the dependences $|\varepsilon(\tau)|$ and $|\rho(\tau)|$ obtained for $\varepsilon_0 = 0$, $b = 0.01$, and $\vartheta = 0$. Initially both values increase exponentially [in dimensionless variables as $\exp(\sqrt{3}\tau/2)$, see (29)]. This is followed by nonlinear stabilization after which the amplitude $|\varepsilon(\tau)|$ exhibits regular oscillations between the maximum and minimum levels. This is caused by trapping of the electron beam by the plasma wave field. At each plasma wavelength the beam electrons form a bunch which oscillates between the humps of the plasma wave potential, sometimes imparting energy to it, sometimes extracting energy from it. Since the bunch collides twice with the potential humps within a single oscillation (once with the front and once with the rear), the oscillations of the amplitude $|\rho(\tau)|$ are twice as frequent as the oscillations of $|\varepsilon(\tau)|$. It can be seen that the maximum is $|\varepsilon(\tau)| \approx 2.34$, i.e., of the order of one. From the definition of the dimensionless quantities (71) we then have an estimate for the maximum amplitude of the dimensionless harmonic of the plasma density:

$$\left|\rho_{p1}^M\right|_{\max} \approx \left(\tilde{q}_1\sqrt{\frac{\tilde{g}_{b1}}{\tilde{g}_{p1}}}\right)^{-1}\frac{|\delta\omega_1|^2}{g_{p1}}. \tag{75}$$

In accordance with inequality (26), the value of (75) is low and condition (63) is satisfied. Thus, the assumption that the plasma is linear can be considered to be justified although the question of nonlinear effects in plasma can only be considered to be definitively solved after making a detailed analysis of the complete nonlinear problem. It can be seen from Fig. 5 that for the electron beam $|\rho|_{\max} \approx 1$, i.e., the beam is completely modulated.

It follows from the expressions (23) that the situation when only the first harmonic of the plasma oscillations is in Cherenkov resonance with the beam and the higher harmonics are not, is typical of the short-wavelength region if $k^2 \gg k_\perp^2$. In the opposite long-wavelength limit as long as $n^2k^2 < k_\perp^2$ all the harmonics are in resonance with the beam and for all n we have $g_{pn} - (nku)^2 \approx 0$ (naturally if resonance is satisfied for at least one harmonic). For this case, as can be seen from (70), the equations (69) and all the following ones are unsuitable. We shall derive and study the equations for the long-wavelength limit.

For all resonant harmonics of the plasma wave we write equations (67) in the form

$$\begin{gathered}2inku\frac{d\rho_{pn}^M}{dt} + [(nku)^2 - g_{pn}]\rho_{pn}^M = \tilde{\omega}_p^2 q_n \rho_{bn}^M,\\ \frac{d^2y}{dt^2} = -\frac{1}{2}i\sum_n\frac{1}{n}[\tilde{\omega}_p^2 q_n \rho_{pn}^M e^{iny} - \text{c.c.}],\end{gathered} \tag{76}$$

and we shall neglect all nonresonant harmonics and the rf space charge of the beam, as before (the number of resonant harmonics is not yet specified). In the first equation (76) the differences $g_{pn} - (nku)^2$ are specially retained because as a result of the nonlinear dispersion law of the plasma waves all these differences cannot go to zero simultaneously. Taking into account only the single-frequency Cherenkov effect, we assume for simplicity that the coefficients are $\tilde{q}_n = 1$ in equations (76) and using the dimensionless variables [see (71)]

$$\begin{gathered}\tau = |\delta\omega_1|t, \quad \varepsilon_n = -i\sqrt{\tilde{g}_{b1}\tilde{g}_{p1}}\,\tilde{\omega}_p^2|\delta\omega_1|^{-2}\rho_{pn}^M,\\ \rho_{bn}^M \equiv \rho_n,\end{gathered} \tag{77}$$

we transform (76) to give

$$\begin{gathered}\frac{d\varepsilon_n}{d\tau} + i\Delta_n\varepsilon_n = -n\rho_n, \quad \rho_n = \frac{1}{\pi}\int_0^{2\pi} e^{-iny}dy_0,\\ \frac{d^2y}{d\tau^2} = \frac{1}{2}\sum_n n(\varepsilon_n e^{iny} + \text{c.c.}),\end{gathered} \tag{78}$$

where

$$\Delta_n = \frac{g_{pn} - (nku)^2}{2n|\delta\omega_1|ku} \tag{79}$$

are quantities characterizing the degree of departure from Cherenkov resonance between the beam and the nth harmonic of the plasma wave. In microwave electronics these quantities are called detunings. In the derivation of (78) we took into account the first of the properties (23) of the coefficients $g_{\alpha n}$. In the linear approximation (78)

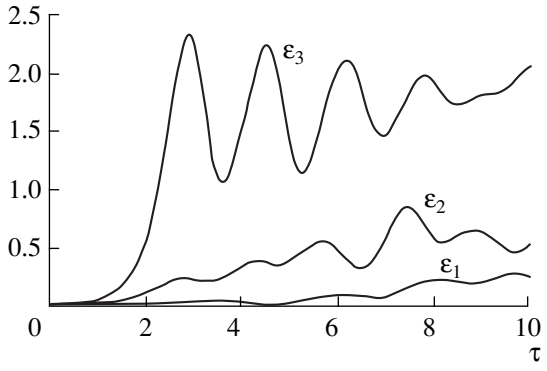


Fig. 6. Time dependences of the amplitudes of the first three harmonics (ε_1, ε_2, ε_3) of the plasma wave in Cherenkov resonance with the electron beam.

gives (for $\Delta_n \longrightarrow 0$) $\varepsilon_n \approx \exp(\sqrt{3}\, n\tau/2)$ which agrees with the growth rate (29) in the limit $n^2k^2 < k_\perp^2$.

Figure 6 gives the dynamics of the amplitudes of the plasma waves for the case when Cherenkov resonance is only satisfied for the first three harmonics ε_1, ε_2, and ε_3 and no resonance occurs for higher harmonics. It can be seen that the highest of the resonant harmonics increases predominantly which can be explained by the fact that this has the highest growth rate. Calculations were made for initial amplitudes of all the plasma wave harmonics of 0.01 and for detunings $\Delta_1 = \Delta_2 = \Delta_3 = 0$, $\Delta_4 = 1$. Although the choice of detunings is extremely arbitrary, it accurately reflects the real dispersion properties of the plasma oscillations in a waveguide.

Hence, the dynamics of instability in the single-particle Cherenkov effect depends strongly on the part of the plasma oscillation spectrum, long-wavelength or short-wavelength, in which it develops. The harmonic of the plasma oscillations near the maximum of the instability growth rate grows rapidly. The beam electrons are trapped by the plasma wave and it becomes completely density modulated. The plasma electrons, at least at the initial nonlinear stage, are described by linear equations. In the collective Cherenkov effect the determining factor for the nonlinear stabilization of the instability is a nonlinear shift of the frequencies of the interacting plasma and beam waves. In this case, the beam and plasma instabilities are weak.

REFERENCES

1. I. N. Onishchenko, A. R. Linetskiĭ, N. G. Matsiborko, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **12**, 407 (1970) [JETP Lett. **12**, 281 (1970)]; I. N. Onishchenko, A. R. Linetski, N. G. Matsiborko, *et al.*, Plasma Phys. **14**, 591 (1972).
2. V. D. Shapiro, in *Problems of Plasma Theory* (Naukova Dumka, Kiev, 1972), p. 257.
3. R. I. Kovtun and A. A. Rukhadze, Zh. Éksp. Teor. Fiz. **58**, 1709 (1970) [Sov. Phys. JETP **31**, 915 (1970)].
4. M. V. Kuzelev and A. A. Rukhadze, Usp. Fiz. Nauk **152**, 285 (1987) [Sov. Phys. Usp. **30**, 507 (1987)].
5. R. Briggs, in *Advances in Plasma Physics*, Ed. by A. Simon and W. B. Thompson (Wiley, New York, 1969–1971; Mir, Moscow, 1974), Vol. 4.
6. D. E. Potter, *Computational Physics* (Wiley, New York, 1973; Mir, Moscow, 1975).
7. M. V. Kuzelev, A. A. Rukhadze, Yu. V. Bobylev, and V. A. Panin, Zh. Éksp. Teor. Fiz. **91**, 1620 (1986) [Sov. Phys. JETP **64**, 956 (1986)].
8. M. V. Kuzelev and A. A. Rukhadze, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **36**, 867 (1993).
9. M. V. Kuzelev and A. A. Rukhadze, *Electrodynamics of Dense Electron Beams in a Plasma* (Nauka, Moscow, 1990).
10. J. Weiland and H. Wilhelmsson, *Coherent Nonlinear Interaction of Waves in Plasmas* (Pergamon Press, Oxford, 1976; Énergoizdat, Moscow, 1981).

Translation was provided by AIP