Journal of Experimental and Theoretical Physics, Vol. 98, No. 2, 2004, pp. 171–180.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 2, 2004, pp. 195–204.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Quantum Nonlocality Effect and Radiative Damping of the Dirac Electron

G. F. Efremov and V. V. Sharkov

Nizhni Novgorod State University, pr. Gagarina 23, Nizhni Novgorod, 603950 Russia
e-mail: efremov@rf.unn.ru

Received April 1, 2002; in final form, June 17, 2003

Abstract—Quantum spacetime nonlocality, i.e., retardation of the interaction between an electron and its own radiation field at distances of about the Compton wavelength, is established. By taking into account a finite variance of the electron-coordinate increment in the intrinsic coordinate system, the radiative damping coefficient is obtained as a divergence-free function of frequency not subject to the well-known paradoxes of the classical theory of radiative damping. A relation between radiative damping, the Lamb shift, and the electromagnetic mass of the electron is found. *© 2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The problems of radiative damping and electromagnetic mass that arose at the beginning of the 20th century continue to attract attention of many researchers in view of the special importance of their solution [1–11]. Radiative damping is the reaction of the self-generated radiation field to the simultaneous effects of motion of a charged particle and the electromagnetic vacuum field. This fundamental effect in electrodynamics is related to issues of paramount importance in physics.

In this paper, we continue the analysis presented in [11] and consider the paradoxes formulated a century ago in the theory of radiative damping. It is well known that the classical expression for radiative damping force,

$$\mathbf{F} = \frac{2}{3}\frac{e^2}{c^3}\frac{d^3}{dt^3}\mathbf{r}(t), \qquad (1)$$

entails instability of free electron motion. Physically, this paradox follows from the fact that the reaction of the self-generated radiation field of a pointlike charged particle does not allow for retardation of the interaction. This leads to acceleration of the particle, and the causality principle is violated. Moreover, the self-energy of a pointlike charged particle is infinite. Therefore, self-action can be consistently described in classical electrodynamics only by introducing a renormalization procedure [12]. Solution of this problem is related to the fundamental problems of locality and nonlocality in modern quantum theory. One is led to the question: How does a local physical theory take into account the retardation of the interaction between a charged particle and its own radiation field?

Despite considerable progress in quantum electrodynamics (QED), the application of its traditional methods to the problem of radiative damping has encountered certain difficulties. We note here that calculation of a friction force of any physical nature is a key problem in quantum statistical physics. The interdependence and mutual influence of methods of quantum field theory and statistical physics are well known. Methods of quantum field theory are successfully applied to solve the fundamental problems in statistical physics and condensed matter physics [13]. At the same time, some methods of statistical physics are most suited for solving certain fundamental problems in quantum field theory, including the problem of radiative damping [4, 11, 14–16].

In [11], one of us (G.F.E.) proposed a solution to the radiative damping problem based on methods of the fluctuation-dissipation theory of nonlinear open quantum systems [17–19]. One fundamental distinction of this theory from the linear theory of Brownian motion in quantum systems developed by Schwinger [20] and Senitzky [21] and from other methods lies in its scope for dealing with fluctuations and retardation of interactions in calculating physical effects. An exact formula for the radiative damping force was used in [11] to calculate the radiative damping coefficient in the nonrelativistic limit for the dynamical variables of the electron. At the same time, a final solution to the problem must include the contribution to physical effects due to large momenta transferred from the radiation field to the electron. Therefore, calculation of the radiative damping coefficient must rely on relativistic dynamics and take into account the internal degrees of freedom of the electron.

Owing to the fundamental fact that the Dirac electron has internal degrees of freedom, the variance of a coordinate increment $\mathbf{r}(t) - \mathbf{r}(t_1)$ is finite in the intrinsic coordinate system. Since the variance $(\mathbf{r}(t) - \mathbf{r}(t_1))^2$ is finite, and the coordinate operators taken at different

instants do not commute, the interaction between the electron and its own radiation field is retarded over a distance of about the Compton wavelength. This quantum spacetime nonlocality solves the problem of a pointlike electron in quantum theory and thus eliminates the paradoxes of self-acceleration and causality violation inherent in classical theory. Moreover, the removal of divergences in calculations of fundamental QED effects makes it possible to relate radiative damping to the Lamb shift and the finite contribution to the electromagnetic mass due to the reaction of the radiation field.

2. STARTING EQUATIONS AND RADIATIVE DAMPING COEFFICIENT

The interaction of the Dirac electron with the electromagnetic vacuum field and its own radiation field determines radiative damping and electromagnetic electron mass, offers a fundamental mechanism of fluctuations, and provides physical explanations for the Lamb shift and the anomalous magnetic moment of the electron. These QED effects can be analyzed in the framework of the one-particle fluctuation-dissipation quantum electrodynamics proposed in [11]. Allowance for retardation of the interaction between the electron and the radiation field not only resolves the radiation-damping paradox, but also rules out any divergence in calculations of the aforementioned effects in electrodynamics.

Consider the interaction of a relativistic electron with a quantized radiation field embedded in an external potential field $V(\mathbf{r}, t)$. The total Hamiltonian of the system is

$$H = c\boldsymbol{\alpha}\cdot\left(\mathbf{p} - \frac{e}{c}\mathbf{A}(\mathbf{r}, t)\right) + \beta mc^2 + eA_0(\mathbf{r}, t) + V(\mathbf{r}, t) + F, \quad (2)$$

where $\mathbf{A}(\mathbf{r}, t)$ and $A_0(\mathbf{r}, t)$ constitute the field potential, F is the Hamiltonian of the quantum radiation field, and $\boldsymbol{\alpha}$ and β are the Dirac matrices. In the transverse gauge for the field potential,

$$\mathrm{div}\mathbf{A}(\mathbf{r}, t) = 0. \quad (3)$$

Under this condition, the scalar potential $A_0(\mathbf{r}, t)$ responsible in this case for the Coulomb interaction does not give rise to observable effects in the one-electron problem and can be omitted in the Hamiltonian of the system.

The starting equations corresponding to Hamiltonian (2) are Maxwell's equations for the field potential and the Lorentz equation for the dynamical variables of the Dirac electron:

$$\frac{d}{dt}\pi_j(t) + \nabla_j V(\mathbf{r}, t) = -\frac{e}{c}\frac{d}{dt}A_j(\mathbf{r}, t) + \frac{e}{c}\nabla_j\dot{r}_\alpha(t)A_\alpha(\mathbf{r}(t), t), \quad (4)$$

where

$$\pi_j(t) = p_j - \frac{e}{c}A_j(\mathbf{r}(t), t)$$

is the canonical momentum of the electron.

The Lorentz force exerted on the electron by the radiation field on the right-hand side of Eq. (4) is determined by the total derivative of the vector potential and the velocity $\dot{r}_\alpha(t)$. In the Heisenberg representation, the components of the field potential in Eq. (4) are functions of the electron coordinate $\mathbf{r}(t)$. Therefore, it is convenient to represent them as Fourier integrals:

$$A_j(\mathbf{r}(t), t) = \int\frac{d\mathbf{k}}{(2\pi)^3}\exp(i\mathbf{k}\cdot\mathbf{r})A_j(\mathbf{k}, t), \quad (5)$$

where $A_j(\mathbf{k}, t)$ does not depend explicitly on the dynamical variables of the electron. The current density components are canonically conjugate to $A_j(\mathbf{k}, t)$ and are derived from Hamiltonian (2):

$$-\frac{\delta H}{\delta A_j(\mathbf{k}, t)} = \frac{e}{c}\dot{r}_j(t)\exp(i\mathbf{k}\cdot\mathbf{r}(t)).$$

Furthermore, expression (2) entails Maxwell's equations for the Fourier components $A_j(\mathbf{k}, t)$ of the field potential:

$$\left(k^2 + \frac{1}{c^2}\frac{d^2}{dt^2}\right)A_j(\mathbf{k}, t) = \frac{4\pi}{c}e\dot{r}_j(t)\exp(i\mathbf{k}\cdot\mathbf{r}(t)). \quad (6)$$

We seek a solution of Eq. (6) in the form

$$A_j(\mathbf{k}, t) = A_j^0(\mathbf{k}, t) + \frac{e}{c}\int_{-\infty}^{\infty}dt_1 D_{jl}(\mathbf{k}, t - t_1)\exp(-i\mathbf{k}\cdot\mathbf{r}(t_1))\dot{r}_l(t_1). \quad (7)$$

Here, the term containing photon Green's function $D_{jl}(\mathbf{k}, t - t_1)$ is the electron's own radiation field, and the absolute term $\mathbf{A}^0(\mathbf{k}, t)$ is the unperturbed electromagnetic vacuum field. In the adopted gauge (3), Green's

function and its Fourier spectrum are expressed as follows [13]:

$$D_{jl}(\mathbf{k}, t) = \frac{4\pi c}{k}\sin(ckt)\left(\delta_{jl} - \frac{k_l k_j}{k^2}\right)\eta(t),$$

$$D_{jl}(\mathbf{k}, \omega) = \int_{-\infty}^{\infty} dt D_{jl}(\mathbf{k}, t)\exp(i\omega t)$$

$$= 4\pi(k^2 + \kappa^2)^{-1}\left(\delta_{jl} - \frac{k_l k_j}{k^2}\right), \quad (8)$$

$$k = |\mathbf{k}|, \quad \kappa^2 = \left(\frac{i\omega}{c}\right)^2(1 + i\varepsilon\,\mathrm{sgn}\,\omega).$$

Owing to $i\varepsilon\,\mathrm{sgn}\,\omega$ ($\varepsilon > 0$), a correct detour is taken around the pole of the retarded Green's function.

We substitute solution (7) into vector potential (5) determining the quantum Lorentz force in Eq. (4). Symmetrizing the product of the commuting operators $A_j(\mathbf{k}, t)$ and $\exp(i\mathbf{k}\cdot\mathbf{r}(t))$, we obtain

$$A_j(\mathbf{r}(t), t) = \frac{e}{c}\int_{-\infty}^{\infty} dt_1 \int \frac{d^3k}{(2\pi)^3} D_{jl}(\mathbf{k}, t - t_1)$$

$$\times \frac{1}{2}[\exp(i\mathbf{k}\cdot\mathbf{r}(t)), \dot{r}_l(t_1)\exp(-i\mathbf{k}\cdot\mathbf{r}(t_1))]_+.$$

Dropping the second term in the expression for the Lorentz force, we arrive at the following equation of motion of the relativistic electron allowing for its own radiation field:

$$\frac{d}{dt}\pi_j(t) + \nabla_j V(\mathbf{r}, t) = -\frac{e^2}{c^2}\frac{d}{dt}$$

$$\times \int_{-\infty}^{\infty} dt_1 \int \frac{d^3k}{(2\pi)^3} D_{jl}(\mathbf{k}, t - t_1) \quad (9)$$

$$\times \frac{1}{2}[\exp(i\mathbf{k}\cdot\mathbf{r}(t)), \dot{r}_l(t_1)\exp(-i\mathbf{k}\cdot\mathbf{r}(t_1))]_+ = F_j(t).$$

Thus, we ignore the effect of the electromagnetic vacuum field on the electron. This considerably simplifies the initial stage of analysis of the interaction retardation effect.

To analyze the radiative damping effect, we consider the electron in a uniform electric field of strength $\mathbf{E}(t)$. The corresponding interaction energy is

$$V(t) = -r_j e E_j(t) = -r_j f_j.$$

The external force $f_j(t) = eE_j(t)$ introduced here is canonically conjugate to the electron coordinate $r_j(t)$. According to time-dependent perturbation theory, the coordinate operator $r_\beta(t)$ of an electron driven by an external force is written as [22]:

$$r_j(t) = r_j^0(t) + \int_{-\infty}^{\infty} dt_1 \hat{\varphi}_{jl}(t, t_1) f_l(t_1),$$

where $r_j^0(t)$ corresponds to free evolution ($f = 0$) and

$$\hat{\varphi}_{jl}(t, t_1) = \frac{i}{\hbar}[r_j^0(t), r_l^0(t_1)]_-\eta(t - t_1) = \frac{\delta r_j(t)}{\delta f_l(t_1)} \quad (10)$$

is the linear random response equal to the quantum Poisson bracket of $r_j^0(t)$ and $r_l^0(t_1)$ taken with the opposite sign and multiplied by the Heaviside step function $\eta(\tau)$. Averaging Eq. (10) over the initial state of the entire system, we obtain

$$\langle r_j(t)\rangle = \int dt_1 \varphi_{jl}(t, t_1) f_l(t_1), \quad (11)$$

where

$$\left\langle \frac{\delta r_j(t)}{\delta f_l(t_1)}\bigg|_{f=0}\right\rangle$$

$$= \left\langle \frac{i}{\hbar}[r_j^0(t), r_l^0(t_1)]_-\eta(t - t_1)\right\rangle = \varphi_{jl}(t, t_1)$$

is the so-called difference-time linear response. It is isotropic:

$$\varphi_{jl}(t, t_1) = \varphi(t, t_1)\delta_{jl}. \quad (12)$$

Under a periodic perturbation

$$f_\alpha(t) = f_\alpha(\omega)\exp(-i\omega t) + \text{c.c.},$$

Eq. (11) yields

$$\langle r_\alpha(t)\rangle = \chi(\omega) f_\alpha(\omega)\exp(-i\omega t) + \text{c.c.},$$

where

$$\chi(\omega) = \int d\tau \exp(i\omega\tau)\varphi(\tau)$$

is the linear susceptibility determining the frequency dependence of energy dissipation and dispersion due to reaction of the radiation field. To calculate the susceptibility modified to allow for the radiative force con-

tained in (9), one must find its increment induced by the perturbation, i.e., calculate the derivative

$$\left\langle \frac{1}{m}\frac{\delta F_j(t)}{\delta f_l(t_2)}\right\rangle = -\frac{d}{dt}\int dt'\left\langle \hat{\gamma}_{jl}(t,t')\frac{d}{dt'}\hat{\varphi}_{jl}(t',t_2)\right\rangle. \quad (13)$$

Here, we make use of isotropy of the linear response and transverse gauge (3) for Green's function.

We assume that the nonlinearity associated with the radiative force is small. Therefore, fluctuations of response can be neglected according to nonlinear fluctuation-dissipation theorems [22]. Thus, the response in Eq. (13) can be replaced by its mean value. Using Eqs. (11) and (12), we obtain

$$\left\langle \frac{1}{m}\frac{\delta F_j(t)}{\delta f_l(t_2)}\right\rangle = -\frac{d}{dt}\int dt'\gamma_{jl}(t-t')\frac{d}{dt'}\varphi(t'-t_2), \quad (14)$$

where

$$\gamma_{jl}(t-t_1) = \frac{e^2}{mc^2}\int\frac{d^3k}{(2\pi)^3}D_{jl}(\mathbf{k},t-t_1) \times\left\langle \frac{1}{2}[\exp(i\mathbf{k}\cdot\mathbf{r}(t)),\exp(-i\mathbf{k}\cdot\mathbf{r}(t_1))]_+\right\rangle. \quad (15)$$

Angle brackets mean averaging over the ground state of the entire system. Both function (15) and its Fourier spectrum,

$$\gamma_{jl}(\omega) = \int d\tau\gamma_{jl}(\tau),$$

will be called the radiative damping coefficient. By virtue of (14) and (15), Eq. (9) takes the form

$$\frac{1}{m}\frac{d}{dt}\pi_j + \frac{1}{m}\nabla_j V(\mathbf{r}) + \frac{d}{dt}\int dt_1\gamma_{jl}(t-t_1)\dot{r}_l(t_1) = \frac{1}{m}f_j(t). \quad (16)$$

In the approximation adopted here, the radiative damping coefficient is calculated without using the second term on the right-hand side of the Lorentz equation. Moreover, we ignore the contribution to Eq. (15) due to the parametric effect of electromagnetic vacuum fluctuations.

In the intrinsic coordinate system, the average coordinate $\langle r_j(t)\rangle$ of the electron in the potential field $V(r) = m\Omega^2 r^2/2$ and uniform electric field is governed by the equation

$$\frac{d^2}{dt^2}\langle r_j(t)\rangle + \Omega^2\langle r_j(t)\rangle = \frac{1}{m}f_j(t) - \frac{d}{dt}\int dt_1\gamma_{jl}(t-t_1)\frac{d}{dt_1}\langle r_l(t_1)\rangle. \quad (17)$$

Now, we use Eq. (17) to find the response to the harmonic force

$$f_j(t) = f_j(\omega)\exp(-i\omega t),$$

assuming that $\gamma_{jl}(\tau) = \delta_{jl}\gamma(\tau)$. By definition,

$$\langle r_j(\omega)\rangle = \chi(\omega)f_j(\omega). \quad (18)$$

Taking the Fourier transform of Eq. (17), we obtain

$$(\Omega^2-\omega^2)\langle r_j(\omega)\rangle = \frac{1}{m}f_j(\omega) + \omega^2\gamma(\omega)\langle r_j(\omega)\rangle.$$

According to definition (18), the linear susceptibility has the form

$$\chi(\omega) = -\frac{1}{m}\frac{1}{\omega^2(1+\gamma(\omega))-\Omega^2}. \quad (19)$$

By the causality principle, susceptibility (19) must be an analytic function of the complex variable ω in the upper half-plane [23].

3. QUANTUM SPACETIME NONLOCALITY

The radiative damping coefficient satisfies the causality principle and allows for the retardation of interaction between the electron and its own radiation field. The latter property is explained by the finite speed of light and the use of the electron-coordinate Heisenberg operator taken at different instants. The essentially nonlinear dependence of (15) on the coordinate operators $\mathbf{r}(t)$ and $\mathbf{r}(t_1)$ is determined by the product of exponential factors

$$\exp(i\mathbf{k}\cdot\mathbf{r}(t))\exp(-i\mathbf{k}\cdot\mathbf{r}(t_1)) = \exp(i\mathbf{k}\cdot\Delta\mathbf{r})\exp(-iB), \quad (20)$$

where $\Delta\mathbf{r} = \mathbf{r}(t) - \mathbf{r}(t_1)$ is the operator of displacement over the time interval $t - t_1$. The unitary operator $\exp(-iB)$ appears in Eq. (20) because the operators $\mathbf{r}(t)$ and $\mathbf{r}(t_1)$ taken at different instants do not commute. It affects the retardation of the interaction between the electron and radiation field. In the nonrelativistic limit, this retardation was taken into account in [11] in order to resolve the self-acceleration paradox. In contrast to classical electrodynamics, QED involves a funda-

mental small parameter, the fine-structure constant $\alpha = e^2/\hbar c$. Therefore, product (20) contained in expression (15) for the radiative damping coefficient can be approximately calculated for the free evolution of the electron coordinate operator $r_j(t)$ governed by the Hamiltonian

$$H = c\boldsymbol{\alpha}\cdot\mathbf{p} + \beta\varepsilon_0, \quad \varepsilon_0 = mc^2. \tag{21}$$

According to the Heisenberg equations,

$$\begin{aligned} \dot{r}_j(t) &= c\alpha_j, \\ \dot{\alpha}_j(t) &= \frac{1}{i\hbar}[\alpha_j, H]. \end{aligned} \tag{22}$$

The momentum $\mathbf{p}$ and Hamiltonian (21) of the electron are integrals of motion. The commutator of the Dirac matrices $\boldsymbol{\alpha}$ and β yields

$$\alpha_j H + H\alpha_j = 2cp_j. \tag{23}$$

Solving Eqs. (22) combined with Eq. (23), we obtain

$$\begin{aligned} r_j(t) &= r_j(0) + \frac{c^2 p_j}{H}\left[t - \frac{\hbar}{2\tilde{H}}\sin\left(\frac{2\tilde{H}}{\hbar}t\right)\right] \\ &+ \alpha_j(0)\frac{\hbar c}{2\tilde{H}}\sin\left(\frac{2\tilde{H}}{\hbar}t\right) \\ &+ \dot{\alpha}_j(0)\frac{\hbar^2 c}{4\tilde{H}^2}\left[1 - \cos\left(\frac{2\tilde{H}}{\hbar}t\right)\right], \end{aligned} \tag{24}$$

where $\tilde{H} = \sqrt{c^2p^2 + \varepsilon_0^2}$, which follows from the condition $\tilde{H}^2 \equiv H^2 = c^2p^2 + \varepsilon_0^2$. Solution (24) was originally obtained by Schrödinger (see [24]).

We use solution (24) to calculate the variance of the operator

$$F_0 = \mathbf{k}\cdot(\mathbf{r}(t) - \mathbf{r}(t_1)) = \mathbf{k}\cdot\Delta\mathbf{r} \tag{25}$$

in the electron's intrinsic coordinate system. Using the commutator of the matrices α_j and $\dot{\alpha}_j$, we derive the following expression for the variance of increment (25):

$$F_0^2 = k^2\lambda_0^2\sin^2(\omega_0 t) = x^2\sin^2(\omega_0 t), \tag{26}$$

where

$$\omega_0 = \frac{mc^2}{\hbar}, \quad x = \frac{k}{k_0}, \quad \lambda_0 = \frac{1}{k_0}.$$

The variance of an increment of the Dirac electron coordinate is finite in the intrinsic coordinate system in contrast to the case of a relativistic spinless particle. Fast oscillation at long times ($t \gg t_0 = 1/\omega_0$) is insignificant, and we can write

$$F_0^2 \approx \frac{1}{2}k^2\lambda_0^2 = \frac{1}{2}x^2. \tag{27}$$

Thus, the variance of $\mathbf{r}(t) - \mathbf{r}(t_1)$ for the Dirac electron is finite in the intrinsic coordinate system by virtue of the internal degrees of freedom defined by the Dirac matrices. As shown below, the finite variance entails retardation of the interaction between the electron and its own radiation field. Thus, the well-known paradoxes of radiative damping and logarithmic divergence in QED are resolved. The retardation of the interaction is determined by the time required for light to travel the Compton wavelength:

$$t_0 = \frac{\lambda_0}{c} = \frac{\hbar}{mc^2} = \frac{1}{\omega_0}.$$

Therefore, the quantum theory of radiative damping must be essentially non-Markovian to allow for retardation of the interaction between electron and radiation field. Since the characteristic retardation time t_0 is very small, the approximation of free evolution of $\mathbf{r}(t)$ in Eqs. (9) and (20) over times on the order of t_0 is justified.

To find the radiative damping coefficient, we must calculate the product in (20). The factor $\exp(i\mathbf{k}\cdot\Delta\mathbf{r})$ is strictly determined by solution (24), where operator (25) has property (26). A nontrivial part of the calculation of the unitary operator $\exp(-iB)$ is associated with the dependence of the electron coordinate $\mathbf{r}(t)$ on momentum and the Dirac matrices dictated by Eq. (24). After some calculations, we have

$$\begin{aligned} &\exp(i\mathbf{k}\cdot\mathbf{r}(t))\exp(-i\mathbf{k}\cdot\mathbf{r}(t_1)) = \exp(i\mathbf{k}\cdot\Delta\mathbf{r}) \\ &\times\exp\left\{-\frac{i\omega_0\tau}{2}[\beta x\arctan x + \mathbf{n}\cdot\boldsymbol{\alpha}(x - \arctan x)]\right\} \end{aligned} \tag{28}$$

and the inverse product of the exponentials

$$\begin{aligned} &\exp(-i\mathbf{k}\cdot\mathbf{r}(t_1))\exp(i\mathbf{k}\cdot\mathbf{r}(t)) \\ &= [\exp(i\mathbf{k}\cdot\mathbf{r}(t))\exp(-i\mathbf{k}\cdot\mathbf{r}(t_1))]^{\dagger}\big|_{\mathbf{k}\to-\mathbf{k}} \\ &= \exp\left\{-\frac{i\omega_0\tau}{2}[-\beta x\arctan x + \mathbf{n}\cdot\boldsymbol{\alpha}(x - \arctan x)]\right\} \\ &\times\exp(i\mathbf{k}\cdot\Delta\mathbf{r}), \end{aligned}$$

where $\mathbf{n}$ is the unit vector along $\mathbf{k}$ and $\tau = t - t_1$. Using these expressions, one can write the anticommutator in formula (15) for the radiative damping coefficient in explicit form. Since the expression for the operator $\Delta\mathbf{r}$

includes the Dirac matrices by virtue of Eq. (24), we can represent $\exp(i\mathbf{k}\cdot\Delta\mathbf{r})$ as

$$\exp(i\mathbf{k}\cdot\Delta\mathbf{r}) = \cos(k\Delta r) + \frac{i\mathbf{k}\cdot\Delta\mathbf{r}}{k\Delta r}\sin(k\Delta r), \quad (29)$$

where $\Delta r = |\Delta\mathbf{r}| = \lambda_0|\sin(\omega_0\tau)|$ does not involve the Dirac matrices. In what follows, the small term $\sin(k\Delta r)$ in Eq. (29) is disregarded. By virtue of expressions (28) and (29) combined with (24) for $\Delta\mathbf{r}$ in the intrinsic coordinate system, the anticommutator in Eq. (15) in the isotropic case is

$$\begin{aligned}&\langle[\exp(i\mathbf{k}\cdot\mathbf{r}(t)), \exp(-i\mathbf{k}\cdot\mathbf{r}(t_1))]_+\rangle\\ &= 2\cos(x|\sin(\omega_0\tau)|)\cos(\omega_0\tau f(x)),\end{aligned} \quad (30)$$

where

$$f(x) = 0.5\sqrt{(x-\arctan x)^2 + x^2\arctan^2 x} \approx 0.5x\arctan x.$$

Substituting (30) into expression (15) for the radiative damping coefficient, averaging over the electron ground state $\langle\beta\rangle = 1$, and calculating the integral over the angular coordinates in the wave-vector space, we obtain

$$\begin{aligned}\gamma(\tau) &= \frac{4\alpha\omega_0}{3\pi}\int_0^\infty dx x\sin(x\omega_0\tau)\\ &\times\cos(f(x)\omega_0\tau)\cos(x|\sin(\omega_0\tau)|)\eta(\tau).\end{aligned} \quad (31)$$

3.1. First, we consider the asymptotic contribution to $\gamma(\tau)$ corresponding to high momentum transfer as $x \longrightarrow \infty$. In this case, we can set $\arctan x = \pi/2$ and rewrite (31) as

$$\begin{aligned}\gamma(\tau) &= \frac{\alpha}{3\pi}\omega_0\int_0^\infty dx x\\ &\times\left\{\sin\left[\left(\left(1+\frac{\pi}{4}\right)\omega_0\tau + |\sin(\omega_0\tau)|\right)x\right]\right.\\ &+\sin\left[\left(\left(1+\frac{\pi}{4}\right)\omega_0\tau - |\sin(\omega_0\tau)|\right)x\right]\\ &+\sin\left[\left(\left(1-\frac{\pi}{4}\right)\omega_0\tau + |\sin(\omega_0\tau)|\right)x\right]\\ &\left.+\sin\left[\left(\left(1-\frac{\pi}{4}\right)\omega_0\tau - |\sin(\omega_0\tau)|\right)x\right]\right\}\eta(\tau).\end{aligned} \quad (32)$$

Since the integrand is even with respect to x, we use the causality principle ($\tau > 0$) and calculate (32) to obtain

$$\gamma(\tau) = \frac{\alpha\omega_0}{3}\delta'\left(\left(1-\frac{\pi}{4}\right)\omega_0\tau - |\sin(\omega_0\tau)|\right)\eta(\tau).$$

To calculate the frequency dependence of the radiative damping coefficient,

$$\begin{aligned}\gamma(\omega) &= \int_{-\infty}^\infty dt\exp(i\omega t)\gamma(t)\\ &= \frac{\alpha\omega_0}{3}\int_{-\infty}^\infty dt\exp(i\omega t)\\ &\times\delta'\left(\left(1-\frac{\pi}{4}\right)\omega_0 t - |\sin(\omega_0 t)|\right)\eta(t),\end{aligned}$$

we set the argument of the δ function equal to zero:

$$\begin{aligned}\sin(\omega_0 t^*) &= \left(1-\frac{\pi}{4}\right)\omega_0 t^* = \theta = 0.56,\\ \omega_0 t^* &= 2.55.\end{aligned} \quad (33)$$

Then, we obtain the asymptotic formula

$$\begin{aligned}\gamma(\omega) &= \frac{\alpha}{3}\left[\frac{i\omega}{\omega_0}\frac{1}{[1-\pi/4-\cos(\omega_0 t^*)]^2}\right.\\ &\left.+\frac{|\sin(\omega_0 t^*)|}{|1-\pi/4-\cos(\omega_0 t^*)|^3}\right]\exp(i\omega t^*),\end{aligned} \quad (34)$$

which satisfies the causality principle and is free of the self-acceleration paradox. The main conclusion that follows from our asymptotic analysis is that the region defined by condition (33) provides a considerable contribution to the radiative damping coefficient. Therefore, we can use (33) to rewrite the starting formula (31) as

$$\begin{aligned}\gamma(\tau) &= \frac{4\alpha\omega_0}{3\pi}\\ &\times\int_0^\infty x dx\sin(x\omega_0\tau)\cos(f(x)\omega_0\tau)\cos(x\theta)\eta(\tau).\end{aligned} \quad (35)$$

Hence, we obtain the frequency dependence of the damping coefficient:

$$\begin{aligned}\gamma(\omega) &= \frac{4\alpha\omega_0}{3\pi}\int_{-\infty}^\infty d\tau\exp(i\omega\tau)\\ &\times\int_0^\infty x dx\sin(x\omega_0\tau)\cos(f(x)\omega_0\tau)\cos(x\theta)\eta(\tau).\end{aligned} \quad (36)$$

In the asymptotic limit, $\arctan x = \pi/2$ and expression (36) reduces to (34).

3.2. Now, we analyze the contribution to Eq. (31) due to low momentum transfer at low frequencies. In this region, the fast oscillation in expression (26) for the electron-coordinate variance can be disregarded; i.e., $F_0^2 \approx x^2/2$ according to Eq. (27). In view of this estimate, expression (31) for $\theta = 1/\sqrt{2}$ entails (35).

The basic property of radiative damping coefficient (36) is that it characterizes the retardation of the interaction between the electron and its own radiation field. To elucidate this property, one can ignore the quantum corrections associated with the factor $\arctan x$ and set $\cos(f(x)\omega_0\tau) \approx 1$ in Eq. (36). Then, the frequency dependence of the radiative damping coefficient is written as

$$\gamma(\omega) = \alpha\frac{4\omega_0}{3\pi}\int_{-\infty}^{\infty} dx x \times \int_{-\infty}^{\infty} d\tau \exp(i\omega\tau)\sin(x\omega_0\tau)\cos\left(\frac{x}{\sqrt{2}}\right)\eta(\tau). \quad (37)$$

Calculating the integral in (37), we obtain

$$\gamma(\omega) = i\frac{2\alpha}{3}\frac{\omega}{\omega_0}\exp\frac{i\omega}{\sqrt{2}\omega_0}. \quad (38)$$

In view of Eq. (38), linear susceptibility (19) at $\Omega = 0$ is

$$\chi(\omega) = -\frac{1}{m\omega^2}\left(1 + i\frac{2\alpha}{3}\frac{\omega}{\omega_0}\exp\frac{i\omega}{\sqrt{2}\omega_0}\right)^{-1}. \quad (39)$$

It is easy to see that susceptibility (39) has no poles in the upper half-plane of the complex variable ω and therefore satisfies the causality principle.

At the same time, the radiative damping coefficient

$$\gamma^0(\omega) = i\omega\tau_0, \quad \tau_0 = \frac{2}{3}\frac{e^2}{mc^3} = \frac{2}{3}\frac{\alpha}{\omega_0},$$

derived from the classical Abraham–Lorentz formula [1], gives rise to a pole of susceptibility (39) in the upper ω half-plane, which leads to the self-acceleration paradox in the classical radiative damping theory [23].

Formula (39) for the linear response implies the following equation of motion for an electron in external electric field:

$$m\ddot{r}_j(t) - m\tau_0\dddot{r}_j(t - t_0) = eE_j(t), \quad \tau_0 = \frac{2}{3}\frac{e^2}{mc^3}, \quad t_0 = \frac{1}{\sqrt{2}}\frac{\hbar}{mc^2}. \quad (40)$$

In particular, in the absence of the external field, Eq. (40) becomes

$$m\ddot{r}_j(t) - m\tau_0\dddot{r}_j(t - t_0) = 0,$$

which describes stable free motion of the electron with $\mathbf{v} = \text{const}$.

Thus, the retardation of the interaction resolves the self-acceleration paradox inherent in classical Abraham–Lorentz theory. Moreover, Eq. (40) provides a rigorous justification of classical approximate formulas for the radiative damping force. Since the radiative damping force "switches on" after a delay t_0, during the delay interval, the electron can be treated as driven by the external force only. Therefore, the equation of motion for this interval has the form

$$m\ddot{r}_j(t) = eE_j(t).$$

Substituting

$$\dddot{r}_j(t) = \frac{1}{m}\frac{d}{dt}eE_j(t)$$

into Eq. (40), we obtain

$$m\ddot{r}_j(t) - eE_j(t) = \tau_0\frac{d}{dt}eE_j(t - t_0) \approx \tau_0\frac{d}{dt}eE_j(t). \quad (41)$$

This equation rigorously justifies the formal procedure of eliminating the self-acceleration paradox commonly used in the classical theory [3, 25].

Finally, we present the expression

$$\gamma''(\omega) = \frac{2\alpha}{3}\int_0^{\infty} dx x\cos\frac{x}{\sqrt{2}} \times \delta\left(\frac{\omega}{\omega_0} - (x + f(x))\right)\operatorname{sgn}\omega \quad (42)$$

for the imaginary part of the radiative damping coefficient obtained in [16] by taking into account electromagnetic vacuum fluctuations. This quantity determines radiative damping proper.

The real part of the radiative damping coefficient determines the so-called dispersion or self-energy of the electron. In particular, the zero-frequency damping coefficient $\gamma(0) = \gamma'(0)$ determines the contribution to the electromagnetic electron mass due to reaction of the radiation field.

4. RADIATIVE DAMPING, ELECTROMAGNETIC MASS OF THE ELECTRON, AND THE LAMB SHIFT

The physical significance of the quantum nonlocality (retardation of interaction) is not restricted to resolution of paradoxes inherent in the classical theory of radiative damping. The retardation of the interaction between the electron and the radiation field removes divergences inherent in traditional QED methods. In particular, the above formulas for $\gamma(\omega)$ imply that the contribution of the reaction of the radiation field to the electromagnetic electron mass is finite. Indeed, setting $\omega = 0$ in asymptotic formula (34), we obtain

$$\delta m_{\rm el} = m_0\gamma(0) = \frac{\alpha m_0}{3}\frac{|\sin(\omega_0 t^*)|}{|1-\pi/4-\cos(\omega_0 t^*)|^3}. \quad (43)$$

Let us find out how physical effects caused by the interaction of the electron with its own radiation field are related to electromagnetic vacuum field.

4.1. By the causality principle, the real part $\gamma'(\omega)$ of the radiative damping coefficient is uniquely determined by its imaginary part. It follows from the Kramers–Kronig relations that

$$\gamma'(\omega) = -\frac{1}{\pi}\int_{-\infty}^{\infty} d\omega'\frac{\gamma''(\omega')}{\omega-\omega'} = -\frac{2}{\pi}\int_0^{\infty} d\omega'\frac{\gamma''(\omega')\omega'}{\omega^2-\omega'^2}.$$

In view of Eq. (42), we have the following expression for the electromagnetic mass of the electron:

$$\delta m_{\rm el} = m_0\gamma'(0) = \frac{2m_0}{\pi}\int_0^{\infty} d\omega'\frac{\gamma''(\omega')}{\omega'}$$

$$= \frac{2m_0\alpha}{\pi}\int_0^{\infty} dx\frac{\cos(x/\sqrt{2})}{1+0.5\arctan x} \approx 0.5\alpha m_0.$$

Thus, we come to the fundamental conclusion that the electron mass is only in part of electromagnetic nature. Calculation of the lepton mass spectrum is an open problem. The relation between radiative damping controlled by $\gamma''(\omega)$ and the electromagnetic electron mass determined by dispersion relations indicates that knowledge of the imaginary part of the radiative damping coefficient over the entire frequency range is indispensable.

4.2. Moreover, the Callen–Welton fluctuation-dissipation theorem can be invoked to determine the contribution of $\gamma'(\omega)$ to the Lamb shift, a phenomenon entirely due to fluctuations.

Suppose that an electron in the spherically symmetric potential field $V(r) = m\Omega_n^2 r^2/2$ is subjected to electromagnetic vacuum fluctuations. By the Callen–Welton fluctuation–dissipation theorem, the spectral density of coordinate-projection fluctuations at zero temperature is determined by the imaginary part of susceptibility (19):

$$(x^2)_\omega = \hbar\chi''(\omega)\,\mathrm{sgn}\,\omega = \frac{\hbar}{m}\frac{\omega^2\gamma''(\omega)\,\mathrm{sgn}\,\omega}{(\omega^2-\Omega_m^2)^2+\omega^4\gamma''^2}.$$

Therefore, the total square of coordinate fluctuations has the form

$$\langle\delta x^2\rangle = \frac{\hbar}{\pi m}\int_0^{\infty} d\omega\frac{\omega^2\gamma''(\omega)}{(\omega^2-\Omega_n^2)^2+\omega^4\gamma''^2}. \quad (44)$$

According to Welton [26], the Lamb shift is due to a fluctuation-induced correction to the potential energy that can be found as follows. Consider the electron in a hydrogen atom with the effective potential

$$V(\mathbf{r}+\delta\mathbf{r}), \quad (45)$$

where $\mathbf{r}$ is the average location of the electron, $\delta\mathbf{r}$ is its fluctuation, and $V(\mathbf{r}) = -e^2/r$. We expand potential (45) into a series in fluctuations of the electron's location and average over the unperturbed vacuum state. Owing to isotropy, this expansion has the form

$$\langle V(\mathbf{r}+\delta\mathbf{r})\rangle = V(\mathbf{r}) + \frac{1}{2}\Delta V(r)\langle\delta x^2\rangle + \ldots.$$

Thus, the fluctuation correction can be written in dimensionless variables as

$$\delta V = \frac{1}{2}\Delta V(r)\langle\delta x^2\rangle = \frac{1}{2}\frac{e^2}{a}4\pi\delta(\boldsymbol{\rho})\langle\delta\rho_x^2\rangle, \quad \boldsymbol{\rho} = \frac{\mathbf{r}}{a},$$

where a is the Bohr radius.

According to perturbation theory, the first correction to the nth level is

$$\delta E_{\rm L} = \langle n|\delta V|n\rangle = \frac{1}{2}\frac{e^2}{a}\langle\delta\rho_x^2\rangle 4\pi|\psi_n(0)|^2$$

$$= \frac{2\alpha^3 e^2}{n^3\lambda}\langle\delta\rho_x^2\rangle. \quad (46)$$

Accurate calculations remain beyond the scope of this paper in view of the approximations made above. For this reason, we calculate the Lamb shift given by Eq. (46) by applying the Callen–Welton fluctuation–dissipation theorem to a very simple oscillator model of the bound electron state in the hydrogen atom with

potential (45). Using expression (42) for the imaginary part of the radiative damping coefficient, we obtain

$$\delta E_{\mathrm{L}} = \frac{2\alpha^4 mc^2}{\pi n^3}\int_0^\infty d\omega \frac{\omega^2\gamma''(\omega)}{(\omega^2-\tilde{\Omega}_n^2)^2+\omega^4\gamma''^2} = \frac{4\alpha^5 mc^2}{3\pi n^3}\int_0^\infty dx \frac{x(x+f(x))^2\cos(x/\sqrt{2})}{[(x+f(x))^2-\tilde{\Omega}_n^2]^2}, \quad (47)$$

where $\tilde{\Omega}_n = \Omega_n/\omega_0$ is a characteristic frequency proportional to α^2 and weakly dependent on the level number n. To calculate the integral in (47), we divide the integration interval into two parts: from 0 to α and from α to infinity. We note that quantum nonlocality eliminates the divergence of integral (47) at the upper limit $x = \infty$. The resulting high-frequency contribution of the range $\alpha < x < \infty$ is calculated accurately, because motion of the electron can be treated as free in this range:

$$\int_\alpha^\infty dx \frac{x\cos(x/\sqrt{2})}{(x+x\arctan(x)/2)^2} \approx 4.14.$$

Calculations in the low-frequency range $0 < x < \alpha$ are less accurate, because we use an oscillator model of the bound state and disregard the change in the radiative damping coefficient $\gamma(\omega)$ due to the bound state of the electron. In this approximation, according to (47),

$$\int_0^\alpha dx \frac{x(x+x^2/2)\cos(x/\sqrt{2})}{((x+x^2/2)^2-\tilde{\Omega}_n^2)^2} = -\frac{1}{2}+\ln\alpha-\ln\tilde{\Omega}_n-\frac{\alpha}{2}.$$

The sum of the two parts is the Lamb shift given by Eq. (47):

$$\delta E_{\mathrm{L}} = \frac{4\alpha^5 mc^2}{3\pi n^3}\left(2\ln\frac{1}{\alpha}+\ln 2-1.3\right) = 0.48mc^2.$$

Thus, the important dissipation characteristic $\gamma''(\omega)$ calculated here by taking into account the interaction retardation simultaneously determines the Lamb shift and the self-energy part including the electromagnetic mass of the electron.

5. CONCLUSIONS

This paper reports a new QED approach free of divergences characteristic for the traditional methods of quantum field theory. The Lamb shift and other QED effects are actively investigated in current experiments. The new methods and results presented in this work may provide a starting point for transcending the limitations of perturbation theory and reducing the gap between theory and experiment.

It is proved that the interaction between the Dirac electron and its own radiation field in local quantum theory is retarded by about the time required for light to travel the Compton wavelength. The quantum space-time nonlocality is directly associated both with the finiteness of the variance of coordinate increment and with the phase factor arising due to the fact that coordinate operators taken at different times do not commute.

Thus, it is rigorously shown how the interaction between the Dirac electron and its own radiation field is retarded. This retardation eliminates the divergences and paradoxes associated with the classical electron as a pointlike entity. The solution to the century-old paradox of self-acceleration is of fundamental importance per se and can be essential for further development of theory. Finally, the relation between radiative damping, finite contribution to the electromagnetic mass of the electron, and the Lamb shift is found.

REFERENCES

1. M. Abraham, *Theorie der Elektrizität*, Vol. 2: *Elektromagnetische Theorie der Strahlung* (Teubner, Leipzig, 1905); H. A. Lorentz, *The Theory of Electrons and Its Applications to the Phenomena of Light and Radiant Heat*, 2nd ed. (Dover, New York, 1952; Gostekhizdat, Moscow, 1953).
2. J. D. Jackson, *Classical Electrodynamics*, 2nd ed. (Wiley, New York, 1975; Inostrannaya Literatura, Moscow, 1965).
3. V. L. Ginzburg, *Theoretical Physics and Astrophysics*, 2nd ed. (Nauka, Moscow, 1981; Pergamon Press, Oxford, 1979).
4. G. W. Ford, J. T. Lewis, and R. F. O'Connell, Phys. Rev. Lett. **55**, 2273 (1985).
5. P. M. Barone and A. O. Caldeira, Phys. Rev. A **43**, 57 (1991).
6. V. Hakim and V. Ambegaokar, Phys. Rev. A **32**, 423 (1985).
7. M. Cetto and L. de la Pena, Phys. Rev. A **37**, 1952 (1988).
8. A. O. Barut and J. F. Van Huele, Phys. Rev. A **32**, 3187 (1985).
9. M. D. Crisp, Phys. Rev. A **42**, 3703 (1990).
10. A. I. Nikishov, Zh. Éksp. Teor. Fiz. **110**, 510 (1996) [JETP **83**, 274 (1996)].
11. G. F. Efremov, Zh. Éksp. Teor. Fiz. **110**, 1629 (1996) [JETP **83**, 896 (1996)].
12. G. F. Efremov, Zh. Éksp. Teor. Fiz. **114**, 1661 (1998) [JETP **87**, 899 (1998)].
13. A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskiĭ, *Methods of Quantum Field Theory in Statistical Physics* (Fizmatgiz, Moscow, 1962; Prentice Hall, Englewood Cliffs, N.J., 1963).

14. G. F. Efremov, M. A. Novicov, and L. G. Mourokh, in *The Present Status of Quantum Theory of Light*, Ed. by S. Jeffers and J. P. Vigier (Kluwer Academic, Dordrecht, 1997), p. 97.
15. G. F. Efremov, M. A. Novicov, and B. B. Ivanov, in *Causality and Locality in Modern Physics* (Kluwer Academic, Dordrecht, 1996), p. 87.
16. G. F. Efremov and V. V. Sharkov, Vestn. Nizhegorod. Univ. im. N. I. Lobachevskogo, Mat. Model. Optim. Upr., No. 25, 141 (2002).
17. G. F. Efremov and V. A. Kazakov, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **22**, 453 (1979).
18. G. F. Efremov and A. Yu. Smirnov, Zh. Éksp. Teor. Fiz. **80**, 1071 (1981) [Sov. Phys. JETP **53**, 547 (1981)].
19. G. F. Efremov, *Stochastic Equations for Open Quantum Systems: A Textbook* (Gork. Gos. Univ., Gorkiĭ, 1982).
20. J. Schwinger, J. Math. Phys. **2**, 407 (1961).
21. R. Senitzky, Phys. Rev. **119**, 670 (1960); **124**, 642 (1961).
22. G. F. Efremov, Zh. Éksp. Teor. Fiz. **55**, 2322 (1968) [Sov. Phys. JETP **28**, 1232 (1968)].
23. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 5: *Statistical Physics*, 3rd ed. (Nauka, Moscow, 1976; Pergamon Press, Oxford, 1980), Part 1.
24. P. A. M. Dirac, *The Principles of Quantum Mechanics*, 4th ed. (Clarendon Press, Oxford, 1958; Fizmatgiz, Moscow, 1960).
25. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, 7th ed. (Nauka, Moscow, 1988; Pergamon Press, Oxford, 1975).
26. T. A. Welton, Phys. Rev. **74**, 1157 (1948).

Translated by R. Tyapaev

Journal of Experimental and Theoretical Physics, Vol. 98, No. 2, 2004, pp. 181–185.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 2, 2004, pp. 205–209.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Corrections to Deuterium Hyperfine Structure due to Deuteron Excitations¶

I. B. Khriplovich and A. I. Milstein

Budker Institute of Nuclear Physics, Russian Academy of Sciences, Novosibirsk, 630090 Russia
Novosibirsk State University, Novosibirsk, 630000 Russia
e-mail: khriplovich@inp.nsk.su, e-mail: milstein@inp.nsk.su

Received April 23, 2003

Abstract—We consider the corrections to the deuterium hyperfine structure originating from the two-photon exchange between an electron and deuteron, with deuteron excitations in the intermediate states. In particular, the motion of the two intermediate nucleons as a whole is taken into account. The problem is solved in the zero-range approximation. The result is in good agreement with the experimental value of deuterium hyperfine splitting. *© 2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Hyperfine splitting in the deuterium ground state has been measured with high accuracy. The most precise experimental result, obtained with an atomic deuterium maser, is given by [1]

$$\nu_{\rm exp} = 327\,384.3525222(17)\ \text{kHz}. \qquad (1)$$

On the other hand, theoretical calculation, including higher order pure QED corrections, gives

$$\nu_{QED} = 327\,339.27(7)\ \text{kHz}. \qquad (2)$$

This value was obtained by using the theoretical result for the hydrogen hyperfine splitting in [2],

$$1\,420\,451.95(14)\ \text{kHz},$$

which does not include the proton structure and recoil radiative correction, and by further combining it with the theoretical ratio of the hyperfine constants in hydrogen and deuterium from [3],

$$4.3393876(8),$$

based on the ratio of the nuclear magnetic moments and including the reduced mass effect in $|\psi(0)|^2$.

It was recognized long ago that the discrepancy

$$\nu_{\rm exp} - \nu_{QED} = 45\ \text{kHz} \qquad (3)$$

is due to the effects caused by the finite size of the deuteron. Such effects are obviously much larger in deuterium than in hydrogen. The corresponding contributions to deuterium hyperfine splitting were discussed long ago with some intuitive arguments [4], and then in more detail in [5–7].

We believe that in the past, the most systematic treatment of such effects, which are due to electron–deuteron interaction of second order in α, was performed in [8]. The effective Hamiltonian of the hyperfine interaction of second order in $\alpha = e^2/4\pi$ was derived therein from the elastic forward scattering amplitude of virtual photons off the deuteron.

In particular, the low-energy theorem for forward Compton scattering [9–12] was generalized in [8] to the case of virtual photons and a target with an arbitrary spin. The corresponding contribution of the momentum transfers k, bounded from above by an inverse deuteron size of $\kappa = 45.7$ MeV, to the relative correction to the deuterium hyperfine structure is

$$\Delta_{el}^{d} = \frac{3\alpha}{8\pi}\left(\mu_d - 2 - \frac{3}{\mu_d}\right)\frac{m_e}{m_p}\ln\frac{\kappa}{m_e}. \qquad (4)$$

Here, m_e and m_p are the electron and proton masses, respectively, and $\mu_d = 0.857$ is the deuteron magnetic moment. The relative corrections Δ are defined here and below as the ratios of the corresponding contributions to the electron-deuteron scattering amplitude to the spin-dependent Born term in this amplitude,

$$T_0 = -\frac{2\pi\alpha}{3m_e m_p}\mu_d(\boldsymbol{\sigma}\cdot\mathbf{s}), \qquad (5)$$

where $\mathbf{s}$ is the deuteron spin.

At larger momentum transfers of $k > \kappa$, the amplitude of the Compton scattering on a deuteron is just the

¶This article was submitted by the authors in English.

coherent sum of those amplitudes on free proton and neutron. This correction is given by

$$\Delta_{in}^{pn} = \frac{3\alpha}{4\pi}\frac{1}{\mu_d}(\mu_p^2 - 2\mu_p - 3 + \mu_n^2)\frac{m_e}{m_p}\ln\frac{m_\rho}{\kappa}, \quad (6)$$

where $\mu_p = 2.79$ and $\mu_n = -1.91$ are the proton and neutron magnetic moments and $m_\rho = 770$ MeV is the usual hadronic scale.

Noteworthy also is the strong numerical cancellation between Δ_{el} and Δ_{in}.

The next correction to the deuterium hyperfine structure, obtained in [8], is induced by the deuteron virtual excitations due to spin currents only. It is given by

$$\Delta_{in}^{(1)} = \frac{3\alpha}{8\pi}\frac{(\mu_p - \mu_n)^2}{\mu_d}\frac{m_e}{m_p}\ln\frac{m_p}{\kappa}. \quad (7)$$

There is also a correction due to a finite distribution of the deuteron charge and magnetic moment.[1] In the zero-range approximation used in [8], this correction is

$$\Delta_f = -\alpha\frac{m_e}{3\kappa}(1 + 2\ln 2). \quad (8)$$

2. LEADING INELASTIC NUCLEAR CORRECTION TO THE DEUTERIUM HYPERFINE STRUCTURE

The leading inelastic nuclear correction is of the relative order $\alpha m_e/\kappa$ (as well as the Zemach correction (8)). The corresponding effect calculated in [8] is additionally enhanced by a large factor:

$$\mu_p - \mu_n = 4.7.$$

In the present paper, we consider two more effects of the same order $\alpha m_e/\kappa$. Although both of them are proportional to

$$\mu_p + \mu_n = 0.88$$

(and are therefore essentially smaller numerically than the effect considered in [8]), we believe that their investigation is worth considering.

We use the gauge $A_0 = 0$, where the photon propagator is given by

$$D_{im}(\omega, \mathbf{k}) = \frac{d_{im}}{\omega^2 - \mathbf{k}^2}, \qquad d_{im} = \delta_{im} - \frac{k_i k_m}{\omega^2}, \quad D_{00} = D_{0m} = 0. \quad (9)$$

[1] In the case of hydrogen, this problem was considered many years ago by Zemach [13].

The electron–deuteron nuclear-spin-dependent scattering amplitude generated by the two-photon exchange is

$$T = 4\pi\alpha i\int\frac{d^4k}{(2\pi)^4}\frac{d_{im}d_{jn}}{k^4}\frac{\gamma_i(\hat{l} - \hat{k} + m_e)\gamma_j}{k^2 - 2lk}M_{mn}, \quad (10)$$

where $l_\mu = (m_e, 0, 0, 0)$ is the electron momentum. The structure $\gamma_i(\hat{l} - \hat{k} + m_e)\gamma_j$ reduces to $-i\omega\epsilon_{ijl}\sigma_l$, where $\boldsymbol{\sigma}$ is the electron spin. We calculate the nuclear matrix elements entering the deuteron Compton amplitude M_{mn} in the zero-range approximation, which allows us to obtain all results in a closed analytical form.

The inelastic $1/\kappa$ contribution to the hyperfine structure is induced by the combined action of the convection and spin currents. Because the convection current is spin-independent, all the intermediate states are triplet ones, as is the ground state. Therefore, the spin current operator

$$\frac{e}{2m_p}i\mathbf{k}\times[\mu_p\boldsymbol{\sigma}_p\exp(i\mathbf{k}\cdot\mathbf{r}_p) + \mu_n\boldsymbol{\sigma}_n\exp(i\mathbf{k}\cdot\mathbf{r}_n)]$$

simplifies to

$$\frac{e}{2m_P}i[\mathbf{k}\times\mathbf{s}][\mu_p\exp(i\mathbf{k}\cdot\mathbf{r}_p) + \mu_n\exp(i\mathbf{k}\cdot\mathbf{r}_n)]. \quad (11)$$

In the initial state $|0\rangle$, the deuteron is at rest, but in the excited state, the system of nucleons moves as a whole with the momentum $\mathbf{k}$ and its wave function is therefore given by $|n\rangle\exp(i\mathbf{k}\cdot\mathbf{R})$, where $|n\rangle$ refers to the deuteron internal degrees of freedom and is a function of $\mathbf{r} = \mathbf{r}_p - \mathbf{r}_n$ and $\mathbf{R} = (\mathbf{r}_p + \mathbf{r}_n)/2$ is the deuteron center-of-mass coordinate. Thus, a typical matrix element of the spin current can be written as

$$\begin{aligned}\frac{e}{2m_p}i\mathbf{k}\times\langle n|\exp(-i\mathbf{k}\cdot\mathbf{R})\mathbf{s}[\mu_p\exp(i\mathbf{k}\cdot\mathbf{r}_p)&\\ + \mu_n\exp(i\mathbf{k}\cdot\mathbf{r}_n)]|0\rangle&\\ = \frac{e}{2m_p}i\mathbf{k}\times\langle n|\mathbf{s}[\mu_p\exp(i\mathbf{k}\cdot\mathbf{r}/2)&\\ + \mu_n\exp(-i\mathbf{k}\cdot\mathbf{r}/2)]|0\rangle.&\end{aligned} \quad (12)$$

A typical matrix element of the convection current transforms as

$$\begin{aligned}\frac{e}{2m_p}\langle n|\exp(-i\mathbf{k}\cdot\mathbf{R})\hat{\mathbf{p}}_p\exp(i\mathbf{k}\cdot\mathbf{r}_p)&\\ + \exp(i\mathbf{k}\cdot\mathbf{r}_p)\hat{\mathbf{p}}_p|0\rangle&\\ = \frac{e}{2m_p}\langle n|\left(\hat{\mathbf{p}} + \frac{\mathbf{k}}{2}\right)\exp(i\mathbf{k}\cdot\mathbf{r}/2) + \exp(i\mathbf{k}\cdot\mathbf{r}/2)\hat{\mathbf{p}}|0\rangle&\\ = \frac{e}{m_p}\langle n|\hat{\mathbf{p}}\exp(i\mathbf{k}\cdot\mathbf{r}/2)|0\rangle,&\end{aligned} \quad (13)$$

where $\hat{\mathbf{p}}$ acts on the relative coordinate $\mathbf{r}$.

We first take the intermediate states $|n\rangle$ in the corresponding nuclear Compton amplitude to be just plane waves, eigenstates of $\hat{\mathbf{p}}$. We thus take into account all states with $l \neq 0$, which are free in our zero-range approximation, and in addition, the $^3S_1$ wave function in the free form

$$\psi_p(r) = \frac{\sin(pr)}{pr}$$

(the deviation of the $^3S_1$ wave function from the free one is considered below). Then, with the zero-range-approximation deuteron wave function

$$\psi_0(r) = \sqrt{\frac{\kappa}{2\pi}}\frac{\exp(-\kappa r)}{r}, \tag{14}$$

the only matrix element entering the amplitude is

$$\langle\psi_0|\exp(\pm i\mathbf{k}\cdot\mathbf{r}/2)|\mathbf{p}\rangle = \frac{\sqrt{8\pi\kappa}}{(\mathbf{p}\pm\mathbf{k}/2)^2+\kappa^2}. \tag{15}$$

Thus the amplitude simplifies to

$$M^{(2)}_{mn} = \left(\frac{e}{2m_p}\right)^2 2\kappa\omega\int\frac{d\mathbf{p}}{\pi^2}\left\{\frac{\mu_p}{[(\mathbf{p}-\mathbf{k}/2)^2+\kappa^2]^2} + \frac{\mu_n}{[(\mathbf{p}-\mathbf{k}/2)^2+\kappa^2][(\mathbf{p}+\mathbf{k}/2)^2+\kappa^2]}\right\} \times\frac{2p_m i\epsilon_{nrs}k_r s_s - 2p_n i\epsilon_{mrs}k_r s_s}{\omega^2-(p^2+k^2/4+\kappa^2)^2/m_p^2}. \tag{16}$$

Because the motion of the system as a whole is taken into account in the intermediate states, this expression differs from the corresponding one in our previous paper [8] in two respects. First, in [8], the operator $\hat{\mathbf{p}}_p$ in (13) was identified with $\mathbf{p}$, and instead of $2p_{m,n}$ we therefore obtained $(2p-k/2)_{m,n}$ in an analog of the present formula (16). At present, the term proportional to μ_n in (16) is an odd function of $\mathbf{p}$ and therefore vanishes after integration over $d\mathbf{p}$. Second, the energy difference in the denominator has acquired the contribution $k^2/4m_p$, which is the kinetic energy of the proton–neutron system as a whole, and thus

$$\frac{p^2+\kappa^2}{m_p}$$

has transformed into

$$\frac{p^2+k^2/4+\kappa^2}{m_p}.$$

We now substitute (16) in (10) and take the integral over ω under the condition $\omega \gg \kappa^2/m$. For the relative correction to the hyperfine structure, we obtain

$$\Delta^{(2)}_{in} = \frac{2\alpha\kappa\mu_p m_e}{\pi^4\mu_d m_p}\int\int\frac{d\mathbf{p}\,d\mathbf{k}}{k^4} \times\frac{\mathbf{pk}}{[(\mathbf{p}-\mathbf{k}/2)^2+\kappa^2]^2}\left[\frac{m_p}{p^2+k^2/4+\kappa^2}-\frac{3}{2k}\right]. \tag{17}$$

The result of integration over $\mathbf{p}$ and then over $\mathbf{k}$ is given by

$$\Delta^{(2)}_{in} = \alpha\frac{\mu_p}{\mu_d}\frac{m_e}{\kappa} - \frac{6\alpha}{\pi}\frac{\mu_p}{\mu_d}\frac{m_e}{m_p}\ln\frac{m_p}{\kappa}. \tag{18}$$

The logarithmic contribution here originates from integration of the term $3/2k$ in square brackets in (17) over the range

$$\kappa^2/m_p \ll k \ll \kappa.$$

The result (18) differs from the corresponding one in [8] by a term proportional to $\mu_p + \mu_n$, which is relatively small numerically. It is only natural that our present account of the motion of the proton–neutron system as a whole in the intermediate states results in a correction proportional to $\mu_p + \mu_n$.

We now calculate the correction $\Delta^{(3)}_{in}$ corresponding to the effect of deviation of the intermediate $^3S_1$ wave function $\Psi_p(r)$ from the free one. In the zero-range approximation, $\Psi_p(r)$ is given by

$$\Psi_p(r) = \frac{\sin(pr)}{pr} - \frac{1}{\kappa+ip}\frac{\exp(ipr)}{r} = \frac{\kappa\sin(pr)-p\cos(pr)}{pr(\kappa+ip)}. \tag{19}$$

This follows, for instance, from the orthogonality to deuteron wave function (14). Below, we use the function

$$\rho_p(r_1,r_2) = \Psi_p(r_1)\Psi^*_p(r_2) - \psi_p(r_1)\psi^*_p(r_2) = \frac{p\cos(p(r_1+r_2)) - \kappa\sin(p(r_1+r_2))}{(\kappa^2+p^2)pr_1r_2}. \tag{20}$$

After integration over ω, the expression for $\Delta^{(3)}_{in}$ becomes

$$\Delta^{(3)}_{in} = \frac{4\alpha(\mu_p+\mu_n)m_e}{\pi^3\mu_d m_p}\int\int dk\,dp\,p^2 \times\int\int d\mathbf{r}_1 d\mathbf{r}_2\psi_0(r_1)\psi_0(r_2)\rho_p(r_1,r_2)\frac{\sin(kr_1)}{kr_1} \times\left[\frac{\sin(kr_2)}{kr_2} - \frac{(1+\kappa r_2)}{(kr_2)^2}\left(\frac{\sin(kr_2)}{kr_2}-\cos(kr_2)\right)\right] \times\left[\frac{m_p}{p^2+k^2/4+\kappa^2}-\frac{3}{2k}\right]. \tag{21}$$

The integral over p is

$$\int_0^\infty dp \frac{p^2 \rho_p(r_1, r_2)}{p^2 + k^2/4 + \kappa^2} = \frac{2\pi}{r_1 r_2 k^2} \times [(Q+\kappa)\exp[-Q(r_1+r_2)] - 2\kappa \exp[-\kappa(r_1+r_2)]], \quad (22)$$

where

$$Q = \sqrt{\kappa^2 + k^2/4}.$$

We now integrate (21) over r_1 and r_2 and then over k. The final result for the discussed correction is given by

$$\Delta_{in}^{(3)} = -\alpha \frac{\mu_p + \mu_n}{\mu_d} \frac{m_e}{\kappa} \frac{1}{3}(2 - 2\ln 2) + \frac{3\alpha}{\pi} \frac{\mu_p + \mu_n}{\mu_d} \frac{m_e}{m_p} \ln \frac{m_p}{\kappa}. \quad (23)$$

Again, it is only natural that due to common selection rules, the contribution of the $^3S_1$ intermediate state is proportional to $\mu_p + \mu_n$. We also note that the first term in (23) is additionally suppressed by the small numerical factor

$$\frac{2 - 2\ln 2}{3} = 0.20.$$

We finally return to the effect due to a finite distribution of the deuteron charge and magnetic moment. The Zemach correction Δ_f in Eq. (8) can also be easily derived in the present approach. Using the identity

$$\langle \psi_0 | \hat{\mathbf{p}} \exp(i\mathbf{k}\cdot\mathbf{r}/2) | \psi_0 \rangle = \frac{\mathbf{k}}{4} \langle \psi_0 | \exp(i\mathbf{k}\cdot\mathbf{r}/2) | \psi_0 \rangle,$$

we obtain the corresponding amplitude

$$M_{mn}^f = \left(\frac{e}{2m_p}\right)^2 (\mu_p + \mu_n)\omega[F^2(k) - 1] \times \frac{k_m i\epsilon_{nrs} k_r s_s - k_n i\epsilon_{mrs} k_r s_s}{\omega^2 - (p^2 + k^2/4 + \kappa^2)^2/m_p^2}, \quad (24)$$

where

$$F(k) = \langle \psi_0 | \exp(i\mathbf{k}\cdot\mathbf{r}/2) | \psi_0 \rangle = \frac{4\kappa}{k} \arctan \frac{k}{4\kappa} \quad (25)$$

is the deuteron form factor in the zero-range approximation. We note that in our approximation, the electric and magnetic form factors, which in the present case enter the convection current and spin current matrix elements, respectively, coincide and are equal to $F(k)$.

Integration over ω leads to the following result for the relative correction Δ_f:

$$\Delta_f = \frac{8\alpha(\mu_p + \mu_n) m_e}{\pi \mu_d} \int_0^\infty \frac{dk}{k^2} [F^2(k) - 1] = -\alpha \frac{\mu_p + \mu_n}{\mu_d} \frac{m_e}{\kappa} \frac{1}{3}(1 + 2\ln 2). \quad (26)$$

There is no logarithmic term in Δ_f because

$$F^2(k) - 1 \sim k^2 \quad \text{for} \quad k \ll \kappa.$$

In fact, result (26) agrees with (8) because $\mu_p + \mu_n = \mu_d$ within our accuracy.

Corrections (18), (23), and (26) combine into the compact result

$$\Delta_c = -\alpha \frac{\mu_n}{\mu_d} \frac{m_e}{\kappa} - \frac{3\alpha}{\pi} \frac{\mu_p - \mu_n}{\mu_d} \frac{m_e}{m_p} \ln \frac{m_p}{\kappa}. \quad (27)$$

We note that the logarithmic part of Δ_c coincides with the corresponding logarithmic term in [8] (see Eq. (27) therein). This is quite natural: the logarithmic contribution is dominated by small k and cannot therefore be influenced by an extra power of k arising from the recoil of the proton–neutron system as a whole.

The leading term in (27) coincides with the result in [5].[2] However, we could not find any correspondence between our arguments and those in [5]. In particular, it is stated explicitly in [5] that the motion of the intermediate proton–neutron system as a whole is neglected there.

3. DISCUSSION OF THE RESULTS

Our total result for the nuclear-structure corrections to the deuterium hyperfine structure, comprising contributions (4), (6), (7), and (27), is

$$\Delta = -\alpha \frac{\mu_n}{\mu_d} \frac{m_e}{\kappa} - \frac{3\alpha}{\pi} \frac{\mu_p - \mu_n}{\mu_d} \frac{m_e}{m_p} \ln \frac{m_p}{\kappa} + \frac{3\alpha}{8\pi} \frac{(\mu_p - \mu_n)^2}{\mu_d} \frac{m_e}{m_p} \ln \frac{m_p}{\kappa} + \frac{3\alpha}{8\pi} \frac{1}{\mu_d} (\mu_d^2 - 2\mu_d - 3) \frac{m_e}{m_p} \ln \frac{\kappa}{m_e} + \frac{3\alpha}{4\pi} \frac{1}{\mu_d} (\mu_p^2 - 2\mu_p - 3 + \mu_n^2) \frac{m_e}{m_p} \ln \frac{m_p}{\kappa}. \quad (28)$$

[2] We are sorry for misquoting the result of [5] in our paper [8].

Numerically, this correction to the hyperfine splitting in deuterium is given by

$$\Delta\nu = 50 \text{ kHz}. \quad (29)$$

It should be compared with the lacking 45 kHz (see (3)). We believe that the agreement is quite satisfactory if one recalls the crude nuclear model (zero-range approximation) used here; in particular, the deuteron form factors calculated in the zero-range approximation are certainly harder than the real ones, and the negative Zemach correction is therefore underestimated in that approximation.

We mention here that in recent paper [14], elastic contributions and the Zemach effect were considered in a quite different theoretical technique, but with a certain phenomenological description of the deuteron form factors. The result is smaller than the corresponding part of ours by 13 kHz.

Clearly, the nuclear effects discussed are responsible for the bulk of the difference between the pure QED calculations and the experimental value of the deuterium hyperfine splitting. Calculation of this correction, including accurate treatment of nuclear effects, would serve as one more sensitive check of detailed models of the deuteron structure.

ACKNOWLEDGMENTS

We are grateful to J.L. Friar for insistently directing our attention to the role of the $^3S_1$ positive-energy state in the inelastic contribution. The investigation was supported in part by the Russian Foundation for Basic Research (project nos. 03-02-17612 and 03-02-16510).

REFERENCES

1. D. J. Wineland and N. F. Ramsey, Phys. Rev. A **5**, 821 (1972).
2. G. T. Bodwin and D. R. Yennie, Phys. Rev. D **37**, 498 (1988).
3. N. F. Ramsey, in *Quantum Electrodynamics*, Ed. by T. Kinoshita (World Sci., Singapore, 1990).
4. A. Bohr, Phys. Rev. **73**, 1109 (1948).
5. F. E. Low, Phys. Rev. **77**, 361 (1950).
6. F. E. Low and E. E. Salpeter, Phys. Rev. **83**, 478 (1951).
7. D. A. Greenberg and H. M. Foley, Phys. Rev. **120**, 1684 (1960).
8. A. I. Milstein, I. B. Khriplovich, and S. S. Petrosyan, Zh. Éksp. Teor. Fiz. **109**, 1146 (1996) [JETP **82**, 616 (1996)].
9. F. E. Low, Phys. Rev. **96**, 1428 (1954).
10. M. Gell-Mann and M. L. Goldberger, Phys. Rev. **96**, 1433 (1954).
11. R. P. Feynman, *Photon–Hadron Interactions* (Benjamin, Reading, Mass., 1972; Mir, Moscow, 1975), p. 219.
12. A. I. Milstein, Sib. J. Phys., No. 1, 43 (1995).
13. A. C. Zemach, Phys. Rev. **104**, 1771 (1956).
14. R. N. Faustov and A. P. Martynenko, hep-ph/0211445.

Journal of Experimental and Theoretical Physics, Vol. 98, No. 2, 2004, pp. 186–196.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 2, 2004, pp. 210–221.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Photon Propagation in a Supercritical Magnetic Field¶

A. E. Shabad

Tamm Department of Theoretical Physics, Lebedev Physical Institute, Russian Academy of Sciences, Moscow, 119991 Russia
e-mail: shabad@lpi.ru
Received August 25, 2003

Abstract—We show that, for the asymptotically strong (super-Schwinger) magnetic field B exceeding the critical value $B_{cr} = m^2c^3/eh = 4.4 \times 10^{13}$ G, the vacuum polarization effects become important not only in the γ-range, but also for softer electromagnetic quanta, including X-rays and optical photons, and for electromagnetic waves of radio frequencies. This is a consequence of the linearly growing term $\asymp B/B_{cr}$ present in the vacuum polarization in an asymptotically strong magnetic field. The results may be essential in studying reflection, refraction, and splitting of X-rays, light and radio waves by magnetic fields of magnetars, and in considering emission of such waves by charged particles. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Although much time has passed since our first understanding of the refracting and birefringing properties of a strong magnetic field in vacuum, the photon splitting effect [1] and the effect of photon capture [2–8] are the only essential consequences of them that have been considered in a realistic astrophysical context. Both effects are currently discussed mostly in application to electromagnetic radiation in the γ-range. They depend crucially on the deviation of the photon dispersion curve from its customary shape in the empty vacuum, $k_0^2 = |\mathbf{k}|^2$, where k_0 is the photon energy and $\mathbf{k}$ is its momentum. For magnetic fields B below the Schwinger critical value,

$$B \le B_{cr} = m^2c^3/eh = 4.4 \times 10^{13}\ \text{Gs},$$

where m and e are the electron mass and charge, the only essential source of this deviation is the singular behavior of the polarization operator $\Pi_{\mu\nu}(k)$ near the thresholds of creation of mutually independent electron and positron on Landau levels n, n' by a photon (the cyclotron resonance) [2–4] or an even stronger singular behavior of $\Pi_{\mu\nu}$ near the points of a mutually bound e^+e^--pair (the positronium atom) formation [5–7, 9]. To reach (at least the lower of) these positions, the photon must belong to the γ-ray range, with its energy above or on the order 1 MeV. For this reason, the effect of photon capture, with its transformation into an electron–positron pair, derived from the singular behavior of $\Pi_{\mu\nu}(k)$, applies mostly to the γ-quanta, as long as their propagation in a pulsar magnetosphere of traditional pulsars is concerned. It was estimated that fields of about $B = 0.1B_{cr}$ are sufficient to provide this effect [4] and to protect the positronium atom into which the captured γ-quantum is transformed against ionization by the accelerating electric field in the polar gap and by thermal photons [5–9].

Also, the Adler effect [1] of photon splitting $\gamma \longrightarrow \gamma\gamma$ in such fields is usually discussed for γ-quanta [10–13]. There are two reasons why, again, the γ-range is important. The first one is that the photon splitting becomes possible in the magnetic field because the deviation of the dispersion curve from the $k_0^2 = \mathbf{k}^2$ law opens a kinematical aperture for this process. The stronger the deviation, the wider the aperture (and the deviation is strong near the thresholds). In addition, there is a strong birefringence for the photons in the γ-range, because only one eigenvalue $\kappa_2(k)$ of the tensor $\Pi_{\mu\nu}$ is singular near the lowest ($n = n' = 0$) threshold, while the other two eigenvalues $\kappa_{1,3}(k)$ remain finite until the next thresholds ($n = 0$, $n' = 1$ or $n = 1$, $n' = 0$) are reached. This implies that the photons of only one polarization mode are substantially affected by the medium. This birefringence leads to polarization selection rules in the photon splitting process, which are well pronounced. The second reason is dynamical. The matrix elements of the photon splitting are subject to the same resonant behavior near the thresholds as the polarization operator. The aforesaid explains why the γ-range is first to be affected by the magnetized vacuum.

The situation changes considerably in super-Schwinger magnetic fields $B \gg B_{cr}$, which are believed to exist in soft γ-ray repeaters and anomalous X-ray pulsars (see, e.g., [14]). In this asymptotic range, a linearly increasing term proportional to B/B_{cr} appears in one of the eigenvalues, κ_2, of the polarization operator [15, 16], thus providing an extra large contribution (in addition to the cyclotron resonance) to the refraction of the vacuum.

¶This article was submitted by the author in English.

In Section 3, we study the consequences of this phenomenon for photon propagation, on the basis on the first three leading contributions to the asymptotic expansion of the polarization-operator eigenvalues for large B, which have been obtained within the one-loop approximation. One of these consequences is a frequency-independent, but direction-sensitive, large refraction index for propagation nonparallel to the magnetic field in one (of three) polarization modes in the kinematical domain far from the threshold. The corresponding strong polarization- and direction-sensitive refraction occurs for electromagnetic radiation of any frequency range, including the X-ray, optical, and radio ranges.

In the preceding Section 2, exact results on electromagnetic radiation propagation in the magnetized vacuum are described. These follow only from the general properties of relativistic, gauge, and charge invariance [17] and the Onsager theorem [18]. The results in Section 2 are valid irrespective of any approximation and the field strength, unless the opposite is explicitly indicated.

In the Appendix, the asymptotic expansion used in Section 3 is derived.

2. ELECTROMAGNETIC EIGENMODES IN AN EXTERNAL MAGNETIC FIELD

There are three propagating eigenmodes corresponding to vacuum excitations with photon quantum numbers in an external magnetic field B. The dispersion law, i.e., the dependence of the energy k_0 of the quantum (or the frequency in the wave) on its momentum $\mathbf{k}$, is given for each mode by the solution of the equation

$$k^2 = \kappa_i(k_0^2 - k_\parallel^2, k_\perp^2), \quad i = 1, 2, 3, \tag{1}$$

where $k_\parallel$ and $k_\perp$ are the respective momentum components parallel and perpendicular to magnetic field $\mathbf{B}$ and k^2 is the photon 4-momentum squared,

$$k^2 = k_\perp^2 + k_\parallel^2 - k_0^2.$$

The κ_i in the right-hand sides in Eqs. (1) are eigenvalues of the polarization operator [2, 3, 17].

A general consequence of the relativistic covariance is that the eigenvalues depend on the two combinations of the momentum specified in (1). This implies that solutions of dispersion equations (1) have the general structure

$$k_0^2 = k_\parallel^2 + f_i(k_\perp^2), \quad i = 1, 2, 3, \tag{2}$$

and that the direction of the group velocity $\mathbf{v} = \partial k_0/\partial \mathbf{k}$ in each mode does not coincide (for $k_\perp \neq 0$) with that of the phase velocity $\mathbf{k}/k_0$. To see this, we calculate the components of the respective group velocities $v_\perp$ and $v_\parallel$ across and along the magnetic field $\mathbf{B}$ on solutions (2) of each dispersion equation (1),

$$v_\perp \equiv \frac{\partial k_0}{\partial k_\perp} = \frac{k_\perp}{k_0}\frac{\partial k_0^2}{\partial k_\perp^2}$$

$$= \frac{k_\perp}{k_0}\frac{1 - \partial\kappa_i/\partial k_\perp^2}{1 + \partial\kappa_i/\partial(k_0^2 - k_\parallel^2)} = \frac{k_\perp}{k_0}\frac{df_i(k_\perp^2)}{dk_\perp^2}, \tag{3}$$

$$v_\parallel \equiv \frac{\partial k_0}{\partial k_\parallel} = \frac{k_\parallel}{k_0}.$$

It follows from (3) that the angle θ between the direction $\mathbf{v}$ of the electromagnetic energy propagation and the external magnetic field satisfies the relation

$$\frac{v_\perp}{v_\parallel} \equiv \tan\theta = \left(1 - \frac{\partial\kappa_i}{\partial k_\perp^2}\right)\left(1 + \frac{\partial\kappa_i}{\partial(k_0^2 - k_\parallel^2)}\right)^{-1}\tan\vartheta, \tag{4}$$

where ϑ is the angle between the photon momentum (phase velocity) and the external field, $\tan\vartheta \equiv k_\perp/k_\parallel$. The following statement holds: If the phase velocity $|\mathbf{k}|/k_0$ exceeds the velocity of light c, i.e., if $k_\perp^2 + k_\parallel^2 > k_0^2$ (or $f_i(k_\perp^2) < k_\perp^2$ in (2)), but the group velocity (3) does not, $v_\perp^2 + v_\parallel^2 \leq 1$, then $\tan\theta < \tan\vartheta$. The conditions of this statement are fulfilled for the dispersion laws found within approximation-dependent calculations of κ_i. For super-Schwinger fields, treated within the one-loop approximation, this fact follows explicitly from the equations in Section 3. Therefore, the photon tends to deviate closer to the magnetic field line.

It follows from the gauge invariance that

$$\kappa_i(0, 0) = 0, \quad i = 1, 2, 3. \tag{5}$$

This property implies that, for each mode, there always exists a dispersion curve with $f_i(0) = 0$, which passes through the origin in the $(k_0^2 - k_\parallel^2, k_\perp^2)$ plane. However, only two of these three solutions may simultaneously correspond to physical massless particles: photons. The third solution is a nonphysical degree of freedom, characteristic of gauge theories: in a magnetic field, a photon has two degrees of freedom, the same as in an empty vacuum. Which of the modes becomes nonphysical depends on the propagation direction and on the specific form of the function $f_i(k_\perp^2)$ in (2). We discuss this point for the super-Schwinger field limit in the next section. Massive branches of solutions of (1), with $f_i(0) > 0$, may also exist despite (5). For these, the number of physical degrees of freedom is three, and, hence, all three equations (1) can simultaneously have physical solutions (see, e.g., the positronium branches obtained in [7, 19, 20]).

The refraction index n_i in mode i is

$$n_i = \frac{|\mathbf{k}|}{k_0} = \left(1 + \frac{\kappa_i}{k_0^2}\right)^{1/2} = \left(1 + \frac{k_\perp^2 - f_i(k_\perp^2)}{k_0^2}\right)^{1/2}. \quad (6)$$

Unlike κ_i, the refractive index n_i is not a Lorentz scalar. It may depend on two energy-momentum variables after the dispersion relation is reduced to (2). Gauge invariance property (5) implies that the refraction index (6) for parallel propagation, $k_\perp = 0$, is exactly equal to unity for the massless ($f_i(0) = 0$) branches in every mode,

$$n_i^{\parallel} = 1. \quad (7)$$

The electromagnetic wave propagating strictly along the external constant and homogeneous magnetic field propagates with the velocity of light c in vacuum, the phase and group velocities coinciding in this case.

If, within a certain approximation, the eigenvalue κ_i is a linear function of its arguments with condition (5) being satisfied, refraction index (6) for the corresponding dispersion relation depends on a single combination of the photon energy and momentum, which is the propagation direction ϑ. This happens in a nonresonant situation, for instance, as described in the next section.

The polarizations of the modes are described in an approximation-independent way [3, 17] by the relations

$$\mathbf{e}^{(1)} = -\frac{\mathbf{k}_\perp}{k_\perp}k_0, \quad \mathbf{h}^{(1)} = \left(\frac{\mathbf{k}_\perp}{k_\perp} \times \mathbf{k}_\parallel\right), \quad (8)$$

$$\mathbf{e}_\perp^{(2)} = \mathbf{k}_\perp k_\parallel, \quad \mathbf{e}_\parallel^{(2)} = \frac{\mathbf{k}_\parallel}{k_\parallel}(k_\parallel^2 - k_0^2),$$
$$\mathbf{h}^{(2)} = -k_0\left(\mathbf{k}_\perp \times \frac{\mathbf{k}_\parallel}{k_\parallel}\right), \quad (9)$$

$$\mathbf{e}^{(3)} = -k_0\left(\frac{\mathbf{k}_\perp}{k_\perp} \times \frac{\mathbf{k}_\parallel}{k_\parallel}\right), \quad \mathbf{h}_\perp^{(3)} = -\frac{\mathbf{k}_\perp}{k_\perp}k_\parallel,$$
$$\mathbf{h}_\parallel^{(3)} = \frac{\mathbf{k}_\parallel}{k_\parallel}k_\perp, \quad (10)$$

where $\mathbf{e}^{(i)}$ and $\mathbf{h}^{(i)}$ are the electric and magnetic fields in the ith mode ($i = 1, 2, 3$); the cross denotes the vector product; and boldface letters with subscripts "$\parallel$" and "$\perp$" denote vectors along the directions parallel and perpendicular to the external magnetic field, respectively. In mode 1, the electric field $\mathbf{e}$ is parallel to $\mathbf{k}_\perp$; in mode 2, it lies in the plane containing the vectors $\mathbf{k}$ and $\mathbf{B}$; and in mode 3, it is orthogonal to this plane, which means that mode 3 is always transversely polarized.

We note that the normalizations in Eqs. (8), (9), and (10) are different, and we can therefore judge the vanishing of some components compared to others within one equation, but not between different equations.

Concerning the direction of propagation, two cases are essentially different. If $k_\perp = 0$, we are talking about longitudinal propagation. Otherwise, a Lorentz boost exists along the external (constant and homogeneous) magnetic field, which does not change the value of the magnetic field; nor does it introduce an extra electric field—it nullifies $k_\parallel$. Hence, the general case of nonparallel propagation $k_\perp \neq 0$, $k_\parallel \neq 0$ is reduced to purely transversal propagation, $k_\parallel = 0$ (in the corresponding reference frame). One should bear in mind, however, that the above transformation changes the photon energy k_0 and should be treated with caution when one considers a field with curved force lines.

For transversal propagation, $\mathbf{k} \perp \mathbf{B}$ ($k_\parallel = 0$), modes 2 and 3 are transversely polarized ($\mathbf{e}^{(2),(3)} \perp \mathbf{k}$) in two mutually orthogonal planes, $\mathbf{e}^{(2)} \perp \mathbf{e}^{(3)}$, while mode 1 is longitudinally polarized ($\mathbf{e}^{(1)} \parallel \mathbf{k}$) with no magnetic field in it, $\mathbf{h}^{(1)} = 0$. It should not correspond to a photon (depending on the dispersion law).

On the contrary, for longitudinal propagation, $\mathbf{k} \parallel \mathbf{B}$ ($k_\perp = 0$), modes 1 and 3 are transversely polarized ($\mathbf{e}^{(1,3)} \perp \mathbf{B}$) and their electric field vectors lie in mutually orthogonal planes, $\mathbf{e}^{(1)} \perp \mathbf{e}^{(3)}$, as they always do, while mode 2 is longitudinally polarized ($\mathbf{e}^{(2)} \parallel \mathbf{B}$) and does not contain a magnetic field, $\mathbf{h}^{(2)} = 0$. Mode 2 is then should not correspond to a photon, whereas mode 1 is a physical electromagnetic wave, which matches the electromagnetic wave of mode 3: together, they may form a circularly polarized transversal wave because of the degeneracy property

$$\kappa_1((k_0^2 - k_\parallel^2), 0) = \kappa_3((k_0^2 - k_\parallel^2), 0). \quad (11)$$

This relation reflects the cylindrical symmetry of the problem of a photon propagating along the external magnetic field.

Another remark of almost general character is in order. One might expect the possibility of Cherenkov radiation by a charged particle moving in an optically dense medium formed by the magnetized vacuum. However, this effect does not take place for Cherenkov photons softer than those with $k_0 = 2m$. We consider emission of a photon by an electron in a magnetic field, unaccompanied by a change in its Landau quantum number, $n = n'$ (otherwise, this would be cyclotron, and not Cherenkov, radiation). According to kinematical analysis of the energy and momentum conservation in [21] (and to the study [21] of analyticity regions of the one-loop photon polarization operator in an electron–positron plasma in a magnetic field, calculated in [18]), the Cherenkov photon with $k_0 < 2m$ can only belong to the right lower sector

$$k_0^2 - k_\parallel^2 \leq 0, \quad k_\perp^2 \geq 0 \quad (12)$$

in the $(k_0^2 - k_\parallel^2, k_\perp^2)$ plane. The essential reason for this is the degeneration of the electron energy with respect

to the center-of-orbit position in the transversal plane. No dynamical calculations hitherto known ensure penetration of photon dispersion curves into this sector. The only exception is a nonphysical situation owing to exponentially strong external fields, which will be mentioned in subsection 3.2. We conclude that no Cherenkov emission of a photon with $k_0 < 2m$ is possible under standard conditions.

3. PHOTON DISPERSION IN A SUPER-SCHWINGER MAGNETIC FIELD

3.1. Asymptotic Expansion of Polarization Tensor Eigenvalues

In the asymptotic region of supercritical magnetic fields $B \gg B_{cr}$ and restricted energy of longitudinal motion,

$$k_0^2 - k_\parallel^2 \ll (B/B_{cr})m^2,$$

the three eigenvalues $\kappa_{1,2,3}(k)$ of the polarization operator (if it is calculated within the one-loop approximation as in [17, 22]) have the following behavior, derived from equations of [3] (see the Appendix),

$$\kappa_1(k_0^2 - k_\parallel^2, k_\perp^2) = \frac{\alpha k^2}{3\pi}\left(\ln\frac{B}{B_{cr}} - C - 1.21\right), \quad (13)$$

$$\kappa_2(k_0^2 - k_\parallel^2, k_\perp^2) = \frac{\alpha B m^2 (k_0^2 - k_\parallel^2)}{\pi B_{cr}} \times \exp\left(-\frac{k_\perp^2}{2m^2}\frac{B_{cr}}{B}\right)\int_{-1}^{1}\frac{(1-\eta^2)d\eta}{4m^2 - (k_0^2 - k_\parallel^2)(1-\eta^2)}, \quad (14)$$

$$\kappa_3(k_0^2 - k_\parallel^2, k_\perp^2) = \frac{\alpha k^2}{3\pi}\left(\ln\frac{B}{B_{cr}} - C\right) - \frac{\alpha}{3\pi}(0.21k_\perp^2 - 1.21(k_0^2 - k_\parallel^2)). \quad (15)$$

Here, $\alpha = 1/137$ is the fine structure constant and $C = 0.577$ is the Euler constant. Equations (13) and (15) are accurate up to terms decreasing with B as $(B_{cr}/B)\ln(B/B_{cr})$ or faster. Equation (14) is accurate up to terms logarithmically increasing with B. In $\kappa_{1,3}$, we also took the limit

$$k_\perp^2 \ll (B/B_{cr})m^2,$$

which is not the case for κ_2, where the factor $\exp(-k_\perp^2 B_{cr}/2m^2 B)$ is kept different from unity, because it is important near the cyclotron resonance, as explained in Section 3.2 below. The integral in (14) can readily be calculated, but we do not need its explicit form here.

The parts increasing with B in $\kappa_{1,2,3}$ were written in [16], and their derivation from equations of [3] is traced in detail in [19, 20]. The linearly increasing term in Eq. (14) was obtained in [15] in a different way using a two-dimensional (one time, one space) diagrammatic technique developed to describe the asymptotic magnetic field regime. The logarithmic terms in the expressions above do not dominate over the constant terms unless exponentially large magnetic fields are included into consideration.[1] The derivation of all terms in Eqs. (13)–(15), including those that do not grow with B, is given in the Appendix using a straightforward method different from the one applied earlier in [19, 20]. The asymptotic expressions used in [13] do not coincide with ours, except for the term linear in B.

The limiting expressions (13)–(15) satisfy the exact properties (11) and (5).

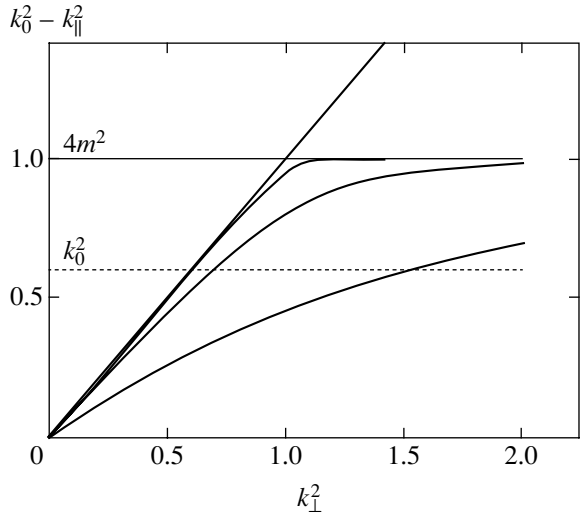
In this paper, we only deal with the transparency region, $k_0^2 - k_\parallel^2 \le 4m^2$ (i.e., with the kinematical domain where $\kappa_{1,2,3}$ are real), because we are interested in photons with $k_0 < 2m$, or even $k_0 \ll 2m$, which never reach the free pair creation threshold $k_0^2 - k_\parallel^2 = 4m^2$. The eigenvalue κ_2 in (14) has a singular branching point in the complex plane of the variable $(k_0^2 - k_\parallel^2)$ near the lowest pair creation threshold $(k_0^2 - k_\parallel^2)_{thr} = 4m^2$. Thresholds of creation of e^+e^--pairs with the electron and the positron on excited Landau levels n, $n' \neq 0$,

$$(k_0^2 - k_\parallel^2)_{thr}^{n,n'} = m^2\left[\left(1 + n\frac{B}{B_{cr}}\right)^{1/2} + \left(1 + n'\frac{B}{B_{cr}}\right)^{1/2}\right]^2 \quad (16)$$

are shifted in the asymptotic regime to the infinitely remote region. Therefore, the eigenvalues $\kappa_{1,3}$, which are responsible for photons of such polarizations that can only create e^+e^- pairs with at least one charged particle in an excited Landau state, do not contain imaginary parts or singular branching points in this regime. On the other hand, eigenvalue κ_2 has only one singular branching point, which corresponds to the possibility that an electron and a positron are created in the lowest Landau states by the photon polarized as in mode 2. The singular threshold behavior of (14) near the point

$$k_0^2 - k_\parallel^2 = 4m^2 - \epsilon, \quad \epsilon > 0, \quad \epsilon \longrightarrow 0$$

[1] That would be unreasonable not only because such fields are hardly expected to exist in nature, but mainly because their consideration is beyond the scope of quantum electrodynamics: the logarithmically increasing terms in (13) and (15) are associated with the absence of asymptotic freedom in QED (cf. analogous asymptotic behavior [23] in the Euler–Heisenberg effective Lagrangian).



A family of dispersion curves for mode 2 (solutions of Eq. (1) with Eq. (14) taken for the right-hand side) below the threshold $k_0^2 - k_\parallel^2 = 4m^2$. The values of the external magnetic field corresponding to the curves are (from left to right) $B = 10B_{cr}$, $100B_{cr}$, and $1000B_{cr}$. The straight line is the light cone dispersion curve for $B = 0$. The dashed horizontal line marks the maximum to which the photon with the energy k_0 may proceed if $k_0 < 2m$. The variables along the axes are plotted in the units of $4m^2 \approx 1$ MeV^2.

is

$$\kappa_2(k) \sim \frac{2\alpha B m^3}{B_{cr}} \exp\left(-\frac{k_\perp^2}{2m^2}\frac{B_{cr}}{B}\right) \times (4m^2 - k_0^2 + k_\parallel^2)^{-1/2}. \quad (17)$$

As could be expected, this is the same as the behavior of the exact one-loop expression for $\kappa_2(k)$ [3] near this threshold, before the limiting transition to large fields has been performed.

3.2. Propagation of Eigenmodes in the Super-Schwinger Field Limit

If Eq. (13) for κ_1 is taken as the right-hand side of Eq. (1), the latter has only one solution, which is the trivial dispersion law $k^2 = 0$. When the relation $k^2 = 0$ is satisfied, however, the 4-potential corresponding to the electromagnetic field of mode 1 becomes proportional to the photon 4-momentum vector k_μ, unless $k_\parallel = 0$ (see [3, 19, 20]). Therefore, for nonparallel propagation, mode 1 corresponds to only the gauge degree of freedom discussed in Section 2, with no real electromagnetic field associated with it.

The solutions of Eq. (1) for the second mode $i = 2$ shown in the figure is dominated by the cyclotron resonance (17), which causes a strong deviation of the dispersion curves from the shape $k^2 = 0$ (the light cone). As $k_\perp^2 \longrightarrow \infty$ near the threshold on the dispersion curves, the quantity $k_\perp^2 B_{cr}/m^2 B$ must be kept different from zero even in the large-field limit under consideration.

The behavior of the dispersion curves of mode 2 near the threshold for super-Schwinger magnetic fields $B \gg B_{cr}$ is the same as for the "moderate" fields $B \le B_{cr}$, and therefore it also presents the photon capture effect for photons harder than $2m$, known for such fields [4]: if we calculate (4) near the threshold $k_0^2 - k_\parallel^2 = 4m^2$ using Eq. (17) as κ_2 to obtain

$$\tan\theta = \frac{k_\perp}{k_\parallel}\frac{B_{cr}}{Bm^2}(4m^2 - k_0^2 + k_\parallel^2), \quad (18)$$

we conclude that the angle θ between the external magnetic field and the direction of the wave packet propagation in mode 2 tends to zero, the faster, the stronger the field. If the photon energy k_0 is slightly less than $2m$, the photon may be close to the threshold when its $k_\parallel$ disappears. At this upper point, the wave packet stops, because the group velocity length

$$v_\perp^2 + v_\parallel^2 = v_\parallel^2(1 + \tan^2\theta),$$

equal to $k_\parallel^2/k_0^2$ according to the second line in (3) and (18), disappears together with $k_\parallel$.

Applied to the conventional pattern of a pulsar magnetosphere, this effect acts as follows [4]. A curvature γ-quantum emitted tangentially to the magnetic force line, i.e., placed initially at the origin in the figure, then evolves along its dispersion curve as it propagates in the dipole magnetic field with its force line curved, because the components $k_\parallel$ and $k_\perp$ are changing. The maximum value of the ordinate $k_0^2 - k_\parallel^2$ occurs at $k_\parallel = 0$, and it is the photon energy squared, k_0^2. If the latter is greater than $4m^2$, the photon may reach the horizontal asymptote in the figure. Here, its group velocity $dk_0/dk_\perp$ across the magnetic field disappears, $dk_0/dk_\perp \longrightarrow 0$, and hence it propagates along the magnetic field and does not cross the threshold, because the other branch of the dispersion curve, which passes above the threshold, is separated from the initial branch by a gap. A mixed state, a photon pair, is actually formed [4], analogous to the polariton known in condensed matter physics. The massless part of its spectrum is represented by the dispersion curves in the figure. The photon gradually turns into the e^+e^- pair and exists mostly in that form when it is finally propagating along the magnetic force lines. This capturing effect is important for the formation of pulsar radiation with fields of $B > 0.1B_{cr}$, because it prevents the screening of the accelerating electric field in the polar gap (if the binding of the electron–positron pair into a positronium atom is taken into account [5–9]). It may also be essential for magnetars with fields of approximately 10^{14}–10^{15} G.

The new features introduced by super-Schwinger fields are that the dispersion curves for mode 2 in the figure already deviate from the light cone far from the resonance region. This means that, although photons softer than $2m = 1$ MeV cannot proceed to the values of the ordinate in the figure higher than their energy squared (corresponding to $k_\parallel = 0$), they can still reach the region where the transversal group velocity $dk_0/dk_\perp$ becomes much less than unity and are therefore captured in a trajectory almost parallel to the magnetic field. This is how the capture effect extends to photon energies below the limit $k_0 = 2m$. The cyclotron singularity at the pair-creation threshold in such fields is so strong that even low-energy photons that are unable to create a pair are sensitive to it provided that they belong to mode 2!

In addition to extension of the photon capture effect to softer photons, taking super-Schwinger fields into consideration produces another impact. It leads to a large direction-dependent refraction of mode 2 electromagnetic waves of low frequency. To see this, we consider the limit

$$k_0^2 - k_\parallel^2 \ll 4m^2 \tag{19}$$

in Eq. (14), which reduces to neglecting $k_0^2 - k_\parallel^2$ in the integrand in (14). Then, (14) becomes

$$\kappa_2(k) = \frac{\alpha}{3\pi}(k_0^2 - k_\parallel^2)\frac{B}{B_{cr}}\exp\left(-\frac{k_\perp^2}{2m^2}\frac{B_{cr}}{B}\right). \tag{20}$$

The exponential factor in (20) cannot be significant within region (19). Dispersion equation (1) for mode 2 ($i = 2$) then has solutions that express the photon energy k_0 as a function of its transversal and longitudinal momentum,

$$k_0^2 = k_\parallel^2 + k_\perp^2\left(1 + \frac{\alpha}{3\pi}\frac{B}{B_{cr}}\right)^{-1}. \tag{21}$$

Equation (21) analytically presents the linear parts of the dispersion curves in the figure adjacent to the origin for various values of B. The components $v_{\perp,\parallel}$ of the group velocity, Eq. (3), calculated from (21) are

$$v_\perp = \frac{k_\perp}{k_0}\left(1 + \frac{\alpha}{3\pi}\frac{B}{B_{cr}}\right)^{-1}, \quad v_\parallel = \frac{k_\parallel}{k_0}. \tag{22}$$

The modulus of the group velocity squared is now given by

$$v_\perp^2 + v_\parallel^2 = \frac{1}{1 + \frac{\alpha}{3\pi}\frac{B}{B_{cr}}} + \frac{\frac{\alpha}{3\pi}\frac{B}{B_{cr}}\cos^2\vartheta}{1 + \frac{\alpha}{3\pi}\frac{B}{B_{cr}}\cos^2\vartheta}, \tag{23}$$

where ϑ is the angle between the photon momentum and the field, $\tan\vartheta = k_\perp/k_\parallel$. Equation (23) has the maximum value of unity for the parallel propagation, $\vartheta = 0$, in accordance with the general statement in Section 2, and is minimum for perpendicular propagation, $\vartheta = \pi/2$.

Expression (4) for the angle θ between the direction of the electromagnetic energy propagation and the external magnetic field in the super-Schwinger limit for mode 2 becomes

$$\frac{v_\perp}{v_\parallel} = \tan\theta = \frac{k_\perp}{k_\parallel}\left(1 + \frac{\alpha}{3\pi}\frac{B}{B_{cr}}\right)^{-1} = \tan\vartheta\left(1 + \frac{\alpha}{3\pi}\frac{B}{B_{cr}}\right)^{-1}. \tag{24}$$

Because $\tan\theta < \tan\vartheta$, the photon emitted tangentially to curved force lines bends towards these lines. This is also related to low-frequency radiation.

The refraction index (6) in mode 2 for $k_0^2 - k_\parallel^2 \ll 4m^2$ and $B \gg B_{cr}$ is given by

$$n_2 = \left(\frac{1 + \frac{\alpha}{3\pi}\frac{B}{B_{cr}}}{1 + \frac{\alpha}{3\pi}\frac{B}{B_{cr}}\cos^2\vartheta}\right)^{1/2}. \tag{25}$$

The refraction index obtained depends on the direction of the photon momentum, characterized by the angle ϑ, but does not depend on its energy. In other words, there is no frequency dispersion in a wide range from slow radio waves up to soft γ-rays with $k_0 \ll 2m$. This is a consequence of the fact that only linear parts in momenta squared were actually left in κ_2 (correspondingly, $f(k_\perp^2)$ in (2) is proportional to $k_\perp^2$ according to (21)).

Refraction index (25) reaches its maximum for transversal propagation ($k_\parallel = 0$, $\vartheta = \pi/2$),

$$n_2^\perp = \left(1 + \frac{\alpha}{3\pi}\frac{B}{B_{cr}}\right)^{1/2} = \left(1 + 7.7\times 10^{-4}\frac{B}{B_{cr}}\right)^{1/2}. \tag{26}$$

For $B \sim 10B_{cr}$, the deviation of refraction index (26) from unity exceeds that value for gases at atmospheric pressure in the optic range by an order of magnitude; for $B \sim 1000B_{cr}$, it reaches a value characteristic of transparent liquids and glass; the refraction index (26) becomes equal to that of diamond ($n = 2.4$) for $B = 27 \times 10^{16}$ G.

Contrary to the case of mode 2 just considered, the polarization tensor eigenvalue κ_3 in (15) contains neither the contribution linearly growing with the external field nor the resonance. For mode 3, dispersion equa-

tion (1) with its right-hand side given as (15) has the solution

$$k_0^2 = k_\parallel^2 + k_\perp^2 \frac{Z - \alpha/3\pi}{Z}, \quad (27)$$

where

$$Z = 1 - \frac{\alpha}{3\pi}\left(\ln\frac{B}{B_{cr}} - C - 1.21\right). \quad (28)$$

The well-known absence of asymptotic freedom in QED manifests itself in the minus before the logarithm in (28). This results in pathological consequences for the fields as large as $B_{cr}\exp(3\pi/\alpha)$. In this domain, the coefficient of $k_\perp^2$ in (27) first becomes less than zero and then greater than unity as the field grows. The corresponding dispersion laws are nonphysical because they lead to a group velocity greater than unity. In the negative slope case in (27),

$$e > \frac{B}{B_{cr}}\exp(-0.21 - C - 3\pi/\alpha) > 1,$$

the dispersion curve enters sector (12), acceptable for Cherenkov radiation. However, this is the Cherenkov emission of tachyons! It is also odd that, in the latter case, electromagnetic waves can only propagate inside the cone

$$0 < \tan\vartheta < -1 + \frac{\alpha}{3\pi Z}$$

with its axis along the external field, irrespective of the way they are produced. This domain of exponentially large external fields is of no interest to us in this paper.

For fields that are not exponentially large, with logarithmic terms on the order of unity, one should treat all the terms marked by the coefficient $\alpha/3\pi$ in (27) as small. Then, finally, the dispersion law for mode 3 becomes

$$k_0^2 = k_\parallel^2 + k_\perp^2\left(1 - \frac{\alpha}{3\pi}\right). \quad (29)$$

Notably, the field-containing logarithmic terms have cancelled at this point. Thus, dispersion law (29) of mode 3 is saturated in the sense that, unlike Eq. (21) for mode 2, it has reached the universal form, independent of the external field in the super-Schwinger limit. The refraction index of mode 3 corresponding to (29) is

$$n_3 = 1 + \frac{\alpha}{6\pi}\sin^2\vartheta. \quad (30)$$

As in (25), the maximum refraction in mode 3 is achieved in perpendicular propagation, $\vartheta = \pi/2$:

$$n_3^\perp = 1 + 3.8\times10^{-4}. \quad (31)$$

This refraction index is of the order of that of gaseous ammonia and cannot be made larger by increasing the external field any further.

4. CONCLUSIONS

We have found that, in the asymptotic case of external magnetic fields B that can be orders of magnitude larger than the Schwinger value 4.4×10^{13} G, the refractive capacity of the magnetized vacuum grows unlimitedly with this field for electromagnetic radiation belonging to polarization mode 2, but reaches saturation at a moderate level of corrections on the order of $\alpha/3\pi$ for mode 3. For the "parallel energy" of the photon not close to the cyclotron resonance, $k_0^2 - k_\parallel^2 \ll 4m^2$, the refraction effects for mode 2 substantially exceed the above small corrections, typical of the nonasymptotic domain, already for $B \sim 10B_{cr}$. In the range of photon frequencies/energies extending from zero to soft γ-rays, a regime is established for which the dispersive properties of the magnetized vacuum are independent of the photon frequency/energy in each mode, but depend on the direction of its propagation. Apart from the fact that the refraction index in mode 2 for the propagation nonparallel to the external field increases numerically with field, it is remarkable that the angle between the group velocity and the direction of the photon momentum also increases, the wave packet being attracted by the force line of the external field. The effect of γ-quantum capture by a strong magnetic field, known to exist due to resonance phenomena associated with free and bound pair creation, is thus extended to lower energy ranges. Therefore, not only hard γ-rays, but also X-rays, light, and radio waves experience the strong dispersive influence of the magnetized vacuum when the magnetic fields are on the order of magnitude of those estimated to exist in magnetars. In view of this, electromagnetic energy canalization phenomena may become important not only within the traditional context described in Section 3.2, but also in application to scattering of electromagnetic waves falling onto the magnetic field from outside [2]. These may be, for instance, the X-rays emitted from the accretion disk or from the pulsar surface outside the region where the magnetic field enters it. The problem of electromagnetic radiation being bent by the dipole magnetic field of a neutron star was recently addressed in [24], and the competition of this process with the effects of gravity was considered.[2] We insist, however, that such effects cannot be adequately treated disregarding the refraction index dependence on the direction of propagation and using the quadratic-in-field expressions for the polarization operator, only valid in the low-field limit, as is the case in [24].

[2] The author is indebted to H. Mosquera Cuesta, who drew his attention to that work.

ACKNOWLEDGMENTS

I am indebted to Prof. H. Pérez Rojas for the hospitality extended to me during the Workshop on Strong Magnetic Fields and Neutron Stars at ICIMAF in Havana and for encouraging me to renew the study of the magnetic optics of a vacuum. I acknowledge the financial support of the Russian Foundation for Basic Research (project no. 02-02-16944) and the President of the Russian Federation Program for Support of Leading Scientific Schools (grant no. LSS-1578.2003.2). My stay in Havana, where the majority of this work was fulfilled, has become possible thanks to the support of the Instituto de Cibernética, Matemática y Física (ICIMAF), Centro Latino-Americano de Física (CLAF), and the Abdus Salam International Centre for Theoretical Physics (ICTP).

APPENDIX

In this Appendix, the asymptotic expansion presented in Section 3.1 is derived from expressions in [3, 19, 20].

The three eigenvalues κ_i, $i = 1, 2, 3$ of the photon polarization operator in the one-loop approximation, calculated using the exact electron propagator in an external magnetic field, can be expressed as linear combinations of the three functions Σ_i,

$$\begin{aligned}\kappa_1 &= -\frac{1}{2}(z_1 + z_2)\Sigma_1,\\ \kappa_2 &= -\frac{1}{2}(z_1\Sigma_2 + z_2\Sigma_1),\\ \kappa_3 &= -\frac{1}{2}(z_2\Sigma_3 + z_1\Sigma_1),\end{aligned} \tag{A.1}$$

where the new notation

$$z_1 = k_\parallel^2 - k_0^2, \quad z_2 = k_\perp^2 \tag{A.2}$$

is introduced for the momentum variables, with $k^2 = z_1 + z_2$. Here, Σ_i are dimensionless functions of the three ratios B_{cr}/B, z_2B_{cr}/m^2B, and z_1B_{cr}/m^2B, given by

$$\Sigma_i = \Sigma_i^{(1)} + \Sigma_i^{(2)}, \tag{A.3}$$

$$\begin{aligned}\Sigma_i^{(1)}\left(\frac{B_{cr}}{B}\right) &= \frac{2\alpha}{\pi}\int_0^\infty dt\exp\left(-\frac{tB_{cr}}{B}\right)\\ &\times\int_{-1}^1 d\eta\left[\frac{\sigma_i(t,\eta)}{\sinh t} - \lim_{t\to 0}\frac{\sigma_i(t,\eta)}{\sinh t}\right],\end{aligned} \tag{A.4}$$

$$\begin{aligned}&\Sigma_i^{(2)}\left(\frac{B_{cr}}{B}, \frac{z_2B_{cr}}{m^2B}, \frac{z_1B_{cr}}{m^2B}\right)\\ &= \frac{2\alpha}{\pi}\int_0^\infty dt\exp\left(-\frac{tB_{cr}}{B}\right)\int_{-1}^1 d\eta\frac{\sigma_i(t,\eta)}{\sinh t}\\ &\times\left[\exp\left(-z_2\frac{M(t,\eta)}{eB} - z_1\frac{1-\eta^2}{4eB}t\right) - 1\right],\end{aligned} \tag{A.5}$$

where

$$M(t,\eta) = \frac{\cosh t - \cosh t\eta}{2\sinh t} \tag{A.6}$$

and

$$\sigma_1(t,\eta) = \frac{1-\eta}{2}\frac{\sinh(1+\eta)t}{2\sinh t}, \tag{A.7}$$

$$\sigma_2(t,\eta) = \frac{1-\eta^2}{4}\cosh t, \tag{A.8}$$

$$\sigma_3(t,\eta) = \frac{\cosh t - \cosh\eta t}{2\sinh^2 t}. \tag{A.9}$$

The notation "lim" in (A.4) stands for the asymptotic limit

$$\lim_{t\to 0}\frac{\sigma_i(t,\eta)}{\sinh t} = \frac{1-\eta^2}{4t}, \quad i = 1, 2, 3. \tag{A.10}$$

The fact that Σ_i is independent of the fourth possible dimensionless variable z_1/z_2 seems to be an approximation-independent manifestation of analyticity properties due to dispersion relations of the Kramers–Kronig nature.

We first consider $\Sigma_i^{(1)}$. It is independent of the photon energy and momentum. With

$$g_i(t) = \int_{-1}^1 \sigma_i(t,\eta)d\eta, \tag{A.11}$$

Eq. (A.4) can be represented as

$$\Sigma_i^{(1)} = \frac{2\alpha}{\pi}\int_0^\infty \exp\left(-\frac{tB_{cr}}{B}\right)\left(\frac{g_i(t)}{\sinh t} - \frac{1}{3t}\right)dt. \tag{A.12}$$

The integrals in (A.11) are explicitly calculated to give

$$\begin{aligned}g_1(t) &= \frac{1}{4t\sinh t}\left(\frac{\sinh 2t}{t} - 2\right), \quad g_2(t) = \frac{\cosh t}{3},\\ g_3(t) &= \frac{1}{\sinh^2 t}\left(\cosh t - \frac{\sinh t}{t}\right).\end{aligned} \tag{A.13}$$

Our goal is now to find the asymptotic behavior of (A.12) as

$$\frac{B}{B_{\rm cr}} \longrightarrow \infty. \qquad \text{(A.14)}$$

The integrands in (A.12) do not contain singularities at $t = 0$, but would diverge as $t \longrightarrow \infty$ if we precisely set the limiting value

$$\exp(-tB_{\rm cr}/B) = 1.$$

We must therefore divide the integration domain into two parts. In addition,

$$\frac{3g_2(t)}{\sinh t} \longrightarrow 1$$

as $t \longrightarrow \infty$, and, hence, we have to add and subtract this limit beforehand in the integrand of $\Sigma_2^{(1)}$. This is not required in treating the cases of $i = 1$ and 3, because

$$\frac{g_{1,3}}{\sinh t} \longrightarrow 0$$

sufficiently fast. Thus, we have

$$\begin{aligned}\frac{3\pi}{2\alpha}\Sigma_i^{(1)} &= \int_0^\infty \exp\left(-\frac{tB_{\rm cr}}{B}\right)\left(\frac{3g_i(t)}{\sinh t} - \frac{1}{t} - \delta_{i2}\right)dt \\ &\quad + \delta_{i2}\int_0^\infty \exp\left(-\frac{tB_{\rm cr}}{B}\right)dt \\ &= \int_0^T \exp\left(-\frac{tB_{\rm cr}}{B}\right)\left(\frac{3g_i(t)}{\sinh t} - \frac{1}{t} - \delta_{i2}\right)dt + \frac{B}{B_{\rm cr}}\delta_{i2} \\ &\quad + \int_T^\infty \exp\left(-\frac{tB_{\rm cr}}{B}\right)\left(\frac{3g_i(t)}{\sinh t} - \delta_{i2}\right)dt \qquad \text{(A.15)} \\ &\quad - \int_T^\infty \exp\left(-\frac{tB_{\rm cr}}{B}\right)\frac{dt}{t} \\ &= \int_0^T \left(\frac{3g_i(t)}{\sinh t} - \frac{1}{t}\right)dt - \delta_{i2}T + \int_T^\infty \left(\frac{3g_i(t)}{\sinh t} - \delta_{i2}\right)dt \\ &\quad + \frac{B}{B_{\rm cr}}\delta_{i2} + \ln\left(\frac{B_{\rm cr}}{B}\right) + C + \ln T,\end{aligned}$$

where T is an arbitrary positive number, δ_{i2} is the Kronecker delta, and C is the Euler constant. We have omitted the exponentials in the first two integrals after the second equality sign in (A.15) because the resulting integrals converge. We then used the asymptotic expansion of the standard exponential–integral function (up to terms linearly decreasing with $B/B_{\rm cr}$) [25]:

$$-\int_T^\infty \exp\left(-\frac{tB_{\rm cr}}{B}\right)\frac{dt}{t} = \ln\frac{B_{\rm cr}}{B} + \ln T + C. \qquad \text{(A.16)}$$

The most slowly decreasing term neglected in

$$3\int_T^\infty dt \exp(-tB_{\rm cr}/B)\frac{g_i(t)}{\sinh t}$$

is

$$\delta_{i1}\frac{3B_{\rm cr}}{4B}\ln\frac{B_{\rm cr}}{B},$$

because

$$\frac{g_1(t)}{\sinh t} \sim \frac{1}{4t^2}$$

as $t \longrightarrow \infty$. Other neglected terms decrease at least as fast as $B_{\rm cr}/B$, because $3g_i(t)/\sinh t - \delta_{i2}$ decreases exponentially, as $\exp(-2t)$, for $i = 2$, 3 when t is large.

Numerical calculations using the MATHCAD code make it possible to evaluate the constants

$$\begin{aligned}h_i &= \int_0^T dt\left(\frac{3g_i(t)}{\sinh t} - \frac{1}{t}\right) \\ &\quad + \int_T^\infty \left(\frac{3g_i(t)}{\sinh t} - \delta_{1,2}\right)dt + \ln T - \delta_{i2}T\end{aligned} \qquad \text{(A.17)}$$

involved in (A.15) as

$$h_1 = 1.21, \quad h_2 = -0.69, \quad h_3 = 0.21. \qquad \text{(A.18)}$$

Finally, in the asymptotic regime $B/B_{\rm cr} \gg 1$, we have

$$\Sigma_i^{(1)} = \frac{2\alpha}{3\pi}\left(\ln\frac{B_{\rm cr}}{B} + C + h_i + \frac{B}{B_{\rm cr}}\delta_{i2}\right) \qquad \text{(A.19)}$$

up to terms decreasing at least as fast as integral powers of the ratio $B_{\rm cr}/B$ and to the slower term

$$\frac{\alpha}{2\pi}\frac{B_{\rm cr}}{B}\ln\frac{B_{\rm cr}}{B},$$

omitted in $\Sigma_1^{(1)}$.

We now turn to $\Sigma_i^{(2)}$ in Eq. (A.5). This depends on the three arguments indicated in (A.5). We are interested in the asymptotic domain described by condition (A.14) and

$$\frac{2eB}{z_1} \longrightarrow \infty. \qquad \text{(A.20)}$$

We keep the ratio $2eB/z_2$ finite whenever it makes sense.

The asymptotic expansion of (A.7)–(A.9) in powers of $\exp(-t)$ and $\exp(\eta t)$ produces an expansion of (A.5) into a sum of contributions coming from thresholds (16), the singular behavior at the threshold points originating from the divergences of the t-integration in (A.5) near $t = \infty$ (see [3, 19, 20] for details). The leading terms in the expansion of (A.7)–(A.9) at $t \longrightarrow \infty$ are

$$\left(\frac{\sigma_1(t,\eta)}{\sinh t}\right)\Bigg|_{t\to\infty} = \frac{1-\eta}{2}\exp(t(\eta-1))), \quad \text{(A.21)}$$

$$\left(\frac{\sigma_2(t,\eta)}{\sinh t}\right)\Bigg|_{t\to\infty} = \frac{1-\eta^2}{4}(1+2\exp(-2t)), \quad \text{(A.22)}$$

$$\left(\frac{\sigma_3(t,\eta)}{\sinh t}\right)\Bigg|_{t\to\infty} = 2\exp(-2t). \quad \text{(A.23)}$$

Changing to $\tau = t/eB$ and taking into account that $M(\infty, \eta) = 1/2$ (see (A.6)), we evaluate (A.5) near the lowest singular thresholds ($n = 0$, $n' = 1$ or $n' = 0$, $n = 1$ for $i = 1$ in (16), $n = n' = 1$ for $i = 3$, and $n = n' = 0$ and $n = n' = 1$ for $i = 2$) as

$$\Sigma_i^{(2)} = \frac{2\alpha eB}{\pi}\int_{-1}^{1} d\eta \int_0^\infty d\tau \exp(-m^2\tau) \times \left(\frac{\sigma_i(eB\tau,\eta)}{\sinh eB\tau}\right)\Bigg|_{t\to\infty} \times\left[\exp\left(-\frac{z_2}{2eB} - \frac{z_1(1+-\eta^2)}{4}\tau\right) - 1\right]. \quad \text{(A.24)}$$

After integration over τ, we obtain, e.g.,

$$\Sigma_1^{(2)} = \frac{4\alpha eB}{\pi}\int_{-1}^{1} d\eta(1-\eta) \times\left(\frac{\exp(-z_2/2eB)}{4m^2 + 4(1-\eta)eB + z_1(1-\eta^2)} - \frac{1}{4m^2 + 4(1-\eta)eB}\right). \quad \text{(A.25)}$$

The pole in the above expression, caused by the integration over t, turns into the inverse square root singularity after the integration over η (cf. the derivation of (17) from (14)). In the limit (A.14), (A.20) when $B \gg m^2$, $B \gg |z_1|$, no singularity remains in this expression (it is shifted to the infinitely remote region) and we are left with

$$\Sigma_1^{(2)} = \frac{2\alpha}{\pi}\left(\exp\left(-\frac{z_2}{2eB}\right) - 1\right).$$

The same situation occurs for $\Sigma_3^{(2)}$ and for higher thresholds (also for contributions into $\Sigma_2^{(2)}$ other than those coming from the first term in (A.22)). The result of the calculation analogous to (A.25) is

$$\Sigma_3^{(2)} = \frac{4\alpha}{\pi}\left(\exp\left(-\frac{z_2}{2eB}\right) - 1\right).$$

We conclude that, in the limit (A.14), (A.20), there are no cyclotron resonances in eigenvalues $\kappa_{1,3}$ according to (A.1), and that $\Sigma_1^{(2)}$ does not introduce a singular contribution into κ_2. Consequently, there is no reason to keep the ratio eB/z_2 finite as $B \longrightarrow \infty$, in $\Sigma_{1,3}$, because z_2 may grow infinitely on the dispersion curve only when there is a resonance.

We must, therefore, consider only the limit when all the three arguments in $\Sigma_{1,3}^{(2)}$ tend to zero. Handling this limit in (A.5) is straightforward:

$$\lim\Sigma_{1,3}^{(2)} = \frac{-2\alpha}{\pi eB}\int_0^\infty dt \times\int_{-1}^{1} d\eta \frac{\sigma_{1,3}(t,\eta)}{\sinh t}\left(z_2 M(t,\eta) + z_1\frac{1-\eta^2}{4}t\right). \quad \text{(A.26)}$$

Both integrations here converge, and, hence, this contribution decreases as z_1/eB and z_2/eB when $B \longrightarrow \infty$. This is to be neglected within our scope of accuracy.

The situation is different with $\Sigma_2^{(2)}$. The resonance behavior is present here due to the contribution of the leading asymptotic term $(1 - \eta^2)/4$ in (A.22). It is responsible for the first threshold at $-z_1 = 4m^2$ (the ground Landau state $n = n' = 0$ in (16)), which remains in place as $B \longrightarrow \infty$. We must, therefore, keep the ratio z_2/eB nonzero in passing to the limit of large fields (because $z_2 \longrightarrow \infty$ near the singular threshold on the dispersion curve) for the contribution of this term to $\Sigma_2^{(2)}$. The contributions of nonleading terms in expansion (A.22) to $\Sigma_2^{(2)}$ are nonsingular and should be treated along the same lines as $\Sigma_{1,3}$ above. They decrease as z_1/eB, z_2/eB and are to be neglected. Finally, for (A.5), we are left in the limit (A.14), (A.20) with

$$\Sigma_2^{(2)} = \frac{2\alpha}{\pi}\int_0^\infty dt \exp\left(-\frac{tB_{\rm cr}}{B}\right)\int_{-1}^{1} d\eta \frac{1-\eta^2}{4} \times\left[\exp\left(-z_2\frac{M(t,\eta)}{eB} - z_1\frac{1-\eta^2}{4eB}t\right) - 1\right]. \quad \text{(A.27)}$$

With $t = eB\tau$ and the asymptotic form $M(eB\tau, \eta) = 1/2$, $eB\tau \gg 1$, we finally obtain the leading contribution to

$\Sigma_2^{(2)}$ in the limit (A.14), (A.20):

$$\Sigma_2^{(2)} = \frac{2\alpha eB}{\pi}\int_{-1}^{1} d\eta(1-\eta^2) \times \frac{\exp(-z_2/2eB)}{4m^2 + z_1(1-\eta^2)} - \frac{2\alpha eB}{3\pi m^2}. \quad \text{(A.28)}$$

Combining Eqs. (A.28) and (A.19) in accordance with (A.1), and bearing (A.2) and (A.3) in mind, we obtain expressions (13)–(15) for the polarization operator eigenvalues if we also neglect the constant and logarithmic terms in κ_2 coming from $\Sigma_{1,2}^{(1)}$ in (A.19) as compared to the terms increasing linearly with B.

REFERENCES

1. S. L. Adler, J. N. Bahcall, G. G. Callan, and M. N. Rosenbluth, Phys. Rev. Lett. **25**, 1061 (1970).
2. A. E. Shabad, Lett. Nuovo Cimento **2**, 457 (1972).
3. A. E. Shabad, Ann. Phys. (N.Y.) **90**, 166 (1975).
4. A. E. Shabad and V. V. Usov, Nature **295**, 215 (1982); Astrophys. Space Sci. **102**, 327 (1984).
5. V. V. Usov and A. E. Shabad, Pis'ma Zh. Éksp. Teor. Fiz. **42**, 17 (1985) [JETP Lett. **42**, 19 (1985)].
6. H. Herold, H. Ruder, and G. Wunner, Phys. Rev. Lett. **54**, 1452 (1985).
7. A. E. Shabad and V. V. Usov, Astrophys. Space Sci. **117**, 309 (1985); **128**, 377 (1986).
8. V. B. Bhatia, Namrata Chopra, and N. Panchapakesan, Astrophys. J. **388**, 131 (1992).
9. V. V. Usov and D. B. Melrose, Aust. J. Phys. **48**, 571 (1995).
10. V. V. Usov, Astrophys. J. Lett. **572**, L87 (2002).
11. M. G. Baring, Astrophys. J. Lett. **440**, L69 (1995).
12. A. C. Harding, M. G. Baring, and P. L. Gonthier, Astrophys. J. **246**, 476 (1997).
13. M. V. Chistyakov, A. V. Kuznetsov, and N. V. Mikheev, Phys. Lett. B **434**, 67 (1998); M. V. Chistyakov and N. V. Mikheev, hep-ph/0107217.
14. V. M. Kaspi, in *Pulsar Astronomy-2000 and Beyond. ASP Conference Series*, Ed. by M. Kramer, N. Wex, and R. Wielebinski (1999), Vol. 202; astro-ph/9912284.
15. V. V. Skobelev, Izv. Vyssh. Uchebn. Zaved., Fiz., No. 10, 142 (1975); Yu. M. Loskutov and V. V. Skobelev, Phys. Lett. A **56**, 151 (1976).
16. A. E. Shabad, Kratk. Soobshch. Fiz. **3**, 13 (1976).
17. I. A. Batalin and A. E. Shabad, Zh. Éksp. Teor. Fiz. **60**, 894 (1971) [Sov. Phys. JETP **33**, 483 (1971)].
18. H. Pérez Rojas and A. E. Shabad, Ann. Phys. (N.Y.) **121**, 432 (1979).
19. A. E. Shabad, Tr. Fiz. Inst. im. P.N. Lebedeva, Akad. Nauk SSSR **192**, 5 (1988).
20. A. E. Shabad, *Polarization of the Vacuum and a Quantum Relativistic Gas in an External Field* (Nova Sci., New York, 1991); Proc. (Tr.) P. N. Lebedev Phys. Inst., Vol. 192.
21. H. Pérez Rojas, Zh. Éksp. Teor. Fiz. **76**, 3 (1979) [Sov. Phys. JETP **49**, 1 (1979)]; H. Pérez Rojas and A. E. Shabad, Ann. Phys. (N.Y.) **138**, 1 (1982).
22. Wu-yang Tsai, Phys. Rev. D **10**, 2699 (1974); V. N. Baier, V. M. Katkov, and V. M. Strakhovenko, Zh. Éksp. Teor. Fiz. **68**, 403 (1975) [Sov. Phys. JETP **41**, 198 (1975)].
23. V. I. Ritus, Zh. Éksp. Teor. Fiz. **69**, 1517 (1975) [Sov. Phys. JETP **42**, 774 (1975)]; Zh. Éksp. Teor. Fiz. **73**, 807 (1977) [Sov. Phys. JETP **46**, 423 (1977)].
24. V. L. Denisov, I. P. Denisova, and S. I. Svertilov, Dokl. Akad. Nauk **380**, 754 (2001) [Dokl. Phys. **46**, 705 (2001)]; astro-ph/0110705.
25. *Handbook of Mathematical Functions*, Ed. by M. Abramowitz and I. A. Stegun (National Bureau of Standards, Washington, 1964; Nauka, Moscow, 1979).

Journal of Experimental and Theoretical Physics, Vol. 98, No. 2, 2004, pp. 197–206.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 2, 2004, pp. 222–232.

ATOMS, SPECTRA, RADIATION

Clustering in Molecular Gases Freely Expanding into Vacuum

A. A. Vostrikov and D. Yu. Dubov

Kutateladze Institute of Thermophysics, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia

e-mail: vostrikov@itp.nsc.ru

Received June 2, 2003

Abstract—The actual enthalpies and entropies of clustering and the properties of clusters characterizing their catalytic effect on the rate of relaxation of vibrational degrees of freedom are used to compute $(N_2O)_N$ and $(CO_2)_N$ cluster formation and growth in gases expanding into a vacuum through a sonic nozzle. Relations are found between characteristics of gaseous N_2O and CO_2 in the nozzle source and in the jet flow containing clusters and those of the molecular cluster beam formed from the axial region of the jet. Calculated and measured values of the following cluster-beam characteristics are compared: intensity, flux density for molecules in the beam, scaling parameters for transition to well-developed condensation in the jet, cluster size distribution function, mean cluster size, and internal cluster temperature. Realistic characterization of cluster properties ensures good agreement between calculated and measured results and provides a basis for adequate description of the mechanisms of molecular cluster formation in supersonic jets issuing from sonic nozzles characterized by extremely rapid decrease in gas temperature and highly nonequilibrium distribution of energy over molecular degrees of freedom. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

We present a model that describes clustering in molecular gases freely expanding into a vacuum and that predicts the characteristics of a cluster beam formed in the axial region of the jet. Even though clustering by bulk condensation in supersonic jets has been a subject of considerable interest [1, 2], the dependence of condensation kinetics on cluster size characteristics and the influence of nonequilibrium in a molecular-gas jet flow on cluster formation have yet to be studied.

Free expansion of gases into vacuum is characterized by extremely high rates of decrease in gas temperature (~10^6 K/s). Despite the rapid drop in gas density along the jet axis ($n_j \sim (d_*/x)^2$, where d_* is the sonic-nozzle exit diameter and x is the distance from the nozzle exit along the jet axis), high degrees of gas supersaturation are reached for certain values of nozzle source pressure P_0 and temperature T_0, leading to condensation [3]. The kinetics of the initial stage of condensation are described by the theory of homogeneous nucleation [4–6]. At high degrees of gas supersaturation, the condensation nuclei are clusters consisting of several molecules. Under these conditions, the liquid droplet theory developed in [4–6] cannot be applied to describe the formation and growth of clusters [7, 8]. Furthermore, the rapid drop in the density n_j gives rise to a highly nonequilibrium distribution of energy over molecular degrees of freedom. The vibrational, rotational, and (during collisionless expansion) translational degrees of freedom are frozen consecutively downstream along the jet [9, 10]. Relaxation of internal degrees of freedom resumes after molecules aggregate into clusters [11–15], impeding condensation in the jet [16].

The nozzle beam method is widely used to study condensation processes and properties of weakly bound molecular clusters. This has motivated studies of relations between gas characteristics in the nozzle source, characteristics of the jet flow of a condensing gas, and cluster properties. In this paper, we propose a model of condensation based on cluster size characteristics and compute the formation of $(N_2O)_N$ and $(CO_2)_N$ clusters in gases freely expanding into vacuum to find relations between the characteristics mentioned above. Numerical results are compared with those obtained in beam measurements.

2. COMPUTATIONAL MODEL

The present model is based on the results obtained by measuring the binding enthalpy h for small $(N_2O)_N$ and $(CO_2)_N$ clusters as a function of the cluster size N [17, 18]. These results can be used to express a key parameter in condensation theory, the work of clustering $\Delta G(N)$ (see [4–6]), in terms of actual physical properties of a cluster. Indeed,

$$\Delta G(N) = -Nh(N) + T\Delta S(N), \qquad (1)$$

in an isothermal process, where $-Nh(N)$ and $\Delta S(N)$ are the changes in enthalpy and entropy associated with

clustering in a gas at temperature T and pressure P (the cluster temperature is $T_c = T$).

The change in entropy associated with aggregation of N molecules into a cluster can be represented as the sum of changes in the entropies corresponding to internal and external degrees of freedom. The change in the entropy of translational and rotational motion is calculated in explicit form, since it corresponds to a change in the number of these degrees of freedom only. The entropy of intermolecular vibrations in a cluster is set equal to the vibration entropy in a liquid, because clusters of size $N < 10$ are "liquid" at typical temperatures of cluster formation in jets (between 100 and 150 K) [19–21]. For larger clusters, the jumps in the enthalpy and entropy of a cluster due to phase transition (solidification) are taken into account. The resulting expression for $\Delta G(N)$ is (see [7, 8])

$$\Delta G = -N(h-L) + (N-1)k_B T \ln\frac{P_l^0}{P} - T\Delta S_c^g - (T_f - T)\Theta N\frac{L}{T}. \quad (2)$$

Here, k_B is the Boltzmann constant; P_l^0 is the constant preexponential factor in the expression for pressure over a flat liquid surface; $T_f(N)$ and $L(N)$ are the cluster melting point and heat of fusion, respectively; $\Theta = 0$ at $T > T_f$; and $\Theta = 1$ at $T < T_f$. The difference in entropy of translational and rotational degrees of freedom between a cluster and a molecule, ΔS_c^g, is due to the difference between their masses and moments of inertia and, for CO_2 and N_2O linear molecules, to the extra rotational degree of freedom of a cluster.

For clusters of size not greater than $N \approx 10$, the heat of fusion can be neglected (see [19–21]), and expression (2) reduces to

$$\Delta G = -Nh + (N-1)k_B T \ln\frac{P_l^0}{P} - T\Delta S_c^g. \quad (3)$$

An analysis of expression (3) leads to the following important conclusions.

1. The dependence of ΔG on N is determined by the first two terms, while the contribution due to ΔS_c^g reduces the value of ΔG, but weakly depends on N.

2. The activation barrier in the formation of small clusters (the maximum of $\Delta G(N)$) arises as a result of competition between the enthalpy term $-Nh$ and the entropy-related term

$$(N-1)k_B T \ln\frac{P_l^0}{P} - T\Delta S_c^g.$$

Aggregation of molecules into a cluster is always advantageous in terms of energy, but clustering is impeded by the decrease in entropy associated with aggregation of molecules.

3. The key entropy-related term in (3) can be rewritten as

$$k_B T(N-1)\ln(v_g/v_l),$$

where $v_g = k_B T/P$ is the average volume per molecule in a gas and $v_l \equiv k_B T/P_l^0$ is the coordinate part of the average phase space volume per particle in a liquid (i.e., the phase space volume per molecule in a small cluster is set equal to the phase space volume per particle in a liquid).

4. The factor $N - 1$ in the entropy-related term in (3) is due to the thermal motion of a cluster as a whole. Thus, the entropy change associated with clustering is substantially reduced by taking into account the external degrees of freedom for a small cluster. However, their contribution to the enthalpy of clustering is on the order of $k_B T$, which is much smaller than Nh. This method for taking into account the thermal motion of a cluster is better justified as compared to those proposed in [4, 5].

The dependence $P_l^0(T)$ is ignored in (2) and (3) because $P_l^0(T)$ varies weakly at the low gas temperatures characteristic of jet flows. In calculations with constant T_0 and varying P_0, the dependence $h(T)$ is also neglected, because its influence on condensation is essential only for the small clusters forming in a narrow jet region where temperature varies within a range of 20 K.

Following the approach proposed in [22], we represent gas condensation in a jet flow as two interrelated processes. One of these is the formation of condensation nuclei of size $N = N_*$. The rate of nucleation is computed by solving classical kinetic equations [23]. The other process is the growth of cluster size starting from $N = N_*$. The rate of cluster growth is determined by the difference in flux between the gas molecules colliding with the cluster and the molecules evaporating from the cluster. The inward flux is calculated in the approximation of free-molecular flow for a spherical cluster. The outward flux of evaporating molecules depends on the saturation pressure P_c over a cluster, which is found by using the condition

$$\left[\frac{\partial \Delta G}{\partial N}\right]_{P = P_c,\, T = T_c} = 0 \quad (4)$$

and expression (2). The cluster temperature T_c is determined by the difference of the inward and outward energy fluxes, i.e., by the energy released as a result of condensation onto a cluster and the energy carried away by evaporating molecules. Since evaporating molecules are in thermal equilibrium with the cluster, the energy

carried away by each molecule released from a cluster is

$$h'(N) - \left(\frac{C_P^C}{N} - \xi k_B\right)T_c, \qquad h'(N) = \frac{d(Nh(N))}{dN}, \tag{5}$$

where $h'(N)$ is the change in enthalpy due to release of a molecule from a cluster and C_P^C is the specific heat associated with the internal cluster degrees of freedom. The factor ξ allows for a non-Maxwellian velocity distribution for molecules evaporating from the cluster. For a molecule having two rotational degrees of freedom, $\xi = 3$ [24]. Similarly, the energy per molecule released by a condensing gas is

$$h'(N) + \xi k_B T - \frac{C_P^C}{N}T_c.$$

Using these assumptions, we obtain the following expression for the rate of cluster size variation:

$$\frac{dN}{dt} = \nu_N \frac{\xi k_B (T_c - T)}{h'(N) - (C_P^C/N - \xi k_B)T}. \tag{6}$$

Here, ν_N is the frequency of molecule–cluster collisions expressed as

$$\nu_N = \sqrt{\frac{N+1}{N}}\sigma_1(1 + rN^{2/3})^2 \frac{4P}{\sqrt{2\pi m k_B T}}, \tag{7}$$

where σ_1 is the gas kinetic cross section of a molecule, m is the molecule mass, and r is a constant introduced to allow for loose packing of molecules in a cluster and calculated by using the density of molecules in the liquid phase. The physical meaning of (6) is as follows. The release of the latent heat of condensation keeps the temperature of a growing cluster above the ambient gas temperature. The flux of evaporating molecules increases exponentially with T_c. As a consequence, the temperature T_c tends to a steady value at which the outward and inward heat fluxes are balanced. Therefore, the cluster growth rate in a supersaturated gas is mainly determined by heat transfer to the gas. We note here that even a small concentration of easily condensing molecules in a carrier gas, such as H_2O in CO_2 or CO_2 in N_2, can initiate premature condensation of the carrier gas, because fast heat transfer accelerates aggregation of the additive into clusters, which play the role of nucleation centers in carrier-gas condensation. As a result, heteromolecular (mixed) clusters are produced in jets [25–27]. The rate of heat transfer is proportional to ν_N and $T_c - T$. Therefore, the second term in (6) determines the fraction of condensing molecules in the total number of molecules colliding with the cluster. In our computations, we assume local equilibrium in the incident-molecule–cluster system, because the fraction of clustered molecules calculated under this assumption agrees with experimental data. For example, it was found in [28] that about 3% of the molecules colliding with a $(CO_2)_{n=2-4}$ cluster in a supersonic expanding flow condense on the cluster. This finding is consistent with our results. Our computations show that the fraction of condensing molecules increases with N, reaching about 8% for $N = 10$.

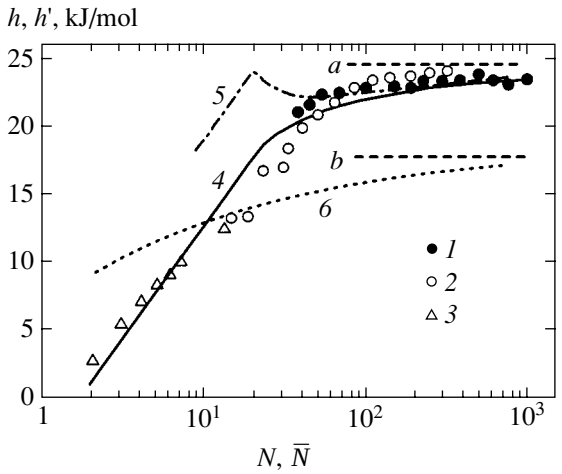


Fig. 1. Size dependence of binding enthalpy $h(N)$ in $(N_2O)_N$ (*1*, *2*, *4*, *6*) and $(CO_2)_N$ (*3*) and enthalpy $h'(N)$ of release of an N_2O molecule from a cluster (*5*).

The functions $L(N)$ and $T_f(N)$ in (2) were constructed by using numerical results [19], and their variation revealed their weak effect on computations of condensation in a jet. The function $h(N)$ in (2) and (3) was determined from the results of bolometric measurements of the enthalpies of bonding for $(CO_2)_N$ and $(N_2O)_N$ clusters [17, 18] and from the results obtained by modeling small clusters of CO_2 molecules [19]. Figure 1 shows $h(\bar{N})$ measured by varying P_0 (*1*) and T_0 (*2*) for gaseous N_2O. Triangles in Fig. 1 represent numerical results borrowed from [19]. The data represented by symbols *1–3* were approximated as follows:

$$h(N) = \begin{cases} B\ln\dfrac{N}{2} + h_2, & N < N_0, \\ h_s^\infty\left[1 - \left(\dfrac{0.1}{N}\right)^{1/3} - \left(\dfrac{\omega}{N}\right)^2\right], & N > N_0. \end{cases} \tag{8}$$

Here, B and h_2 are empirical parameters and h_s^∞ is the heat of sublimation. The constant ω and the value of N_0 at which the functions in (8) are matched were determined by the continuity conditions for $h(N)$ and its derivative at $N = N_0$. The approximation (8) with $B = 7064$ J/mol and $h_2 = 831$ J/mol obtained for $(N_2O)_N$

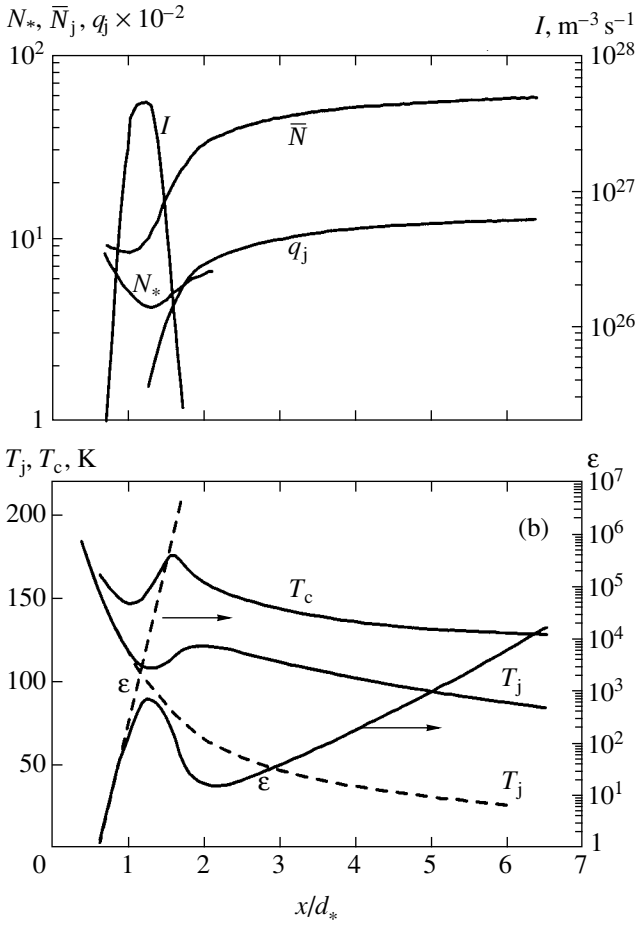


Fig. 2. Variation of critical cluster size N_*, mean cluster size $\bar{N}_j$, condensate mass fraction q_j, nucleation rate I, gas supersaturation ε, gas temperature T_j, and cluster temperature T_c along an N_2O jet.

clusters corresponds to curve *4* in Fig. 1. This curve was used in (5) to find the enthalpy $h'(N)$ (curve *5*). Curves *a* and *b* are the heat of sublimation and heat of vaporization for liquid N_2O, respectively. Relying on the droplet theory of nucleation [4] based on the concept of cluster surface tension, we calculated curve *6* for $h(N)$, setting the surface tension equal to 2.4×10^{-2} N m^{-1}. This value was obtained by using Eq. (154.9) from [29] in the low-temperature (T_c) limit. It is clear that curve *6* disagrees with experimental data.

The condensation equations were combined with the system of equations describing one-dimensional expansion of an inviscid non-heat-conducting perfect gas with isentropic exponent $\gamma_0 = 1.3$ upstream of the nozzle exit and $\gamma = 1.4$ in the supersonic flow region. The contribution of thermal motion of clusters to the gas pressure in the jet flow was neglected (see below). The streamtube geometry was prescribed by using the results calculated in [30] for $\gamma = 1.4$, which are consistent with the experiment described in [31] in the case of free expansion of CO_2 and N_2O without condensation. Vibrational relaxation of molecules and the heat of condensation released in clustering increase the jet angle and have a certain effect on cluster properties (see below).

The cluster beam intensity J (molecular flux density in the beam) was computed in the continuum approximation upstream of the skimmer (a hollow conical diaphragm with an opening that does not distort the flow in a near-axis region of the jet). Downstream of the skimmer, particles were assumed to move without collisions with equal velocities U along the beam axis. Under these conditions, the expression for J corresponding to a small angle of observation has the form (see [8])

$$J = J_m + J_c = 0.5 n_j U \varphi^2 \gamma M^2 \times \left[(1 - q_j) + q_j \int_2^\infty N f_j(N) dN \right]. \quad (9)$$

Here, J_m and J_c are the free-molecule and cluster partial beam intensities, respectively; $n_j U$ is the flux density for molecules entering the skimmer; 2φ is the angle spanned by the skimmer opening at the point where J is measured; q_j is the mass fraction of the condensate (clustered molecules) entering the skimmer;

$$M = \sqrt{\frac{mU^2}{\gamma k_B T}}$$

is the Mach number for molecules; and $f_j(N)$ is the cluster size distribution normalized to unity at the skimmer entrance.

The function $f_j(N)$ was computed by partitioning the nucleation region (about $10d_*$ along the jet axis) into 100 cells. The growth of the clusters forming in each cell was computed separately.

3. NUMERICAL RESULTS AND COMPARISON WITH EXPERIMENT

The starting parameters required to compute condensation of N_2O and CO_2 in jets issuing from a sonic nozzle, B and h_2 in (8), were determined by comparing the computed and measured J and $\bar{N}$ (mean cluster size in a beam) as functions of P_0 at $T_0 = 285$ K. The best agreement was achieved with $B = 7064$ and 7479 J/mol and $h_2 = 831$ and 997 J/mol for N_2O and CO_2, respectively. The functions $h(N)$ and $h'(N)$ used in computations of N_2O condensation are shown in Fig. 1 (curves *4* and *5*, respectively).

Figure 2 shows how the basic characteristics of the gas and cluster phases vary along the axis of an N_2O jet when $P_0 = 7 \times 10^4$ Pa, $T_0 = 285$ K, and $d_* = 1.0$ mm. Under these conditions, the clustering begins near the

nozzle exit (at $x/d_* < 1$). The highest nucleation rate, $I \approx 5 \times 10^{27}$ m^{-3} s^{-1}, and the smallest nucleus size, $N_* \approx 4$, are observed at $x/d_* \approx 1.2$. These extremely high and low values of I and N_*, respectively, are explained by a high gas supersaturation in the jet, $\varepsilon = P_j/P_s \approx 1000$, due to an extremely high rate of decrease in gas temperature T_j in a free jet. Here, P_j is the static gas pressure at a point x_j on the jet axis, and P_s is the gas pressure on the saturation curve that corresponds to T_j. The function $\varepsilon(x/d_*)$ computed for an isentropically expanding gas with $\gamma_0 = 1.3$ upstream of the nozzle exit and $\gamma = 1.4$ downstream of the nozzle is shown in Fig. 2 by a dashed curve, and an analogous function computed for condensing gas is shown by a solid curve. The rapid drop in ε within the interval $1.2 < x/d_* < 2$ is caused by the release of heat of condensation associated with steep growth of the condensate fraction q_j in the jet. The curves representing q_j and $\bar{N}_j$ in Fig. 2 have similar shapes at $x/d_* > 1.5$; i.e., the increase in q_j is due to an increase in $\bar{N}$ rather than in the number of clusters. The limit values approached at large x/d_* are $q_j = 15\%$ and $\bar{N}_j = 68$ for the values of P_0, T_0, and d_* used in the computations. The corresponding concentration of clusters of size $\bar{N}_j$ in the jet is about 0.2%. The cluster concentration rapidly decreases with increasing P_0, because the positive slope of $\bar{N}_j(P_0)$ is greater than that of $q_j(P_0)$. The heat release due to condensation in the jet causes the variation of T_j to deviate from an isentropic curve. The function $T_j(x/d_*)$ is shown in Fig. 2 by a solid curve; the isentropic curve extended to $x/d_* > 1$, by a dashed curve. Since the latent heat of condensation is absorbed by clusters, we have $T_c > T_j$. The curve of $T_c(x/d_*)$ shown in Fig. 2 was computed for clusters of size $N = \bar{N}_j$. The difference $T_c - T_j$ is tens of kelvins. However, it is obvious that the difference $T_c - T_j$ for strongly bound atomic clusters forming in a jet can be as high as hundreds of kelvins, and "hot" clusters of this kind can be detected by analyzing thermal radiation spectra [32].

Note that T_c depends on the enthalpy of vaporization $h'(N)$ per molecule released from a cluster (see Eq. (5) and Fig. 1). Figure 1 demonstrates that the function $h'(N)$ has a sharp maximum. The maximum value of $h'(N)$ is close to the enthalpy of sublimation (shown by curve *a* in Fig. 1). This implies the existence of a range of cluster size in which clusters are characterized by higher stability with respect to vaporization of molecules. We call such clusters hyperstable. According to our computations, the presence of hyperstable clusters manifests itself by peaks in curves of $T_c(N)$ at $N \approx 20$ and 16 for $(N_2O)_N$ and $(CO_2)_N$ clusters, respectively. The temperature T_c corresponding to $N > 30$ in the beam far from the nozzle depends weakly on N; in particular, $T_c = 94 \pm 2$ and 105 ± 2 K for $(N_2O)_N$ and $(CO_2)_N$ clusters, respectively. Electron diffraction measurements have shown that the nanosized $(CO_2)_N$ clusters produced in a jet are solid and their temperature is $T_c = 108 \pm 10$ K [33], which agrees well with the values of T_c computed and $h(N)$ measured in the present study (see Figs. 1 and 2).

In our computations, the decrease in N due to vaporization of molecules does not exceed one percent when the cluster residence time in the beam is about 10^{-3} s and the temperature of clusters entering the skimmer is about 1000 K. Under these conditions, T_c decreases to 78 and 88 K for $(N_2O)_{10}$ and $(N_2O)_{100}$ clusters, respectively. In the general case, the effect of radiative heat transfer on N and T_c may also play an important role. The values of N and T_c for weakly bound low-temperature clusters can be affected by the radiation coming from the heated walls of the vacuum chamber. When dealing with strongly bound atomic clusters characterized by a high initial temperature, one must take into account the radiation emitted by these clusters. In particular, we detected the thermal radiation of C_{60} clusters and found that the corresponding rate of radiative loss is proportional to T_c^5 [34].

Important characteristics of gas condensation and clustering in a jet include the cluster size distributions in the jet and beam, $f_j(N)$ and $f(N)$, respectively. These functions are required to obtain correct cross sections for various processes involving clusters [35].

Figure 3 shows $f_j(N)$ (curves *1–3*) and $f(N)$ (curves *3*' and *4*) for $(N_2O)_N$ clusters. Curves *1–3* and *3*' were obtained for $T_0 = 285$ K, $d_* = 1$ mm, and various values of x/d_* and P_0. Curves *1* and *2* correspond to $x/d_* = 3$ and 100, respectively, at $P_0 = 3.7 \times 10^4$ Pa. Curves *3* and *3*' were obtained for $P_0 = 8.5 \times 10^4$ Pa at $x/d_* = 100$ and at a large distance from the skimmer (i.e., for $\varphi \ll 1$ in the beam, see Eq. (9)), respectively. Curve *4* is the bell-shaped $f(N)$ measured in [36] by means of mass spectrometry at $P_0 = 3.2 \times 10^5$ Pa, $T_0 = 215$ K, and $d_* = 0.35$ mm.

Curve *1* in Fig. 3 demonstrates that $f_j(N)$ is a monotonically decreasing function of N near the nozzle exit. This behavior is characteristic of steady-state condensation [23]. As the distance from the nozzle increases, the curve of $f_j(N)$ tends to become bell-shaped and exhibits a small local maximum when $\bar{N}_j < 60$ (see curve *2*). The bimodal form of curve *2* is explained by the presence of hyperstable clusters (see curve *5* in Fig. 1). With increasing $\bar{N}_j$, the bell-shaped portions of $f_j(N)$ and $f(N)$ tend to have a more regular form (see

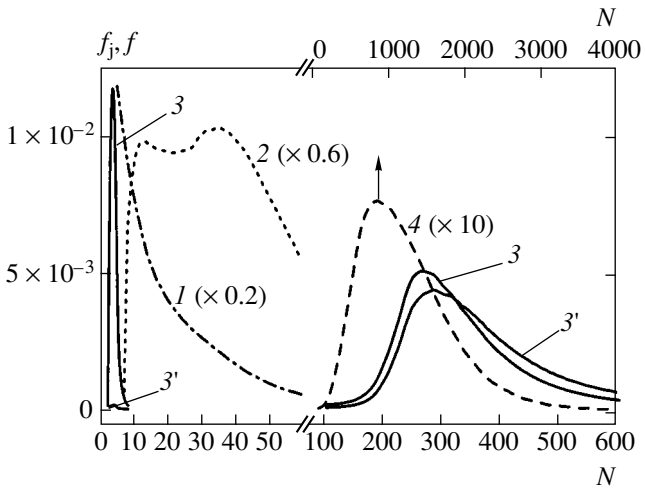


Fig. 3. Size distribution functions for $(N_2O)_N$ clusters computed at $x/d_* = 3$ (*1*) and $x/d_* = 100$ (*2*, *3*) for $P_0 = 3.7 \times 10^4$ (*1*, *2*) and $P_0 = 8.5 \times 10^4$ Pa (*3*, *3*') on the jet axis (*1*–*3*) and in the beam (*3*'). Curve *4* is $f(N)$ measured in an $(N_2O)_N$ beam [36].

curves *3* and *3*' in Fig. 3). However, the functions $f_j(N)$ and $f(N)$ tend to have "tails" corresponding to small clusters (see curves *3* and *3*'). This change in $f_j(N)$ and $f(N)$ correlates with the behavior of I, ε, T_c (see Fig. 2), and gas density as functions of x/d_*. Figure 2 shows that I vanishes with increasing x/d_* while $q_j/\bar{N}_j \approx$ const. This means that the cluster size increases, whereas the number of clusters remains nearly constant. The condensation nuclei that form far from the nozzle cannot grow because the ambient gas density rapidly decreases. These condensation nuclei are responsible for the tails of $f_j(N)$ and $f(N)$. Comparing the tails of $f_j(N)$ and $f(N)$ (curves *3* and *3*' in Fig. 3), one finds that the tail almost disappears in the beam and the function $f(N)$ shifts toward larger N. These changes are caused by the increase in the concentration of heavier clusters due to their narrower transverse spread (the transverse particle velocity in the beam is $v_\perp \propto (mN)^{-1/2}$). Expression (9) implies that $f(N) \propto Nf_j(N)$ and the increase in $\bar{N}$ depends on the width of $f_j(N)$: $\bar{N} - \bar{N}_j = \delta_j \bar{N}_j$, where δ_j is the relative variance. According to our calculations of $\delta_j(N_j)$, the value of δ_j is a nonmonotonic function varying between 0.3 and 0.7 at $x/d_* = 100$ [7, 8].

The behavior of the computed $f(N)$ agrees with experimental observations [2, 36, 37]. However, they were interpreted without taking into account the evolution of $f_j(N)$ in jets. In particular, the model proposed in [37] explained the development of a bell-shaped $f(N)$ by coalescence of small clusters in a jet. However, even if the probability of cluster–cluster aggregation in collisions is high [38], the effect of coalescence on $f(N)$ is weak because the frequency of collisions between them in a jet is low:

$$\nu_{N-N} \approx \frac{\nu_N N^{1/6} q_j}{\bar{N}}.$$

Aggregation of condensation nuclei cannot cause any significant increase in cluster size either. Because h strongly depends on N (see Fig. 1), cluster–cluster aggregation results in a sharp increase in T_c, which stimulates vaporization. Our estimates show that coalescence does not lead to a faster increase in $\bar{N}$, as compared to molecule-by-molecule condensation.

Note that the small-cluster tail in a measured $f(N)$ should mainly be attributed to experimental conditions. For example, a tail is observed when $f(N)$ is measured at the skimmer exit, i.e., in the jet (see curves *3* and *3*' in Fig. 3). In mass spectrometry of clusters ionized by electron impact, the tail appears as a result of cluster fragmentation and ejection of small cluster ions from large clusters [36, 39, 40].

In Fig. 3, the width of a computed $f(N)$ (curve *3*') is noticeably smaller than that of a measured one (curve *4*). This is an expected discrepancy, because the additional jet expansion due to the heat release associated with cluster nucleation was not taken into account in the gasdynamic part of our computations.

We also performed a comparative numerical analysis of cluster properties in a jet and in a beam, using the simplified approach developed in [22]. It was assumed in [22] that the cluster growth rate scales with $N^{2/3}$, i.e., with the cluster surface area. Under this assumption, $f_j(N)$ can be disregarded and computations are simplified. We found that this simplification leads to poor agreement between computed and measured results, which is explained by the complex form of the function $f_j(N)$ (see Fig. 3). For this reason, the results presented below were obtained in computations taking into account the evolution of $f_j(N)$.

Beam intensity is the most reliably measured characteristic of cluster beams [41], because its measurement does not require primary ionization of clusters. Expression (9) implies that beam intensity is sensitive to the onset and development of condensation in a jet. Accordingly, we compared the computed and measured results in terms of J as a function of P_0 and T_0 in the first place. Note that weakly bound molecular clusters characterized by a wide range of $\bar{N}$ must be produced in a large-scale apparatus providing a high rate of evacuation of the gas issuing from the nozzle. Our experiments were conducted in a large-scale beam generator equipped with cryogenic vacuum pumps [42], which made it possible to study clustering of molecular gases in a wide range of $\bar{N}$ for various P_0, T_0, and d_* [18, 25, 39, 43]. In what follows, the data obtained by using this apparatus are compared with computed results.

In Fig. 4, the solid curves of q_j and M were computed at $x/d_* = 100$ for N_2O jets, whereas the solid curves of J, J_m, $\bar{N}$, and condensate fraction $q = (J - J_m)/J$ correspond to N_2O beams at a large distance from the skimmer (where $\varphi \ll 1$, see (9)). The dashed curve is a measured $J(P_0)$. These results were obtained for $T_0 = 285$ K and $d_* = 1$ mm. The leftmost portions of the $J(P_0)$ curves are merged in Fig. 4. Since similar curves were obtained for CO_2, we present only the dash-and-dot curve of $\bar{N}(P_0)$ computed for CO_2 at $T_0 = 300$ K and $d_* = 1.9$ mm. In Fig. 4, open and closed symbols represent the mean cluster size measured for $(N_2O)_N$ and $(CO_2)_N$, respectively. Symbols labeled *1*, *2*, and *3* correspond, respectively, to the $(N_2O)_N^+$, $(CO_2)_N^+$, and $(N_2O)_N^-$ ion sizes measured with a retarding potential analyzer [35] (the ions were created in the region of intersection of cluster and electron beams). Symbols labeled *4* represent the lowest pressure at which $(N_2O)_2$ and $(CO_2)_2$ dimers were detected by means of mass spectrometry in [18] and [42], respectively. The results of measurements performed by other authors are shown by symbols *5*–*9* in Fig. 4. These data were matched with computations (in terms of P_0, T_0, and d_*) by using the scaling parameters for condensation downstream of a sonic nozzle considered below. The data represented by symbols *5* and *6* were obtained in the mass spectrometric studies of $(N_2O)_N^+$ and $(CO_2)_N^+$ ion currents reported in [36] and [37], respectively. Symbol *7* denotes the values of $\bar{N}$ obtained by measuring the $(CO_2)_N^+$ ion current with a retarding potential analyzer [44]. Symbols *8* and *9* correspond to the electron diffraction measurements of $\bar{N}$ for $(CO_2)_N$ clusters reported in [33] and [45], respectively. The fact that the computed values of $\bar{N}$ systematically exceed the measured ones is explained as follows. In experiments, inelastic collisions of clusters with electrons reduce N because of cluster evaporation and fragmentation [2, 36, 39]. Size $\bar{N}$ was overestimated since the computations did not take into account the decrease in the overall molecular collision frequency caused by the additional jet expansion due to vibrational relaxation in molecule–cluster collisions and heat release by condensation. Nevertheless, the computed values of $\bar{N}$ are in fair agreement with the measured ones. Moreover, the computed results agree with experiment in terms of behavior of $J(P_0)$. This observation will be discussed now in some detail.

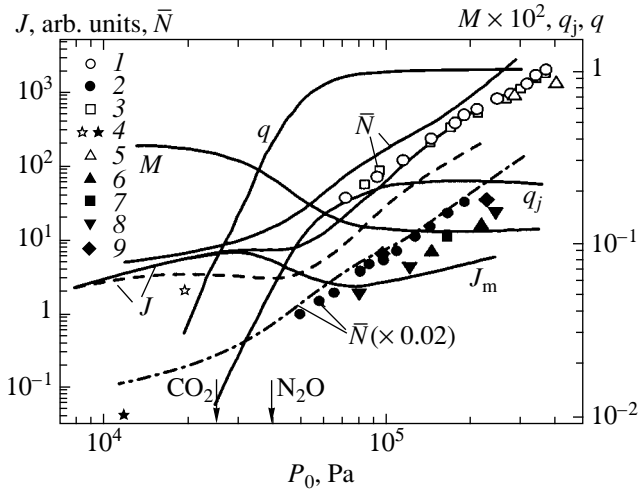


Fig. 4. Mach number M, condensate fractions q_j (in jets) and q (in beams), total beam intensity J, intensity J_m of the free-molecule beam component, mean cluster size $\bar{N}$ in N_2O beam (solid curves), and mean $(CO_2)_N$ cluster size (dash-and-dot curve) computed as functions of nozzle source pressure. Also shown are $J(P_0)$ measured for N_2O beams (dashed curve) and $\bar{N}$ measured for $(N_2O)_N$ and $(CO_2)_N$ clusters (open and closed symbols, respectively). Vertical arrows indicate the threshold pressures for well-developed condensation in N_2O and CO_2 jets.

Both the bending of curves in Fig. 4 and the decrease in $J_m(P_0)$ are associated with gas condensation in a jet. An increase in $q_j(P_0)$ leads to a rise of gas temperature in the jet (see Fig. 2), which reduces the Mach number (see Fig. 4). The decrease in $M(P_0)$ in the interval of P_0 where $\bar{N}$ is small reduces the slope of $J(P_0)$ and causes $J_m(P_0)$ to decrease (see (9)). According to (9), intensity increases with $\bar{N}$ as $q_j \bar{N}_j M$. Our computations showed that the bending curves of $J(P_0)$ flatten out for a gas characterized by a low latent heat of condensation and a slowly varying $\bar{N}(P_0)$, in agreement with experiment [46]. The fact that the computed J is higher than the measured one is explained by the use of simplifying assumptions in the gasdynamic part of the computations. The numerical model ignores the "background" deceleration and scattering of molecules due to their interactions with those scattered by the skimmer and those penetrating into the low-density parts of the jet and the beam from the vacuum chamber [3]. These effects are more pronounced for lower M, i.e., for higher rates of heat release by condensation. Clusters are less sensitive to background deceleration as compared to molecules because of difference in mass. Therefore, the measured $J_m(P_0)$ decreases more rapidly, as compared to the computed one, with increasing gas flow rate (i.e., with increasing background gas density in the vacuum chamber). The resulting curve of $J(P_0)$ has a sharp minimum [3, 47], and the corresponding beam tends to contain more clusters as P_0 increases. In effect, the curve of $q(P_0)$ becomes even steeper (see Fig. 4). Computations are further simplified by using

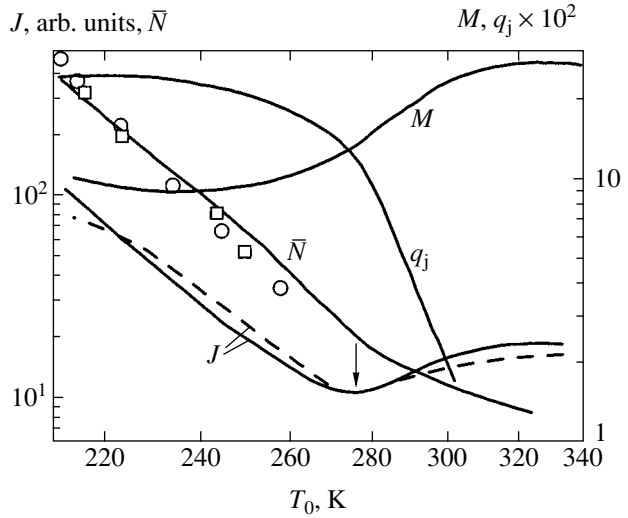


Fig. 5. Mach number M, condensate fraction q_j in jets, total beam intensity J, and mean cluster size $\bar{N}$ in N_2O beams (solid curves) computed as functions of nozzle source temperature. Also shown are $J(T_0)$ (dashed curve) and $\bar{N}$ (symbols corresponding to *1* and *3* in Fig. 4) measured for N_2O beams. The vertical arrow indicates the threshold temperature for well-developed condensation in N_2O jets.

the Mach number $M_c(N) = \sqrt{N}M$ in (9) for all clusters entering the skimmer, with M calculated by assuming continuum-flow expansion upstream of the skimmer. However, a real jet flow characterized by a low value of $P_0 d_*$ involves a transition to collisionless gas expansion and "frozen" M at a certain $x_f/d_* = \bar{x}_f$. The ensuing effect on the properties of a free-molecule beam was analyzed in [10]. In cluster beams, the value of $\bar{x}_f(N)$ decreases with increasing N. This explains why the downstream acceleration of particles due to heat release by condensation was observed only for molecules or clusters with $N \le 4$ [46]. The cluster velocity U does not increase when the cluster size is large [48]. The effects described here affect the measured $J(P_0)$ and explain their disagreement with the computed curves shown in Fig. 4.

We set $\gamma = 1.4$ in computing the expansion downstream of the nozzle; i.e., we assumed that the vibrational degrees of freedom of molecules were frozen. This assumption is valid when expansion is not associated with condensation [9]. However, when the degree of gas supersaturation in a jet reaches a critical level, one must allow for transfer of vibrational energy to a cluster in almost every molecule–cluster collision and subsequent rapid thermalization of the transferred energy in the cluster [11–13]. According to the experimental observations reported in [16], even a low degree of vibrational excitation impedes the formation of condensation nuclei in a jet and a higher gas density in the nozzle source is required to reach the condensation threshold. Instead of using the size-dependent rates of V–T relaxation in molecule–cluster collisions obtained for N_2O and CO_2 [13] in our condensation computations performed for various T_0, we relied on a simpler computational procedure. Good agreement between numerical results and experiment was achieved when a change in T_0 was combined with a certain correction of the dimer binding enthalpy h_2 in (8): when T_0 was increased, the enthalpies h_2 for $(N_2O)_2$ and $(CO_2)_2$ were reduced relative to their values at $T_0 = 285$, and vice versa. This procedure took into account the impeding effect of vibrational energy on the formation of condensation nuclei and cluster growth. Good agreement of computed functions $J(T_0)$ and $\bar{N}(T_0)$ with experiment was achieved when the change in h_2 was equal to the change in the specific vibrational enthalpy of gas in the nozzle source. This is illustrated by the numerical results presented in Fig. 5.

Solid curves in Fig. 5 represent the beam intensity J, the mean cluster size $\bar{N}$, the Mach number M, and the condensate fraction q_j as functions of temperature at $x/d_* = 100$. Both computed and measured $J(T_0)$ are shown in arbitrary units, and their minima are matched. The existence of a minimum in $J(T_0)$ is attributed to the mechanisms responsible for the sharp increase in the slope of the $J(P_0)$ curves in Fig. 4. Figure 5 also shows the values of $\bar{N}$ corresponding to our experimental results represented by symbols *1* and *3* in Fig. 4. The values of T_0 used to compute the curves of $\bar{N}$ plotted here were obtained by invoking a scaling parameter for gas condensation in jets issuing from sonic nozzles [49]. The constants involved in the scaling parameter are found in the present study (see below). The computed results are in good agreement with experiment. The slightly higher values of the measured $\bar{N}(T_0)$ at low T_0, as compared to the numerical results, are explained by the lower accuracy of the scaling parameter as applied to a wider interval of T_0 (see below).

The parameter values $P_0 = P_{0,c}$ and $T_0 = T_{0,c}$ corresponding to $J(P_0, T_0)$ increasing with the condensate fraction in the jet (indicated by arrows in Figs. 4 and 5) were called the threshold parameters for well-developed condensation in the jet [18]. Performing computations for various P_0, T_0, and d_*, we obtained relations between $P_{0,c}$, $T_{0,c}$, and d_* that are consistent with experimental results and with the scaling parameter proposed in [49]:

$$A_c = P_{0,c} T_{0,c}^{-\alpha} d_*^{\beta}, \qquad (10)$$

where A_c, α, and β are empirical constants. Varying P_0, T_0, and d_*, we found that $\beta = 0.7$ and 0.6, $\alpha = 4.85$ and 4.45, $Z_c = 5.2 \times 10^{-8}$ Pa mm$^{0.7}$ K$^{-4.85}$ and 3.5×10^{-7} Pa mm$^{-0.6}$ K$^{-4.45}$ for N_2O and CO_2, respectively. It

can readily be shown that scaling parameter (10) is associated with a constant molecular collision frequency in a jet under conditions of gas supersaturation [3]. Therefore, Eq. (10) provides an accurate relation between the condensate characteristics obtained for different values of P_0, T_0, and d_*.

According to (10), the threshold parameters for well-developed condensation obey the quasi-isentropic relation

$$P_0 T_0^{-\gamma_0/(\gamma_0-1)} = P_j T_j^{-\gamma_0/(\gamma_0-1)}$$

between gas characteristics in the nozzle source and those in the jet. Therefore, one can define a quasi-isentropic exponent γ_{eff} for expansion of a condensing gas by the relation $\alpha = \gamma_{eff}/(\gamma_{eff} - 1)$. Computations and experiments in which T_0 and d_* were varied from 240 to 320 K and from 0.5 to 5.0 mm, respectively, showed that $\gamma_{eff} = 1.26$ for N_2O and 1.29 for CO_2. For these gases, $\gamma_0 \approx 1.3$ at $T = 290$ K [50]. The fact that the values of γ_{eff} were found to be slightly lower than γ_0 is explained by the heat release associated with the condensate nucleation and vibrational relaxation preceding the onset of well-developed condensation as P_0 is increased. Because of the lower slope of the P–T phase transition curve for N_2O as compared to CO_2, the supersaturation conditions in the N_2O jet are reached at a higher gas density in the nozzle source. Accordingly, gaseous N_2O carries a greater initial amount of vibrational energy, as compared to CO_2, and $\gamma_{eff}(N_2O) < \gamma_{eff}(CO_2)$. It should also be noted that the constants in (10) are valid for a narrow interval of T_0. This clearly explains why experimental data lie above the computed $\bar{N}(T_0)$ curve in Fig. 5.

4. CONCLUSIONS

Overall, the proposed computational model based on a realistic physical characterization of clusters provides a basis for acquiring comprehensive and reliable information about the mechanisms of cluster formation and growth in supersonic free jets of molecular gases. The model can be used to find relations between gas characteristics in the nozzle source, the jet, and the beam formed from an axial part of the jet.

A natural extension of the model of gas condensation in supersonic jets should rely on a more accurate description of the nucleation kinetics. The inherent difficulties of the classical kinetic theory (which ignores temperature dependence, nonequilibrium effects, and discrete variation of size) can be dealt with by simulating directly a sequence of elementary acts of condensation of C molecules to a C_N cluster ($C_N + C \longleftrightarrow C_{N+1}$). The number of such reactions is restricted because of the small size of condensation nuclei, which will simplify their simulation.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project no. 01-02-17372.

REFERENCES

1. *Atomic and Molecular Beams—the State of the Art 2000*, Ed. by R. Campargue (Springer, Berlin, 2001).
2. C. Bobbert, S. Schuette, C. Steinbach, *et al.*, Eur. Phys. J. D **19**, 183 (2002).
3. A. A. Vostrikov, Yu. S. Kusner, A. K. Rebrov, *et al.*, Prikl. Mekh. Tekh. Fiz., No. 2, 34 (1975).
4. Ya. I. Frenkel, *Kinetic Theory of Liquids* (Nauka, Leningrad, 1975; Dover, New York, 1955).
5. J. Feder, J. P. Hirth, J. Lothe, *et al.*, in *Heterogeneous Combustion*, Ed. by H. G. Wolfhard, I. Glassman, and L. Green, Jr. (Academic, New York, 1964; Mir, Moscow, 1967).
6. A. A. Lushnikov and A. G. Sutugin, Usp. Khim. **45**, 385 (1976).
7. A. A. Vostrikov and D. Yu. Dubov, Pis'ma Zh. Tekh. Fiz. **10**, 31 (1984) [Sov. Tech. Phys. Lett. **10**, 13 (1984)].
8. A. A. Vostrikov and D. Yu. Dubov, Preprint No. 112-84, IT SO AN SSSR (Inst. of Thermal Physics, Siberian Division, USSR Academy of Sciences, Novosibirsk, 1984).
9. A. E. Beylich, Phys. Fluids **14**, 898 (1971).
10. A. L. Stasenko, Inzh.–Fiz. Zh. **16**, 9 (1969).
11. A. A. Vostrikov, S. G. Mironov, and B. E. Semyachkin, Zh. Tekh. Fiz. **53**, 81 (1983) [Sov. Phys. Tech. Phys. **28**, 46 (1983)].
12. A. A. Vostrikov and S. G. Mironov, Zh. Tekh. Fiz. **54**, 417 (1984) [Sov. Phys. Tech. Phys. **29**, 251 (1984)].
13. A. A. Vostrikov and S. G. Mironov, Chem. Phys. Lett. **101**, 583 (1983).
14. A. A. Vostrikov, D. Yu. Dubov, and V. P. Gilyova, Z. Phys. D **20**, 205 (1991).
15. A. A. Vostrikov and V. P. Gileva, Pis'ma Zh. Tekh. Fiz. **20** (15), 40 (1994) [Tech. Phys. Lett. **20**, 625 (1994)].
16. I. M. Beterov, Yu. V. Brzhazovskiĭ, A. A. Vostrikov, *et al.*, Kvantovaya Élektron. (Moscow) **7**, 2443 (1980) [Sov. J. Quantum Electron. **10**, 1422 (1980)].
17. A. A. Vostrikov, S. G. Mironov, A. K. Rebrov, *et al.*, Zh. Tekh. Fiz. **49**, 2680 (1979) [Sov. Phys. Tech. Phys. **24**, 1513 (1979)].
18. A. A. Vostrikov, Zh. Tekh. Fiz. **54**, 327 (1984) [Sov. Phys. Tech. Phys. **29**, 191 (1984)].
19. R. D. Etters, K. Flurchik, R. P. Ran, *et al.*, J. Chem. Phys. **75**, 929 (1981).
20. Yu. A. Alekseev, V. N. Bogomolov, and V. A. Egorov, Pis'ma Zh. Éksp. Teor. Fiz. **36**, 384 (1982) [JETP Lett. **36**, 463 (1982)].
21. S. V. Drozdov and A. A. Vostrikov, Pis'ma Zh. Tekh. Fiz. **26** (9), 81 (2000) [Tech. Phys. Lett. **26**, 397 (2000)].
22. L. M. Davydov, Izv. Akad. Nauk SSSR, Mekh. Zhidk. Gaza, No. 3, 66 (1971).
23. E. M. Lifshitz and L. P. Pitaevskiĭ, *Physical Kinetics* (Nauka, Moscow, 1979; Pergamon Press, Oxford, 1981).

24. F. O. Goodman and H. Y. Wachman, *Dynamics of Gas–Surface Scattering* (Academic, New York, 1976; Mir, Moscow, 1980).
25. A. A. Vostrikov and D. Yu. Dubov, Z. Phys. D **20**, 429 (1991).
26. A. A. Vostrikov and D. Yu. Dubov, J. Aerosol Sci. **24**, S175 (1993).
27. A. A. Vostrikov and D. Yu. Dubov, Zh. Tekh. Fiz. **64** (12), 137 (1994) [Tech. Phys. **39**, 1277 (1994)].
28. W. G. Dorfeld and J. B. Hudson, J. Chem. Phys. **59**, 1253 (1973).
29. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 5: *Statistical Physics*, 3rd ed. (Nauka, Moscow, 1976; Pergamon Press, Oxford, 1980), Part 1.
30. V. A. Zhokhov and A. A. Khomutskiĭ, *Atlas of Supersound Flows of Free-Expanding Ideal Gas Flowing out an Axis-Symmetric Jet* (TsAGI, Moscow, 1966).
31. H. Ashkenas and F. Sherman, in *Proceedings of 4th International Symposium on Rarefied Gas Dynamics* (Academic, New York, 1966), Vol. 2, p. 84.
32. A. A. Vostrikov, D. Yu. Dubov, and A. A. Agarkov, Pis'ma Zh. Éksp. Teor. Fiz. **63**, 915 (1996) [JETP Lett. **63**, 963 (1996)].
33. G. Torchet, H. Bouchier, J. Farges, *et al.*, J. Chem. Phys. **81**, 2137 (1984).
34. A. A. Vostrikov, D. Yu. Dubov, and A. A. Agarkov, Teplofiz. Vys. Temp. **39**, 26 (2001) [High Temp. **39**, 22 (2001)].
35. A. A. Vostrikov, D. Yu. Dubov, and M. R. Predtechenskiĭ, Preprint No. 150-86, IT SO AN SSSR (Inst. of Thermal Physics, Siberian Division, USSR Academy of Sciences, Novosibirsk, 1986).
36. O. Echt, M. Knapp, K. Sattler, *et al.*, Z. Phys. B **53**, 71 (1983).
37. J. M. Soler, N. Garcia, O. Echt, *et al.*, Phys. Rev. Lett. **49**, 1857 (1982).
38. S. V. Drozdov and A. A. Vostrikov, Pis'ma Zh. Tekh. Fiz. **26** (11), 90 (2000) [Tech. Phys. Lett. **26**, 493 (2000)].
39. A. A. Vostrikov, D. Yu. Dubov, and M. R. Predtechenskiĭ, Zh. Tekh. Fiz. **57** (4), 760 (1987) [Sov. Phys. Tech. Phys. **32**, 459 (1987)].
40. A. A. Vostrikov, D. Yu. Dubov, and V. P. Gileva, Zh. Tekh. Fiz. **59** (8), 52 (1989) [Sov. Phys. Tech. Phys. **34**, 872 (1989)].
41. A. A. Vostrikov, Yu. S. Kusner, A. K. Rebrov, *et al.*, Prib. Tekh. Éksp., No. 1, 177 (1975).
42. A. A. Vostrikov, S. G. Mironov, and B. E. Semyachkin, Fluid Mech.–Sov. Res. **11** (4), 98 (1982).
43. A. A. Vostrikov, D. Yu. Dubov, and I. V. Samoĭlov, Zh. Tekh. Fiz. **64** (12), 120 (1994) [Tech. Phys. **39**, 1267 (1994)].
44. O. F. Hagena and W. Obert, J. Chem. Phys. **56**, 1793 (1972).
45. G. D. Stein and J. A. Armstrong, J. Chem. Phys. **58**, 1999 (1973).
46. D. Golomb, R. E. Good, A. B. Balley, *et al.*, J. Chem. Phys. **57**, 3844 (1972).
47. A. A. Vostrikov, V. V. Grigor'ev, D. Yu. Dubov, *et al.*, Pis'ma Zh. Tekh. Fiz. **18** (19), 25 (1992) [Sov. Tech. Phys. Lett. **18**, 620 (1992)].
48. D. Dreyfuss and H. Y. Wachman, J. Chem. Phys. **76**, 2031 (1982).
49. O. F. Hagena, Z. Phys. D **4**, 101 (1987).
50. *Handbook of Chemistry and Physics*, Ed. by C. D. Hodgman (Rubbers, Cleveland, USA, 1960).

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 98, No. 2, 2004, pp. 207–219.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 2, 2004, pp. 233–246.

ATOMS, SPECTRA, RADIATION

Nonergodic Nuclear Depolarization in Nanocavities¶

E. B. Fel'dman and M. G. Rudavets

Institute of Problems of Chemical Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia

e-mail: feldman@icp.ac.ru

Received June 17, 2003

Abstract—Recently, it has been observed that the effective dipolar interactions between nuclear spins of spin-carrying molecules of gas in closed nanocavities are independent of the spacing between all the spins. We derive an exact time-dependent polarization for all spins in the spin-1/2 ensemble with spatially independent effective dipolar interactions. If the initial polarization is on a single (first) spin, $P_1(0) = 1$, then the exact spin dynamics of the model exhibits periodic short pulses of the polarization of the first spin, typical of systems having a large number N of spins. If $N \gg 1$, then within the period $4\pi/g$ ($2\pi/g$) for odd (even) N-spin clusters, with g standing for spin coupling, the polarization of spin 1 switches quickly from unity to the time-independent value of 1/3 over a time interval of about $(g\sqrt{N})^{-1}$. Thus, spin 1 spends almost the entire time in the time-independent condition $P_1(t) = 1/3$. The period and the width of the pulses determine the volume and the form factor of the ellipsoidal cavity. The formalism is adapted to the case of time-varying nanofluctuations of the volume $V(t)$ of cavitating nanobubbles. If the coupling $g(V(t))$ is varied by the Gaussian-in-time random noise due to the variation of the volume $V(t)$, then the envelope of the polarization peaks passes irreversibly to 1/3. The polarization dynamics of a single spin exhibit a Gaussian (exponential) time dependence when the correlation time of fluctuations of the nanovolume is larger (smaller) than $\langle(\delta g)^2\rangle^{-1/2}$, where $\langle(\delta g)^2\rangle$ is the variance of the $g(V(t))$ coupling. Finally, we report exact calculations of the NMR line shape for the N-spin gaseous aggregate. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The nature of ergodicity as a fundamentally important element for a consonant description of statistical mechanics is currently being discussed in the context of NMR [1]. Spin dynamics is ergodic if the initial polarization prepared at a single (first) spin is spread over the system, leading over time to the spatially uniform distribution of polarization, as expected on the basis of simple physics. On the other hand, nonergodic behavior, which was recently observed numerically in the nuclear spin-1/2 1*D* chains with the general *XYZ* spin Hamiltonian [2], enters such that the time-averaged polarization of the first spin turns out to be several times larger than the time-averaged polarization of any other spin in the chain. This observation of nonergodicity has been extended to 1*D* chains and rings with the *XY* Hamiltonian [3], demonstrating analytically that the time-averaged polarization of the first spin differs by a factor of 1.5–2 from the time-averaged polarization of all other spins in the chain. These considerations in 1*D* spin clusters address the problem of the nature of ergodicity for different spin Hamiltonians. Motivated by the study of nonergodic spin dynamics and because an exact solution is a fortuitous exception in statistical mechanics, we assume in this paper that spin interactions can be considered independent of the spacing between the spins rather than having an r^{-3} dependence.

Recently, a spin Hamiltonian with space-independent spin couplings has been applied for exploring the NMR spectra of the gas of spin-carrying molecules undergoing fast thermal motion within nonspherical cavities [4]. In that report, the authors arrived at the space-independent effective spin couplings by motionally averaging the exact dipolar Hamiltonian over uniformly distributed spatial coordinates of the spins in nanometer-size cavities. This technique is expected to have promising application for determining pore shapes and sizes [5] by NMR spectra.

With regard to the effective nuclear spin Hamiltonian with infinite-range couplings, it is noteworthy that this type of interactions has also been proposed in the theory of nanoelectrodes [6, 7]. There, the infinite-range dipolar nuclear interactions are induced indirectly due to the fast energy transfer between the electron and nuclear spins. On the coarse-grain time scale of the fast electron spin dynamics, the slow effective nuclear spin dynamics is governed by an effective nuclear spin Hamiltonian with infinite-range interaction. Quite apart from its importance as a physical model in NMR experiments for the many-spin aggregate in a confined volume [4, 6, 7] and few proton molecules [8], the model with infinite-range spin interac-

¶This article was submitted by the authors in English.

tions is of fundamental interest in its own right because it makes it possible to treat the three-dimensional case exactly, without any reference to 1*D* spin ordering. It represents the quantum nonequilibrium version of the exactly solvable equilibrium spin model [9], has a mapping to the BCS pairing Hamiltonian of superconductivity [10], and has long been considered a test for many-body problems in higher spatial dimensions, $D \gg 1$. The objective of this paper is to present the exact solution of nonergodic dynamics with an infinite-range spin Hamiltonian in the *N*-spin ensemble.

To our knowledge, the only result reported on this model is that of Waugh [11], who announced (without proof) that the time-averaged polarization of the first spin is exactly equal to $(N + 2)/3N$ and the polarization of any other spin is exactly $2/3N$ for an odd numbered, *N*, spin cluster. To clarify the problem of spin dynamics, the present paper reports a detailed analytic theory on the average polarization for both odd- and even-numbered spin clusters; it also gives a theory on spin dynamics that is entirely missing in [11]. A condensed form of this paper was published in [12]. A brief overview of the present paper is as follows. In Section 2, we construct the effective nuclear spin Hamiltonian of spin-carrying molecules in a nanocavity. Section 3 gives the formalism required to obtain the exact time-dependent polarization. This is followed by Section 4, which discusses three issues of the polarization dynamics that are amenable to the techniques of Section 2: the nonergodicity of the polarization dynamics of a single spin in the nanocavity, the polarization dynamics of a single spin within a fluctuating nanobubbles, and the spectral line shape of the nuclear spin ensemble. Finally, Section 5 summarizes the results of the calculations and compares the results obtained with the known analytic results for the *XY* Hamiltonian.

2. EFFECTIVE NUCLEAR SPIN HAMILTONIAN IN A NANOCAVITY

The purpose of this section is to construct an effective spin Hamiltonian $\bar{H}$ that governs spin dynamics of spin-carrying molecules in a nanosized cavity on the coarse-grain temporary scale on the order of 10 ps. At these space-time scales, the effective spin Hamiltonian differs from the exact dipolar Hamiltonian; in particular, the many-body spin Hamiltonian $\bar{H}$ has a high symmetry that permits the exact solution for the spectrum and, as a result, the exact derivation of the polarization dynamics of the gas within the nanocavity. In this section, we summarize the main ideas of [4]; however, in deriving the effective spin Hamiltonian $\bar{H}$ by averaging over spin spatial coordinates, we generalize the effective spin coupling to the case of a nonperfect gas in the nanocavity.

The starting point of the derivation of operator $\bar{H}$ is the expression

$$\begin{aligned}&\rho(t, \{\mathbf{I}_n, \mathbf{r}_n(t), \mathbf{p}_n(t)\}_{n=1}^N)\\ &= U(t)\rho(0, \{\mathbf{I}_n, \mathbf{r}_n(0), \mathbf{p}_n(0)\}_{n=1}^N)U^{-1}(t),\end{aligned} \tag{1}$$

for the density matrix with completely specified coordinates $\{\mathbf{r}_n(t)\}_{n=1}^N$ and momenta $\{\mathbf{p}_n(t)\}_{n=1}^N$ of *N* spin-carrying molecules. The propagator $U(t)$ is associated with the time-dependent exact dipolar Hamiltonian (in frequency units)

$$\begin{aligned}H(t) &= \sum_{1 \le i \le j}^{N} h_{i,j}(t),\\ h_{i,j}(t) &= \gamma^2 \hbar P_2(\cos\theta_{ij}(t)) r_{ij}^{-3}(t)(\mathbf{I}_i\mathbf{I}_j - 3I_{iz}I_{jz}),\end{aligned} \tag{2}$$

where γ stands for the gyromagnetic ratio, $I_{n\alpha}$ ($\alpha = x, y, z$) specify the spin-1/2 operators, and $\theta_{ij}(t)$ is the instant polar angle between the vector $r_{ij}(t)$ from $r_i(t)$ to $r_j(t)$ and the external magnetic field *B*.

A cornerstone fact for the construction of an effective spin Hamiltonian is the significant difference between the time scale of the relaxation in the phase space $r^N - p^N$ and the time scale of the spin dynamics under the Hamiltonian in Eq. (2). Actually, for hydrogen gas at room temperature and atmospheric pressure, the following estimations hold. The average concentration

$$\bar{n} \approx 2.7 \times 10^{19} \text{ molecules/cm}^3,$$

and the mean free path

$$\lambda = (\bar{n}\pi a^2)^{-1} \approx 10^{-4} \text{ cm}$$

for the radius of a molecule $a \approx 10^{-8}$ cm and the thermal velocity $\bar{v} \approx 10^5$ cm/s. Then, a simple order-of-magnitude calculation leads us to expect that for a gas in the cavity with a size of $\ell \approx 10$ nm, the Knudsen diffusion coefficient is

$$\mathcal{D} \approx \bar{v}\,\ell \approx 10^{-1} \text{ cm}^2/\text{s},$$

the characteristic time scale of the spatial relaxation of the gas (due to the diffusive reflection from the wall of the pore) to the spatially homogeneous distribution is

$$t_{\text{dif}} \approx \ell^2/\mathcal{D} \approx 10^{-11} \text{ s},$$

and the characteristic time scale of the velocity relaxation towards the Maxwell distribution is

$$t_v \approx \ell/\bar{v} \approx 10^{-11} \text{ s}.$$

These time scales t_v and t_{dif} are well separated from the NMR time scale $t_{nmr} = 10^{-4}$–10^{-3} s associated with the

dipolar interaction in Eq. (2). The smallness of the parameter

$$\varepsilon = \frac{t_{\rm rel}}{t_{nmr}} = 10^{-7} \ll 1, \quad (3)$$

where

$$t_{\rm rel} = \max(t_v, t_{\rm dif}),$$

allows one to determine the average nuclear spin Hamiltonian governing the behavior of the nuclear spins over coarse-grain time intervals Δt obeying

$$t_{\rm rel} \ll \Delta t \ll t_{nmr}. \quad (4)$$

Averaging the exact Hamiltonian over time Δt is performed to the zeroth order in the perturbation expansion in powers of the parameter ε, yielding an average (or effective) Hamiltonian [13]

$$\overline{h_{i,j}} = \frac{1}{\Delta t}\int_0^{\Delta t} h_{i,j}(r_i(t'), r_j(t'))dt', \quad (5)$$

with corrections on the order of $\mathbb{O}(\varepsilon^1)$.

The decisive point of the following treatment is the replacement of time integration in Eq. (5), with integration over spatial coordinates within the confined region. Equating the temporal averaging with the spatial averaging makes sense under the ergodic hypotheses [14],

$$\frac{\delta t(dr^N, dp^N)}{t} = Z^{-1}\exp\left(-\frac{E}{kT}\right)dr^N dp^N, \quad (6)$$

where the notation implies that a representative point lying in the whole phase space $r^N - p^N$, while moving over a time interval t ($t_{\rm rel} \ll t \ll t_{nmr}$), spends only a fraction,

$$Z^{-1}\exp\left(-\frac{E}{kT}\right)dr^N dp^N$$

of the entire time t within the volume $dr^N dp^N$, with E being the total energy. Relation (6) incorporates the Gibbs stochastic level of description into the dynamical treatment of molecular collisions. The phase space degrees of freedom are assumed to relax to their equilibrium distribution at a given temperature T.

We introduce the equilibrium pair distribution function for molecules 1 and 2,

$$D_2(r_1, r_2) = \frac{\int_V d^3r_3 \ldots \int_V d^3r_N \exp\left(-\frac{U(r^N)}{kT}\right)}{\int_V d^3r_1 \ldots \int_V d^3r_N \exp\left(-\frac{U(r^N)}{kT}\right)}, \quad (7)$$

where $U(r^N)$ denotes intermolecular electrostatic interactions (recall that $\|U\|/\|H\| \approx 10^7$ [15]). Then, taking ergodicity (6) for granted, the evolution of the spin degrees of freedom on the coarse-grain time scale Δt in Eq. (4) is governed by the static (time-independent) effective Hamiltonian

$$\overline{H} = \sum_{1 \le i < j}^{N} \overline{h_{i,j}}, \quad \overline{h_{i,j}} = g \sum_{1 \le i < j}^{N} (\mathbf{I}_i\mathbf{I}_j - 3I_{iz}I_{jz}) \quad (8)$$

with space-independent pair couplings g for any pair of spins i and j,

$$g = \gamma^2\hbar\int_V\int_V d^3r_i d^3r_j D_2(r_i, r_j)P_2(\cos\theta_{ij})r_{ij}^{-3}. \quad (9)$$

The effective operator $\overline{H}$ involves only the (slow) spin operators, whereas the (fast) spatial coordinates of the nucleus (labeled by indices i and j) are integrated out. On the coarse-grain scale Δt, any nuclear spin "feels" the field that is independent of the spatial coordinates of all the other spins flying within the nanocavity, but it depends on the quantum states of those spins.

The effective spin coupling g encodes the information about the shape and size of the nanocavity. The primary objective of the preceding discussion is to present the expression for the coupling g in Eq. (9) for an ellipsoidal nanocavity.

For perfect hard-sphere molecules within the nanocavity, the pair distribution function is given by

$$D_2(r_i, r_j) = V^{-2}$$

for the molecules i and j in the cavity. Hence, the averaging in Eq. (9) gives

$$g = \frac{\gamma^2\hbar F}{V}, \quad F = \frac{1}{V}\int_V\int_V d^3r_1 d^3r_2 \theta(|r_1 - r_2| - \sigma)P_2(\cos\theta_{12})r_{12}^{-3}, \quad (10)$$

where the function $\theta(x)$ (= 1(0) for $x > 0(< 0)$) excludes the intersection of two hard spheres having a diameter σ. In this paper, we use the remarkable fact [4] that the volume V of the nanocavity enters the expression for the effective coupling g in Eq. (10), which itself enters the polarization (defined below in Section 3), giving rise to a dependence of the polarization on the volume of the nanocavity that is by no means as trivial as simply proportional to the volume.

The transformation of coordinates r_1 and r_2 to the relative coordinate $r_{12} = r_1 - r_2$ and the coordinate of the

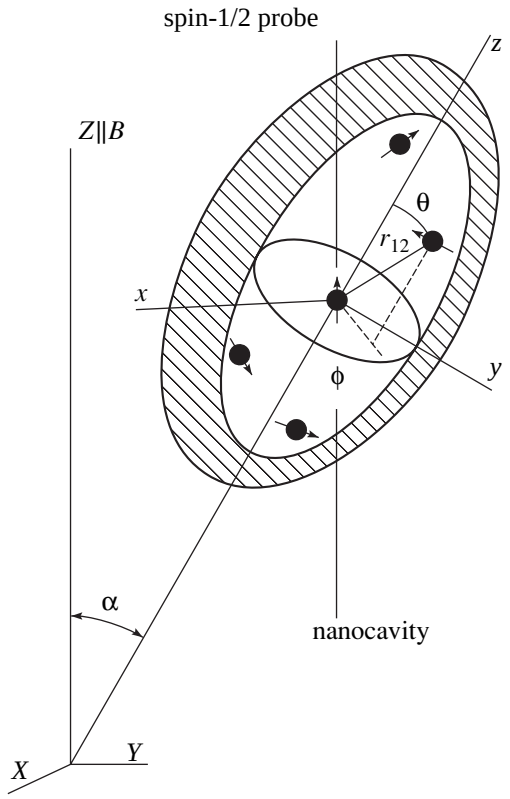


Fig. 1. Schematic representation of a nanocavity partially occupied by spin-carrying molecules undergoing rapid thermal motion.

center of gravity $r = (r_1 + r_2)/2$ (see Fig. 1) reduces form factor F in Eq. (10) to the form

$$F = \int_V \theta(|r_1 - r_2| - \sigma) d^3 r_{12} P_2(\cos\theta_{12}) r_{12}^{-3}. \qquad (11)$$

It is convenient to assume that the initial point of the vector $r_{12} = r_1 - r_2$ starts at the origin of the frame of reference xyz connected with the ellipsoid (see Fig. 1). Straightforward integration over r_{12} in Eq. (11), see, e.g., [16], yields the sought-for form factor:

$$F = I\pi P_2(\cos\alpha),$$

$$I = \begin{cases} \frac{2}{3} + 2\left(\frac{1}{\varepsilon^2} - 1\right)\left(1 - \frac{1}{3}\operatorname{arctanh}\varepsilon\right), & a \geq b, \\ \frac{2}{3} - 2\left(\frac{1}{|\varepsilon|^2} + 1\right)\left(1 - \frac{1}{|\varepsilon|}\operatorname{arctanh}|\varepsilon|\right), & a \leq b. \end{cases} \qquad (12)$$

Here, ε is the eccentricity of the ellipsoid with major axes a and $b = c$. Equation (12) shows that the dependence of the form factor on the angle α between the z axis of the reference frame of the ellipsoid and the Z axis of the laboratory frame of reference is factored out. For $a \gg b$, we have $\varepsilon = 1$ and $I = 2/3$. For $a = b$, we have $\varepsilon \longrightarrow 0$, and Eq. (12) gives $I = 0$. For $a \ll b$, $|\varepsilon| \longrightarrow \infty$ and $I = 2/3 - 2 = -4/3$. These limiting cases confirm the result reported in [4].

3. POLARIZATION DYNAMICS

We consider the spin Hamiltonian $\bar{H}$ of an N-spin cluster in a uniform external magnetic field B parallel to the Z axis of a fixed reference frame XYZ and the space-independent spin couplings g in Eq. (10),

$$\bar{H} = \omega \sum_{n=1}^{N} I_{nz} + \frac{g}{2} \sum_{m \neq n}^{N} \{\zeta I_{mz} I_{nz} - I_{mx} I_{nx} - I_{my} I_{ny}\}, \qquad (13)$$

where $\omega = \gamma B$ denotes the Zeeman frequency and ζ is an arbitrary factor.

The standard way of approaching the N-spin aggregate is to find the polarization at the nth spin at time instant t given the initial polarization at the first spin,

$$P_n(t) = \frac{\operatorname{tr}\{\exp(i\bar{H}t) I_{1z} \exp(-i\bar{H}t) I_{nz}\}}{\operatorname{tr}\{I_{1z}^2\}}. \qquad (14)$$

The Hamiltonian in Eq. (13) can be rewritten as (up to the constant $gN(1 - \zeta/2)/4$)

$$\bar{H} = \omega I_z + \frac{g}{2}(\zeta + 1) I_z^2 - \frac{g}{2} I^2, \qquad (15)$$

where

$$I = \sum_{n=1}^{N} I_n$$

is the total spin and

$$I_z = \sum_{n=1}^{N} I_{nz}$$

is its projection on the z axis. The polarization $P_n(t)$ in Eq. (14) is unchanged if we modify the Hamiltonian $\bar{H}$ in Eq. (15) to the effective one,

$$H' = -\frac{g}{2} I^2. \qquad (16)$$

In addition, the equivalence of polarizations $P_n(t)$ of all spins except for the first one, as well as time conserva-

tion of the total polarization $\sum_{n=1}^{N} P_n(t)$, makes it possible to focus on the first spin only,

$$P_1(\tau) = \frac{\mathrm{tr}\{\exp(i\tau I^2) I_{1z} \exp(-i\tau I^2) I_{1z}\}}{\mathrm{tr}\{I_{1z}^2\}}, \qquad (17)$$

where the dimensionless time scale is defined as $\tau = gt/2$.

A powerful tool for investigating the problem in Eq. (17) is the theory of coupling of angular momenta [17, 18]. To describe it, we consider the total spin cluster composed of two subsystems A and B. The subsystem A has only the spin $I_1 = I_A$, and the remaining fragment B of the spin cluster has the spin I_B, with the total spin being $I = I_A + I_B$. The states of the two subsystems A and B are coupled together within the state of the whole system $A \otimes B$ via the Clebsch–Gordan (CG) coefficients

$$|I_A, I_B, I, m\rangle = \sum_{\substack{m_A = \pm 1/2 \\ m_B = m - m_A}} C^{I,m}_{I_A, m_A; I_B, m_B} |I_A m_A\rangle |I_B m_B\rangle, \qquad (18)$$

where $I_A = 1/2$ and $m_A = \pm 1/2$ are, respectively, the spin and its magnetic quantum numbers for the first spin, and I_B and $m_B = m - m_A$ are the spin and its magnetic quantum numbers of fragment B. For $I_B = 0$, only $I = 1/2$ and $m = \pm 1/2$ are allowed. For $I_B \geq 1/2$, the allowed I and m are as follows: $I = I_B \pm 1/2$, $-I \leq m \leq I$. The CG coefficients are given by (see, e.g., [19])

$$\begin{aligned} C^{I_B+1/2,\,m}_{1/2,\,1/2;\,I_B,\,m-1/2} &= C^{I_B-1/2,\,m}_{1/2,\,-1/2;\,I_B,\,m+1/2} \\ &= \left(\frac{I_B + 1/2 + m}{2I_B + 1}\right)^{1/2}, \\ C^{I_B+1/2,\,m}_{1/2,\,-1/2;\,I_B,\,m+1/2} &= C^{I_B-1/2,\,m}_{1/2,\,1/2;\,I_B,\,m-1/2} \\ &= \left(\frac{I_B + 1/2 - m}{2I_B + 1}\right)^{1/2}. \end{aligned} \qquad (19)$$

The two pairs of independent variables (I_B, m_B) and $(I_A = 1/2, m_A = \pm 1/2)$ are used for determining the trace in Eq. (17) for the whole N-spin system $A \otimes B$,

$$\mathrm{tr}\{\ldots\} = \sum_{I_B = I_B^{\min}}^{N_B/2} w(I_B) \times \sum_{I = |I_B - 1/2|}^{I_B + 1/2} \sum_{m=-I}^{I} \langle I_A, I_B, I, m| \ldots |I_A, I_B, I, m\rangle, \qquad (20)$$

where $N_B = N - 1$ is the number of spins in fragment B; the minimal value of I_B is $I_B^{\min} = 0$ for even N_B and $I_B^{\min} = 1/2$ for odd N_B. The factor

$$w(I_B) = \frac{2I_B + 1}{N_B + 1} \binom{N_B + 1}{\frac{1}{2}N_B + I_B + 1} \qquad (21)$$

is the number of ways of grouping N_B individual spins-1/2 into the total spin I_B. The factor $w(I_B)$ satisfies the relation [17, 18, 20],

$$\sum_{I_B \geq |m_B|}^{N_B/2} w(I_B) = \binom{N_B}{\frac{1}{2}N_B + I_B}. \qquad (22)$$

The right-hand side of Eq. (22) is the number of states for each allowed eigenvalue m_B of the fragment B.

To deal with the diagonal evolution matrices in Eq. (17), we introduce additional bases of the bra, $\langle I_A, I_B, I', m'|$, and ket, $|I_A, I_B, I', m'\rangle$, vectors in the Hilbert space $\mathcal{H}(I_A) \otimes \mathcal{H}(I_B)$ for fixed values $I_A = 1/2$ and I_B; we then use the completeness of the $2(2I_B + 1)$ orthonormal basis vectors belonging to the space $\mathcal{H}(I_A) \otimes \mathcal{H}(I_B)$,

$$1_{\mathcal{H}(I_A) \otimes \mathcal{H}(I_B)} = \sum_{I' = |I_B - 1/2|}^{I_B + 1/2} \sum_{m' = -I}^{I} |I_A, I_B, I', m'\rangle \langle I_A, I_B, I', m'|, \qquad (23)$$

and, finally, insert the representation of unity in Eq. (23) in front of the rightmost operator I_{1z} in Eq. (17), whose matrix elements are given by

$$\langle I_A, I_B, I', m'|I_{1z}|I_A, I_B, I, m'\rangle = \delta_{m,m'} \times \sum_{m_A = \pm 1/2} m_A C^{I',m'}_{1/2,\,m_A;\,I_B,\,m'-m_A} C^{I,m}_{1/2,\,m_A;\,I_B,\,m-m_A}. \qquad (24)$$

With these algebraic steps, we immediately obtain the polarization $P_1(\tau)$ in terms of the CG coefficients as

$$P_1(\tau) = 2^{-(N_B-1)} \sum_{I_B = I_B^{\min}}^{N_B/2} w(I_B) \times \sum_{\substack{|I_B - 1/2| \leq I \leq I_B + 1/2 \\ |I_B - 1/2| \leq I' \leq I_B + 1/2}} \sum_{\substack{-I \leq m \leq I \\ -I' \leq m' \leq I'}} \delta_{m,m'} \times \exp(i\tau\{I(I+1) - I'(I'+1)\}) \times \left(\sum_{m_A = \pm 1/2} m_A C^{I',m'}_{1/2,\,m_A;\,I_B,\,m'-m_A} C^{I,m}_{1/2,\,m_A;\,I_B,\,m-m_A} \right)^2. \qquad (25)$$

For the term $I_B = 0$, only the single pair ($I = 1/2$, $I' = 1/2$) is allowed in the sum in Eq. (25), and for $I_B \geq 1/2$, the four pairs of (I, I') must be distinguished in this sum depending on the (+) or (–) sign in the expressions,

$$(I, I') = \left(I = I_B \pm \frac{1}{2}, I' = I_B \pm \frac{1}{2}\right). \quad (26)$$

Armed with the polarization $P_1(\tau)$ in Eq. (25), we now decompose it into the time-independent part $\overline{P_1}$ and the oscillating part $P_1^{\text{osc}}(\tau)$,

$$P_1(\tau) = \overline{P_1} + P_1^{\text{osc}}(\tau). \quad (27)$$

The time-independent contribution $\overline{P_1}$ to the function $P_1(\tau)$ is provided by the quantum numbers m, m' belonging to the states $I = I' = I_B \pm 1/2$ if $I_B \geq 1/2$, and by the quantum numbers m, m' belonging to the states $I = I' = 1/2$ if $I_B = 0$,

$$\overline{P_1} = 2^{-(N_B-1)} \sum_{I_B = I_B^{\min}}^{N_B/2} w(I_B) \times \sum_{I = |I_B - 1/2|}^{I_B + 1/2} \sum_{m = -I}^{I} \left(\sum_{m_A = \pm 1/2} m_A (C_{1/2, m_A; I_B, m - m_A})^2 \right)^2. \quad (28)$$

Our aim is now to sum over the indices m and I for a fixed value of I_B in Eq. (28). To this end, we start with the state $I_B = 0$ that arises for even N_B (see the comments to Eq. (20)). For $I_B = 0$, only $I = 1/2$ is allowed and the partial polarization $\overline{P_1}(I_B)$ in Eq. (28) is given by

$$\overline{P_1}(I_B = 0) = 2^{-(N_B-1)} w(0) \sum_{m = -1/2}^{1/2} m^2. \quad (29)$$

Next, we consider the contribution to $\overline{P_1}$ in Eq. (28) from the spin $I_B \geq 1/2$. In this situation, $I = I_B \pm 1/2$ are allowed and, invoking the CG coefficients in Eq. (19), the contributions $\overline{P_1}(I_B)$ to $\overline{P_1}$ can be conveniently written as

$$\overline{P_1}(I_B) = 2^{-N_B} w(I_B) \sum_{\mu = -I_B}^{I_B} \frac{2\mu + 1}{2I_B + 1}. \quad (30)$$

Combining $\overline{P_1}(I_B = 0)$ in Eq. (29) and $\overline{P_1}(I_B)$ in Eq. (30) results in

$$\overline{P_1} = 2^{-N_B} \sum_{I_B = I_B^{\min}}^{N_B/2} w(I_B) \sum_{\mu = -I_B}^{I_B} \frac{2\mu + 1}{2I_B + 1}. \quad (31)$$

The sum over μ in Eq. (31) easily yields

$$\sum_{\mu = -I_B}^{I_B} (2\mu + 1)^2 = (2I_B + 1)\left(1 + \frac{4}{3} I_B(I_B + 1)\right), \quad (32)$$

and substituting $w(I_B)$ from Eq. (21), we arrive at the sought-for result,

$$\overline{P_1} = \frac{2^{-N_B}}{N_B + 1} \sum_{I_B = I_B^{\min}}^{N_B/2} \binom{N_B + 1}{\frac{1}{2}N_B + I_B + 1} \times \left(1 + \frac{4}{3} I_B(I_B + 1)\right). \quad (33)$$

The remaining sum over I_B in Eq. (33) depends on whether N_B is an even or odd number. If N_B is an even number, then $I_B^{\min} = 0$ and straightforward summation over I_B in Eq. (33), with the known sums involving the binomial coefficients

$$\sum_{I_B = 0}^{N_B/2} \binom{N_B + 1}{\frac{1}{2}N_B + I_B + 1} = 2^{N_B},$$
$$\sum_{I_B = 0}^{N_B/2} I_B(I_B + 1) \binom{N_B + 1}{\frac{1}{2}N_B + I_B + 1} = N_B 2^{N_B - 2}, \quad (34)$$

yields the polarization

$$\overline{P_1} = \frac{N + 2}{3N} \quad (35)$$

for an odd $N = N_B + 1$ spin cluster [11]. If N_B is an odd number, then $I_B^{\min} = 1/2$ and some simple algebra gives the polarization,

$$\overline{P_1} = \frac{N + 2 - 2^{1-N} \binom{N}{N/2}}{3N} \quad (36)$$

for an even $N = N_B + 1$ spin cluster.

When $N \gg 1$, the polarization $\overline{P_1}$ in Eq. (36) behaves as

$$\frac{N + 2 - 2(\pi N/2)^{-1/2}}{3N}.$$

Equations (35) and (36) give the sought-for time-independent contributions $\overline{P_1}$ to the total polarization $P_1(\tau)$ in Eq. (27) for odd- and even-numbered spin clusters, respectively.

It remains to find the time-dependent contribution $P_1^{osc}(\tau)$ to the total polarization $P_1(\tau)$ in Eq. (25). Among the four pairs (I, I') in Eq. (26), only the pairs (I, I') with $I \neq I'$ contribute to the time-dependent part of the function $P_1(\tau)$ in Eq. (25). This occurs for $I_B \geq 1/2$ only, because otherwise, i.e., for $I_B = 0$, the allowed values $I = I' = 1/2$ are already encountered in the time-independent polarization $\overline{P_1}$. Thus, among the four pairs (I, I') in Eq. (26), only the two pairs $(I = I_B + 1/2, I' = I_B - 1/2)$ and $(I = I_B - 1/2, I' = I_B + 1/2)$ are allowed and provide complex conjugate contributions to the real-valued function $P_1^{osc}(\tau)$. It suffices to deal with the first pair, $(I = I_B + 1/2, I' = I_B - 1/2)$. The polarization becomes

$$P_1^{osc}(\tau) = 2^{-(N_B-1)} \sum_{I_B = 1/2}^{N_B/2} w(I_B)$$

$$\times \sum_{m=-(I_B+1/2)}^{I_B+1/2} \sum_{m'=-(I_B-1/2)}^{I_B-1/2} \delta_{m,m'} 2\cos(2\tau(I_B+1/2)) \quad (37)$$

$$\times \left(\sum_{m_A = \pm 1/2} m_A C^{I_B-1/2, m'}_{1/2, m_A; I_B, m'-m_A} C^{I_B+1/2, m}_{1/2, m_A; I_B, m-m_A} \right)^2.$$

To complete the derivation of the function $P_1^{osc}(\tau)$, we use the expression for the factor $w(I_B)$ in Eq. (21), the CG coefficients in Eq. (19), and sum over m and m' in Eq. (37) for a fixed value of I_B. This gives

$$P_1^{osc}(\tau) = \frac{2^{-N_B+3}}{3(N_B+1)} \sum_{I_B=1/2}^{N_B/2} \binom{N_B+1}{\frac{1}{2}N_B + I_B + 1} \quad (38)$$
$$\times I_B(I_B+1)\cos(2\tau(I_B+1/2)).$$

Finally, by gathering the expressions for $\overline{P_1}$ in Eqs. (35) and (36) and the expression for $P_1^{osc}(\tau)$ in Eq. (38), we pass (with the substitution $k = I_B - 1/2$ for even N and $k = I_B$ for odd N) to the total polarization at the first spin,

$$P_1(\tau) = \frac{N + 2 - 2^{1-N}\binom{N}{N/2}}{3N}$$
$$+ \frac{2^{4-N^{N/2-1}}}{3N} \sum_{k=0} A_k(N)\cos(\tau(N-2k)) \quad (39)$$

for an even N-cluster, and

$$P_1(\tau) = \frac{N+2}{3N}$$
$$\times \frac{2^{4-N^{(N-1)/2}}}{3N} \sum_{k=0} A_k(N)\cos(\tau(N-2k)) \quad (40)$$

for an odd N-cluster, with the coefficient

$$A_k(N) = \left(\frac{N+1}{2} - k\right)\left(\frac{N-1}{2} - k\right)\binom{N}{k}$$

arising in both cases. Formulas (39) and (40) are the central result of the paper. They are used to describe a variety of systems in the next section.

4. DISCUSSION

4.1. Nonergodic Spin Dynamics

As Eq. (40) states, for large odd-N clusters, the time-averaged polarization $\langle P_1(\tau)\rangle$ of spin 1 tends to 1/3, while the time-averaged polarization $2/3N$ of any other spin tends to 0; i.e., polarization of spin 1 does not spread uniformly over the N-spin cluster. We call this behavior the nonergodic spin dynamics, to compare it with the ergodic spin dynamics providing the $1/N$ polarization for all spins in an N-spin ensemble. Figure 2a shows the behavior of the polarization $P_1(\tau)$ for a series of odd-N clusters. The principle features of the periodic pulses of the polarization are determined by two factors: first, time reversibility of the dynamics affects the exact reentrance of the polarization to the prepared value $P_1(0) = 1$ after each period $4\pi/g$, and, second, gives rise to the time interval with the time-independent polarization of spin 1 (see the Appendix for details). For large-N clusters, the total period $4\pi/g$ can be partitioned into the switching time

$$t^{sw} = 4\pi\frac{\mathbb{C}(1)}{g\sqrt{N}}$$

and the stopping time

$$t^{st} = \frac{4\pi}{g}\left(1 - \frac{\mathbb{C}(1)}{\sqrt{N}}\right);$$

recall that $\tau = gt/2$. As shown in the Appendix, the polarization $P_1(\tau)$ peaks at the instants $t = 0, 2\pi/g, 4\pi/g, \ldots$. The profile of the function $P_1(\tau)$, e.g., around $\tau = 0$, is

$$P_1(\tau) = \frac{1}{3} + \frac{2}{3}(1 - \tau^2 N)\exp\left(-\frac{\tau^2 N}{2}\right). \quad (41)$$

The function $P_1(\tau)$ has the same profile around all the instants $\tau = m\pi$ for all integers m. The interval between

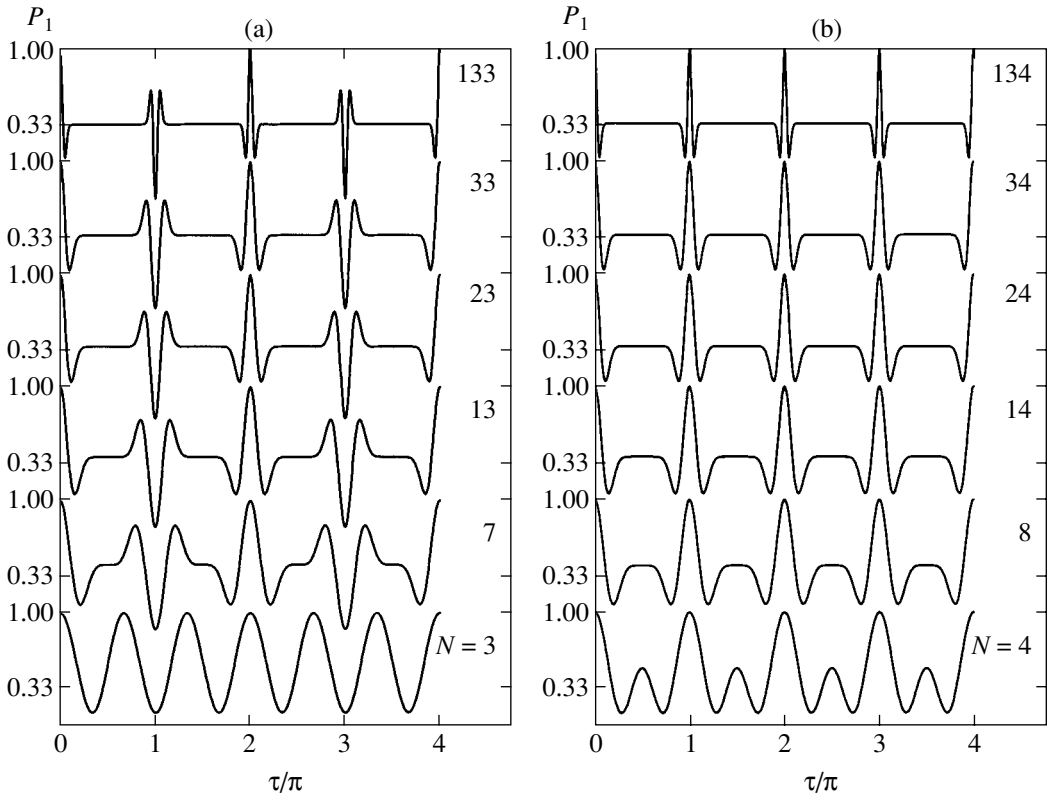


Fig. 2. a—Polarization $P_1(\tau)$ in Eq. (40) of the first spin is varied with the dimensionless time $\tau = gt/2$ for the series of an odd total number N of spins. b—Polarization $P_1(\tau)$ in Eq. (39) for the series of even N.

the successive peaks and their widths are

$$T = \frac{2\pi}{g}, \quad \Delta_T = 4\pi\frac{\mathbb{O}(1)}{g\sqrt{N}}, \tag{42}$$

respectively. In other words, for large-N clusters, the polarization of spin 1 stays at the fixed value $\overline{P_1} = 1/3$ almost the whole time. The oscillating part of $P_1(\tau)$ is an odd function of time with respect to the instants $\tau = \pi/2, 3\pi/2, \ldots$, as apparent from Eq. (40).

Figure 2b shows the profiles of the polarization for N-spin clusters with even N. For large even values of N, the polarization at spin 1 stays fixed over a long time interval

$$t^{st} = \frac{2\pi}{g}\left(1 - \frac{\mathbb{O}(1)}{\sqrt{N}}\right)$$

within each period $2\pi/g$. Unlike odd-N clusters, the profiles of $P_1(\tau)$ for even-N clusters are even functions of time with respect to time instants $\tau = \pi/2, 3\pi/2, \ldots$.

Using the experimental values of the time interval T and the width of the pulses Δ_T in Eq. (42) together with the expressions for the coupling g in Eqs. (10) and (12), we find that the volume and form factor are given by

$$V = \frac{4}{c}\left(\frac{T}{\Delta_T}\right)^2,$$
$$F\left(\frac{a}{b}\right)P_2(\cos\alpha) = \frac{8}{c}\frac{T}{\gamma^2\hbar\Delta_T^2}, \tag{43}$$

where $c = N/V$ denotes the concentration of the molecules carrying spin 1/2 and the angle α is assumed to be known.

4.2. Polarization Dynamics in Fluctuating Nanobubbles

Equations (39) and (40) can be adapted to account for the time dependence of the volume of the nanocavity, thereby providing a means to explore NMR imaging of cavitation bubbles in water [21], blood [22], etc., along with conventional high-speed photography. The dynamics of the surface of a typical behavior of bubbling occurs on a millisecond time scale [21], i.e., at the same time scale relevant for the nuclear spin dynamics. It is therefore jusitifiable to ask how the dynamics of a nanosize volume affects the nuclear spin dynamics. Our intention in this section is to show that fluctuations of the nanovolume (governed either by external inputs or

by inherently thermal noise) drive the polarization to the nonergodic value 1/3 irreversibly, and therefore, time-periodic pulsation of the polarization breaks down as time proceeds.

The formulation in Section 3 is easily extended to the case of a time-varying volume V because the coupling $g(V(t))$ enters Hamiltonian (13) as a common factor in front of the operator part. The form function of the polarization $P_1^{\text{osc}}(\tau)$ in Eq. (38), which has been derived for time-independent coupling g, is generalized to the case of a function $g(t)$ provided that the time $\tau = gt/2$ in Eq. (38) is replaced with a new time,

$$\tau = \frac{1}{2}gt \longrightarrow \frac{1}{2}\int_0^t g(t')dt'. \tag{44}$$

We are interested in transformation (44),

$$g(t) = \langle g\rangle + \delta g(t), \tag{45}$$

where $\delta g(t)$ is Gaussian random noise characterized by the first two moments

$$\begin{aligned}\langle \delta g(t)\rangle &= 0,\\ \langle \delta g(t_1)\delta g(t_2)\rangle &= \langle (\delta g)^2\rangle \gamma(|t_1 - t_2|),\end{aligned} \tag{46}$$

where $\langle(\delta g)^2\rangle$ is the variance and $\gamma(t)$ denotes the correlation function, for example, $\gamma(t) = \exp(-t/t_c)$, with t_c being the correlation time. In accordance with the comment before Eq. (44), we replace the factor

$$\cos(2\tau(I_B + 1/2))$$

in Eq. (38) with

$$\cos\left[(I_B + 1/2)\left(\langle g\rangle t + \int_0^t \delta g(t')dt'\right)\right].$$

The Gaussian averaging of this factor over the random function $\delta g(t)$ is performed as follows (see, e.g., [15]):

$$\begin{aligned}&\left\langle \exp\left(i(I_B + 1/2)\int_0^t \delta g(t')dt'\right)\right\rangle_{\delta g}\\ &= \exp(-(I_B + 1/2)^2\langle(\delta g)^2\rangle T^2),\end{aligned} \tag{47}$$

where

$$T^2 = \int_0^t (t - t')\gamma(t')dt'. \tag{48}$$

We first restrict our attention to the polarization for even N, with $N \gg 1$, and then close the section with the final result for odd N, $N \gg 1$. We write polarization (38) averaged by using (47) as

$$\begin{aligned}P_1(t) &= \overline{P_1} + \frac{2^{-N_B+3}}{3(N_B+1)}\\ &\times \sum_{I_B = 1/2}^{N_B/2}\binom{N_B+1}{\frac{1}{2}N_B + I_B + 1} I_B(I_B+1)\Phi(t),\\ \Phi(t) &= \exp(-(I_B + 1/2)^2\langle(\delta g)^2\rangle T^2)\\ &\times \cos(\langle g\rangle t(I_B + 1/2)).\end{aligned} \tag{49}$$

The exponent in Eq. (49) tell us that successive peaks of the time-dependent part of $P_1(t)$ reduce to zero as $t \longrightarrow \infty$, and therefore only the time-independent part of $P_1(t)$, i.e., $\overline{P_1} = 1/3$ in Eq. (33), survives as $t \longrightarrow \infty$ after the Gaussian averaging over the function $\delta g(t)$. The integral over t' in the constant T^2 in Eq. (48) can be evaluated in the two asymptotic cases, for the large and the small correlation times [15],

$$T^2 = \begin{cases} t^2/2 & \text{if } t_c^2\langle(\delta g)^2\rangle \gg 1,\\ t_c t & \text{if } t_c^2\langle(\delta g)^2\rangle \ll 1.\end{cases} \tag{50}$$

To find function $P_1(t)$ in Eq. (49) for $N \gg 1$, we can replace the sum in Eq. (49) with Gaussian averaging, as in Eq. (A.3) in the Appendix, which gives

$$\begin{aligned}P_1(t) &= \frac{1}{3} + \frac{16}{3N^{3/2}\sqrt{\pi/2}}\\ &\times \sum_{n=1}^{N/2}\cos(\langle g\rangle tn)\left(n^2 - \frac{1}{4}\right)\exp(-an^2),\end{aligned} \tag{51}$$

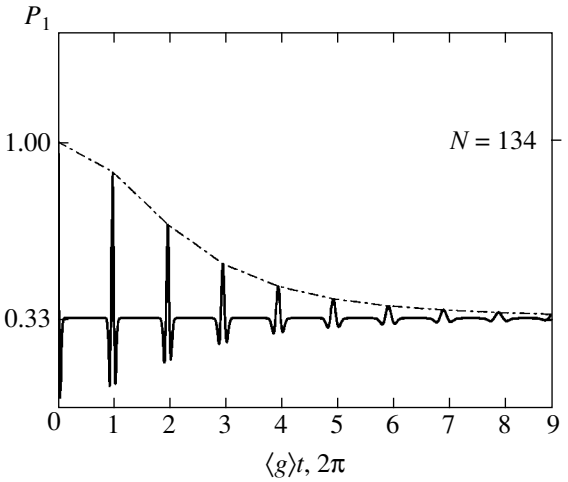


Fig. 3. Polarization dynamics of a single spin, $P_1(t)$, within the $N = 134$ spin aggregate when the volume of the nanocavity fluctuates, providing relative variance of the g coupling equal to $\langle(\delta g)^2\rangle/\langle g\rangle^2 = 10^{-4}$ [see Eqs. (51) and (52)].

where

$$a = \frac{2}{N} + \langle(\delta g)^2\rangle T^2. \tag{52}$$

Figure 3 shows the polarization dynamics of a single spin within the $N = 134$ spin aggregate for

$$\frac{\langle(\delta g)^2\rangle}{\langle g\rangle^2} = 10^{-4}.$$

Using formulas given in the Appendix, we simplify the sum over n in Eq. (51) via the Poisson resummation formula. Defining the partial sums entering Eq. (51) by

$$\begin{aligned} S_1(t) &= \sum_{n=1}^{N/2} \cos(\langle g\rangle tn)\exp(-an^2) \\ &= -\frac{1}{2} + \frac{1}{2}\sqrt{\frac{\pi}{a}}\sum_{q=-\infty}^{\infty} \exp\left(-\pi^2\left(q + \frac{\langle g\rangle t}{2\pi}\right)^2/a\right), \end{aligned} \tag{53}$$

and

$$\begin{aligned} S_2(t) &= \sum_{n=1}^{N/2} \cos(\langle g\rangle tn) n^2 e^{-an^2} = -\frac{\partial S_1(t)}{\partial a} \\ &= \frac{\sqrt{\pi}}{4a^{3/2}}\sum_{q=-\infty}^{\infty}\left(1 - \frac{2\pi^2}{a}\left(q + \frac{\langle q\rangle t}{2\pi}\right)^2\right) \\ &\quad \times \exp\left(-\frac{\pi^2}{a}\left(q + \frac{\langle g\rangle t}{2\pi}\right)^2\right), \end{aligned} \tag{54}$$

we obtain

$$P_1(t) = \frac{1}{3} + \frac{16}{3N^{3/2}\sqrt{\pi/2}}\left(S_2(t) - \frac{1}{4}S_1(t)\right). \tag{55}$$

To find the envelope of the successive peaks of the function $P_1(t)$ in Eq. (55), we substitute the time $t = 2\pi m/\langle g\rangle$ in Eqs. (53–55), with m running over integer numbers. This gives polarization at discrete values m,

$$\begin{aligned} S_1(m) &= \sum_{n=1}^{N/2} \exp(-an^2) \\ &= -\frac{1}{2} + \frac{1}{2}\sqrt{\frac{\pi}{a}}\sum_{k=-\infty}^{\infty} \exp\left(-\frac{\pi^2 k^2}{a}\right), \end{aligned} \tag{56}$$

$$\begin{aligned} S_2(m) &= \sum_{n=1}^{N/2} n^2 \exp(-an^2) \\ &= \frac{\sqrt{\pi}}{4a^{3/2}}\sum_{k=-\infty}^{\infty}\left(1 - \frac{2}{a}\pi^2 k^2\right)\exp\left(-\frac{\pi^2 k^2}{a}\right). \end{aligned} \tag{57}$$

The functions $S_1(m)$ and $S_2(m)$ inherit their dependence on the "time" m through the constant a in Eqs. (52) and (50),

$$a = \begin{cases} \dfrac{2}{N} + 2\pi^2 m^2 \dfrac{\langle(\delta g)^2\rangle}{\langle g\rangle^2} & \text{if } t_c^2\langle(\delta g)^2\rangle \gg 1, \\ \dfrac{2}{N} + 2\pi m t_c \dfrac{\langle(\delta g)^2\rangle}{\langle g\rangle} & \text{if } t_c^2\langle(\delta g)^2\rangle \ll 1, \end{cases} \tag{58}$$

where we substitute $t = 2\pi m/\langle g\rangle$ in Eqs. (52) and (50). For $N \gg 1$ and $m \gg 1$, we drop the summand $2/N$ in Eq. (58) assuming that $a \gg 1$,

$$a \approx \begin{cases} 2\pi^2 m^2 \dfrac{\langle(\delta g)^2\rangle}{\langle g\rangle^2} & \text{if } t_c^2\langle(\delta g)^2\rangle \gg 1, \\ 2\pi m t_c \dfrac{\langle(\delta g)^2\rangle}{\langle g\rangle} & \text{if } t_c^2\langle(\delta g)^2\rangle \ll 1. \end{cases} \tag{59}$$

For $a \gg 1$, we find the sums over k in Eqs. (56) and (57) by again using the Poisson resummation formula, Eq. (A.5), which accelerates convergence of the sums for $a \gg 1$. Reading Eq. (A.4) backwards, from the right-hand side to the left-hand side, we obtain

$$\begin{aligned} I(a) &= \sqrt{\frac{\pi}{a}}\sum_{k=-\infty}^{\infty} \exp\left(-\frac{\pi^2 k^2}{a}\right) \\ &= \sum_{\ell=-\infty}^{\infty} \exp(-a\ell^2) = 1 + 2e^{-a} + \mathbb{O}(e^{-4a}). \end{aligned} \tag{60}$$

Thus, Eqs. (53) and (54) become

$$S_1(m) = -\frac{1}{2} + \frac{1}{2}I(a) = e^{-a}, \tag{61}$$

$$S_2(m) = -\frac{1}{2}\frac{\partial I(a)}{\partial a} = e^{-a}, \tag{62}$$

and Eq. (55) therefore gives the polarization of the first spin,

$$P_1(m) = \frac{1}{3} + \frac{4\sqrt{2}}{N^{3/2}\sqrt{\pi}}e^{-a} \tag{63}$$

with a in Eq. (59).

We conclude this section with the result for the total polarization for an odd total number N of spins. Due to

alternating peaks of the polarization $P_1(t)$ in Eq. (40) (see also Fig. 2a), we obtain

$$P_1(m) = \begin{cases} \frac{1}{3} - \frac{4\sqrt{2}}{N^{3/2}\sqrt{\pi}} e^{-a} & \text{for large odd } m, \\ \frac{1}{3} + \frac{4\sqrt{2}}{N^{3/2}\sqrt{\pi}} e^{-a} & \text{for large even } m. \end{cases} \quad (64)$$

Equation (64), with a in Eq. (59), shows that the polarization peaks $P_1(m)$ of a spin-carrying gas have a Gaussian and an exponential time dependence for large and small correlation times, respectively, of nanobubble fluctuations.

4.3. NMR Line Shape

To calculate the NMR line shape exactly, we use the same effective Hamiltonian (13) as described in Section 3. The NMR line shape is the Fourier transform of the free induction decay (FID), $F(t)$, of an N-spin ensemble [15]. It was the NMR line shape on the protons in hydrogenated thin silicon films provided the first experimental evidence for the validity of effective Hamiltonian (13) in nanocavities [4]. We are interested in the FID signal

$$F(t) = \frac{\mathrm{tr}\{e^{iHt} I_+ e^{-iHt} I_-\}}{\mathrm{tr}\{I_+ I_-\}} \quad (65)$$

$$I_\alpha = \sum_{n=1}^{N} I_{n\alpha}, \quad I_\pm = I_x \pm i I_y, \quad \alpha = x, y, z.$$

The reason for there being an exact solution for the FID in Eq. (65) is that the total Hamiltonian in Eq. (13) can be expressed in terms of the three collective spin operators I_α just as in Eq. (15). Because $[\mathbf{I}^2, I_\alpha] = 0$, we can rewrite Eq. (65) as

$$F(t) = \frac{\mathrm{tr}\left\{e^{iGtI_z^2} I_+ e^{-iGtI_z^2} I_-\right\}}{\mathrm{tr}\{I_+ I_-\}}, \quad (66)$$

with $G = 3g/2$ for dipolar interactions in the effective Hamiltonian (13) with $\zeta = 2$. The Heisenberg equation of motion for the operator

$$I_+(t) = \exp(iGtI_z^2) I_+ \exp(-iGtI_z^2)$$

is solved exactly as

$$I_+(t) = \exp(iGt(2I_z - 1)) I_+(0), \quad I_+(0) = I_+.$$

The averaging in Eq. (65) is performed in an $N!/(N_\uparrow! N_\downarrow!)$-fold degenerate basis of states $(N_\uparrow, N_\downarrow)$ with $N_\uparrow$ ($N_\downarrow$) spins up (down), such that $N_\uparrow + N_\downarrow = N$ and $I_z = (N_\uparrow - N_\downarrow)/2$. The averaging gives the FID

$$F(t) = \left[\cos\left(\frac{3}{2}gt\right)\right]^{N-1}. \quad (67)$$

The effect of dephasing of proton spins within the nanocavity due to interactions with the protons at the surface of the nanocavity is introduced phenomenologically as

$$F(t) = \left[\cos\left(\frac{3}{2}gt\right)\right]^{N-1} \exp\left(-\frac{t}{T_2}\right), \quad (68)$$

where the time T_2 relevant for the experiments [4] is $T_2 \approx 1$–3 ms. The moments of the line shape are

$$M_n = \int_{-\infty}^{\infty} d\omega \omega^n \Im(\omega) = \left[\frac{d^n F(t)}{d(it)^n}\right]\Big|_{t=0}, \quad (69)$$

where $\Im(\omega)$ enters through the Fourier transformation of the FID,

$$F(t) = \int_{-\infty}^{\infty} d\omega \Im(\omega) e^{i\omega t}. \quad (70)$$

The function $\Im(\omega)$ is meaningful for the frequencies $0 < \omega < \gamma^2\hbar/\sigma^3$, with σ standing for the diameter of hard-sphere spin-carrying molecules. The upper cutoff of the frequency provides finite second and fourth moments for $gT_2 \gg 1$,

$$M_2 = (N-1)(3g/2)^2, \quad M_4 = (N-1)(3N-5)(3g/2)^4. \quad (71)$$

The moment M_2 derived in [4] by the Van Vleck formula coincides with M_2 in Eq. (70), as it should. The line shape in the nanocavity volume appears to be volume-dependent (via the coupling g in Eq. (10)), allowing one to determine the volume of the nanopores in a hydrogenated silicon film [4].

5. CONCLUSIONS

We have presented the exact time-dependent description of spin-1/2 dynamics with infinite-range spin interactions and the initial polarization prepared on a single spin 1, i.e., $P_1(0) = 1$. Spin dynamics for odd- and even-numbered clusters demonstrates periodic pulses of polarization $P_1(\tau)$ on spin 1. For large-N clusters with odd N, the polarization on spin 1 has pulses over the time interval

$$t^{sw} = 4\pi \frac{\mathbb{O}(t)}{g\sqrt{N}},$$

from $P_1(0) = 1$ to the time-independent polarization,

which therefore lasts

$$t^{st} = \frac{4\pi}{g}\left(1 - \frac{\mathbb{O}(1)}{\sqrt{N}}\right)$$

within any period $4\pi/g$. For large-N clusters with even N, the switching time is

$$t^{sw} = 2\pi\frac{\mathbb{O}(1)}{g\sqrt{N}}$$

and the period equals $2\pi/g$. The stationary polarization on spin 1 is nonergodic, because its value tends to 1/3 (instead of tending to the ergodic value $1/N$) as N tends to infinity. The profiles of the polarizations within the series of odd (even) large clusters are remarkably similar.

The specific polarization profile in clusters with infinite-range spin interactions is in sharp contrast with the polarization profiles in 1D clusters with the nearest-neighbor XY Hamiltonian [3]. Two differences can be drawn from the results.

1. The overall behavior of the polarization $P_1(t)$ in the system with an infinite-range interaction is strictly reversible and periodic with a period $4\pi/g$ for any N, whereas on large 1D chains ($N \gg 1$) with the XY Hamiltonian, the polarization $P_1(t)$ on spin 1 moves in an irregular fashion.

2. For N-spin clusters with $N \gg 1$, the polarization $P_1(t)$ of spin 1 exhibits a plateau region at the nonergodic value $\overline{P_1} = 1/3$; the pulses of the polarization $P_1(t)$ have a short time span of about $4\pi\mathbb{O}(1)/(g\sqrt{N})$. This is in contrast to the behavior of the polarization $P_1(t)$ in 1D spin chains with the XY Hamiltonian, where polarization on spin 1 depends on time in an irregular fashion with $t^{st} = 0$.

Finally, this paper demonstrates the sensitivity of the polarization dynamics (reversibility and ergodicity in many-spin systems) to the radius of the interaction. Incorporation of the real dipolar interactions into the theory is the most challenging task of dynamics theory and an accurate answer has not yet been attained, although the general picture of the spin dynamics is known to be diffusional [23].

We thank D.E. Fel'dman and S.V. Iordanskii for helpful discussions, A. K. Khitrin for sending the report [4], and S.I. Doronin and I.I. Maximov for help in preparing the manuscript. This paper was supported in part by the Russian Foundation for Basic Research (project no. 01-03-33273).

APPENDIX

Derivation of Eq. (41)

We want to prove that the function $P_1^{osc}(\tau)$ in Eq. (38) for $N \gg 1$ has the form of periodic pulses, each having a width of

$$\Delta_T = 4\pi\frac{\mathbb{O}(1)}{g\sqrt{N}}$$

at equidistant instants $\tau = 0, 2\pi, 4\pi, \ldots$, such that the profile of $P_1^{osc}(\tau)$, e.g., at the time instant $\tau = 0$, is

$$P_1^{osc}(\tau) = \frac{2}{3}(1 - \tau^2 N)\exp(-\tau^2 N/2), \quad \text{for } N \gg 1. \tag{A.1}$$

To prove Eq. (A.1), we introduce the new variable

$$n = I_B + \frac{1}{2}$$

in Eq. (38), such that the function $P_1^{osc}(\tau)$ takes the form (recall that the total number of spins is equal to $N = N_B + 1$)

$$P_1^{osc}(\tau) = \frac{16}{3N}\sum_{n=1}^{N/2} 2^{-N}\binom{N}{N/2+n} \times \left(n^2 - \frac{1}{2}\right)\cos(2\tau n). \tag{A.2}$$

Next, we use the asymptotic formula for the binomial distribution,

$$2^{-N}\binom{N}{N/2+n} = \frac{1}{\sqrt{\pi N/2}} \times \exp\left(-\frac{n^2}{N/2}\right)\left(1 + n^3\frac{\mathbb{O}(1)}{\sqrt{N}}\right). \tag{A.3}$$

Equation (A.3) allows us to consider the summation in Eq. (A.2) as averaging over the Gaussian distribution function. To simplify the calculation of Eq. (A.2) further, we apply the Poisson identity [24]

$$\sum_{\ell=-\infty}^{\infty}\cos(2\pi\epsilon\ell)\exp(-a\ell^2) = \sqrt{\frac{\pi}{a}}\sum_{k=-\infty}^{\infty}\exp\left(-\frac{\pi^2(k+\epsilon)^2}{a}\right). \tag{A.4}$$

In many circumstances, including the present ones, the resulting sum over k in the right-hand side of Eq. (A.4)

converges much faster than the original sum over ℓ in the left-hand side of Eq. (A.4). To apply Eqs. (A.4) to Eq. (A.2), we can expand the sum in Eq. (A.2) up to $n = \infty$ because the terms in the sum in Eq. (A.2) practically vanish for $n > N/2$ and $N \gg 1$. Thus, by Poisson identity (A.4), we introduce the sum (a partial contribution to the sum in Eq. (A.2))

$$\begin{aligned} S_1(\tau) &= \sum_{n=1}^{N/2} \cos(2\tau n)\exp\left(-\frac{n^2}{N/2}\right) \\ &= -\frac{1}{2} + \frac{1}{2}\sqrt{\frac{\pi N}{2}}\sum_{k=-\infty}^{\infty} \exp\left(-\pi^2\left(k+\frac{\tau}{\pi}\right)^2\frac{N}{2}\right). \end{aligned} \tag{A.5}$$

To show that the function $S_1(\tau)$ has the form of Gaussian peaks at the equidistant instants $\tau = 0, \pm\pi, \pm 2\pi, \ldots$, it suffices to analyze the function $S_1(\tau)$ around the point $\tau = 0$. The leading contribution to the sum in Eq. (A.5) is then made by the term $k = 0$. We note that if we analyze the peak around $\tau = m\pi$, where m is an integer, then the leading contribution to $S_1(\tau)$ comes from the term $k = -m$. Thus, in considering $N \gg 1$, we can drop all terms in Eq. (A.5) except the leading term $k = 0$, which yields

$$S_1(\tau) = -\frac{1}{2} + \frac{1}{2}\sqrt{\frac{\pi N}{2}}\exp\left(-\frac{\tau^2 N}{2}\right). \tag{A.6}$$

Analogously, we determine the partial sum

$$\begin{aligned} S_2(\tau) &= \sum_{n=1}^{N/2} \cos(2\tau n) n^2 \exp\left(-\frac{n^2}{N/2}\right) \\ &= -\frac{\partial}{\partial(2/N)} S_1(\tau) \\ &= \frac{N^{3/2}\sqrt{\pi}}{8\sqrt{2}}(1-\tau^2 N)\exp\left(-\frac{\tau^2 N}{2}\right). \end{aligned} \tag{A.7}$$

At $N \gg 1$, the function $S_1(\tau)$ in Eq. (A.6) makes a negligible contribution to the function

$$P_1^{\mathrm{osc}}(\tau) = \frac{16}{3N^{3/2}\sqrt{\pi/2}}\left(S_2(\tau) - \frac{1}{4}S_1(\tau)\right)$$

in comparison with the contribution of the function $S_2(\tau)$ in Eq. (A.7), yielding the sought-for result in Eq. (A.1).

In general, the function $P_1^{\mathrm{osc}}(\tau)$ for an arbitrary τ has pulses at the moments $\tau = k\pi$ with integer k,

$$\begin{aligned} P_1^{\mathrm{osc}}(\tau) &= \sum_{k=-\infty}^{\infty} \frac{2}{3}\left(1 - \pi^2\left(k+\frac{\tau}{\pi}\right)^2 N\right) \\ &\times \exp\left(-\pi^2\left(k+\frac{\tau}{\pi}\right)^2\frac{N}{2}\right). \end{aligned} \tag{A.8}$$

REFERENCES

1. F. S. Dzheparov, Zh. Éksp. Teor. Fiz. **116**, 1398 (1999) [JETP **89**, 753 (1999)].
2. R. Brüschweiler and R. R. Ernst, Chem. Phys. Lett. **264**, 393 (1997).
3. E. B. Fel'dman, R. Brüschweiler, and R. R. Ernst, Chem. Phys. Lett. **294**, 297 (1998); E. B. Fel'dman and M. G. Rudavets, Chem. Phys. Lett. **311**, 453 (1999).
4. J. Baugh, A. Kleinhammes, D. Han, *et al.*, Science **294**, 1505 (2001).
5. S. Inagaki, S. Guan, T. Ohsuna, and O. Terasaki, Nature **416**, 304 (2002).
6. B. E. Kane, Nature **393**, 133 (1998).
7. D. Mozyrsky, V. Privman, and M. L. Glasser, Phys. Rev. B **63**, 085313 (2001).
8. C. S. Yannoni, J. Am. Chem. Soc. **92**, 5237 (1970).
9. M. Kac, in *Statistical Physics, Phase Transitions and Superfluidity*, Ed. by M. Chrétien, E. P. Gross, and S. Deser (Gordon and Breach, New York, 1968), Vol. 1, p. 241.
10. P. W. Anderson, Phys. Rev. **112**, 1900 (1958); Y. Wada, F. Takano, and N. Fukuda, Prog. Theor. Phys. **19**, 597 (1958); R. W. Richardson and N. Sherman, Nucl. Phys. **52**, 221 (1964); D. J. Thouless, *The Quantum Mechanics of Many-Body Systems*, 2nd ed. (Academic, New York, 1972), Chap. 7, Sect. 4, Eq. (7.26).
11. J. S. Waugh, Mol. Phys. **95**, 731 (1998).
12. M. G. Rudavets and E. B. Fel'dman, Pis'ma Zh. Éksp. Teor. Fiz. **75**, 760 (2002) [JETP Lett. **75**, 635 (2002)].
13. U. Haeberlen, *High Resolution NMR in Solids, Selective Averaging* (Academic, New York, 1976).
14. A. Ya. Khinchin, *Mathematical Foundations of Statistical Mechanics* (Gostekhizdat, Moscow, 1943; Dover, New York, 1949).
15. A. Abragam, *The Principles of Nuclear Magnetism* (Clarendon Press, Oxford, 1961; Inostrannaya Literatura, Moscow, 1963).
16. L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Pergamon, New York, 1960), p. 43.
17. L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Pergamon, New York, 1960), Chap. 9, Sect. 63, Ex. 1.
18. A. Messiah, *Quantum Mechanics* (Wiley, New York, 1964; Nauka, Moscow, 1978), Vol. 2, Chap. 13.
19. H. A. Bethe, *Intermediate Quantum Mechanics* (Benjamin, New York, 1964; Mir, Moscow, 1965), Chap. 8, Eqs. (8.33) and (8.34).
20. A. Messiah, *Quantum Mechanics* (Wiley, New York, 1964; Nauka, Moscow, 1978), Vol. 1, Chap. 12, Sect. 15; Vol. 2, Chap. 13, Eq. (105).
21. M. P. Brenner, S. Hilgenfeldt, and D. Lohse, Rev. Mod. Phys. **74**, 425 (2002).
22. E. A. Brujan, Europhys. Lett. **50**, 437 (2000).
23. W. Zhang and D. G. Cory, Phys. Rev. Lett. **80**, 1324 (1998).
24. P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953; Inostrannaya Literatura, Moscow, 1958 and 1959), Vols. 1 and 2.

Journal of Experimental and Theoretical Physics, Vol. 98, No. 2, 2004, pp. 220–226.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 2, 2004, pp. 247–255.

ATOMS, SPECTRA, RADIATION

Second- and Third-Harmonic Generation by Carbon Nanotubes Irradiated with Femtosecond Laser Pulses

D. A. Akimov[a], M. V. Alfimov[b], S. O. Konorov[a], A. A. Ivanov[b], S. Botti[c], A. A. Podshivalov[d], R. Ciardi[c], L. De Dominicis[c], L. S. Asilyan[c], R. Fantoni[c], and A. M. Zheltikov[a,d]

[a]*Department of Physics, Moscow State University, Vorob'evy gory, Moscow, 119992 Russia*

[b]*Center of Photochemistry, Russian Academy of Sciences, ul. Novatorov 7a, Moscow, 117421 Russia*

[c]*Advanced Physics Technologies Department, C.R. Frascati, ENEA, Via Enrico Fermi 45, Frascati, I-00044 Italy*

[d]*International Laser Center, Moscow State University, Vorob'evy gory, Moscow, 119992 Russia*

e-mail: zheltikov@top.phys.msu.su

Received June 18, 2003

Abstract—Femtosecond pulses of a Cr:forsterite laser are used to study second- and third-harmonic generation in a layer of single-wall carbon nanotubes produced by low-velocity spraying. The harmonic amplitude in our experiments scales as $(I_p)^n$ as a function of the pump intensity I_p, with $n = 2$ and 3 for the second and third harmonics, respectively. This scaling law holds up to pump intensities on the order of 10^{12} W/cm^2. The ratio of the maximum signal to the averaged background in the spectra of the second and third harmonics is estimated as 50 and 30, respectively. The second and third harmonics produced by a linearly polarized pump field are also linearly polarized, with their polarization vectors oriented along the polarization direction of the pump field. The capabilities of nonlinear-optical methods for structural and morphological analysis of carbon nanotubes are discussed, as well as ways to create solid-state carbon-nanotube generators of optical harmonic. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The nonlinear optics of carbon nanotubes (CNTs) is a new growing field of research that integrates advances in solid-state physics, laser science, photonics, physics of low-dimensional structures, nanoscale optics, and nanotechnologies. Research in this field is strongly motivated by the rapid progress in CNT technologies, opening ways of creating new materials with unique properties, including ultrahigh strength and broadly tunable electric conductivity [1–4]. Theoretical studies predict strong optical nonlinearities of CNTs [5–7], suggesting the possibility of using these nanoscale systems for the generation of optical harmonics, including harmonics of high orders [9, 10]. Results of experiments devoted to the investigation of nonlinear-optical phenomena in CNT systems and analysis of nonlinear-optical properties of CNTs are briefly summarized in the table. Until recently experimental efforts were mainly focused on optical limiting in CNT suspensions and CNT–polymer composite materials [11–16]. Experiments on degenerate four-wave mixing [17, 18] show that CNTs offer much promise for the creation of new nonlinear-optical materials and the development of switching and limiting photonic devices. Experiments performed in the past few years have demonstrated (see table) a high potential of CNTs as a nonlinear material for ultrafast optics and photonics. Investigation of the optical Kerr effect on the femtosecond time scale in CNT solutions [19] has revealed strong fast-response optical nonlinearities of CNTs. Chen *et al.* [20] have demonstrated ultrafast optical switching in CNT–polymer composite materials. Stanciu *et al.* [21] report third-harmonic generation (THG) in reflection of femtosecond Cr:forsterite-laser pulses from a CNT sample. The third-harmonic signal displayed saturation under conditions of experiments [21] with sufficiently low pump intensities on the order of 10^{10} W/cm^2.

In this work, we experimentally study generation of the second and third harmonics by femtosecond pulses of a Cr:forsterite laser in a layer of single-wall carbon nanotubes produced by low-velocity spraying. Based on the results of these experiments, we will discuss new possibilities of structural and topological analysis of CNTs using optical harmonic generation.

2. PREPARATION OF CARBON-NANOTUBE SAMPLES

Carbon-nanotube samples were produced with the use of the recently developed technique of low-velocity spraying [22, 23]. One of the main advantages of this technique is that carbon nanoparticles produced by

Catalog of nonlinear-optical phenomena experimentally demonstrated in CNT materials

Process	Nonlinear susceptibility	CNT samples	Laser	Pulse duration	References
Optical limiting	$\chi^{(3)}(\omega; \omega, -\omega, \omega)$	suspensions and polymer composites	mostly Nd:YAG laser	nano- and picosecond pulses	[11–16]
Degenerate four-wave mixing	$\chi^{(3)}(\omega; \omega, -\omega, \omega)$	suspension	Nd:YAG laser	8 ns, 30 ps	[17]
		solid-state sample	Nd:YAG laser	nanosecond pulses	[18]
Optical Kerr effect	$\chi^{(3)}(\omega; \omega, -\omega, \omega)$	solution	Ti:sapphire laser	120 fs	[19]
Optical switching	$\chi^{(3)}(\omega; \omega, -\omega, \omega)$ $\chi^{(3)}(\omega_2; \omega_2, -\omega_1, \omega_1)$	polymer composites	Fiber laser	150 fs	[20]
Third-harmonic generation	$\chi^{(3)}(3\omega; \omega, \omega, \omega)$	solid-state sample	Cr:forsterite	160 fs	[21]
				75 fs	this work
Second-harmonic generation	$\chi^{(3)}(2\omega; \omega, \omega)$	solid-state sample	Cr:forsterite	75 fs	this work

laser-induced pyrolysis at the first stage of this process are then used as a solid precursor for CNT growth with no metal catalysts, which are usually employed in standard CNT technologies [24, 25]. No additional purification is required for CNT samples produced with the use of this technique.

Technologically, the procedure of CNT-sample preparation by low-velocity spraying consists of two steps. The first step involves the generation of carbon nanoparticles by laser-induced pyrolysis of an ethylene–acetylene mixture. This process, initiated by CO_2-laser radiation, yields carbon nanoparticles with a mean size of 50 ± 20 nm. At the second step, a beam of carbon nanoparticles passes through a nozzle in an argon gas jet, expanding to a vacuum chamber evacuated to a pressure of 10^{-6} atm. The gas flow rate was kept constant at a level of 30 m/s. In the vacuum chamber, carbon nanoparticles were deposited on a (100) surface of a silicon substrate. The substrate was then heated, which resulted in a self-organized growth of single-wall carbon nanotubes.

The morphology and the composition of CNT samples were analyzed with the use of a scanning electron microscope (Jeol 5400) with a spatial resolution of about 3 nm and by means of Raman scattering. The diameter of single-wall CNTs in the samples under study ranged from 0.9 up to 1.5 nm. The CNT length exceeded 1 μm. Analysis of scanning electron microscope images (Fig. 1) shows that CNTs tend to form bundles in our samples with a typical diameter of about 30 nm.

3. THE LASER SYSTEM

The laser system employed in our experiments (Fig. 2) consisted of a Cr^{4+}:forsterite master oscillator, a stretcher, an optical isolator, a regenerative amplifier, and a compressor. The master oscillator, pumped with a fiber ytterbium laser, generated 30–50-fs light pulses with a repetition rate of 120 MHz. The central wavelength of this laser radiation was 1250 nm with a bandwidth of 26 nm and the mean power of about 180 mW.

Horizontally polarized 30–50-fs pulses were then stretched up to 700 ps in a grating stretcher (Fig. 2). Upon passing through a Faraday isolator and a λ/4 plate, the light pulses became vertically polarized. These pulses were then transmitted through a broadband polarizer to be injected in the regenerative amplifier at the moment of time corresponding to maximum population inversion, created by pump pulses with a repetition rate of 1 kHz. A switch was used to set a horizontal polarization of pulses injected into the cavity of the amplifier. An amplified pulse with an energy of 100 μJ was coupled out of the amplifier through the switch, triggered at the moment of time corresponding to optimal amplification. Radiation coming out of the amplifier was vertically polarized again. The amplified pulse was returned to the isolator along the same optical path. Radiation passing through the isolator in the backward direction experienced no change in its polarization since polarization rotations introduced by the λ/4 plate and the Faraday isolator compensate for each other. The pulses coupled out of the isolator through the broadband polarizer were transmitted through a λ/2 plate and compressed to a 75-fs duration in a grating compressor. Approximately 50% of the pulse energy was lost at this stage. Radiation generated by the Cr:forsterite laser system was focused onto a CNT film deposited onto a glass substrate (Fig. 3).

4. INTERACTION OF LASER RADIATION WITH CARBON NANOTUBES

The quasi-one-dimensional structure of carbon nanotubes is the key to understanding the regimes of inter-

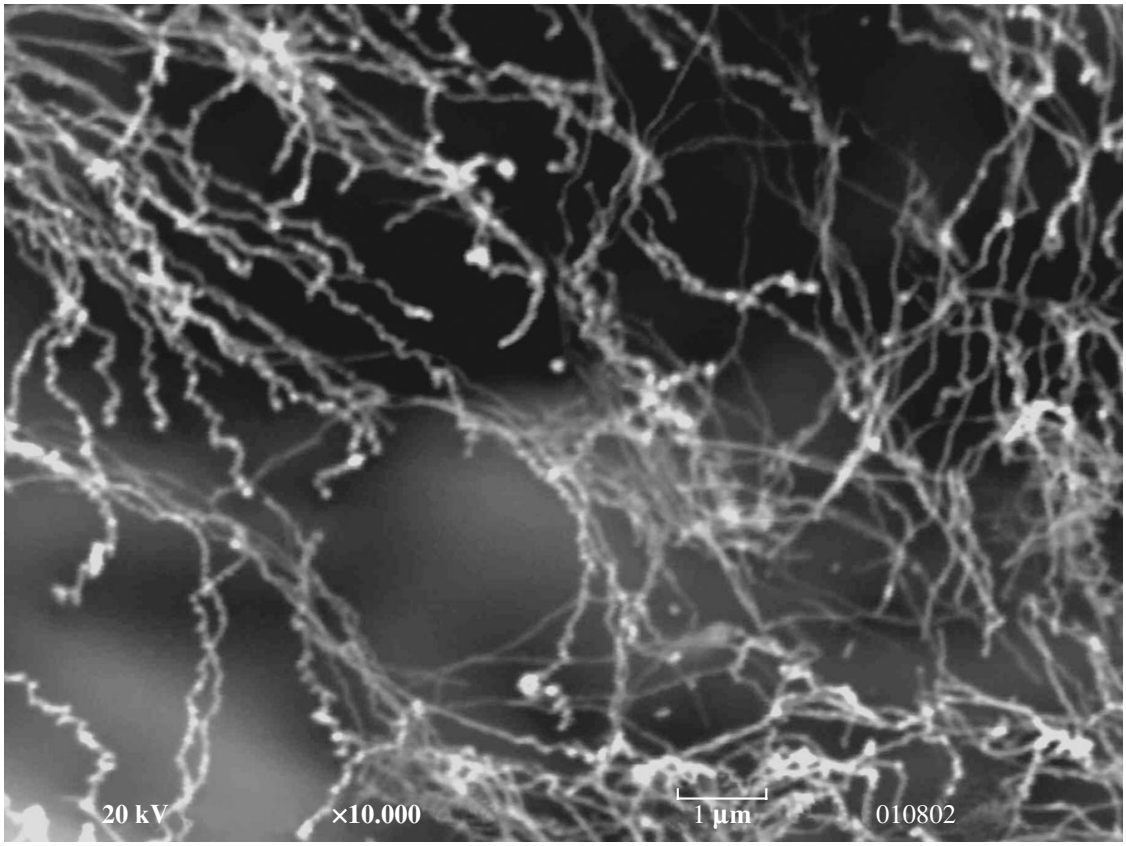


Fig. 1. Scanning electron microscope image of a carbon-nanotube sample. The spatial scale marker corresponds to 1 μm.

action of laser radiation with CNT systems. Quantum confinement of electrons in directions perpendicular to the nanotube axis is manifested in the density spectrum of electron states [26], displaying van Hove singularities (Fig. 4). Fluorescence and Raman studies [26, 27] indicate a strongly resonant character of the interaction of laser radiation with CNTs, revealing the significance of optical van Hove transitions between electron states in the valence and conduction bands (Fig. 4).

The use of strong optical nonlinearities of CNT systems for the generation of reliably detectable harmonic signals, resulting from the nonlinear coherent scattering of pump radiation from CNTs and carrying information on the structure and physical properties of CNTs, is at the heart of our harmonic-generation experiments. Optical harmonics were generated in our experiments using pump radiation with a wavelength of 1.25–1.27 μm produced by a femtosecond Cr:forsterite laser. Within this wavelength range, scattering of radiation by CNT systems can be enhanced due to a series of one-photon resonances corresponding to $v_1 \longrightarrow c_1$ transitions between the electron states in the valence and conduction bands (Fig. 4) with CNT indices (10, 3), (10, 5), (11, 1), (8, 7), (13, 2), and (9, 5), as well as two-photon resonances corresponding to $v_2 \longrightarrow c_2$ transitions (Fig. 4) with CNT indices (10, 3), (7, 5), (11, 1), and others.

Absorption spectra of CNT samples employed in our experiments display clearly resolved peaks at 1.28 and 2.2 eV (Figs. 5a, 5b). These peaks are attributed to optical van Hove transitions. Pump radiation used in our experiments is thus red-detuned from both one- and two-photon resonances with the frequency of optical transitions characteristic of the predominant type of CNTs in the samples. However, the optical density of our CNT systems at the energy of 2 eV, corresponding to the exact two-photon resonance under conditions of our experiments (Fig. 5b), is only a few percent lower than the optical density at the maximum of the absorption spectrum. Our samples are, therefore, characterized by a sufficiently high content of CNTs with electron-state spectra meeting conditions of a two-photon resonance with the frequency of a Cr:forsterite laser (Figs. 4, 5).

The efficiency of nonlinear-optical processes, including second-harmonic generation (SHG) and THG, increases with the growth in the pump radiation intensity. Optical breakdown, however, imposes a limitation on the pump intensity. In the case of shorter pulses, the efficiency of nonlinear-optical interactions in solids can often be increased due to higher intensities corresponding to laser fluences at the threshold of optical breakdown. To illustrate this possibility, we use the following qualitative arguments.

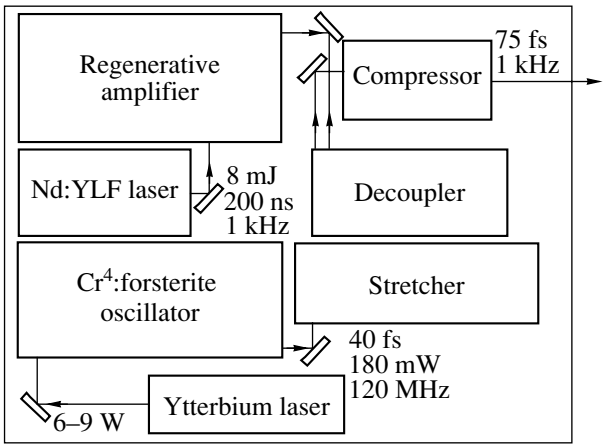


Fig. 2. Diagram of a Cr:forsterite femtosecond laser system with a regenerative amplifier.

Let us represent the intensities of the second and third harmonics, I_{SH} and I_{TH}, perturbatively generated by a pump pulse with an intensity right below the breakdown threshold as

$$I_{SH} \propto |\chi^{(2)}|^2 [F_{\text{th}}(\tau)/\tau]^2, \quad I_{TH} \propto |\chi^{(3)}|^2 [F_{\text{th}}(\tau)/\tau]^3,$$

where $\chi^{(n)}$ is the nth-order nonlinear-optical susceptibility, $n = 2, 3$; $F_{\text{th}}(\tau)$ is the breakdown threshold fluence; and τ is the pump pulse duration. With the scaling law

$$F_{\text{th}} \propto \tau^{1/2},$$

which is typical of a broad class of optical materials irradiated with pulses having durations $\tau > 10$ ps [28–30], the intensities of the second and third harmonics scale as

$$I_{SH} \propto 1/\tau \quad \text{for} \quad \text{SHG}$$

and

$$I_{TH} \propto 1/\tau^{3/2} \quad \text{for} \quad \text{THG}.$$

For shorter pulses, including pulses in the femtosecond range of durations, the dependence of the threshold fluence on the pulse duration becomes even slower than $F_{\text{th}} \propto \tau^{1/2}$ [29], allowing even higher intensities of the second and third harmonics to be achieved around the breakdown threshold [31]. The use of femtosecond pulses in our experiments thus provides a substantial increase in the yield of optical harmonics relative to the picosecond regime. This argument agrees well with the results of earlier experiments on four-wave mixing in CNTs [32], performed with the use of nanosecond pump pulses.

5. RESULTS AND DISCUSSION

Amplified Cr:forsterite-laser pulses were focused on a CNT film to generate the second and third harmonics. Harmonic signals were detected in our experiments in transmission geometry (Fig. 3). We investigated the spectral and polarization properties of optical harmonics and measured the harmonic yields as functions of the pump radiation intensity. The second- and third-harmonic yields scaled (Figs. 6, 7) as $(I_p)^n$, where I_p is the pump intensity and n is the harmonic number, within the range of pump intensities up to at least 10^{12} W/cm^2, indicating the perturbative regime of nonlinear-optical interactions and suggesting a convenient calibration for the second and third harmonics employed as spectroscopic probes:

$$I_n \propto |\chi^{(n)}|^2 I_p^n.$$

THG in reflection from multiwall CNTs, as demonstrated by Stanciu *et al.* [21], is saturated at pump intensities on the order of 10^{10} W/cm^2, resulting in a scaling law of $(I_p)^q$, $q < 3$, for the third-harmonic yield. The results of our measurements show that the intensity range of unsaturated increase in the harmonic yield can be extended under certain conditions up to 10^{12} W/cm^2.

The spectra of the second and third harmonics generated by 75-fs pulses of a Cr:forsterite laser passing through a CNT sample are shown in the insets to Figs. 6 and 7, respectively. Under the conditions of the experiments reported in [21], THG signal was detected against intense nonresonant background. It is important that the ratio η of the harmonic signal at the center of harmonic spectral lines to the background is 30 for

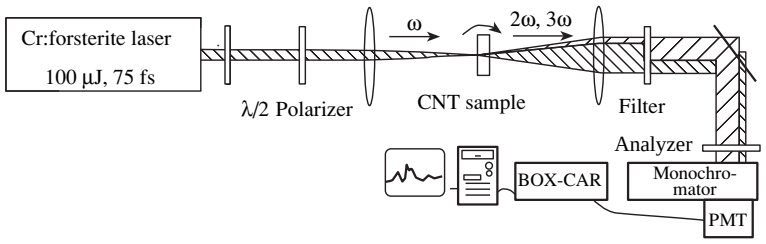


Fig. 3. Diagram of the experimental setup for second- and third-harmonic generation in carbon-nanotube samples using amplified 75-fs pulses of a Cr:forsterite laser.

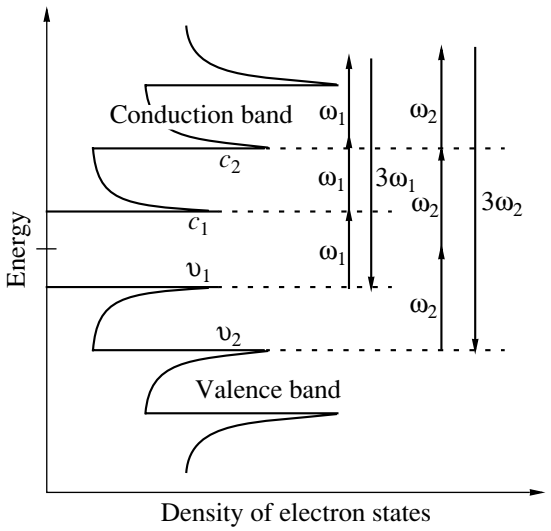


Fig. 4. Diagram of the density of electron states for a carbon nanotube, featuring van Hove singularities. Nonlinear-optical scattering, including harmonic generation, can be enhanced in a system of carbon nanotubes due to one- and two-photon resonances at the frequencies of the first ($v_1 \longrightarrow c_1$) and second ($v_2 \longrightarrow c_2$) optical van Hove transitions between the states in the valence and conduction bands.

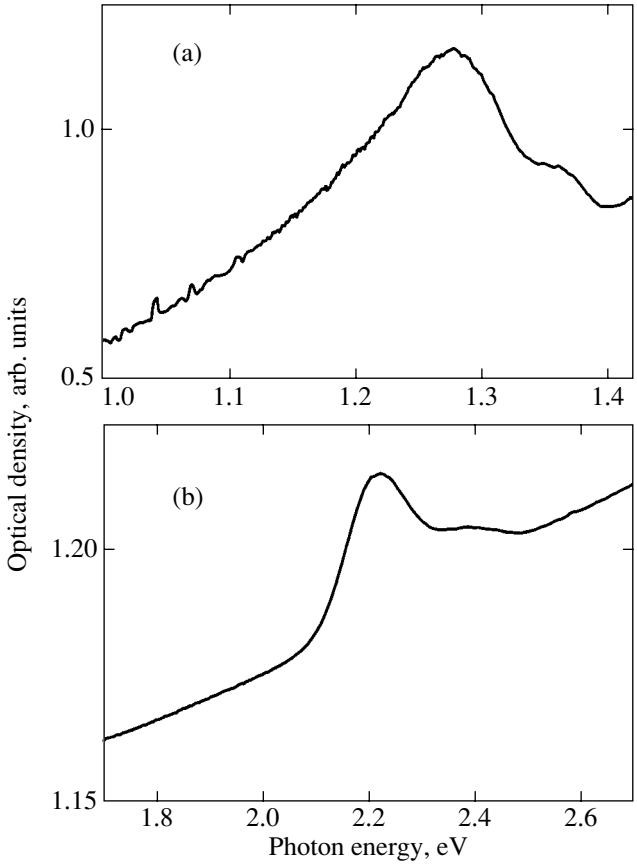


Fig. 5. Spectra of optical density measured for CNTs synthesized by low-velocity spraying within the range of photon energies (a) from 1.0 up to 1.4 eV and (b) from 1.7 up to 2.7 eV.

THG and 50 for SHG under conditions of our experiments, offering THG and SHG as sensitive techniques for CNT detection in transparent materials. The second and third harmonics produced by a linearly polarized pump field were also linearly polarized, with their polarization vectors oriented along the polarization direction of the pump field. The fact that linearly polarized pump radiation gives rise to linearly polarized second and third harmonics with a very low depolarization degree opens the ways to analyze the structure properties of CNTs via polarization measurements on the second and third harmonics.

Our experimental data, however, give no access to the absolute values of the quadratic and cubic nonlinear susceptibilities responsible for SHG and THG. The measurement of nonlinear-optical susceptibilities usually involves a calibration against the harmonic yield from a reference sample with known nonlinear susceptibility. Such a procedure, which was earlier employed to estimate the nonlinear susceptibility of CNTs in suspensions [17], becomes inapplicable in the case of CNTs on a substrate, leading to considerable errors because of spatial inhomogeneities and strong scattering in the CNT sample.

In view of an amazing diversity of carbon nanotubes, featuring different point-group symmetries and broadly tunable, structure-sensitive band gaps [33], nonlinear-optical techniques offer much promise as a tool for local probing of CNTs capable for detecting the band gap and identifying the spatial structure of nanotubes. In particular, second-harmonic generation is governed by the second-order nonlinear susceptibility $\chi^{(2)}(2\omega; \omega, \omega)$, which vanishes for centrosymmetric media. Second-harmonic generation thus allows the detection of nanotubes in a host made of centrosymmetric material with no background related to the nonlinearity of the host. The second harmonic can be generated in CNT materials through surface nonlinear-optical interactions, as well as due to nondipole nonlinear terms or the chirality of some types of CNTs. Methods of measurements distinguishing between these SHG mechanisms would allow SHG to be used to identify the type of CNTs and to detect chiral nanotubes in a sample. Ensembles of chiral CNTs are of special interest for practical implementation of concepts related to the nonlinear optics of media with broken mirror symmetry and observation of a new class of nonlinear-optical phenomena inherent in chiral materials [34–38].

The results of our experiments show that quasi-one-dimensional CNT structures offer new possibilities for optical harmonic generation. Quantum confinement gives rise to singularities in the density spectrum of electron states. Such systems possess strong nonlinearities, which can be enhanced due to one- or multiphoton resonances (Fig. 4), suggesting the ways of creating solid-state generators of optical harmonics. An illuminating and comprehensive overview of the role of one-

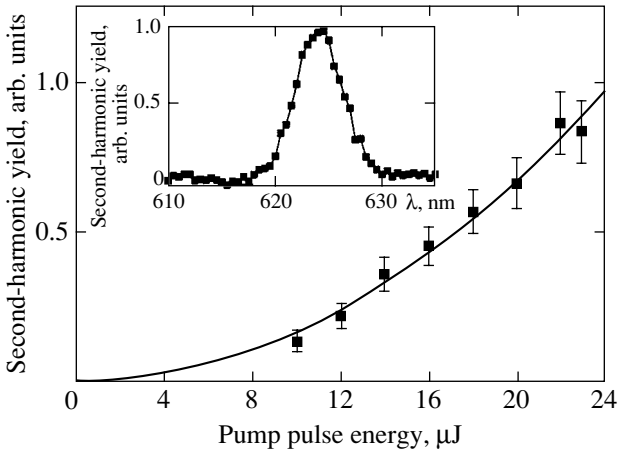


Fig. 6. The yield of the second harmonic generated in a layer of single-wall CNTs on a glass substrate in transmission as a function of the energy of a 75-fs Cr:forsterite laser pump pulse. The diameter of the focused pump beam on the surface of the CNT sample is 120 μm. The solid line represents the quadratic scaling law of the pump pulse energy, which is typical of the perturbative regime of second-harmonic generation. The inset shows the spectrum of the second harmonic.

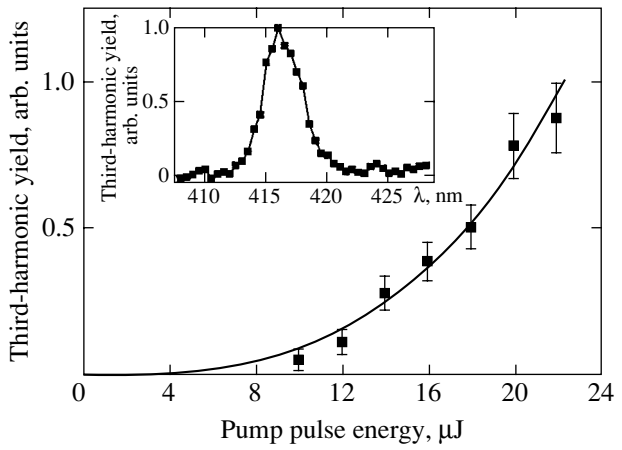


Fig. 7. The yield of the third harmonic generated in a layer of single-wall CNTs on a glass substrate in transmission as a function of the energy of a 75-fs Cr:forsterite laser pump pulse. The diameter of the focused pump beam on the surface of the CNT sample is 120 μm. The solid line represents the cubic scaling law of the pump pulse energy, which is typical of the perturbative regime of third-harmonic generation. The inset shows the spectrum of the third harmonic.

and multiphoton resonances in optical-harmonic generation is provided in classical textbooks on nonlinear optics [39, 40]. In our experimental situation, we do not expect any effects related to the interference of one- and multiphoton excitation pathways [41–44] since the conditions of one- and multiphoton resonances for the same pump frequency can be simultaneously satisfied in our case only for physically different CNTs (CNTs of different structure). The frequencies of optical van Hove transitions are determined by the CNT diameter d and the chiral angle α. In particular, the frequencies ν_i of the first ($i = 1$) and second ($i = 2$) van Hove transitions $v_1 \longrightarrow c_1$ and $v_2 \longrightarrow c_2$ (Fig. 4) are given by the following approximate expression [26]:

$$\nu_i = \frac{a}{b_i + c_i d} + \frac{A_{nm}\cos 3\alpha}{d^2},$$

where a is a constant, b_i and c_i are parameters varying for different optical transitions, and A_{nm} is the parameter depending on the CNT indices m and n. Methods of nonlinear spectroscopy can thus be employed to probe the structure and determine the sizes of CNTs.

6. CONCLUSIONS

The table puts the experimental results presented in this work in the context of earlier studies on the nonlinear optics of CNTs. Amplified 75-fs pulses of a Cr:forsterite laser were employed in our experiments to generate the second and third harmonics in a system of single-wall CNTs. The second- and third-harmonic yields in our experiments scaled as $(I_p)^n$, where I_p is the pump intensity and n is the harmonic number, within the range of pump intensities up to at least 10^{12} W/cm^2, indicating the perturbative regime of nonlinear-optical interactions and suggesting a convenient calibration for the second and third harmonics employed as spectroscopic tools in CNT systems. High-contrast second- and third-harmonic signals observed in our experiments suggest optical harmonics as a highly sensitive probe for the diagnostics of carbon nanotubes. The fact that linearly polarized pump radiation gives rise to linearly polarized second and third harmonics with a very low depolarization degree opens ways to analyze the structural properties of CNTs via polarization measurements on the second and third harmonics.

ACKNOWLEDGMENTS

This study was supported in part by the President of Russian Federation (grant MD-42.2003.02), the Russian Foundation for Basic Research (project no. 03-02-16929), and the Volkswagen Foundation (grant no. I/76 869). This material is also based upon work supported by the European Research Office of the US Army under Contract no. 62558-03-M-0033. The research of the ENEA group was supported by the ASI-ENEA contract NRIDR500110.

REFERENCES

1. M. S. Dresselhaus, G. Dresselhaus, and P. C. Eklund, *Science of Fullerenes and Carbon Nanotubes* (Academic, San Diego, 1996).
2. *Carbon Nanotubes*, Ed. by M. Endo, S. Iijima, and M. S. Dresselhaus (Pergamon, Oxford, 1996).

3. R. Saito, G. Dresselhaus, and M. S. Dresselhaus, *Physical Properties of Carbon Nanotubes* (Imperial College, London, 1998).
4. *Carbon Nanotubes: Synthesis, Structure, Properties, and Applications*, Ed. by M. S. Dresselhaus, G. Dresselhaus, and P. Avouris (Springer, Berlin, 2001).
5. R. Xie and J. Jiang, Appl. Phys. Lett. **71**, 1029 (1997).
6. X. Wan, J. Dong, and D. Y. Xing, Phys. Rev. B **58**, 6756 (1998).
7. J. Jiang, J. M. Dong, and D. Y. Xing, Phys. Rev. B **59**, 9838 (1999).
8. Vl. A. Margulis, E. A. Gaiduk, and E. N. Zhidkin, Diamond Relat. Mater. **8**, 1240 (1999).
9. G. Ya. Slepyan, S. A. Maksimenko, V. P. Kalosha, *et al.*, Phys. Rev. A **60**, R777 (1999).
10. G. Ya. Slepyan, S. A. Maksimenko, V. P. Kalosha, *et al.*, Phys. Rev. A **63**, 053808 (2001).
11. L. Vivien, E. Anglaret, D. Richl, *et al.*, Chem. Phys. Lett. **307**, 317 (1999).
12. Z. X. Jin, X. Sun, G. Q. Xu, *et al.*, Chem. Phys. Lett. **318**, 505 (2000).
13. J. E. Riggs, D. B. Walker, D. L. Carroll, and Y. P. Su, J. Phys. Chem. B **104**, 7071 (2000).
14. L. Vivien, D. Riehl, E. Anglaret, and F. Hache, IEEE J. Quantum Electron. **36**, 680 (2000).
15. L. Vivien, D. Riehl, F. Hache, and E. Anglaret, J. Nonlinear Opt. Phys. Mater. **9**, 297 (2000).
16. Xuan Sun, Yuenan Xiong, Ping Chen, *et al.*, Appl. Opt. **39**, 1998 (2000).
17. X. Liu, J. Si, B. Chang, *et al.*, Appl. Phys. Lett. **74**, 164 (1999).
18. L. De Dominicis, S. Botti, R. Fantoni, and L. Asilyan, in *Abstracts of European Conference on Nonlinear Optical Spectroscopy, Besancon, France* (2003), p. 15.
19. S. Wang, W. Huang, H. Yang, *et al.*, Chem. Phys. Lett. **320**, 411 (2000).
20. Y.-C. Chen, N. R. Raravikar, L. S. Schadler, *et al.*, Appl. Phys. Lett. **81**, 975 (2002).
21. C. Stanciu, R. Ehlich, V. Petrov, *et al.*, Appl. Phys. Lett. **81**, 4064 (2002).
22. S. Botti, R. Ciardi, M. L. Terranova, *et al.*, Appl. Phys. Lett. **80**, 1441 (2000).
23. S. Botti, R. Ciardi, M. L. Terranova, *et al.*, Chem. Phys. Lett. **355**, 395 (2002).
24. J. H. Hafner, M. J. Bronikowski, B. R. Azamian, *et al.*, Chem. Phys. Lett. **296**, 195 (1998).
25. H. M. Cheng, F. Li, X. Sun, *et al.*, Chem. Phys. Lett. **289**, 602 (1998).
26. S. M. Bachilo, M. S. Strano, C. Kittrell, *et al.*, Science **298**, 2361 (2002).
27. A. M. Rao, E. Richter, S. Bandow, *et al.*, Science **275**, 187 (1997).
28. D. Du, X. Liu, and G. Mourou, Appl. Phys. B **63**, 617 (1996).
29. An-Chun Tien, S. Backus, H. Kapteyn, *et al.*, Phys. Rev. Lett. **82**, 3883 (1999).
30. M. Lenzner, J. Kruger, S. Sartania, *et al.*, Phys. Rev. Lett. **80**, 4076 (1998).
31. S. O. Konorov, A. B. Fedotov, A. A. Ivanov, *et al.*, Opt. Commun. **224**, 309 (2003).
32. R. Fantoni, L. De Dominicis, and S. Botti, in *Book of Abstracts of 12th International Laser Physics Workshop (LPHYS'03)* (Hamburg, 2003), p. 292.
33. M. Damnjanović, I. Milosević, T. Vuković, and R. Sredanović, Phys. Rev. B **60**, 2728 (1999).
34. A. P. Shkurinov, A. V. Dubrovskii, and N. I. Koroteev, Phys. Rev. Lett. **70**, 1085 (1993).
35. P. Fischer, D. S. Wiersma, R. Righini, *et al.*, Phys. Rev. Lett. **85**, 4253 (2000).
36. M. A. Belkin, T. A. Kulakov, K.-H. Ernst, *et al.*, Phys. Rev. Lett. **85**, 4474 (2000).
37. M. A. Belkin, S. H. Han, X. Wei, and Y. R. Shen, Phys. Rev. Lett. **87**, 113001 (2001).
38. Th. Verbiest, S. Van Elshocht, M. Kauranen, *et al.*, Science **282**, 913 (1998).
39. J. F. Reintjes, *Nonlinear Optical Parametric Processes in Liquids and Gases* (Academic, New York, 1984; Mir, Moscow, 1987).
40. P. N. Butcher and D. Cotter, *The Elements of Nonlinear Optics* (Cambridge Univ. Press, Cambridge, 1990).
41. É. A. Manykin and A. M. Afanas'ev, Zh. Éksp. Teor. Fiz. **52**, 1246 (1967) [Sov. Phys. JETP **25**, 828 (1967)].
42. A. I. Maĭmistov, É. A. Manykin, and L. B. Khodulev, Zh. Éksp. Teor. Fiz. **76**, 856 (1979) [Sov. Phys. JETP **49**, 431 (1979)].
43. N. Tsukada, R. Tsujinishi, M. Nagano, and K. Tomishima, Phys. Rev. A **21**, 1281 (1980).
44. Ce Chen, Yi-Yian Yin, and D. S. Elliott, Phys. Rev. Lett. **64**, 507 (1990).

Translated by A. Zheltikov

Journal of Experimental and Theoretical Physics, Vol. 98, No. 2, 2004, pp. 227–230.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 2, 2004, pp. 256–260.

ATOMS, SPECTRA, RADIATION

Four-Photon Correlations in Parametric Down-Conversion

O. A. Ivanova and M. V. Chekhova*

Department of Physics, Moscow State University, Vorob'evy gory, Moscow, 119992 Russia
**e-mail: masha@qopt.phys.msu.su*

Received June 26, 2003

Abstract—Four-photon correlations of the output radiation of a parametric amplifier with a vacuum at the input are considered for an arbitrary parametric gain coefficient. Such states are interpreted in the literature as four-photon states. It is shown that the fourth-order correlation function for such states in the limit of a small number of photons has an asymptotics typical of two-photon states. Nevertheless, even in the "classical" limit of high intensities, the level of four-photon correlations, i.e., the value of the normalized fourth-order correlation function, is substantially greater than that for coherent and even thermal fields. *© 2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Most experiments in quantum optics are related to the generation of nonclassical light of different types, i.e., light whose properties can be described only within the framework of a consistent quantum-mechanical approach. However, there exist only a few types of nonclassical light that can be prepared experimentally at present. First of all, there is one-photon light, which is obtained via one-photon transitions in atoms [1], via luminescence of quantum dots [2, 3], and with the help of some transformation performed with two-photon light [4]. In turn, two-photon light can be obtained through two-photon transitions in atoms [5], but much more efficiently, due to spontaneous parametric down-conversion [6]. In the limit of a large number of photons, two-photon light is transformed to squeezed light, which is also nonclassical [7]. Recently, two-photon light was generated due to hyperparametric down-conversion [8]. Note also that both one-photon and two-photon states of light (belonging to the Fock states) are generated in all the cases mentioned above only in superposition with the vacuum state.

Generation of other types of nonclassical light, for example, higher order Fock states, is of interest first of all from the fundamental point of view. The applications of such states have been discussed in connection with the problem of quantum information [9] and the concept of quantum lithography [10]; however, these problems are far from being realized at present. Attempts to experimentally prepare three- and four-photon states are mainly stimulated by the so-called Greenberger–Horne–Zeilinger (GHZ) paradox [11]. The paradox appears when one attempts to classically describe the results of an interference experiment with the state having the form

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|+\,+\,+\,+\rangle + |-\,-\,-\,-\rangle) \quad (1)$$

(the four-photon GHZ state) or

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|+\,+\,+\rangle + |-\,-\,-\rangle) \quad (2)$$

(the three-photon GHZ state). Here, the symbol $|+\,+\,+\,+\rangle$ denotes the state of four photons with the right circular polarization, the symbol $|-\,-\,-\,-\rangle$ denotes the state of four photons with the left circular polarization, etc.

In [12, 13], the GHZ state was obtained from a group of four photons formed due to a random overlap of photon pairs upon parametric down-conversion. One of the photons serves as "trigger," while the three remaining photons form a state with the polarization part of type (2). Groups of three photons obtained in this way were called in many papers three-photon states. Similarly, groups of four photons appearing due to a random overlap of photon pairs are called four-photon states [14]. It is affirmed in [14] that such groups of four photons represent four-photon entangled states. Observation of four-photon interference was reported, i.e., the dependence of the counting rate of four-photon coincidences on the phase introduced between different groups of four photons (more exactly, between the pump pulses generating these groups). Note here that the presence of the interference pattern observed in coincidences of photocounts for four photons is well explained by the interference observed in coincidences of two photons and typical of two-photon light.

The question arises: can the states obtained in this way can be treated as "true four-photon states?" Obviously, the answer depends on the experiment for which the four-photon states are prepared. For example, the method described above is suitable for realization of the conditions of the three-photon GHZ paradox because it makes it possible to prepare polarization

state (2). However, it seems that parametric down-conversion cannot be used to solve the problem of observation of four-photon interference [14]. Finally, of interest is the character of four-photon correlations, i.e., a set of fourth-order intensity correlation functions or, in experimental terms, the number of coincidences of photocounts for four photons. It is from this point of view that we analyze the parametric down-conversion of light. Because it was emphasized in [14] that parametric down-conversion was stimulated, we consider the case of an arbitrary coefficient of parametric gain. Therefore, perturbation theory, which is commonly used for the description of multiphoton correlations, proves to be inapplicable.

Three cases can be distinguished in the description of parametric down-conversion, which differ from each other from the point of view of photon statistics. These are the cases of the single-mode (collinear and frequency-degenerate) regime, two-mode (nondegenerate in frequency, angle of scattering, or polarization), and four-mode regime (when, for example, scattered radiation has two frequency and two polarization modes). The latter regime was used in [14]. It is this regime that leads to the generation of two-photon Bell states, i.e., states of the type

$$\begin{aligned}\Phi^{\pm} &\equiv \frac{1}{\sqrt{2}}(|H_1H_2\rangle \pm |V_1V_2\rangle),\\ \Psi^{\pm} &\equiv \frac{1}{\sqrt{2}}(|H_1V_2\rangle \pm |V_1H_2\rangle).\end{aligned} \tag{3}$$

Here, H and V are photon states with horizontal and vertical polarizations, respectively, and subscripts 1 and 2 denote the frequency (or spatial) modes.

2. THE ONE-MODE REGIME

In this case, the parametric interaction Hamiltonian has the form

$$H = \frac{1}{2}i\hbar\Gamma(a^{\dagger 2} - a^2), \tag{4}$$

where Γ is the parametric gain, and $a^\dagger$ and a are the photon creation and annihilation operators. In the first order of perturbation theory, the state vector of the field emitted via parametric down-conversion is a superposition of the vacuum and two-photon states

$$|\psi\rangle = C_0|0\rangle + C_1|2\rangle. \tag{5}$$

However, an exact solution is given by the vector of the state that also contains—except for the two-photon state—the four-photon, six-photon states, etc.:

$$\begin{aligned}|\psi\rangle &= C_0|0\rangle + C_1|2\rangle + C_2|4\rangle + C_3|6\rangle + \ldots\\ &= \sum_{n=0}^{\infty} C_n|2n\rangle.\end{aligned} \tag{6}$$

It is convenient to characterize the number of groups of four photons by the fourth-order correlation function

$$g^{(4)} = \frac{\langle a^{\dagger 4}a^4\rangle}{\langle a^\dagger a\rangle^4}. \tag{7}$$

This function is measured by the number of coincidences of photocounts from four detectors in the experiment similar to the Brown–Twiss experiment (Fig. 1a), with the normalization to the product of the average detected intensities. The normalized fourth-order correlation function characterizes the radiation efficiency for detection of four-photon effects [15]. For example, $g^{(4)} = 1$ for coherent radiation, and $g^{(4)} = 4! = 24$ for thermal (Gaussian) radiation. In the context of this paper, of interest is the value of $g^{(4)}$ for four-photon radiation (in superposition with the vacuum), which could be obtained through the parametric decay of pump photons into groups of four photons. The state vector for such radiation has the form

$$|\Psi\rangle = C_0|0\rangle + C_1|4\rangle.$$

The fourth-order correlation function for this state is

$$g^{(4)} = \frac{6}{N^3}, \tag{8}$$

where $N \equiv \langle a^\dagger a\rangle$ is the average number of photons.

Consider now a state generated upon parametric down-conversion with the Hamiltonian (4). Within the framework of the Heisenberg approach, the correlation functions can be found exactly for any parametric gain Γ. By writing the Heisenberg equations for the creation and annihilation operators, we obtain the solution in the form

$$a(t) = a_0\cosh(\Gamma t) + a_0^\dagger\sinh(\Gamma t)$$

and similarly for the creation operator. Here, $a_0^\dagger$ and a_0 are the creation and annihilation operators at the instant $t = 0$ (or neglecting parametric interaction). The fourth-order correlation function is determined by expression (7), where $a^\dagger \equiv a^\dagger(t)$, $a \equiv a(t)$, and averaging is performed over the vacuum state. The second-order correlation function can be found similarly. As a result, we obtain

$$g_a^4(t) = 24 + 72\coth^3(\Gamma t) + 9\coth^4(\Gamma t), \tag{9}$$

$$g_a^{(2)}(t) = 2 + \coth^2(\Gamma t). \tag{10}$$

The average number of photons is $N = \sinh^2(\Gamma t)$. One can see that both fourth-order and second-order correlation functions increase infinitely at small parametric gain coefficients

$$g_a^{(4)}(t) \propto \frac{9}{(\Gamma t)^4} \propto \frac{9}{N^2}, \quad g_a^{(2)}(t) \propto \frac{1}{(\Gamma t)} \propto \frac{1}{N}$$

when $\Gamma t \longrightarrow 0$. Such asymptotics means that only the pair correlation of photons takes place: according to (8), the four-photon states should result in a faster increase in $g^{(4)}$ at small N. The asymptotics at large parametric gain coefficients ($\Gamma t \gg 1$) gives $g_a^{(4)} \longrightarrow 105$ and $g_a^{(2)} \longrightarrow 3$. Therefore, in the limit of large gains, the output radiation of a degenerate parametric amplifier should have super-Poissonian or even super-Gaussian statistics.

The statistics of the output radiation of a degenerate parametric amplifier with a vacuum at the input was studied earlier in [16], where the correlations functions of all orders were obtained in the general form, including results (9) and (10).

3. THE TWO-MODE REGIME

Consider now parametric down-conversion in the nondegenerate (two-mode) case. This case is realized through collinear frequency-nondegenerate parametric down-conversion, through noncollinear frequency-degenerate scattering, or through type II parametric down-conversion in the collinear frequency-degenerate regime. Correspondingly, photons from one pair belong to two different frequency, spatial, or polarization modes. Let us denote the creation and annihilation operators in these modes by $a^\dagger$, a and $b^\dagger$, b. Then, the interaction Hamiltonian has the form

$$H_{ab} = i\hbar\Gamma(a^\dagger b^\dagger - ab). \tag{11}$$

By solving the Heisenberg equation, we obtain

$$\begin{aligned} a(t) &= a_0\cosh(\Gamma t) + b_0^\dagger\sinh(\Gamma t), \\ b(t) &= b_0\cosh(\Gamma t) + a_0^\dagger\sinh(\Gamma t). \end{aligned} \tag{12}$$

The exact expressions for the fourth- and second-order correlation functions are

$$\begin{aligned} g_{ab}^{(4)}(t) &\equiv \frac{\langle (a^\dagger)^2 (b^\dagger)^2 a^2 b^2\rangle}{\langle a^\dagger a\rangle^2 \langle b^\dagger b\rangle^2} \\ &= 4 + 16\coth^2(\Gamma t) + 4\coth^4(\Gamma t), \end{aligned} \tag{13}$$

$$g_{ab}^{(2)}(t) = \frac{\langle a^\dagger b^\dagger a b\rangle}{\langle a^\dagger a\rangle\langle b^\dagger b\rangle} = 1 + \coth^2(\Gamma t) \tag{14}$$

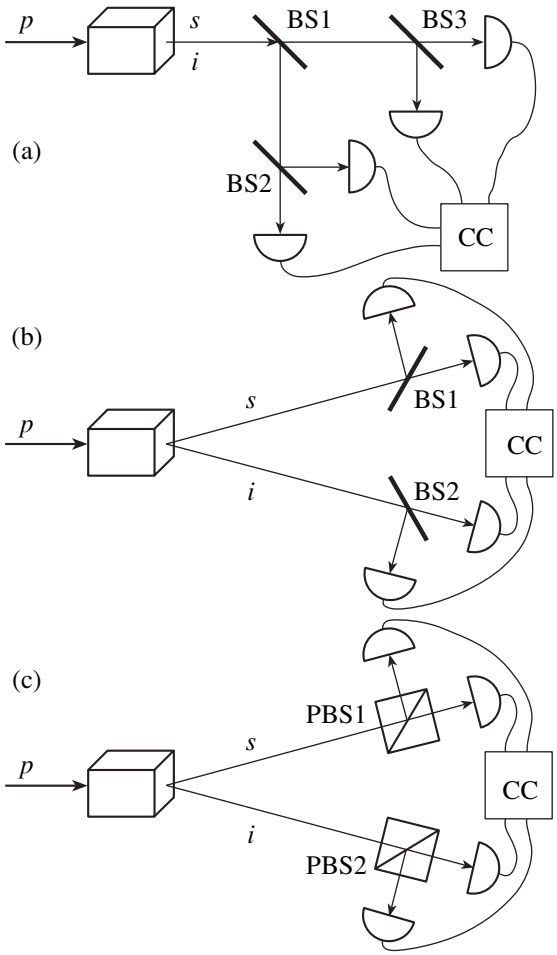


Fig. 1. Scheme of the experiment for measuring $g^{(4)}$ for the output radiation of a parametric converter. Only pump radiation (p) is incident on a nonlinear crystal; the signal (s) and idle (i) modes correspond to a vacuum. (a) The one-mode regime: the signal and idle photons belong to one spatial and one frequency mode. The fourth-order correlation function is measured using three nonpolarizing beamsplitters BS1, BS2, and BS3; four photodetectors; and the fourfold photocount coincidence scheme CC. (b) The two-mode regime: the signal and idle photons belong to different modes (in this case, different spatial modes). The fourth-order correlation function is measured using two nonpolarizing beamsplitters BS1 and BS2, four photodetectors, and the fourfold photocount coincidence scheme CC. (c) The regime of generation of the Bell states. The signal and idle photons are emitted to two spatial and two polarization modes. The fourth-order correlation function is measured using two polarizing beamsplitters PBS1 and PBS2, four photodetectors, and the fourfold photocount coincidence scheme CC.

(the scheme for measuring $g_{ab}^{(4)}$ is shown in Fig. 1b). One can see that the asymptotics of correlation functions at small parametric gains also has a two-photon character; i.e., $g_{ab}^{(4)}$ on the order of $1/N^2$ and $g_{ab}^{(4)}(t)$ on the order of $1/N$. For large parametric gains, we obtain $g_{ab}^{(4)} \longrightarrow 24$ and $g_{ab}^{(2)} \longrightarrow 2$. Such statistics could be

inherent in thermal radiation, but we should take into account that in this case two different modes *a* and *b* are involved. Therefore, the statistics can be called super-Gaussian in the two-mode case as well.

4. GENERATION OF BELL STATES

Consider now the regime of parametric down-conversion that was used in [14] and that leads, in the limit of a small parametric gain coefficient, to the generation of one of the Bell states (3), namely, state Ψ^- in superposition with a vacuum. As above, we consider the case of an arbitrary parametric gain coefficient. The interaction Hamiltonian has the form

$$H_{a_H b_V} = i\hbar\Gamma(a_H^\dagger b_V^\dagger - a_V^\dagger b_H^\dagger) + \text{H.c.} \tag{15}$$

Here, $a^\dagger$ and $b^\dagger$ are, as before, the creation and annihilation operators for the two modes, which can be frequency or spatial modes, and the subscripts *H* and *V* denote vertical and horizontal polarizations, respectively.

The solution for the creation and annihilation operators has the form

$$\begin{aligned} a_H(t) &= a_{H0}\cosh(\Gamma t) + b_{V0}^\dagger\sinh(\Gamma t),\\ b_V(t) &= b_{V0}\cosh(\Gamma t) - a_{H0}^\dagger\sinh(\Gamma t),\\ a_V(t) &= a_{V0}\cosh(\Gamma t) - b_{H0}^\dagger\sinh(\Gamma t),\\ b_H(t) &= b_{H0}\cosh(\Gamma t) + a_{V0}^\dagger\sinh(\Gamma t). \end{aligned} \tag{16}$$

For the fourth- and second-order correlation functions, we obtain

$$\begin{aligned} &g^{(4)}_{a_H b_V a_V b_H}(t)\\ &= \frac{\langle a_H^\dagger b_V^\dagger a_V^\dagger b_H^\dagger a_H b_V a_V b_H\rangle}{\langle a_H^\dagger a_H\rangle\langle b_V^\dagger b_V\rangle\langle a_V^\dagger a_V\rangle\langle b_H^\dagger b_H\rangle}\\ &= 1 + 2\coth^2(\Gamma t) + \coth^4(\Gamma t),\\ g_{a_H b_V}(t) &= \frac{\langle a_H^\dagger b_V^\dagger a_H b_V\rangle}{\langle a_H^\dagger a_H\rangle\langle b_V^\dagger b_V\rangle} = 1 + \coth^2(\Gamma t). \end{aligned} \tag{17}$$

One can see that the statistics in this case is characterized by an even lower value of four-photon correlations than in the two-mode case; nevertheless, it is super-Gaussian: $g^{(4)}_{a_H b_V} \longrightarrow 4$ and $g^{(2)}_{a_H b_V} \longrightarrow 2$ for $\Gamma t \longrightarrow 0$. The scheme for measuring the corresponding fourth-order correlation function is shown in Fig. 1c. We can also consider the case when the fourth-order moment is measured in the same regime of parametric down-conversion (Fig. 1c):

$$g^{(4)}_{a_H b_V a_H b_V}(t) = \frac{\langle a_H^{\dagger 2} b_V^{\dagger 2} a_H^2 b_V^2\rangle}{\langle a_H^\dagger a_H\rangle^2\langle b_V^\dagger b_V\rangle^2}.$$

This correlation function is equal to $4 + 16\coth^2(\Gamma t) + 4\coth^4(\Gamma t)$, as in the two-mode case.

Therefore, the state generated via parametric down-conversion is characterized by substantially weaker four-photon correlations than the true four-photon state, which could be obtained, for example, due to the decay of pump photons into groups of four photons in a medium with the fourth-order nonlinearity (of course, the probability of such a process is extremely low). Nevertheless, four-photon correlations for this state even in the limit of large parametric gains are substantially stronger than those for classical sources with Poisson or Gaussian statistics. In this sense, the strongest correlations should be observed in the degenerate regime of parametric down-conversion.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project no. 02-02-16664), the program for supporting scientific schools (NSh-166.2003.02), and INTAS (grant no. 01-2122).

REFERENCES

1. H. J. Kimble, M. Dagenais, and L. Mandel, Phys. Rev. Lett. **39**, 691 (1977).
2. E. Moreau, I. Robert, L. Manin, *et al.*, Phys. Rev. Lett. **87**, 183601 (2001).
3. A. Kiraz, S. Falth, C. Becher, *et al.*, Phys. Rev. B **65**, 161303 (2002).
4. J. G. Rarity, P. R. Tapster, and E. Jakeman, Opt. Commun. **62**, 201 (1987).
5. C. A. Kocher and E. D. Commins, Phys. Rev. Lett. **18**, 575 (1967).
6. D. N. Klyshko, Pis'ma Zh. Éksp. Teor. Fiz. **6**, 490 (1967) [JETP Lett. **6**, 23 (1967)].
7. D. N. Klyshko, Usp. Fiz. Nauk **166**, 613 (1996) [Phys. Usp. **39**, 573 (1996)].
8. P. Kumar, X. Li, M. Fiorentino, *et al.*, quant-ph/0209112; IEEE Photonics Technol. Lett. **14**, 983 (2002).
9. *Physics of the Quantum Information*, Ed. by D. Bouwmeester, A. Ekert, and A. Zeilinger (Springer, Berlin, 2000; Postmarket, Moscow, 2002).
10. A. N. Boto, P. Kok, D. S. Abrams, *et al.*, Phys. Rev. Lett. **85**, 2733 (2000).
11. D. M. Greenberger, M. A. Horne, A. Shimony, and A. Zeilinger, Am. J. Phys. **58**, 1131 (1990).
12. J.-W. Pan, D. Bouwmeester, M. Daniell, *et al.*, Nature **403**, 515 (2000).
13. J.-W. Pan, M. Daniell, S. Gasparoni, *et al.*, Phys. Rev. Lett. **86**, 4435 (2001).
14. A. Lamas-Linares, J. C. Howell, and D. Bouwmeester, Nature **412**, 887 (2001).
15. R. Glauber, Phys. Rev. **130**, 2529 (1963).
16. J. Jansky and Y. Yushin, Phys. Rev. A **36**, 1288 (1987).

Translated by M. Sapozhnikov

Journal of Experimental and Theoretical Physics, Vol. 98, No. 2, 2004, pp. 231–239.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 2, 2004, pp. 261–271.

ATOMS, SPECTRA, RADIATION

Quantum Tomography as a New Approach to Simulating Quantum Processes

A. S. Arkhipov and Yu. E. Lozovik*

Institute of Spectroscopy, Russian Academy of Sciences, Troitsk, Moscow oblast, 142190 Russia
**e-mail: lozovik@isan.troitsk.ru*

Received July 10, 2003

Abstract—A new method is proposed for *ab initio* calculations of nonstationary quantum processes on the basis of a probability representation of quantum mechanics with the help of a positive definite function (quantum tomogram). The essence of the method is that an ensemble of trajectories associated with the characteristics of the evolution equation for the quantum tomogram is considered in the space where the quantum tomogram is defined. The method is applied for detailed analysis of transient tunneling of a wave packet. The results are in good agreement with the exact numerical solution to the Schrödinger equation for this system. The probability density distributions are obtained in the coordinate and momentum spaces at consecutive instances. For transient tunneling of a wave packet, the probability of penetration behind the barrier and the time of tunneling are calculated as functions of the initial energy. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Computer simulation of quantum processes is one of the important trends in contemporary quantum physics. In spite of considerable advances made in this direction in recent years, serious difficulties also exist. In particular, the quantities used in most simulation techniques (such as the wave function and Wigner function) are not positive definite (see, for example, reviews [1, 2]). This gives rise to problems associated with convergence of integrals, which is most important for Fermi systems ("problem of sign"). This difficulty can probably be overcome by using a positive definite function describing the quantum state of the system.

Indeed, there exist representations of quantum mechanics in which the state of a system is described with the help of positive definite functions in the phase space ([3–5]; see also [1]) or in the space of extended and rotated coordinate systems. The corresponding function is known as the marginal distribution or quantum tomogram [6–11]; a remarkable feature of this function is that it is a probability distribution describing the quantum state completely [12, 13]. A quantum tomogram is a function depending on variables $\{X, \mu, \nu\}$, where $X = \mu q + \nu p$, q and p being the coordinates and momenta of the system, respectively, and μ and ν, the parameters of rotation and extension of the system in the phase space. The quantum tomogram is positive definite and normalized in X; consequently, it is a distribution function for quantity X.

Here, we propose a new method for computer simulation of nonstationary quantum processes, which is based on the tomographic formulation of quantum mechanics. In the framework of this method, the evolution of the system is described by introducing an ensemble of trajectories in space $\{X, \mu, \nu\}$, which obey the dynamic equations derived from the evolution equation for the quantum tomogram. In contrast to the simulation technique based on the numerical solution of the Schrödinger equations on a net, the methods using trajectories (in particular, in the phase space) make it possible to consider a relatively large number of degrees of freedom since there is no need in this case to store in the computer memory huge arrays of numbers representing, say, a wave function. For example, the method of "Wigner trajectories" in the phase space, which is based on the Wigner formulation of quantum mechanics [14], has been recently applied successfully for analyzing tunneling of a wave packet [15–17] as well as several identical interacting particles [17].

The proposed technique is used for solving the problem of transient tunneling of a wave packet through a potential barrier. The evolution of the probability density distribution is determined in the coordinate and momentum spaces. Transient tunneling is analyzed by calculating the probability of passage of a wave packet behind the barrier and the tunneling time. The results are compared with the exact numerical solution of the Schrödinger equation. The method is also tested by comparing with the exact solution for a wave packet in a harmonic potential (in particular, for a coherent state). The method is described in Section 2. The main results are given in Section 3 (motion in an oscillator) and Section 4 (transient tunneling). Conclusions are given in Section 5.

1063-7761/04/9802-0231$26.00 © 2004 MAIK "Nauka/Interperiodica"

2. SIMULATION TECHNIQUE

2.1. Description of Quantum Evolution with the Help of Trajectories in Space $\{X, \mu, \nu\}$

A quantum tomogram $w(X, \mu, \nu)$ is connected with a density matrix $\rho(q, q')$ via the relation [18, 19]

$$\rho(q, q') = \frac{1}{2\pi}\int w(X, \mu, q - q') \times \exp\left[i\left(X - \frac{\mu(q + q')}{2}\right)\right]d\mu dX, \tag{1}$$

$$w(X, \mu, \nu) = \frac{1}{4\pi^2} \times \int \exp\{-i[k(X - \mu q - \nu p) + p(q - q')]\} \times \rho(q, q')dpdkdqdq'. \tag{2}$$

Let us consider a particle of mass m in a one-dimensional space. If the system Hamiltonian is given by

$$H = \frac{p^2}{2m} + V(q), \tag{3}$$

we can apply integral transformation (2) to the time-dependent equation describing the evolution of the density matrix to obtain [6]

$$\dot{w} - \frac{\mu}{m}\frac{\partial w}{\partial \nu} - 2\frac{\partial V(\tilde{q})}{\partial q}\left(\frac{\nu}{2}\frac{\partial}{\partial X}\right)w + 2\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{(2n+1)!}\frac{\partial^{2n+1}(\tilde{q})}{\partial q^{2n+1}}\left(\frac{\nu}{2}\frac{\partial}{\partial X}\right)^{2n+1} w = 0, \tag{4}$$

where we assume that $\hbar = 1$ and $\tilde{q}$ is expressed as follows:

$$\tilde{q} = -\left(\frac{\partial}{\partial X}\right)^{-1}\frac{\partial}{\partial \mu}. \tag{5}$$

Equation (4) can be written in the form

$$\frac{\partial w}{\partial t} + \frac{\partial w}{\partial X}G_X(X, \mu, \nu) + \frac{\partial w}{\partial \mu}G_\mu(X, \mu, \nu) + \frac{\partial w}{\partial \nu}G_\nu(X, \mu, \nu) = 0, \tag{6}$$

where functions G depend on the quantum tomogram, its derivatives, and primitives (the primitives correspond to the terms containing $(\partial/\partial X)^{-1}$ in Eq. (4)). Functions G for the problem considered here will be given in explicit form below. Evolution equation (6) has the form of the continuity equation for the quantum tomogram:

$$\frac{dw}{dt} = \frac{\partial w}{\partial t} + \frac{\partial w}{\partial X}\dot{X} + \frac{\partial w}{\partial \mu}\dot{\mu} + \frac{\partial w}{\partial \nu}\dot{\nu} = 0. \tag{7}$$

This equation is similar to the continuity equation for the classical distribution function or to the Liouville equation whose characteristics are classical trajectories in the phase space, which obey the Hamilton equations of motion. The quantum tomogram is positive definite and will be used as the trajectory distribution function in space $\{X, \mu, \nu\}$. Comparison of Eqs. (6) and (7) shows that the equations of motion for the trajectories in the given case have the form

$$\dot{X} = G_X(X, \mu, \nu), \quad \dot{\mu} = G_\mu(X, \mu, \nu), \quad \dot{\nu} = G_\nu(X, \mu, \nu). \tag{8}$$

The ensemble of trajectories is introduced to avoid direct computation of the distribution function. For example, finite difference methods employ a net on which, for example, the wave function is specified when the Schrödinger equation is solved numerically. Such an approach often makes it possible to attain high accuracy; however, the number of elements in the array representing the wave function and the number of elementary operations corresponding to a transition to the next time "stratum" increases exponentially with the number of degrees of freedom. For this reason, it is difficult to employ the net methods of simulation even for studying a system with four or five degrees of freedom. At the same time, the methods using trajectories make it possible to consider systems consisting of thousands and more particles (e.g., in classical molecular dynamics). In modeling based on the classical molecular dynamics method, independent trajectories in the phase space are used. Since we are dealing with quantum-mechanical problems, the right-hand side of the equation of motion (8) for quantum trajectories contains a function depending on the distribution function for these trajectories (i.e., on the density of trajectories at a given point). For this reason, the trajectories in this method are generally not independent. An ensemble of trajectories is used to avoid direct calculation of the quantum tomogram on a net. However, the quantum tomogram appears in the equation of motion for the trajectories. Since we refuse to calculate the quantum tomogram in explicit form, a certain approximation of this function must be used to obtain the right-hand side of equations of motion (8). We will use the local exponential approximation (see also an analogous approximation for the Wigner function from [17] and the literature cited therein)

$$w(X, \mu, \nu) = w_0\exp\{-[(y - y_a(t))A_a(t)(y - y_a(t)) + b_a(t)(y - y_a(t))]\}, \tag{9}$$

where $y = \{X, \mu, \nu\}$, y_a is the point under study, and matrix A_a and vector b_a are the parameters of this approximation. Instead of derivatives and primitives of the quantum tomogram, evolution equation (4) now contains certain combinations of these parameters. Matrix A_a and vector b_a can be determined by calculating the mean values of X, μ, ν and their mean products in the vicinity of the given point. After this, functions G become known and Eq. (8) can be solved numerically.

Approximation (9) is not universal, but provides good results if the average distance between trajectories is not very large, and the corresponding quantum tomogram (distribution function) is smooth. For instance, this approximation might be inapplicable if we consider a plane wave with a wave vector k; in this case,

$$w(X, \mu = 0, \nu = 1) = \delta(X - k).$$

This approximation should be used with care in studying an infinite motion as well since the distance between the trajectories increases with time in this case. If the number of trajectories in the vicinity of the point at which approximation (9) is applied is too small, this approximation will not reconstruct the quantum tomogram adequately due to insufficient statistics.

We consider the motion of a wave packet in a harmonic oscillator and its tunneling from a potential well through a barrier. A comparison of the results of simulation in the quantum tomography representation with the results of exact numerical calculation shows that approximation (9) is valid for the given problems (see Sections 3 and 4). In the case of tunneling, there exist both the region of finite motion (in the potential well) and the region of infinite motion (behind the barrier); however, the approximation can be used. If the value of the initial energy is high enough, the packet tunnels through the barrier almost completely; consequently, the evolution of most trajectories corresponds to infinite motion, and approximation (9) does not provide correct solutions to quantum equations of motion any longer (see the end of Section 4). However, local approximation (9) turns out to be quite satisfactory on the whole even for quite large time intervals (see Fig. 5).

2.2. Computation of Mean Values

To analyze the problems under investigation, the mean values of several quantities are required. In the quantum tomography representation, the mean value $\langle A \rangle$ of a physical quantity (the corresponding operator is $A(\hat{q}, \hat{p})$) is calculated by the formula [20]

$$\langle A \rangle = \int A(\mu, \nu) e^{iX} w(X, \mu, \nu) dX d\mu d\nu, \qquad (10)$$

where $A(\mu, \nu)$ is the Fourier transform of the Weyl symbol $A^W(q, p)$ of operator $A(\hat{q}, \hat{p})$:

$$A(\mu, \nu) = \int A^W(q, p) \exp[-i(\mu q + \nu p)] \frac{dqdp}{4\pi^2}. \qquad (11)$$

The Weyl symbol of operator $A(\hat{q}, \hat{p})$ can be expressed as follows:

$$A^W(q, p) = \frac{\hbar}{2\pi} \int d\xi d\eta \operatorname{Tr}[A(\hat{q}, \hat{p}) e^{i\xi\hat{q} + i\eta\hat{p}}] e^{-i\xi q - i\eta p}. \qquad (12)$$

We will calculate the mean values by using the following approximation of the quantum tomogram:

$$w(X, \mu, \nu, t) = \sum_{j=1}^{J} \delta(X - X_j(t)) \delta(\mu - \mu_j(t)) \delta(\nu - \nu_j(t)); \qquad (13)$$

here, summation is carried out over all J trajectories and $X_j(t)$, $\mu_j(t)$, and $\nu_j(t)$ are the coordinates of the jth trajectory in space $\{X, \mu, \nu\}$ at instant t. Such an approximation exactly corresponds to the use of an ensemble of trajectories: in the regions where function $w(X, \mu, \nu)$ is small, the trajectories are sparse, while the density of trajectories is higher in the regions where function $w(X, \mu, \nu)$ is large. The larger the number of trajectories used, the more successful the application of approximation (13). If the wave function of a particle has the form of a compact wave packet (even if it consists of several individual parts) in the course of simulation, approximation (13) can be employed since in this case we have compact sets of trajectories in space $\{X, \mu, \nu\}$, and rich statistics can be gathered. A similar situation takes place for the problems considered here; consequently, the use of the given approximation does not distort the results significantly as compared to the exact quantum computations (see Sections 3 and 4).

We consider the quantities that depend either only on q or only on p (see Section 4). If operator $A(\hat{q})$ depends only on $\hat{q}$, the expression for $\langle A \rangle$ assumes the form

$$\langle A \rangle = \int A(X) w(X, \mu = 1, \nu = 0) dX, \qquad (14)$$

where $A(X)$ is a function corresponding to operator $A(\hat{q})$ in the coordinate representation: $A(X) = A(q = X)$. Analogously, the mean values of the corresponding physical quantities for operators $B(\hat{p})$ are given by the formula

$$\langle B \rangle = \int B(X) w(X, \mu = 0, \nu = 1) dX. \qquad (15)$$

For example, the probability densities in the coordinate

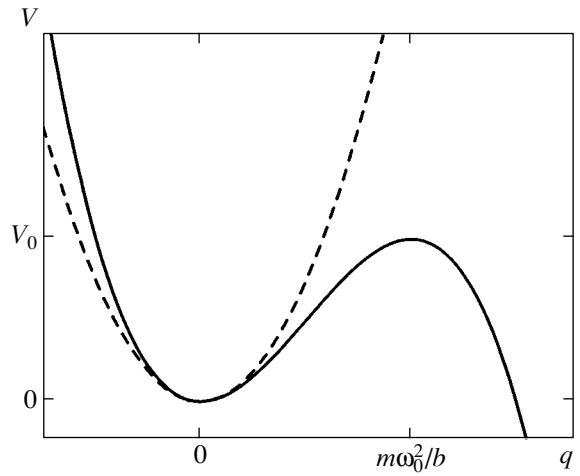


Fig. 1. External potential $V(q)$ (19): dashed and solid curves correspond to a harmonic oscillator with $b = 0$ and $b > 0$, respectively. The barrier height is $V_0 = m^3\omega_0^3/(6b^2)$.

and momentum spaces can be expresses as $w(X, \mu = 1, \nu = 0)$ and $w(X, \mu = 0, \nu = 1)$, respectively. To analyze tunneling, we will use the probability of a reaction and the tunneling time. We define the reaction probability as

$$\int_{q_a}^{\infty} |\psi(x, t)|^2 dx, \tag{16}$$

where $q_a = 0.6709$ (the point corresponding to the potential maximum). In accordance with this definition, the maximal value of the reaction probability is equal to unity. The reaction probability shows the part of the packet behind the barrier at a given instant. Let us suppose that the wave packet is localized on the left of the barrier at the initial instant and moves towards the barrier. Then the initial value of the reaction probability is zero; subsequently, as a result of passage above the barrier and tunneling, part of the packet component appears on the right of the barrier and the reaction probability increases.

The technique used here makes it possible to determine the tunneling time of a wave packet, which is an important characteristic of tunneling. The number of theoretical methods for determining the tunneling time is quite large [21–32]. We will calculate the tunneling time as the difference between the times of emergence (see [16, 33]) of particles at points x_a and x_b locates on different sides of the barrier:

$$t_T(x_a, x_b) = \langle t(x_b)\rangle - \langle t(x_a)\rangle. \tag{17}$$

The time of emergence of a particle at point x_0 is defined by the formula

$$\langle t(x_0)\rangle = \frac{\int_0^{\infty} t|\psi(x_0, t)|^2 dt}{\int_0^{\infty} |\psi(x_0, t)|^2 dt}. \tag{18}$$

2.3. Model Physical Problem

We will now consider (see also Sections 3 and 4) the motion of a wave packet in a one-dimensional space in the external potential

$$V(q) = \frac{m\omega_0^2 q^2}{2} - \frac{bq^3}{3}. \tag{19}$$

For $b = 0$, this potential corresponds to a harmonic oscillator. If $b > 0$, we have a potential well with an infinitely high wall on the left and a potential barrier on the right (Fig. 1). Since formula (19) contains the coordinate only to the second and third powers, all derivatives of the potential of an order higher than third are equal to zero and the evolution equation has the form ($\hbar = 1$)

$$\frac{\partial w}{\partial t} - \frac{\mu}{m}\frac{\partial w}{\partial \nu} + 2\left[-\frac{\partial V(\tilde{q})}{\partial q}\left(\frac{\nu}{2}\frac{\partial}{\partial X}\right) + \frac{1}{6}\frac{\partial^3 V(\tilde{q})}{\partial q^3}\left(\frac{\nu}{2}\frac{\partial}{\partial X}\right)^3\right] w = 0. \tag{20}$$

For potential (19), the evolution equation can be written as

$$\frac{\partial w}{\partial t} - \frac{\mu}{m}\frac{\partial w}{\partial \nu} + m\omega_0^2\nu\frac{\partial w}{\partial \mu} - \frac{b\nu^3}{12}\frac{\partial^3 w}{\partial X^3} + b\nu\left(\frac{\partial}{\partial X}\right)^{-1}\frac{\partial^2 w}{\partial \mu^2} = 0. \tag{21}$$

Dynamic equations (8) assume the form

$$\frac{\partial X}{\partial t} = \frac{b\nu^3}{12}\frac{1}{w}\frac{\partial^2 w}{\partial X^2},$$
$$\frac{\partial \mu}{\partial t} = m\omega_0^2\nu - \frac{b\nu}{w}\left(\frac{\partial}{\partial X}\right)^{-1}\frac{\partial w}{\partial \mu}, \tag{22}$$
$$\frac{\partial \nu}{\partial t} = -\frac{\mu}{m}.$$

3. WAVE PACKET IN A HARMONIC OSCILLATOR

Let us consider the motion of a wave packet in a harmonic oscillator. The external potential is described by formula (19) with $b = 0$. We will use the system of units $\hbar = m = \omega_0 = 1$.

For a quadratic potential, dynamic equations (22) do not contain terms with quantum tomogram w. This considerably simplifies the simulation and reduces time expenditures. As a result, the number of dynamic trajectories can be increased (as compared to the simulation of motion in a potential with $b \neq 0$) and high accuracy can be achieved.

In Figs. 2 and 3, the analytic solution of the problem on the motion of a wave packet in a harmonic oscillator (dashed curves) is compared with the results of modeling in the quantum tomography representation (solid curves); this simulation technique will be referred to as the quantum-tomographic dynamics (QTD) method. We consider the motion of a Gaussian wave packet with an initial mean coordinate of $q_0 = -1$ and a momentum of $p_0 = 0$. The dispersion of the wave packet in the coordinate space is $\sigma_x = \langle x^2 \rangle - \langle x \rangle^2 = 0.5$ (see Fig. 2), while its dispersion in the momentum space is $\sigma_p = 0.5/\sigma_x$. The wave packet vibrates in the oscillator with a period of $T = 2\pi$, expands in the coordinate space, and contracts to the initial width over a time period of $T/2$. The expansion and compression in the momentum space occur in antiphase to this process in the coordinate space. Figure 3 illustrates the evolution of the coherent state. The dispersion in the coordinate space is equal approximately to 0.7. The width of such a packet remains unchanged during its motion in the oscillator.

The results of QTD and analytic solutions are very close (see Figs. 2 and 3). Small differences appear only due to the fact that the number of trajectories and the time step for QTD are finite. Such a coincidence indicates that QTD might be an appropriate technique for a number of problems, which provides results to a high degree of accuracy. It turned out that this technique could also be used for solving more complicated problems such as those involving transient tunneling. An example of such a simulation will be considered in the next section.

4. SIMULATION OF WAVE PACKET TUNNELING

We study the tunneling of a wave packet through barrier $V(q)$ described by formula (19). We will now use the atomic units of measurement, $\hbar = m_e = |e| = 1$, where m_e and e are the mass and charge of a free electron. We consider a particle of mass $m = 2000$. The parameters of the potential are $\omega_0 = 0.01$ and $b = 0.2981$. Such a potential has a minimum at point $q = 0$ ($V(0) = 0$) and a maximum at point $q = 0.6709$ ($V(0.6709) = 0.015$). Consequently, the particle moves in a potential well with an infinitely high left wall and a barrier with a height of 0.015 at point $q = 0.6709$. This model provides a simplified description, for example, of transient tunneling of a hydrogen atom from a trap. Similar problems are important for atomic optics. For example, in order to obtain coherent beams of atoms (atomic laser), an attempt can be made to obtain a Bose condensate of atoms. Experimental methods exist for obtaining a Bose condensate in traps. A transition from a Bose condensate at rest to a moving Bose condensate is possible in principle if the atoms of the condensate tunnel beyond the trap boundaries. A theoretical description of the phenomena involved in this case can be obtained after solving the problem of transient tunneling of atoms.

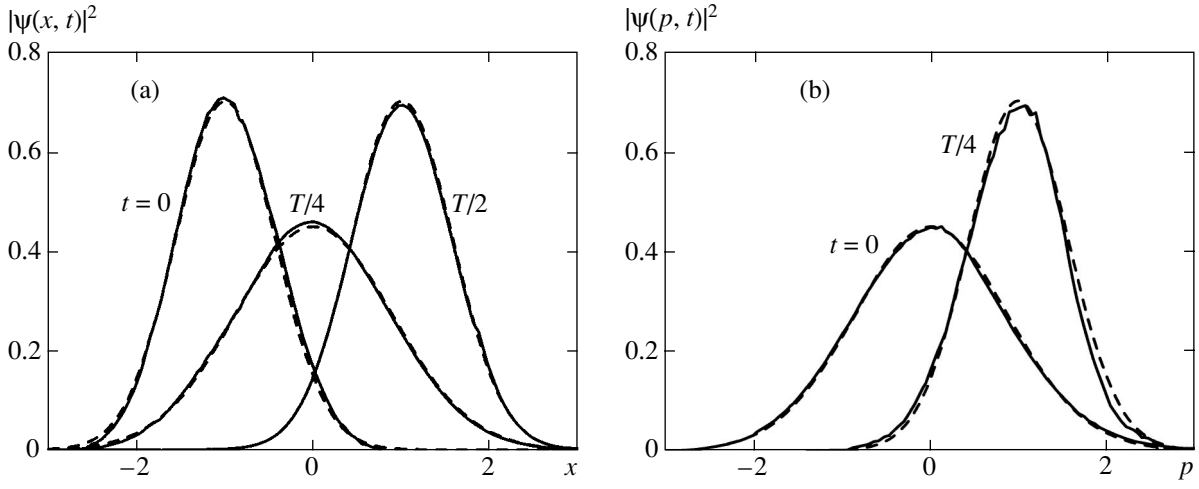


Fig. 2. Motion of a wave packet in a harmonic oscillator. The probability density in the (a) coordinate and (b) momentum spaces at instants $t = 0$, $t = T/4$, and $t = T/2$: solid curves correspond to simulation by the QTD method and dashed curves represent the analytic solution ($\hbar = m = \omega_0 = 1$).

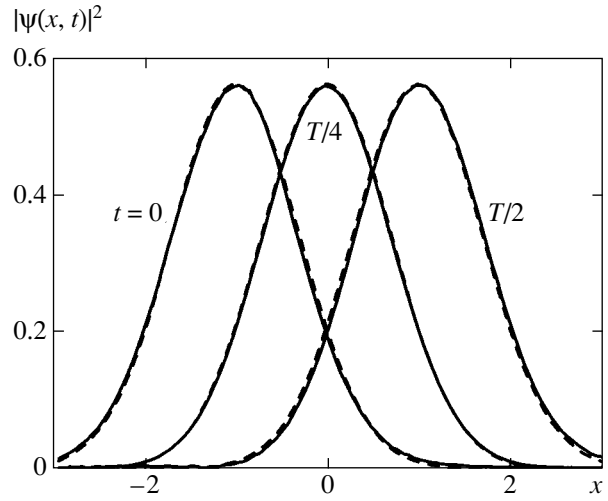


Fig. 3. Evolution of a coherent state of a harmonic oscillator. The probability density in the coordinate space at instants $t = 0$, $t = T/4$, and $t = T/2$: solid curves correspond to simulation by the QTD method and dashed curves represent the analytic solution ($\hbar = m = \omega_0 = 1$).

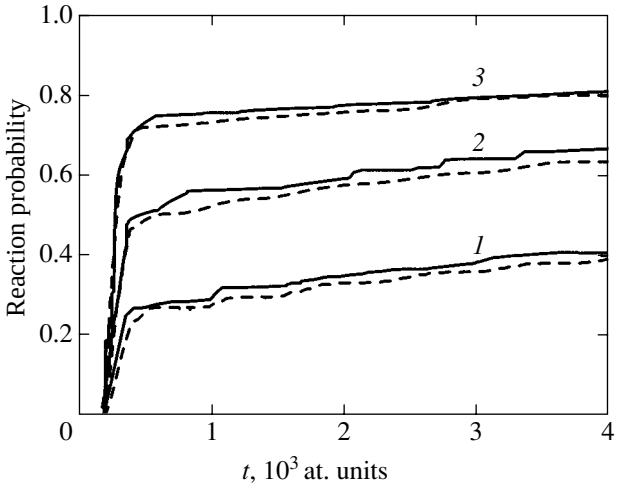


Fig. 4. Dimensionless reaction probabilities, defined by formula (16), for three values of the initial mean coordinate of a wave packet: curves *1*, *2*, and *3* correspond to $q_0 = -0.2$, -0.3, and -0.4 at. units. Solid curves correspond to simulation by the QTD method and dashed curves represent the exact numerical solution.

At the initial instant, the wave packet is on the left of point $q = 0$ and its mean momentum is zero. The particle can either vibrate in the potential well or find itself behind the barrier, having tunneled through the barrier or passed above it. The probabilities of these processes are determined by the initial energy of the wave packet. We are dealing with the problem in which all parameters are fixed except the initial mean coordinate q_0 of the wave packet (the initial mean momentum is equal to zero, and the dispersions of the wave packet in the coordinate and momentum spaces are equal approximately to 0.15 and 3.3, respectively).

4.1. Reaction Probability

Figure 4 shows the time dependences of the reaction probabilities defined by formula (16) for three values of the initial average coordinate of the wave packet: $q_0 = -0.2$, -0.3, and -0.4 with an initial mean energy of $0.75V_0$, $1.25V_0$, and $2.0V_0$, respectively. Solid curves represent the results of simulation by the QTD method, while dashed curves correspond to numerical solution of the Schrödinger equation (exact quantum-mechanical calculations). The number of wave packet components with a high energy increases due to an increase in the mean energy with $|q_0|$. As a result, the number of components behind the barrier (which either pass above or tunnel through it) increases. Consequently, with increasing $|q_0|$, the reaction probability also increases (the curves in Fig. 4 corresponding to different values of q_0 are located above one another). Qualitatively, the reaction probability exhibits the same time variation for all values of q_0 considered here. The components passing behind the barrier cannot return since the potential for $q > 0.6709$ decreases upon an increase in the coordinate; consequently, the reaction probability cannot decrease with time. First, it increases rapidly due to the above-barrier passage of component with energies larger than the barrier height (this can be verified by analyzing the classical solution of this problem, when only the above-barrier passage is possible). Then, the reaction probability continues to increase gradually due to tunneling. All these features can be seen on the curves corresponding both to QTD and to the numerical solution of the Schrödinger equation.

The reaction probability for QTD turns out to be slightly higher than in the case of exact numerical solution. A slight difference can also be noted in the modes of increase in the reaction probability for QTD and for the exact solution: in the former case, the curves are less smooth. These differences are due to the finite number of trajectories used for QTD: for a smaller number of trajectories (this case is not shown in the figure), the reaction probability resembles ladder steps even more strongly. (This is associated with a slight exaggeration of the role of wave packet oscillations in a quantum well when a finite number of trajectories are used.) The quantitative difference between QTD and exact calculation also becomes stronger. However, for a large number of trajectories (as in the case represented in Fig. 4), the QTD results are quite close to the results of exact quantum-mechanical calculations (good matching with the results obtained by the method of Wigner trajectories is also observed [15]).

4.2. Evolution of a Wave Packet and Tunneling Time

In addition to the reaction probability, we also obtained a number of new qualitative and quantitative results, which describe in detail the behavior of a wave packet in the process of tunneling.

Here and below, we consider tunneling of a wave packet with an initial mean value of coordinate of $q_0 = -0.2$. Normalized probability densities in the coordinate ($|\psi(x)|^2$) and momentum ($|\psi(p)|^2$) spaces are shown in Figs. 5 and 6, respectively. Several consecutive instants are considered. Continuous solid curves illustrate the shape of the wave packet, obtained from exact quantum-mechanical calculations. Histograms correspond to the results of a QTD run. For the same number of trajectories, several such runs can be executed and the probability density can be averaged over these runs; in this case, the curves become smoother. We will compare here the results a single QTD run with the results of exact quantum-mechanical calculations. The correspondence of the histograms (QTD) to the continuous curves (exact solution) in Fig. 5 is not ideal, but the similarity is obvious.

Let us first consider the evolution of a packet in the coordinate space (Fig. 5). Initially, the wave packet has a Gaussian shape (Fig. 5a). It moves towards the potential minimum at point $x = 0$ (the initial mean momentum is equal to zero, but the potential decreases in the direction of $x = 0$), passes through this point, and impinges on the barrier. During its motion, the wave packet expands (due to dispersion in the momentum space; cf. the left and right curves in Fig. 5a); as a result of the collision with the barrier, the shape of the wave packet changes even more strongly ($t = 300$ in Fig. 5b and $t = 400$ in Fig. 5c). The wave packet is slightly compressed and some of its component pass through the barrier so that a part of the wave packet appears on the right of the barrier ($x = 0.6709$). Since the probability of overcoming the barrier is the higher the larger the energy of the impinging particle, the transmitted part of the wave packet must be enriched by high-energy components (see below).

All characteristic details described above are present both for the exact solution and for the QTD. The histograms in Figs. 5 are in better agreement with the continuous curves representing the exact solution for earlier instants; however, even after the interaction with the barrier (Fig. 5c), the similarity remains quite obvious. This means that approximations (9) and (13) give satisfactory results for the given problem.

Let us now consider the evolution of a wave packet in the momentum space. In order to confirm the conclusion concerning the acceleration of the transmitted part of the wave packet, we will compare the probability densities ($|\psi(p)|^2$) in the momentum space at instants $t = 0$ and $t = 400$ (Fig. 6). Since the wave function first has the shape of a Gaussian wave packet, the initial probability distribution is Gaussian both in the coordinate and in the momentum space (cf. Figs. 5a and 6a)). However, after the interaction of the wave packet with the barrier, the probability distribution in the momentum space is modified considerably (Fig. 6b). The number of high-momentum components in the transmitted part of the wave packet is larger (as compared to the entire packet in the initial state) since the probability of transmission behind the barrier increases with energy. This means that the barrier serves as an energy selector

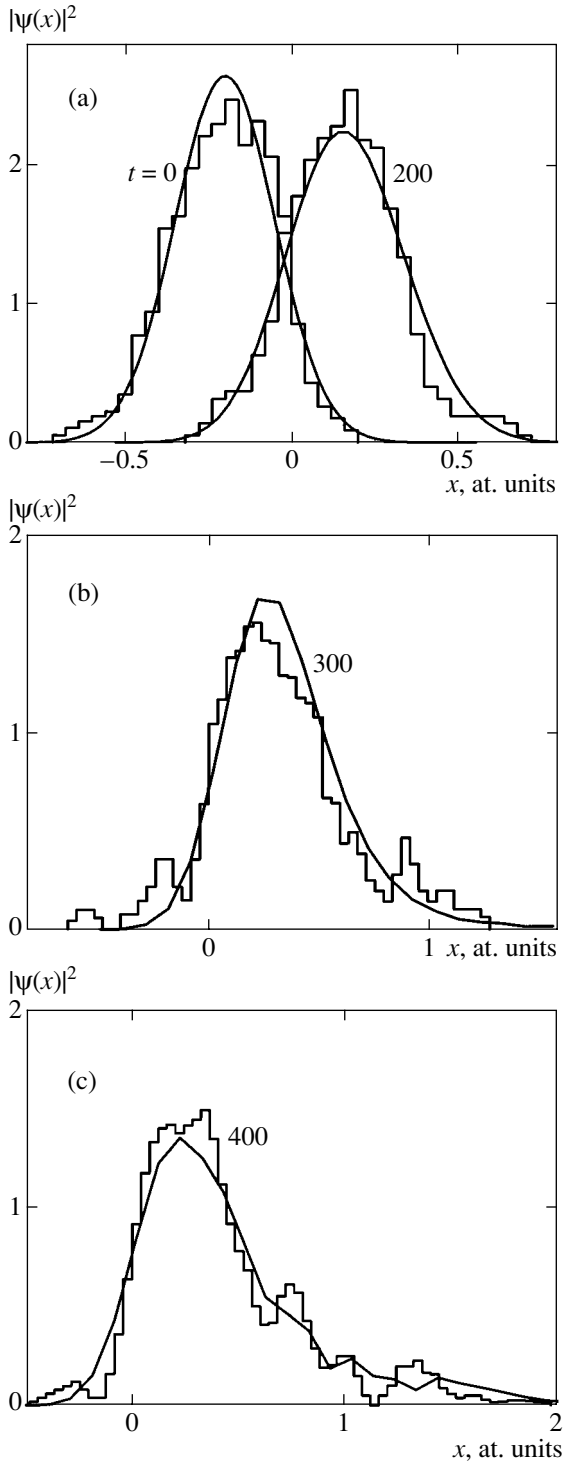


Fig. 5. Probability density in the coordinate space for QTD (histograms) and exact solution (solid curves) at instants $t = 0$, 200 at. units (a); 300 at. units (b), and 400 at. units (c). The top of the barrier is at point 0.6709 at. unit, $q_0 = -0.2$ at unit.

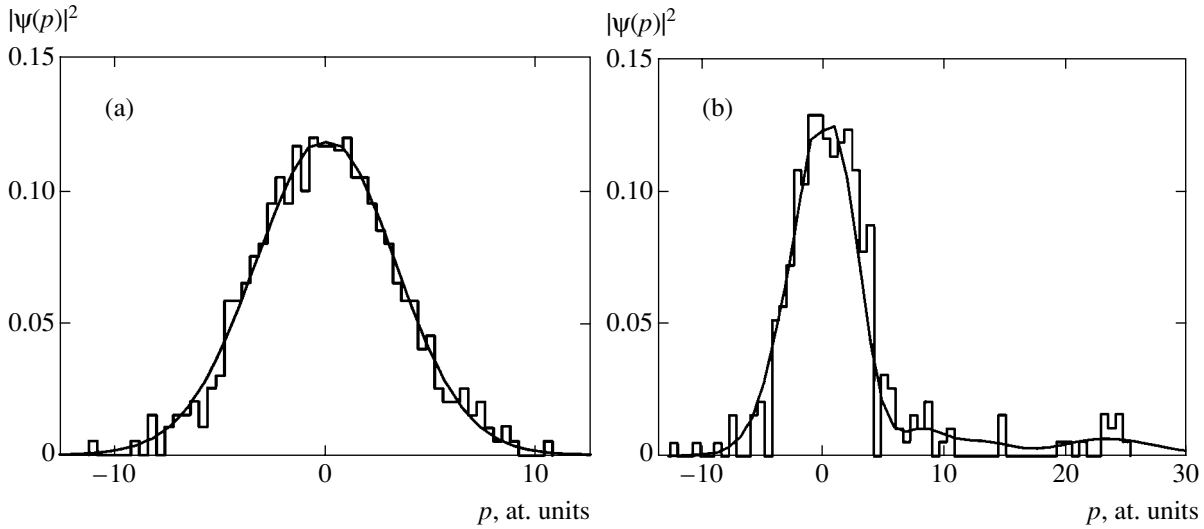


Fig. 6. Probability density in the momentum space for QTD (histograms) and exact solution (solid curves) at $t = 0$ (a) and 400 at. units (b).

and the interaction with the barrier enriches the transmitted part with high-energy components. The coincidence of the histogram (QTD) and the solid curve (exact solution) for the momentum distribution over a relatively long period of time ($T = 400$; see Fig. 6b) is much worse than for the distribution in the coordinate space (see Fig. 5c). This is due to the fact that we operate with a limited number of trajectories, while the region considered for $t = 400$ in the momentum space should be large since the transmitted part of the packet is accelerated. As a result, the momentum distribution strongly expands with time, and the number of trajectories with values of μ and ν close to $\mu = 0$ and $\nu = 1$ becomes too small for a given momentum p (see Section 2.2). For the initial value of the mean coordinate $q_0 = -0.2$ we are dealing with, the initial energy of the wave packet is not very high (approximately $0.75V_0$, where V_0 is the barrier height); consequently, its larger part remains in the quantum well (at instant $t = 400$, only 20% of the wave packet is behind the barrier; see Fig. 4). For this reason, the distribution in the coordinate space turns out to be more compact.

Figure 7 shows the dependence of the tunneling time of a wave packet on the initial value of its mean coordinate. The tunneling time is defined by formula (17) as the difference between the instants at which the packet is present at points $x_a = 0.5 \times 0.6709$ and $x_b = 2.0 \times 0.6709$ (the potential attains its maximum value at point $x = 0.6709$). For $q_0 < 0$, the larger its absolute value, the higher the initial energy $\langle \psi | \hat{H} | \psi \rangle$ of the wave packet. Usually, tunneling is the stronger, the higher the energy. An increase in the wave packet energy also corresponds to an increase in the average velocity of the transmitted part and of the entire wave packet as a whole. In this case, the part of the wave packet behind the barrier passes through the barrier region (the distance between points x_a and x_b) during a shorter time; consequently, an increase in the value of $|q_0|$ reduces the tunneling time. The discrepancy between the QTD results (squares in Fig. 7) and the results of exact computations is within the computational error. The maximum deviation from the exact result is observed for large values of $|q_0|$. This is probably due to the fact that the packet in this case passes behind the barrier almost completely (see Fig. 4) so that the evolution of most trajectories corresponds to an infinite accelerated motion. In such a situation, the trajectories diverge from one another to large distances and approximation (13) represents quantum tomogram less accurately than for smaller values of $|q_0|$.

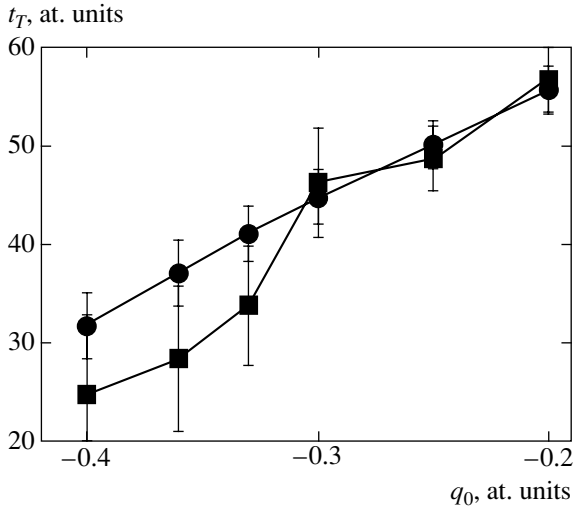


Fig. 7. Tunneling time t_T with errors for several values of the initial mean coordinate q_0 of a wave packet. The QTD results (squares) are compared with the results of exact numerical calculation (circles).

5. CONCLUSIONS

We have developed a new method for numerical simulation of nonstationary quantum-mechanical processes and applied it for studying the problem of tunneling of a wave packet through a potential barrier. The method is based on the tomographic formulation of quantum mechanics. A quantum tomogram is used, in a sense, as a distribution function for an ensemble of trajectories in space $\{X, \mu, \nu\}$, where $X = \mu q + \nu p$ is the coordinate measured in a rotated and extended frame of reference, q and p being the coordinate and momentum of the system, respectively. The trajectories evolve in time in accordance with equations resembling the Hamilton equations of motion; consequently, an analog of the molecular dynamics method can be used.

The results of simulation of the motion of a wave packet in an oscillator reproduce the analytic solution to a high degree of accuracy. The new method is used in the problem of wave packet tunneling for obtaining various quantities characterizing tunneling (probability of passage behind the barrier, tunneling time, etc.). These results are in good agreement with the results of simulation with the help of Wigner trajectories and with the results of exact quantum-mechanical calculations.

The technique developed here has two advantages in describing a quantum state since (i) it employs a real positive definite function and (ii) the computation of this function in the course of simulation is replaced by analysis of evolution of an ensemble of trajectories. The first advantage might be useful for simulating Fermi systems, when considerable difficulties ("sign problem") are associated with the sign variation of the function used for describing the quantum state. The second advantage will probably make it possible to consider a larger number of degrees of freedom as compared to net methods. For this reason, we hope that the quantum-tomographic dynamics method will become a convenient and powerful tool for simulating quantum processes.

ACKNOWLEDGMENTS

The authors are grateful to INTAS, the Russian Foundation for Basic Research, and the Ministry of Science for financial support. One of the authors (A.S.A.) thanks the "Dynasty" Foundation for a scholarship. Thanks are also due to V.I. Man'ko for fruitful discussions.

REFERENCES

1. H. W. Lee, Phys. Rep. **259**, 147 (1995).
2. D. M. Ceperley, Rev. Mod. Phys. **67**, 279 (1995).
3. K. Husimi, Proc. Phys. Math. Soc. Jpn. **22**, 264 (1940).
4. R. J. Glauber, Phys. Rev. Lett. **10**, 84 (1963).
5. E. C. G. Sudarshan, Phys. Rev. Lett. **10**, 277 (1963).
6. V. I. Man'ko, quant-ph/9902079.
7. S. Mancini, V. I. Man'ko, and P. Tombesi, Phys. Lett. A **213**, 1 (1996).
8. V. V. Dodonov and V. I. Man'ko, Phys. Lett. A **229**, 335 (1997).
9. V. I. Man'ko and O. V. Man'ko, Zh. Éksp. Teor. Fiz. **112**, 796 (1997) [JETP **85**, 430 (1997)].
10. K. Vogel and H. Risken, Phys. Rev. A **40**, 2847 (1989).
11. S. Mancini, V. I. Man'ko, and P. Tombesi, Found. Phys. **27**, 801 (1997).
12. S. Mancini, V. I. Man'ko, and P. Tombesi, J. Opt. B **7**, 615 (1995).
13. G. M. D'Ariano, S. Mancini, V. I. Man'ko, and P. Tombesi, J. Opt. B **8**, 1017 (1996); S. Mancini, V. I. Man'ko, and P. Tombesi, Europhys. Lett. **37**, 79 (1997).
14. E. Wigner, Phys. Rev. **40**, 749 (1932).
15. A. Donoso and C. C. Martens, Phys. Rev. Lett. **87**, 223202 (2001).
16. Yu. E. Lozovik and A. V. Filinov, Zh. Éksp. Teor. Fiz. **115**, 1872 (1999) [JETP **88**, 1026 (1999)].
17. Yu. E. Lozovik, A. V. Filinov, and A. S. Arkhipov, Phys. Rev. E **67**, 026707 (2003).
18. S. Mancini, V. I. Man'ko, and P. Tombesi, J. Mod. Opt. **44**, 2281 (1997).
19. V. I. Man'ko, L. Rosa, and P. Vitale, Phys. Rev. A **57**, 3291 (1998).
20. O. Man'ko and V. I. Man'ko, J. Russ. Laser Res. **18**, 407 (1997).
21. M. Büttiker and R. Landauer, Phys. Rev. Lett. **49**, 1739 (1982).
22. R. Landauer and Th. Martin, Solid State Commun. **84**, 115 (1992).
23. D. Sokolovski and J. N. L. Connor, Phys. Rev. A **47**, 4677 (1993).
24. A. I. Baz', Yad. Fiz. **5**, 229 (1967) [Sov. J. Nucl. Phys. **5**, 161 (1967)].
25. V. F. Rybachenko, Yad. Fiz. **5**, 895 (1966) [Sov. J. Nucl. Phys. **5**, 635 (1967)].
26. M. Büttiker, Phys. Rev. B **27**, 6178 (1983).
27. E. H. Hauge and J. A. Stovneng, Rev. Mod. Phys. **61**, 917 (1989).
28. M. Büttiker and R. Landauer, IBM J. Res. Dev. **30**, 451 (1986).
29. Y. Aharonov and D. Bohm, Phys. Rev. **122**, 1649 (1961).
30. J. J. Włodarz, Phys. Rev. A **65**, 044103 (2002).
31. A. D. Baute, I. L. Egusquiza, and J. G. Muga, Phys. Rev. A **65**, 032114 (2002).
32. S. Longhi, P. Laporta, M. Belmonte, and E. Recami, Phys. Rev. E **65**, 046610 (2002).
33. V. Delgado and J. G. Muga, Phys. Rev. A **56**, 3425 (1997).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 98, No. 2, 2004, pp. 240–247.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 2, 2004, pp. 272–280.

ATOMS, SPECTRA, RADIATION

An Analog of the Migdal–Kohn Singularity and the Radiation Width of the High-Frequency Branch of the Polariton Spectrum for a Bounded Crystal of the *J*-Aggregate Type

O. A. Dubovsky

Institute of Physics and Power Engineering, State Research Center of the Russian Federation, pl. Bondarenko 1, Obninsk, Kaluzhskaya oblast, 249020 Russia

e-mail: dubov@ippe.obninsk.ru

Received July 10, 2003

Abstract—The excitation spectra of crystalline ensembles of coherently emitting interacting quantum electric dipole oscillators are investigated. The system of dynamic equations derived for a one-dimensional crystal of the *J*-aggregate type can be used in various limiting cases for studying optical photons as well as X-ray and gamma quanta. An exact analytic solution to the dispersion equation is obtained for polaritons (mixed states of Frenkel excitons and transverse photons). It is shown that the high-frequency polariton branch with anomalously high radiation broadening has the limiting wave vector corresponding to the spectral edge not because the broadening becomes comparable to the frequency (as was generally accepted earlier), but due to smooth joining of this polariton branch with the other (nonphysical) branch determined from the dispersion equation. At this point, the derivative of the dispersion curve goes to infinity, which is an analog of the well-known Migdal–Kohn singularity in the phonon spectra of metals. It is shown that the low-frequency polariton branch also exhibits slight broadening due to the fact that the proper radiation width is taken into account exactly. *© 2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Spectral characteristics of coherent electromagnetic dipole radiation emitted by crystalline systems of atoms and molecules with allowed electric dipole transitions have been studied for a long time [1–4]. Coherent gamma radiation emitted by a crystalline chain of "aligned" nuclei in atoms was analyzed in [1], while the shift and broadening of the Mössbauer line in a three-dimensional crystal was investigated in [2]. In the low-frequency optical spectral region, analogous spectral characteristics for various types of crystals were studied for polaritons (coherent mixed states of Coulomb-type Frenkel excitons and transverse photons) [3, 4]. In spite of a considerable difference in the ratios of the radiation wavelength λ to the lattice constant a ($\lambda/a \ll 1$ for gamma quanta and $\lambda/a \gg 1$ for optical photons), the corresponding coherent states are limiting wave solutions to the same dispersion equation. For this reason, spectral singularities in one optical limit make it possible to judge the possibility of finding an analogous singularity in the other limit of gamma radiation. In the optical range, experimental and theoretical analysis of the polariton spectrum in bounded crystals (*J*-aggregates, molecular crystalline chains of infinitely large [5] and finite lengths [6], two-dimensional crystals of a monolayer type, stacks of planes [7, 8], superlattices [9], etc.) occupies a special place. The interest in these studies is due to the fact that, in the case of bounded crystals formed by monomers with dipole optical transitions, one of the two polariton branches determined earlier in [3, 5] for one- and two-dimensional crystals (namely, the high-frequency branch) has, due to the delayed interaction, a large radiation width determined by the probability of emission of photons into the space outside the crystal. In the model of an unbounded three-dimensional crystal, such a radiation width was not found for polaritons of the high-frequency branch [3].

It was shown in [3, 5] that the radiation half-widths $\hbar\Gamma$ for polaritons in one- and two-dimensional crystals are given by

$$\hbar\Gamma^{(1)} \approx \hbar\gamma(\lambda/a), \quad \hbar\Gamma^2 \approx \hbar\gamma(\lambda/a)^2,$$

respectively, where $\hbar\gamma$ is the natural half-width of the line emitted by an isolated dipole oscillator. For optical transitions with dipole transition energies in a monomer, which are typical of molecular crystals ($E_0 \approx 5$ eV, $\lambda \approx 10^{-6}$ cm, $a \approx 10^{-8}$ cm), the value of Γ exceeds the value of γ by several orders of magnitude. For optical transitions with a dipole transition moment of $P \approx 2e \times 10^{-8}$ cm and $\hbar\gamma = E_0^3 P^2/\hbar^3 c^3 \approx 10^{-2}$ MeV, the value of $\hbar\Gamma$ is relatively high (1–100 MeV). This effect of anomalously large broadening, which was discovered in [5], was later observed in experiments with picosecond

lasers [10, 11]. At present, this effect is known as superradiance [12].

An analysis of the polariton spectrum for one-dimensional crystals revealed [5] that, in the range of wave vectors $k < E_0/\hbar c$, for the same k, there exist two polariton branches with energies $E_k^{(r,n)}$, viz., the high-frequency branch with an energy of $E_k^{(r)} > k\hbar c$ and with the large radiation half-width $\Gamma_k^{(r)}$ mentioned above, and the low-frequency branch with an energy of $E_k^{(n)} < k\hbar c$ with zero radiation width. For $k = 0$, energy $E_k^{(r)}$ is approximately equal to the energy of a Coulomb exciton; as the value of k increases, energy $E_k^{(r)}$ decreases, approaching the optical axis $E = \hbar kc$. In this case, half-width $\Gamma_k^{(r)}$ (for the angle of inclination of dipoles to the crystal axis considered in [3, 5]) increases significantly in the vicinity of this axis.

Due to convergence of the decreasing frequency and the increasing broadening in the vicinity of the optical axis, it was assumed earlier that the concept of a quasiparticle becomes meaningless; this region of the spectral edge was not studied in detail since the dispersion curve broke.

Upon an increase in k (both in the region of $k < E_0/\hbar c$ and for $k > E_0/\hbar c$), the energy $E_k^{(n)}$ of the low-frequency branch increases monotonically, approaching E_0. It should be noted that the natural radiation half-width $\hbar\gamma$ was not taken into account explicitly in [3, 5]; for this reason, an additional parameter (the wave vector of the cutoff for one of the diverging integrals) had to be introduced in computations.

In the present study, the polariton spectrum for a one-dimensional crystal (as the first in the series of bounded crystals) will be analyzed in greater detail without resorting to any additional parameters, but taking into account natural damping of an isolated monomer for various angles of inclination of the dipole moment of the transition to the crystal axis. The corresponding dispersion equation will be derived. For the critical angle θ of inclination of the dipole moment of the transition to the crystal axis ($\theta = \theta_0 = \arccos(1/\sqrt{3})$), when the Coulomb interaction between dipoles is identically equal to zero, three exact analytic solutions to this dispersion relation will be obtained. These solutions reveal most clearly the following characteristic singularities of the polariton spectrum. The solutions with the highest and lowest frequencies correspond to polariton branches $E_k^{(r)}$ and $E_k^{(n)}$. The third solution, with a frequency intermediate between the first two frequencies, defines the branch of nonphysical states. This nonphysical branch is interesting only to the extent to which its dispersion relation determines the behavior of real polariton branches since all these branches are solutions to the same dispersion equation. The solution of the dispersion equation shows that, with increasing k, branch $E_k^{(r)}$ smoothly joins with the branch of nonphysical states for a certain $k = k_c$. At the linking point, which is the termination point of the $E_k^{(r)}$ branch, the derivative

$$\left|\frac{dE_k^{(r)}}{dk}\right| = \infty,$$

i.e., a singularity is present at the edge of the spectrum. This singularity is an analog of the Migdal–Kohn singularity in the phonon spectra of metals with a strong screening of the ion–ion interaction by the electron subsystem [13]. The anomalously large radiation half-width $\hbar\Gamma_k^{(r)}$ (superradiance) increases with k, but has a definite value at the branch termination point, which is considerably smaller than the branch energy at this point. Consequently, branch termination is determined not by the fact that the broadening becomes comparable to the energy, but by the smooth joining of this branch with that of nonphysical states, and an analog of the Migdal–Kohn singularity is observed at the linking point. The presence of this new singularity in the polariton spectrum might be of interest for subsequent experimental investigations similar to numerous experimental studies of the Migdal–Kohn singularity. It should be noted that, in contrast to [3, 5], the low-frequency polariton branch $E_k^{(n)}$ also exhibits a certain weak but finite broadening $\Gamma_k^{(r)} < \gamma$ depending on k in view of the inclusion of γ. For angles of inclination differing from θ_0, the dispersion equation obtained in this way was solved numerically because some of the Coulomb dipole–dipole sums could not be represented analytically. At a certain critical point $k = k_c$, the termination of the radiation branch with an analog of the Migdal–Kohn singularity is observed in this case as before. In general, the behavior of polariton branches and their radiation broadenings are quite similar to those mentioned above. Some quantitative changes will be discussed below. It would be definitely interesting to further investigate possible manifestations of analogs of the observed singularities in polariton spectra for bounded two-dimensional crystals in development of [7, 8] and three-dimensional finite-volume crystals in development of [2] for γ quanta.

2. DERIVATION OF THE DISPERSION EQUATION

Let us consider a one-dimensional crystal with monomers at sites na ($n = 0, \pm1, \pm2, \ldots$), oriented so that the dipole moment of the optical transition in each monomer forms an angle θ with the crystal axis. As in the case of an isolated oscillator [14], we will analyze

natural dipole oscillations of this system first in the classical approach and then in quantum theory in the representation of secondary quantization. In the classical theory of electromagnetic radiation emitted by a system of dipole oscillators, taking into account the retarded interaction at frequency ω with the time dependence $\exp(-i\omega t)$, the electric field strength $\mathbf{E}_{m\omega}(n)$ generated by the dipole at the nth site with the Fourier component $\mathbf{d}_{n\omega}$ and acting on the dipole at the mth site is given by [15]

$$\mathbf{E}_{m\omega}(n) = \Big((\mathbf{d}_{n\omega} - 3\mathbf{s}_c(\mathbf{s}_c\mathbf{d}_{n\omega})) \times\Big(-\frac{1}{|n-m|^3a^3} + \frac{i\omega/c}{|n-m|^2a^2}\Big) + (\mathbf{d}_{n\omega} - \mathbf{s}_c(\mathbf{s}_c\mathbf{d}_{n\omega}))\frac{\omega^2/c^2}{a|n-m|}\Big)\exp\Big(i\frac{\omega a}{c}|n-m|\Big). \quad (1)$$

Here, s_c is the unit vector in the direction of the crystal axis. The first term with $|n-m|^{-3}$ defines the field in the nearest Coulomb nonwave zone, the second term with $|n-m|^{-2}$ describes the field in the intermediate zone, and the third term with $|n-m|^{-1}$ defines the field in the wave zone. The dynamic equation of motion for a classical dipole oscillator as a pair of charges $(-e, e)$ with masses $(M \longrightarrow \infty, \mu)$, coupled by an elastic force with frequency ω_0 of natural oscillations, has the form [14]

$$(\omega^2 - \omega_0^2 + i2\gamma\omega_0)\mathbf{d}_{m\omega} = -T\sum_n \mathbf{E}_{m\omega}(n), \quad (2)$$

$$T = \frac{e^2}{\mu}, \quad \gamma = \frac{1}{3}T\frac{\omega_0^2}{c^3},$$

where γ represents the half-widths of the line $(\gamma < \omega_0)$ emitted by the isolated dipole oscillator, which determines its attenuation due to radiative friction [14], and T is the corresponding coupling constant determined by the same constant γ. Constants T and γ will be renormalized below in the quantum approach. Since all dipole moments are oriented in the same direction $\mathbf{d}_{n\omega} = d_{n\omega}\mathbf{s}_d$ with the unit vector $\mathbf{s}_d$ along the direction of the dipoles, the substitution of Eq. (1) into (2) after projecting Eq. (2) onto $\mathbf{s}_c$ leads to the following system of secular equations for amplitude $d_{n\omega}$:

$$(\omega^2 - \omega_0^2 + i2\gamma\omega_0)d_{m\omega} = T\sum_{n\neq m}\Big(\alpha\Big(\frac{1}{|n-m|^3a^3} - \frac{i\omega/c}{|n-m|^2a^2}\Big) - \beta\frac{\omega^2/c^2}{|n-m|a}\Big)\exp\Big(i\frac{\omega a}{c}|n-m|\Big)d_{n\omega}, \quad (3)$$

$$\alpha = 1 - 3\cos^2\theta, \quad \beta = \sin^2\theta.$$

In quantum theory, in the representation of secondary quantization, the corresponding Hamiltonian has the form [3, 16]

$$H = H_0 + H_f + H_{\text{int}}, \quad H_0 = \sum_n \hbar\omega_0 B_n^+ B_n,$$

$$H_f = \sum_{\mathbf{q}j}\hbar qc a_{\mathbf{q}j}^+ a_{\mathbf{q}j},$$

$$H_{\text{int}} = -\sum_n \mathbf{P}_n\mathbf{E}_n, \quad \mathbf{P}_n = \mathbf{s}_d P(B_n^+ + B_n), \quad P = \sqrt{\frac{\hbar e^2}{2\mu\omega_0}}, \quad (4)$$

$$\mathbf{E}_n = i\sum_{\mathbf{q}j}\mathbf{e}_{\mathbf{q}}jC_q(a_{\mathbf{q}j} - a_{-\mathbf{q}j}^+)\exp(ina(\mathbf{q}\mathbf{s}_c)), \quad C_q = \sqrt{\frac{2\pi\hbar cq}{V}}.$$

In these relations, B_n^+, B_n, $a_{\mathbf{q}j}^+$, and $a_{\mathbf{q}j}$ are the Bose operators of creation and annihilation for excitations of quantum harmonic oscillators and for photons with wave vector $\mathbf{q}$ and with two $(j = 1, 2)$ unit vectors of polarization $\mathbf{e}_{\mathbf{q}j}$ orthogonal to $\mathbf{q}$; P is the corresponding matrix element of the dipole moment operator for a transition between the state of zero-point oscillations and the first excited state; and V is the main quantization volume for photons. Quadratic Hamiltonian (4) can be diagonalized by the standard Bogoliubov–Tyablikov u, v transformation [3]. In this case, the system of equations for functions u_n, v_n, $u_{\mathbf{q}j}$, and $v_{\mathbf{q}j}$, which execute a transition from operators B_n^+, B_n, $a_{\mathbf{q}j}^+$, and $a_{\mathbf{q}j}$ to the polariton creation and annihilation operators ξ^+, ξ [3] with energy E, has the form

$$(E - \hbar\omega_0)u_n = -iP\sum_{\mathbf{q}j}C_q(\mathbf{s}_d\mathbf{e}_{\mathbf{q}j})(u_{\mathbf{q}j} - v_{-\mathbf{q}j}) \times\exp(ina(\mathbf{q}\mathbf{s}_c)), \quad (5a)$$

$$(E - \hbar qc)u_{\mathbf{q}j} = iPC_q(\mathbf{s}_d\mathbf{e}_{\mathbf{q}j}) \times\sum_m(u_m + v_m)\exp(-ima(\mathbf{q}\mathbf{s}_c)), \quad (5b)$$

$$v_n = \frac{\hbar\omega_0 - E}{\hbar\omega_0 + E}u_n, \quad v_{-\mathbf{q}j} = \frac{E - \hbar cq}{E + \hbar cq}u_{\mathbf{q}j}. \quad (5c)$$

Substituting Eq. (5b) into (5c) and then into (5a) leads

to the following equation for u_m:

$$(E^2 - \hbar^2\omega_0^2)u_n = \frac{8P^2\pi\hbar\omega_0}{V} \times \sum_{m\mathbf{q}j} (\mathbf{s}_d \cdot \mathbf{e}_{\mathbf{q}j})^2 \frac{(\hbar cq)^2 \exp(i(n-m)a(\mathbf{q}\mathbf{s}_c))}{E^2 - (\hbar cq)^2} u_m. \quad (6)$$

Passing in this equation from summation over $\mathbf{q}$ to integration,

$$V^{-1}\sum \longrightarrow (2\pi)^{-3}\int d\mathbf{q},$$

choosing retarded waves by an appropriate circumvention of a pole in the phonon Green function,

$$\frac{1}{E^2 - (\hbar cq)^2 + i\varepsilon} = P\frac{1}{E^2 - (\hbar cq)^2} - i\pi\delta(E^2 - \hbar^2c^2q^2), \quad \varepsilon \longrightarrow +0, \quad (7)$$

and calculating the corresponding integrals, we arrive at the equation

$$(E^2 - \hbar^2\omega_0^2 + i\hbar^2 2\gamma\omega_0)u_n = T\hbar^2 \sum_{m\neq n} \left(\alpha\left(\frac{1}{|n-m|^3 a^3} - \frac{iE/\hbar c}{|n-m|^2 a^2}\right) - \beta\frac{E^2/\hbar^2c^2}{a|n-m|}\right)\exp\left(i\frac{Ea}{\hbar c}|n-m|\right)u_m, \quad (8)$$

$$T = \frac{2P^2\omega_0}{\hbar}, \quad \gamma = \frac{2P^2\omega_0^3}{3\hbar c^3}.$$

This equation is a quantum analog of Eq. (3) with redefined constants γ and T. Passing from Eq. (6) to (8), we assumed for the self-action term with $n = m$ in Eq. (6) that the diverging real part defines an infinite field mass included in μ in accordance with the general approach [14]. It is exactly this procedure that was carried out implicitly in [3, 5] by introducing an additional cutoff parameter for the corresponding diverging integral. This determined the reason for terminating the radiation branch, which was formulated earlier. The imaginary component in relation (7) for the term with $n = m$ determines the natural redefined half-width $\hbar\gamma$.

In view of translation invariance, Eq. (8) has a solution in the form of a plane wave

$$u_n = u\exp(ikan)$$

with the wave vector k. For this solution, the substitution of this solution into Eq. (8) leads to the following dispersion equation for wave vector k and frequency $\omega = E/\hbar$:

$$\omega^2 - \omega_0^2 + i2\gamma\omega_0 = 2T\sum_{r=1}^{\infty}\left(\alpha\left(\frac{1}{r^3a^3} - \frac{i\omega/c}{r^2a^2}\right) - \beta\frac{\omega^2/c^2}{ra}\right) \times \cos(kra)\exp\left(\frac{i\omega ra}{c}\right). \quad (9)$$

It can easily be seen that, in the cases when the retarded interaction is ruled out (i.e., for $c \longrightarrow \infty$), only the Coulomb dipole–dipole sum determining the energy of Coulomb excitons is preserved on the right-hand side of Eq. (9) [3].

In the dimensionless variables $\Omega = \omega a$ and $K = ka$, five of the six sums,

$$S_{\omega k}^{r,m} = \sum_{r=1}^{\infty} r^{-m}\cos(\Omega r)\cos(Kr), \quad S_{\omega k}^{i,m} = \sum_{r=1}^{\infty} r^{-m}\sin(\Omega r)\cos(Kr), \quad m = 1, 2, 3, \quad (10)$$

appearing on the right-hand side of Eq. (9) can be represented analytically after certain calculations [17]:

$$S_{\omega k}^{r,1} = -\frac{1}{4}\ln[1 - \cos(\Omega - K)(1 - \cos(\Omega + K))], \quad (11a)$$

$$S_{\omega k}^{i,1} = \frac{\pi}{2}\theta(\Omega - K) - \frac{\Omega}{2}, \quad (11b)$$

$$S_{\omega k}^{r,2} = \frac{1}{2}\left(\frac{\pi^2}{3} + \frac{\Omega^2 + K^2}{2}\right) - \frac{\pi}{2}(\Omega\theta(\Omega - K) + K\theta(K - \Omega)), \quad (11c)$$

$$S_{\omega k}^{r,3} = \frac{1}{2}\left(\frac{\pi^2\Omega}{3} + \frac{\Omega^2 + 3\Omega K^2}{6}\right) - \frac{\pi}{2}\left(\frac{\Omega^2 + K^2}{2}\theta(\Omega - K) + K\Omega\theta(K - \Omega)\right), \quad (11d)$$

$$S_{\omega k}^{r,3} = -\Omega\ln 2 - L\left(\frac{\pi - \Omega - K}{2}\right) + \left(2L\left(\frac{\pi}{2}\right) - L\left(\frac{\pi - \Omega + K}{2}\right)\right)\theta(\Omega - K) + L\left(\frac{\pi - K + \Omega}{2}\right)\theta(K - \Omega). \quad (11e)$$

In these relations, the step function $\theta(x) = 1$ for $x > 0$ and 0 for $x < 0$, while $L(x)$ is the known special Lobachevsky function with the corresponding integral

representation convenient for numerical calculations [17]. It should be noted that relations (11a) and (11b) were used in [1] as the basic relations for analyzing coherent emission of gamma quanta. There exists no analytical representation for function $S_{\omega k}^{r,3}$ (Coulomb dipole-dipole sum); however, this is a rapidly converging sum which can easily be calculated numerically.

For the Green function $G_{\omega k}$ corresponding to system of equations (9), which describes the response of the system to an external action, we obviously have the following representation:

$$G_{\omega k} = \frac{1}{\omega^2 - \omega_0^2 + i2\gamma\omega_0 - W'_{\omega k} + iW''_{\omega k}}. \quad (12)$$

In accordance with Eq. (9), quantities $W'_{\omega k}$ and $W''_{\omega k}$ in this relation, which are equal to the real and imaginary parts of the self-energy component to within dimensional coefficients, have the form

$$W'_{\omega k} = -2T\left(\beta\frac{\omega^2}{c^2 a}S_{\omega k}^{r,1} - \alpha\frac{1}{a^3}S_{\omega k}^{r,3} - \alpha\frac{\omega}{ca^2}S_{\omega k}^{i,2}\right), \quad (13a)$$

$$W''_{\omega k} = 2T\left(\beta\frac{\omega^2}{c^2 a}S_{\omega k}^{i,1} - \alpha\frac{1}{a^3}S_{\omega k}^{i,3} + \alpha\frac{\omega}{ca^2}S_{\omega k}^{r,2}\right). \quad (13b)$$

It should be noted that $W''_{\omega k}$ contains the quantities $S_{\omega k}^{i,1}$, $S_{\omega k}^{i,3}$, and $S_{\omega k}^{r,2}$ having analytic representation in Eqs. (11). Substituting Eqs. (11) into (13b), we obtain

$$W''_{\omega k} = 2T\frac{1}{a^3}\left(\left(\beta\frac{\Omega^2(\pi-\Omega)}{2} - \alpha\frac{1}{2}\left(\frac{\pi(\Omega^2-K^2)}{2} - \frac{2\Omega^3}{3}\right)\right)\theta(\Omega-K) + \frac{\Omega^2}{3}\theta(K-\Omega)\right). \quad (14)$$

The dependence $\omega = \omega(k)$ of frequency ω on the wave vector for polaritons as poles of $G_{\omega k}$ is determined as a solution to the dispersion equation

$$\omega^2 - \omega_0^2 - W'_{\omega k} = 0. \quad (15)$$

The dependences $\Gamma(k)$ of radiation half-widths on k for polariton branches with a fixed k can be determined in the standard way from the corresponding frequency dependence of the imaginary part of the Green functions that determine the density of polariton states as well as the cross sections of scattering, absorption, etc.:

$$\sigma_k(\omega) = \frac{2\gamma\omega_0 + iW''_{\omega k}}{(\omega^2 - \omega_0^2 - W'_{\omega k})^2 + (2\gamma\omega_0 + W''_{\omega k})^2}. \quad (16)$$

Such an approach is most convenient when there are two converging resonances instead of one isolated resonance, and it is possible to single out one zero and the nonzero derivative corresponding to it in the expansion of G^{-1}.

3. DISPERSION RELATIONS FOR POLARITONS

In order to prove the existence of the third branch in principle, we consider the special case of the angle

$$\theta_0 = \arccos(1/\sqrt{3}) = 54°45'.$$

In this case, $\alpha \equiv 0$, $\beta = \beta_0 = 2/3$, and the Coulomb component of interaction between monomers, as well as the interaction in the intermediate region, is absent. The Eq. (15) determining the polariton frequency has the form

$$\omega^2 - \omega_0^2 - \frac{1}{2}T\beta_0\frac{\omega^2}{c^2 a} \times \ln\left(4\left(1-\cos\left(\frac{\omega a}{c} - ka\right)\right)\right) \times\left(1-\cos\left(\frac{\omega a}{c} + ka\right)\right)\Bigg) = 0. \quad (17)$$

We will seek the inverted dependence $k = k(\omega)$ instead of the dependence $\omega = \omega(k)$. We transform Eq. (17) to

$$\left(1-\cos\left(\frac{\omega a}{c} - ka\right)\right)\left(1-\cos\left(\frac{\omega a}{c} + ka\right)\right) = \frac{1}{4}\exp\left(\frac{\omega^2-\omega_0^2}{A\omega^2}\right), \quad A \equiv \frac{T\beta_0}{2c^2 a}. \quad (18)$$

The left-hand side of this equation is identically equal to $(\cos(ka) - \cos(\omega a/c))^2$, and two solutions to Eq. (18) for $k = k(\omega)$ have the form

$$k = \frac{1}{a}\arccos\left(\cos\left(\frac{\omega a}{c}\right) + \frac{1}{2}\exp\left(\frac{\omega^2-\omega_0^2}{2A\omega^2}\right)\right), \quad (19a)$$

$$k = \frac{1}{a}\arccos\left(\cos\left(\frac{\omega a}{c}\right) - \frac{1}{2}\exp\left(\frac{\omega^2-\omega_0^2}{2A\omega^2}\right)\right). \quad (19b)$$

Pay attention to the fact that the function of frequency in the argument of arccos in expression (19a) has a dip at point $\tilde{\omega}$, which can be determined from the equation

$$\frac{a}{c}\sin\left(\frac{\tilde{\omega}a}{c}\right) = \frac{\omega_0^2}{2A\tilde{\omega}^3}\exp\left(\frac{\tilde{\omega}^2-\omega_0^2}{2A\tilde{\omega}^2}\right), \quad (20)$$

since the first component in expression (19a) decreases with increasing frequency, while the second component

increases therewith. Consequently, for a fixed k smaller than a certain value of k_c determined by parameter A, there exist two frequencies $\omega_r(k)$ and $\omega_c(k)$ with $k < \omega_c(k)/c < \omega_r(k)/c$, which satisfy Eq. (19a), while no such branches exist for $k > k_c$. The highest of these frequencies, $\omega_r(k)$, corresponds to the polariton branch with energy $E_k^{(r)}$, which was found earlier. Frequency $\omega_c(k)$ with $k < \omega_c(k)/c$ is the nonphysical solution discussed above. It can be seen from formulas (19) that the pole of G for this branch has the form $\omega_c(k) - \omega + i\Gamma$, which corresponds to amplitude $\exp(-i\omega_c(k)t + \Gamma t)$ increasing with time due to the sign reversal in the corresponding derivative in solutions (19a) and (17). In addition, as will be shown below, the group velocity for this nonphysical branch exceeds the velocity of light. The third polarization branch, $\omega_n(k)$ with $k > \omega_n(k)/c$, can be determined from Eq. (19b) and is the second low-frequency branch $E_k^{(n)}$ [3, 5].

The $k = k(\omega)$ dependences (19) are shown in Fig. 1 in dimensionless variables $\nu = \omega/\omega_0$, $q = kc/\omega_0$ in the form of the direct dependence $\nu = \nu(q)$ for two values of base parameters $\omega_0 a/c = 0.01$ and $2T/ac^2 = 0.1$, which are close to experimental values. Figure 1a shows the main quadrant $0 < \nu < 1$, $0 < q < 1$ on the phase plane, where all polariton branches are present and exhibit the strongest changes. Curve *1* corresponding to the high-frequency component $\omega_r(k)$ of solution (19a) represents the radiative branch. On the given scale, curve *2* corresponding to the low-frequency component of solution (19a) almost coincides with the optical axis $\nu = q$ on the high-frequency side and strongly deviates from this axis only in the square region shown by the dashed contour, where it joins with curve *1* for infinitely large derivatives. Curve *3* representing solution (19b) corresponds to the nonradiative polariton branch $\omega_n(k)$. For small values of q, it coincides with the optical axis on the low-frequency side, but as the value of $q > 1$ increases, it tends asymptotically to $\nu = 1$ as it approaches the boundary $k = \pi/a$ of the Brillouin zone. Figure 1b clearly shows on a larger scale that curve *2* gradually splits from the optical axis upon an increase in q starting from $q = 0.8$. Since the main changes in the dispersion curves occur in the region shown by the dashed contour in Fig. 1a, it is this region that is represented in Fig. 1b on a larger scale. The numbering of branches *1*, *2*, and *3* is the same as in Fig. 1a. Straight line *4* represents the optical axis. It can be clearly seen from Fig. 1b that curve *1* corresponding to radiative polaritons and curve *2* corresponding to nonphysical states are linked at point C with infinitely large derivatives $|d\nu/dq| = \infty$. It is interesting to note that, because of the termination of branch 1 at point C, states *1* cannot be effectively excited by photons with the transverse wave vector

$$k_{tr} < \sqrt{(\omega_c/c)^2 - k^2},$$

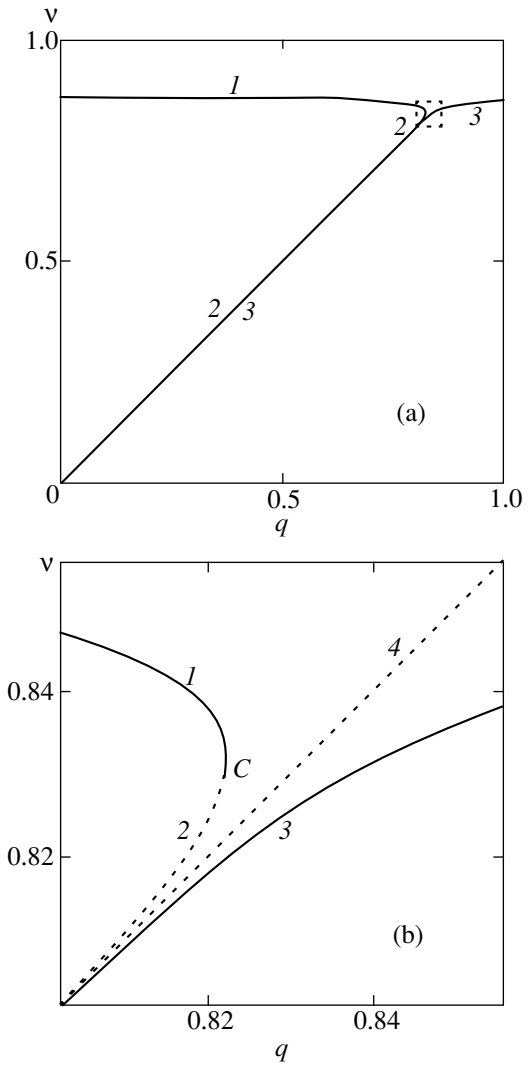


Fig. 1. Dispersion curves for polaritons: (a) dispersion curve for $\theta = \theta_0$; (b) analog of the Migdal–Kohn singularity in the radiative branch.

since the branch of such photons does not intersect curve *1*.

It was noted above that dependence $\Gamma(k)$ was determined by using the frequency dependence $\sigma_k(\omega)$. Figure 2a shows this dependence for $q = 0.81$. Broad high-frequency resonance *1* (solid curve) corresponds to radiative branch *1* in Fig. 1. Narrower resonance *2* with a lower frequency (dashed curve) corresponds to nonphysical states. The high peak (reduced by a factor of 30) with the lowest frequency and a small (but finite) width corresponds to branch *3*. Naturally, two resonances *1* and *2* should be separated and nonphysical resonance *2* should be eliminated. However, the separation of two resonances *1* and *2* at this stage in the same way as in experimental data processing is unnecessary in our opinion. The radiation half-width Γ of branch *1* was determined in the standard manner as the difference between the frequency corresponding to the peak

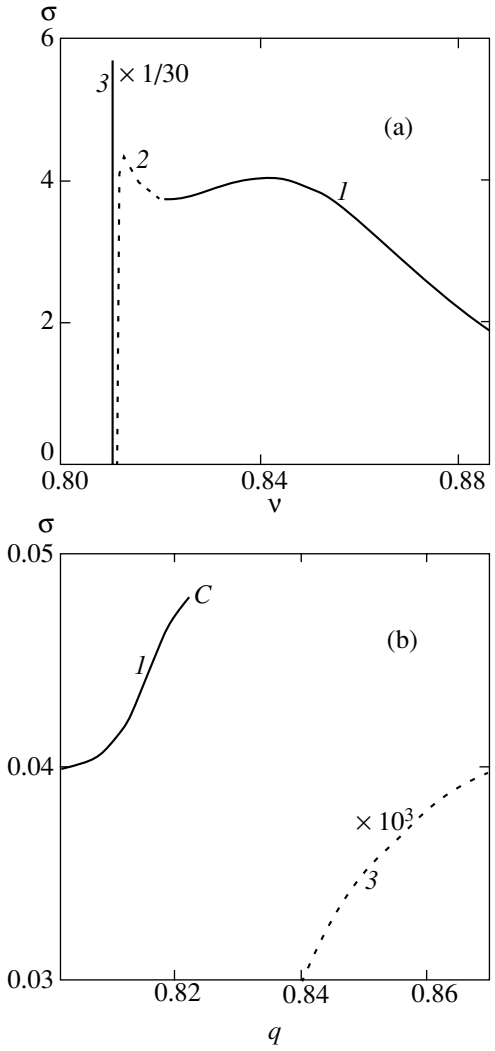


Fig. 2. (a) Frequency dependence of the density of polariton states and (b) the corresponding dependences of the radiation widths of polaritons on the wave vector.

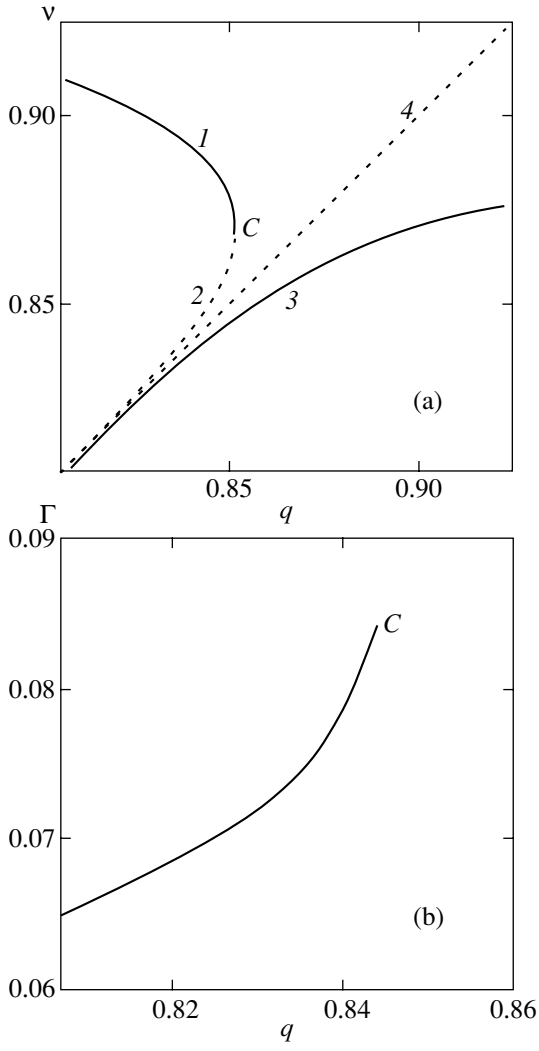


Fig. 3. (a) Dispersion curves describing polariton branches with an analog of the Migdal–Kohn singularity (b) and the corresponding radiation width of polaritons for $\theta = \pi/2$.

of resonance *1* and the frequency on the high-frequency wing of resonance *1*, at which the value of σ decreases by half as compared to the maximal value in resonance. Dependence $\Gamma(k)/\omega_0$ for the radiative branch is represented by curve *1* in Fig. 2b. It can be seen that the value of Γ increases with k. The anomalously large half-width $\Gamma_c = 4.8 \times 10^{-2}\omega_0$ at point *C*, as a manifestation of superradiance, considerably (by two orders of magnitude) exceeds the natural half-width $\gamma = 1.67 \times 10^{-4}\omega_0$. At the same time, it is small as compared to the frequency of the polariton branch at point *C*, which is equal to $\omega_r(k_C) = 0.843\omega_0$; i.e., $\gamma \ll \Gamma_c \ll \omega_r(k_c)$. Curve *3* in Fig. 2b shows the dependence of the half-width of peak *3* (conditionally speaking, nonradiative branch *3*) on k, enlarged by a factor of 10^3. The value of this half-width was defined as the difference between the frequency corresponding to the maximum of peak *3* in Fig. 2a and the frequency corresponding to the point on the low-frequency wing of the peak, at which the value of σ decreases by half. It is clear even from general considerations that the width of this branch is equal to zero for $k = 0$.With increasing k, this branch acquires an insignificant half-width smaller than γ. The value of the half-width in Fig. 3 amounts to $3.5 \times 10^{-5}\omega_0 < \gamma$; as the value of k increases, the half-width asymptotically tends to γ. The fact that polariton states with a width smaller than γ exist follows even from an analysis of a system of two oscillators [1], for which the high-frequency term from two split terms, corresponding to $k = 0$, has a width larger than γ, while the low-frequency term corresponding to $k = \pi/a$ is smaller than γ.

The analytic expressions derived above make it possible to trace the evolution of the spectra upon variation of angle θ. Figure 3 shows the dispersion curves for a crystal with $\theta = \pi/2$, studied in [3, 5]. We used the same values of the main parameters as in the previous figures. Since the energy of the Coulomb interaction has

increased considerably, the frequency of a Coulomb exciton (with $c \longrightarrow \infty$) for $k = 0$ amounts to, in accordance with relation (9),

$$\omega_{\text{Coul}} = (\omega_0^2 + 2T\zeta(3)/a^3)^{1/2} = 0.347c/a,$$

where ζ is the Rieman zeta function ($\zeta(3) = 1.202$). Accordingly, in absolute units, all dispersion curves are shifted to the region of higher frequencies and larger wave vectors as compared to the previous values. For convenience of comparison with previous results, the curves in Fig. 3 are plotted in variables normalized to ω_{Coul}; i.e., dimensionless variables $\nu = \omega/\omega_{\text{Coul}}$ and $q = kc/\omega_{\text{Coul}}$ are used. It can be seen from Fig. 3a that the situation has not changed qualitatively as compared to Fig. 1b. However, the curves in Fig. 3a display a certain quantitative shift of the entire system of dispersion branches towards higher frequencies and larger wave vectors. Figure 3b shows the dependence of q on the half-width of the radiative branch. The form of this dependence is generally the same as in Fig. 2b, but the half-width is doubled as compared to Fig. 2b. This is natural since the retarded interaction attains its maximal value for $\theta = \pi/2$, while in the wave zone, for $\theta = 0$, this interaction is absent altogether. As the value of θ decreases relative to θ_0, changes in the dispersion curves and broadening in the opposite direction as compared to Fig. 3 were observed. Obviously, as the coupling constant (corresponding to the oscillator strength of a dipole transition in an isolated monomer) decreases, the polariton branches are "pressed against" lines $\nu = 1$ and $\nu = q$, and the broadenings decrease accordingly so that the broadening at the critical point becomes much smaller than the critical frequency. Consequently, it is the above-mentioned mechanism of termination of the spectrum with an analog of the Migdal–Kohn singularity, and not the comparability of the broadening and the frequency of the radiative branch, that plays a decisive role in this case. This previously unknown singularity of the polariton spectrum might be of interest for experimental investigations similar to numerous experiments carried out for studying the Migdal–Kohn singularity.

ACKNOWLEDGMENTS

The author is obliged to express his sincere gratitude to V.M. Agranovich and Yu.E. Lozovik for valuable remarks.

This study was supported by the Russian Foundation for Basic Research (project no. 03-02-96344), the administration of Kaluzhskaya oblast, and the Russian State Scientific and Technical Program "Current Trends in the Physics of Condensed Media."

REFERENCES

1. M. I. Podgoretskiĭ and I. I. Roĭzen, Zh. Éksp. Teor. Fiz. **39**, 1473 (1960) [Sov. Phys. JETP **12**, 1023 (1960)].
2. Ch. Muzikarzh, Zh. Éksp. Teor. Fiz. **41**, 1168 (1961) [Sov. Phys. JETP **14**, 833 (1961)].
3. V. M. Agranovich, *The Theory of Excitons* (Nauka, Moscow, 1968).
4. V. M. Agranovich and V. L. Ginzburg, *Crystal Optics with Spatial Dispersion, and Excitons* (Nauka, Moscow, 1965; Springer, New York, 1984).
5. V. M. Agranovich and O. A. Dubovskiĭ, Pis'ma Zh. Éksp. Teor. Fiz. **3**, 345 (1966) [JETP Lett. **3**, 223 (1966)].
6. O. A. Dubovskiĭ, Fiz. Tverd. Tela (St. Petersburg) **40**, 2136 (1998) [Phys. Solid State **40**, 1937 (1998)].
7. V. M. Agranovich, O. A. Dubovsky, T. A. Leskova, *et al.*, Chem. Phys. **203**, 11 (1996).
8. V. M. Agranovich, O. A. Dubovsky, and D. Basko, J. Chem. Phys. **106**, 3896 (1997).
9. J. Knoester, Phys. Rev. Lett. **68**, 654 (1992).
10. Ya. Aaviksoo, Ya. Lippmaa, and T. Reinot, Opt. Spektrosk. **62**, 419 (1987) [Opt. Spectrosc. **62**, 419 (1987)].
11. B. Deveaud, F. Clerot, N. Roy, *et al.*, Phys. Rev. Lett. **67**, 2355 (1991).
12. V. M. Agranovich and S. Mukamel, Phys. Lett. **147**, 155 (1990).
13. N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Holt, Rinehart, and Winston, New York, 1976; Mir, Moscow, 1979), Vol. 2.
14. W. Heitler, *The Quantum Theory of Radiation*, 3rd ed. (Clarendon Press, Oxford, 1954; Inostrannaya Literatura, Moscow, 1956).
15. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, 6th ed. (Nauka, Moscow, 1973; Pergamon Press, Oxford, 1975).
16. A. S. Davydov, *Quantum Mechanics*, 2nd ed. (Fizmatgiz, Moscow, 1963; Pergamon Press, Oxford, 1976).
17. I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 4th ed. (Fizmatgiz, Moscow, 1963; Academic, New York, 1980).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 98, No. 2, 2004, pp. 248–253.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 2, 2004, pp. 281–287.

ATOMS, SPECTRA, RADIATION

The Contribution of the Central Region of the Electron Energy Spectrum to the Total Double Photoionization Cross Section of Helium in the Asymptotic Nonrelativistic Energy Region

A. I. Mikhaĭlov and I. A. Mikhaĭlov

Institute of Nuclear Physics, Russian Academy of Sciences, Gatchina, Leningradskaya oblast, 188300 Russia
e-mail: A.Mikhailov@thd.pnpi.spb.ru

Received July 21, 2003

Abstract—The double photoionization of helium at high photon energies is considered using a nonrelativistic approach. The central region of the energy spectrum and its contribution to the total process cross section and to the ratio between the double and single ionization cross sections are studied. Interelectronic interaction in the initial state is included exactly, whereas the interaction between the fast outgoing electrons is calculated by perturbation theory. A detailed derivation of the expression for the cross section ratio between double and single ionizations is given. The corresponding results obtained by other authors are analyzed and corrected. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The double ionization of an atom by a single photon, which opens up the possibility of effectively studying electron correlations in the atom, has attracted attention of many researchers. Because of the one-particle character of electron–photon interactions, simultaneous knocking-out of two electrons by one photon is fully determined by interelectronic interaction. In the asymptotic nonrelativistic energy region ($I \ll \omega \ll m$, where I is the atom ionization energy, ω is the photon energy, and m is the mass of the electron; we use the $\hbar = c = 1$ relativistic system of units), several simple formulas for the ratio between the double and single ionization cross sections and for the energy distribution of photoelectrons were obtained [1, 2]. As distinguished from the single photoeffect when electrons are ejected with a certain energy, the electron energy in the double photoeffect can assume any value because the excess photon energy is distributed between two electrons. The energy spectrum of electrons extends from zero to $E_{\max} = \omega - I^{++}$, where I^{++} is the double ionization energy of the atom. The spectrum is symmetrical with respect to the middle of the energy interval, $E_0 = E_{\max}/2$. The largest contribution to the total cross section is made by the edge spectrum regions where one electron is slow and the other fast ($E_1 \sim E_{\max}$, $E_2 \sim I$ or $E_1 \sim I$, $E_2 \sim E_{\max}$). The contribution of the edge spectrum region is proportional to $\omega^{-7/2}$ at high photon energies. The single photoeffect cross section behaves similarly. For this reason, the ratio R between the total cross sections of double and single photoionization is independent of ω in the asymptotic region of photon energies. The central photoelectron spectrum region ($E_1 \sim E_2$), where both electrons are fast, also makes a noticeable (about 5–10% [2]) contribution to the total cross section. This contribution, however, decreases more slowly (as $\omega^{-5/2}$) as the photon energy increases. As a result, the ratio R between the total cross sections includes a contribution R' that increases as ω grows ($R' \sim \omega$). This contribution is nevertheless smaller than the edge contribution in the nonrelativistic region. The other spectrum regions make substantially smaller contributions and are usually not considered.

In the edge spectrum region, the momentum transferred to the nucleus is larger. One electron should therefore be near the nucleus, whereas the second one may be situated anywhere within the atom. In the central spectrum region, two electrons can absorb one photon by dividing its energy and momentum between themselves. The process occurs at large distances from the nucleus (on the order of the size of the atom) and is accompanied by the transfer of a small momentum $q \sim \eta$ (η is the mean bound electron momentum) to the nucleus. Interelectronic distances should then be small to ensure fairly strong electron–electron interaction, which causes both electrons to acquire large momenta in the opposite directions ($p_{1,2} \approx \sqrt{m\omega} \gg \omega$, $\mathbf{p}_1 \approx -\mathbf{p}_2$). Such a quasi-free double ionization mechanism was considered for the first time in [1] and, in more detail, in [2]. The central double photoionization spectrum region was also studied at relativistic energies [3, 4]. It was shown in [4] that the contribution of the central region was two times smaller in the nonrelativistic limit than that obtained in [2] for the initial state wave func-

tion[1] used in [4]. Recently, two more works have been published [5, 6] in which the quasi-free mechanism was studied. The results obtained in [5] were also approximately two times smaller than those reported in [2]. The authors of more recent work [6], who used a very accurate two-electron initial state wave function [7, 8], however, calculated the contribution of the central spectrum region to the ionization cross section by the equation obtained in [2]. Accordingly, their results exceeded those reported by Kornberg and Miraglia [5] approximately twofold.

The purpose of this work was to give a detailed derivation of the formula that described the central electron energy spectrum region. We only consider asymptotic nonrelativistic photon energies, $I \ll \omega \ll m$, at which the central spectrum region makes the most noticeable contribution.

2. THE AMPLITUDE OF THE DOUBLE PHOTOEFFECT IN THE CENTRAL ENERGY SPECTRUM REGION

We consider the double photoionization of helium and helium-like ions on the assumption that the momenta of both photoelectrons satisfy the condition $p_1, p_2 \gg \eta$. Such electrons belong to the central energy spectrum region. In the nonrelativistic approximation, the total wave function of a two-electron system is given by the product of the spin and coordinate wave functions. As the principal part of electron–photon interaction at nonrelativistic energies does not contain spin operators, the spin function of the atom does not change and drops out of the problem. By virtue of the Pauli principle, the symmetry properties of the coordinate wave function with respect to the permutation of electrons also remain unchanged. When the ground state of helium is ionized, the coordinate wave function is symmetrical. In what follows, the wave function of the system is identified with its coordinate part.

The general equation for the amplitude of double photoionization has the form

$$\begin{aligned} M^{++} &= \langle \Psi_f(1,2)|\hat{\gamma}(1)+\hat{\gamma}(2)|\Psi_i(1,2)\rangle \\ &= 2\langle \Psi_f|\hat{\gamma}|\Psi_i\rangle. \end{aligned} \tag{1}$$

The numbers 1 and 2 in parentheses stand for the arguments of the electrons. The $\hat{\gamma}$ electron–photon interaction operator is a one-particle operator. In the coordinate representation, we have

$$\hat{\gamma} = N_\gamma\left(-\frac{i\mathbf{e}\nabla}{m}\right)e^{i\mathbf{k}\cdot\mathbf{r}}, \quad N_\gamma = \sqrt{\frac{2\pi\alpha}{\omega}}. \tag{2}$$

We exclude the N_γ multiplier from the amplitude and transfer it into the equation for the cross section. Here, $\alpha = e^2 = 1/137$ and $\mathbf{k}$, ω, and $\mathbf{e}$ are the photon momentum, energy, and polarization vector, respectively.

We use the exact two-electron functions of helium and helium-like ions as the initial state wave function Ψ_i. The particular form of these functions is only necessary for obtaining numerical results. The wave function Ψ_f of the final state is constructed using perturbation theory with respect to interelectronic interaction. In a zeroth approximation, Ψ_f is the product of one-particle Coulomb functions ψ_p. Both photoelectrons have high energies $E_{1,2} \sim \omega/2 \gg I$ and large momenta $p_{1,2} \gg \eta$ in the central spectrum region. On the other hand, the Coulomb parameter ξ is small ($\xi_{1,2} = \eta/p_{1,2} \ll 1$) and the wave functions ψ_p of the continuous spectrum can be expanded in powers of this parameter. We only consider the low-order terms of this expansion that correspond to plane waves. With the first two expansion terms, we obtain

$$\Psi_f = \Psi_f^{(0)} + \Psi_f^{(1)}, \tag{3}$$

$$\begin{aligned} \Psi_f^{(0)} &= \frac{1}{\sqrt{2}}[\psi_{p_1}(\mathbf{r})\psi_{p_2}(\mathbf{r}') + \mathbf{p}_1 \longleftrightarrow \mathbf{p}_2] \\ &\approx \frac{1}{\sqrt{2}}[\exp(i\mathbf{p}_1\cdot\mathbf{r})\exp(i\mathbf{p}_2\cdot\mathbf{r}') + \mathbf{p}_1 \longleftrightarrow \mathbf{p}_2], \end{aligned} \tag{4}$$

$$\begin{aligned} \Psi_f^{(1)} &= G_{12}V_{12}\Psi_f^{(0)}, \\ G_{12} &= (E_f - H_1 - H_2 + i0)^{-1}, \end{aligned} \tag{5}$$

$$E_f = E_1 + E_2 = \omega - I^{++} \approx \omega. \tag{6}$$

Here, I^{++} is the double ionization energy of the K shell, V_{12} is the interelectronic interaction operator, G_{12} is the two-particle Green operator, and H_1 and H_2 are the one-particle Hamiltonians. In the plane-wave approximation, H_1 and H_2 are the Hamiltonians of free particles. Equality (6) is the energy conservation law for the process under consideration.

Substituting (3) and (4) into (1) yields

$$M^{++} = \sqrt{2}(A^{(0)} + A^{(1)} + \mathbf{p}_1 \longleftrightarrow \mathbf{p}_2), \tag{7}$$

$$\begin{aligned} A^{(0)} &= \langle \psi_{p_1}\psi_{p_2}|\hat{\gamma}|\Psi_i\rangle, \\ A^{(1)} &= \langle \psi_{p_1}\psi_{p_2}|V_{12}G_{12}\hat{\gamma}|\Psi_i\rangle. \end{aligned} \tag{8}$$

2.1. Calculation of the Amplitude $A^{(0)}$

The matrix element that determines the amplitude $A^{(0)}$ is calculated in the coordinate representation,

$$\begin{aligned} A^{(0)} &= \frac{1}{m}\int d\mathbf{r}d\mathbf{r}'\exp(-i\mathbf{p}_1\cdot\mathbf{r} - i\mathbf{p}_2\cdot\mathbf{r}')(-i\mathbf{e}\nabla) \\ &\times \exp(i\mathbf{k}\cdot\mathbf{r})\Psi_i(\mathbf{r},\mathbf{r}') = \frac{\mathbf{e}\mathbf{p}_1}{m}\int d\mathbf{r}d\mathbf{r}' \end{aligned} \tag{9}$$

[1] The calculations in [2] were performed for various initial state wave functions.

$$\times \exp(-i\boldsymbol{\kappa}_1 \cdot \mathbf{r} - i\mathbf{p}_2 \cdot \mathbf{r}')\Psi_i(\mathbf{r}, \mathbf{r}'), \quad \boldsymbol{\kappa}_1 = \mathbf{p}_1 - \mathbf{k}.$$

The integrand in (9) contains two rapidly oscillating functions and is inconvenient for numerical purposes. Numerical estimates can easily be obtained in terms of new variables,

$$\mathbf{R} = \frac{1}{2}(\mathbf{r} + \mathbf{r}'), \quad \boldsymbol{\rho} = \mathbf{r}' - \mathbf{r}. \tag{10}$$

The $d\mathbf{r}d\mathbf{r}'$ element is then replaced by $d\mathbf{R}d\boldsymbol{\rho}$. It follows from (10) that

$$\mathbf{r} = \mathbf{R} - \boldsymbol{\rho}/2, \quad \mathbf{r}' = \mathbf{R} + \boldsymbol{\rho}/2. \tag{11}$$

Then,

$$\boldsymbol{\kappa}_1\mathbf{r} + \mathbf{p}_2\mathbf{r}' = \mathbf{q}\mathbf{R} + \mathbf{a}\boldsymbol{\rho}, \quad \mathbf{q} = \mathbf{p}_1 + \mathbf{p}_2 - \mathbf{k}, \quad \mathbf{a} = \mathbf{p}_2 - \mathbf{q}/2, \tag{12}$$

$$\Psi_i(\mathbf{r}, \mathbf{r}') = \Phi_i(\mathbf{R}, \boldsymbol{\rho}). \tag{13}$$

The amplitude $A^{(0)}$ takes the form

$$A^{(0)} = \frac{\mathbf{e} \cdot \mathbf{p}_1}{m}\int d\mathbf{R}e^{-i\mathbf{q} \cdot \mathbf{R}}\int d\boldsymbol{\rho}e^{-i\mathbf{a} \cdot \boldsymbol{\rho}}\Phi_i(\mathbf{R}, \boldsymbol{\rho}). \tag{14}$$

We consider the central spectrum region, where the momentum transferred to the nucleus is small ($q \sim \eta$). The integral in $\mathbf{R}$ contains a feebly oscillating function and is saturated at large (on the order of atomic size) $R \sim q^{-1} \sim \eta^{-1}$. Conversely, the integral in $\boldsymbol{\rho}$ contains a rapidly oscillating function ($a \equiv |\mathbf{a}| \gg q$) and is saturated at small $\rho \sim a^{-1} \sim p_2^{-1}$. This integral can be recast as follows:

$$\begin{aligned} I(\mathbf{a}, \mathbf{R}) &\equiv \int d\boldsymbol{\rho}e^{-i\mathbf{a} \cdot \boldsymbol{\rho}}\Phi_i(\mathbf{R}, \boldsymbol{\rho}) \\ &= -\frac{1}{a^2}\int d\boldsymbol{\rho}\Phi_i(\mathbf{R}, \boldsymbol{\rho})\Delta_\rho e^{-i\mathbf{a} \cdot \boldsymbol{\rho}} \\ &= -\frac{1}{a^2}\int d\boldsymbol{\rho}e^{-i\mathbf{a} \cdot \boldsymbol{\rho}}\Delta_\rho\Phi_i(\mathbf{R}, \boldsymbol{\rho}) \\ &\approx -\frac{1}{a^2}\int d\boldsymbol{\rho}e^{-i\mathbf{a} \cdot \boldsymbol{\rho}}[\Delta_\rho\Phi_i(\mathbf{R}, \boldsymbol{\rho})]_{\rho \to 0}. \end{aligned} \tag{15}$$

Here, $\Delta_\rho \equiv \nabla_\rho^2$ is the Laplace operator in variable $\boldsymbol{\rho}$. The term in square brackets can be found from the Schrödinger equation for the helium atom (helium-like ion) written in the variables $\mathbf{R}$ and $\boldsymbol{\rho}$,

$$\left(-\frac{\Delta_R}{4m} - \frac{\Delta_\rho}{m} - \frac{\alpha Z}{|\mathbf{R} - \boldsymbol{\rho}/2|} - \frac{\alpha Z}{|\mathbf{R} + \boldsymbol{\rho}/2|} + \frac{\alpha}{\rho} - \varepsilon_i\right) \times \Phi_i(\mathbf{R}, \boldsymbol{\rho}) = 0, \tag{16}$$

where ε_i is the initial state energy.

Leaving the principal terms only in (16) when $\rho \longrightarrow 0$, we obtain

$$[\Delta_\rho\Phi_i(\mathbf{R}, \boldsymbol{\rho})]_{\rho \to 0} = \frac{m\alpha}{\rho}\Phi_i(\mathbf{R}, 0). \tag{17}$$

Substituting (17) into (15) yields

$$\begin{aligned} I(\mathbf{a}, \mathbf{R}) &\approx -\frac{m\alpha}{a^2}\Phi_i(\mathbf{R}, 0)\int d\boldsymbol{\rho}\frac{e^{-i\mathbf{a} \cdot \boldsymbol{\rho}}}{\rho} \\ &= -\frac{4\pi\alpha}{a^4}m\Phi_i(\mathbf{R}, 0), \end{aligned} \tag{18}$$

$$A^{(0)} = -\frac{4\pi\alpha}{a^4}\mathbf{e} \cdot \mathbf{p}_1 S(\mathbf{q}), \tag{19}$$

$$\begin{aligned} S(\mathbf{q}) &= \int d\mathbf{R}e^{-i\mathbf{q} \cdot \mathbf{R}}\Phi_i(\mathbf{R}, 0) \\ &= \int d\mathbf{r}e^{-i\mathbf{q} \cdot \mathbf{r}}\Psi_i(\mathbf{r}, \mathbf{r}). \end{aligned} \tag{20}$$

2.2. Calculation of the Amplitude $A^{(1)}$

The $A^{(1)}$ amplitude is also calculated in the coordinate representation,

$$A^{(1)} = \frac{1}{m}\int d\mathbf{r}d\mathbf{r}'\Psi_f^{(1)}(\mathbf{r}, \mathbf{r}')(-i\mathbf{e}\nabla)e^{i\mathbf{k} \cdot \mathbf{r}}\Psi_i(\mathbf{r}, \mathbf{r}'), \tag{21}$$

$$\begin{aligned} \Psi_f^{(1)}(\mathbf{r}, \mathbf{r}') &= \langle\Psi_f^{(0)}|V_{12}G_{12}|\mathbf{r}, \mathbf{r}'\rangle \\ &= \int d\mathbf{r}_1 d\mathbf{r}_2 \exp(-i\mathbf{p}_1 \cdot \mathbf{r}_1 - i\mathbf{p}_2 \cdot \mathbf{r}_2) \\ &\times \frac{\alpha}{|\mathbf{r}_1 - \mathbf{r}_2|}\langle\mathbf{r}_1, \mathbf{r}_2|G_{12}|\mathbf{r}, \mathbf{r}'\rangle, \end{aligned} \tag{22}$$

$$\begin{aligned} \langle\mathbf{r}_1, \mathbf{r}_2|G_{12}|\mathbf{r}, \mathbf{r}'\rangle &= \int \frac{d\mathbf{f}}{(2\pi)^3}\frac{d\mathbf{s}}{(2\pi)^3} \\ &\times \frac{\exp[i\mathbf{f} \cdot (\mathbf{r}_1 - \mathbf{r}) + i\mathbf{s} \cdot (\mathbf{r}_2 - \mathbf{r}')]}{\omega - (f^2 + s^2)/2m + i0}. \end{aligned} \tag{23}$$

An analysis of the equation for $A^{(1)}$ shows that the principal contribution on the order of $\alpha\eta/p_1^4$ to the amplitude is acquired in the region $f \sim s \sim \eta$, whereas the contribution of the region $f \sim s \sim p_1$ is $(p_1/\eta)^3$ times smaller. For this reason, we can use a simpler expression for G_{12}, which is obtained from (23) by ignoring the $(f^2 + s^2)/2m$ term compared with ω. We then have

$$\langle\mathbf{r}_1, \mathbf{r}_2|G_{12}|\mathbf{r}, \mathbf{r}'\rangle = \frac{1}{\omega}\delta(\mathbf{r}_1 - \mathbf{r})\delta(\mathbf{r}_2 - \mathbf{r}'), \tag{24}$$

$$\Psi_f^{(1)}(\mathbf{r}, \mathbf{r}') = \frac{\alpha}{\omega}\frac{\exp(-i\mathbf{p}_1 \cdot \mathbf{r} - i\mathbf{p}_2 \cdot \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}. \tag{25}$$

Substituting (25) into (21) and integrating by parts yields

$$A^{(1)} = \frac{\alpha}{m\omega}\int d\mathbf{r}d\mathbf{r}'\left(\frac{\mathbf{e}\cdot\mathbf{p}_1}{\rho}+\frac{i\mathbf{e}\cdot\boldsymbol{\rho}}{\rho^3}\right) \times \exp(-i\mathbf{p}_1\cdot\mathbf{r}-i\mathbf{p}_2\cdot\mathbf{r}'+i\mathbf{k}\cdot\mathbf{r})\Psi_i(\mathbf{r},\mathbf{r}'). \quad (26)$$

Further calculations are easy to perform in terms of the $\mathbf{R}$ and $\boldsymbol{\rho}$ variables. Leaving the principal term of the expansion in q/p_1 only, we find the simple equation for $A^{(1)}$:

$$\begin{aligned} A^{(1)} &= \frac{\alpha}{m\omega}\int d\mathbf{R}d\boldsymbol{\rho}\left(\frac{\mathbf{e}\cdot\mathbf{p}_1}{\rho}+\frac{i\mathbf{e}\cdot\boldsymbol{\rho}}{\rho^3}\right) \\ &\times \exp(-i\mathbf{q}\cdot\mathbf{R}-i\mathbf{a}\cdot\boldsymbol{\rho})\Phi_i(\mathbf{R},\boldsymbol{\rho}) \\ &\approx \frac{\alpha}{m\omega}\int d\mathbf{R}\exp(-i\mathbf{q}\cdot\mathbf{R})\Phi_i(\mathbf{R},0) \\ &\times \int d\boldsymbol{\rho}\exp(-i\mathbf{a}\cdot\boldsymbol{\rho})\left(\frac{\mathbf{e}\cdot\mathbf{p}_1}{\rho}+\frac{i\mathbf{e}\cdot\boldsymbol{\rho}}{\rho^3}\right) \\ &= \frac{2\pi\alpha}{m\omega}\frac{\mathbf{e}\cdot\mathbf{q}}{a^2}S(\mathbf{q}). \end{aligned} \quad (27)$$

2.3. The Derivation of the Equation for the Amplitude M^{++} of the Process

To find the amplitude M^{++}, we must add exchange terms to $A^{(0)}$ and $A^{(1)}$. Applying the law of conservation of momentum yields

$$\begin{aligned} A^{(0)}+\mathbf{p}_1\longleftrightarrow\mathbf{p}_2 &= -4\pi\alpha S(\mathbf{q})\left(\frac{\mathbf{e}\cdot\mathbf{p}_1}{a^4}+\frac{\mathbf{e}\cdot\mathbf{p}_2}{b^4}\right) \\ &= -4\pi\alpha S(\mathbf{q})\left[\mathbf{e}\cdot\mathbf{p}_1\left(\frac{1}{a^4}-\frac{1}{b^4}\right)+\frac{\mathbf{e}\cdot\mathbf{q}}{b^4}\right], \end{aligned} \quad (28)$$

$$a = |\mathbf{p}_2-\mathbf{q}/2|, \quad b = |\mathbf{p}_1-\mathbf{q}/2|, \quad (29)$$

$$A^{(1)}+\mathbf{p}_1\longleftrightarrow\mathbf{p}_2 = \frac{2\pi\alpha}{m\omega}\mathbf{e}\cdot\mathbf{q}S(\mathbf{q})\left(\frac{1}{a^2}+\frac{1}{b^2}\right). \quad (30)$$

A comparison of (28) and (30) shows that, because of the cancellation of the principal terms in the total (including exchange terms $\mathbf{p}_1\longleftrightarrow\mathbf{p}_2$) $A^{(0)}$ amplitude, this amplitude becomes of the same order of smallness as the total $A^{(1)}$ amplitude, which is the reason why these amplitudes are considered jointly.

Let us express the energies and momenta of photoelectrons via $\Delta = E_1 - E_2$,

$$\begin{aligned} E_1 &= \frac{\omega+\Delta}{2}, \quad E_2 = \frac{\omega-\Delta}{2}, \\ p_1^2 &= m(\omega+\Delta), \quad p_2^2 = m(\omega-\Delta). \end{aligned} \quad (31)$$

The momentum q transferred to the nucleus is small in the central spectrum region. The laws of conservation of energy and momentum can therefore be written as

$$E_1+E_2 = \omega, \quad \mathbf{p}_1+\mathbf{p}_2 = \mathbf{k}. \quad (32)$$

Solving these equations simultaneously yields the constraint on $|\Delta|$,

$$|\Delta| \le \omega\sqrt{\frac{\omega}{m}} \ll \omega. \quad (33)$$

Condition (33) determines the boundaries of the central spectrum region. In the lowest order in the $(\omega/m)^{1/2}$ parameter, we have

$$E_1 = E_2 = \frac{\omega}{2}, \quad p_1 = p_2 = \sqrt{m\omega}, \quad (34)$$

$$b^4-a^4 = 4m\omega(m\Delta-\mathbf{p}_1\cdot\mathbf{q}) \simeq 4m\omega\mathbf{p}_1\cdot\mathbf{k}. \quad (35)$$

In summing (28) and (30), we notice that the terms proportional to $\mathbf{e}\cdot\mathbf{q}$ are cancelled[2] and we arrive at the simple equation for M^{++}

$$\begin{aligned} M^{++} &= -4\pi\alpha\sqrt{2}F, \\ F &= S(\mathbf{q})\frac{4(\mathbf{e}\cdot\mathbf{n}_1)(\mathbf{k}\cdot\mathbf{n}_1)}{(m\omega)^2}, \quad \mathbf{n}_1 = \frac{\mathbf{p}_1}{p_1}. \end{aligned} \quad (36)$$

3. THE CONTRIBUTION OF THE CENTRAL SPECTRUM REGION TO THE TOTAL DOUBLE IONIZATION CROSS SECTION

The differential double ionization cross section is determined by the equation

$$d\sigma^{++} = N_\gamma^2|M^{++}|^2\frac{d\mathbf{p}_1 d\mathbf{p}_2}{(2\pi)^6}2\pi\delta(E_1+E_2-\omega). \quad (37)$$

Substituting (36) into (37) yields the contribution of the central spectral region to the cross section,

$$d\sigma_{\text{c.r.}}^{++} = (4\pi\alpha)^3|F|^2\frac{d\mathbf{p}_1 d\mathbf{p}_2}{\omega(2\pi)^5}\delta(E_1+E_2-\omega). \quad (38)$$

To calculate the total cross section, we must change from $d\mathbf{p}_2$ to $d\mathbf{q}$ and replace $d\mathbf{p}_1$ with $mp_1dE_1d\Omega_1$. It follows from the definition of $\mathbf{q}$ [Eq. (12)] that the energy E_2 is a function of E_1 and $\mathbf{q}$,

$$E_2 = E_1-\frac{\boldsymbol{\kappa}\cdot\mathbf{p}_1}{m}+\frac{\kappa^2}{2m}, \quad \boldsymbol{\kappa} = \mathbf{k}+\mathbf{q}. \quad (39)$$

[2] The cancellation of the terms proportional to $\mathbf{e}\cdot\mathbf{q}$ in the amplitude of the double photoeffect in the central spectrum region was noticed for the first time in [9].

Taking this dependence into account, the removal of the δ function in (38) yields

$$\int dE_1 \delta(E_1 + E_2 - \omega) = \left(2 - \frac{\boldsymbol{\kappa}\cdot\mathbf{p}_1}{p_1^2}\right)^{-1} \approx \frac{1}{2}. \quad (40)$$

The integration in q can be performed from zero to infinity. Indeed, the $q > \eta$ region makes a small contribution to the cross section because the $|S(\mathbf{q})|^2$ value rapidly decreases as q increases.[3] We then have

$$\frac{d\sigma_{\text{c.r.}}^{++}}{d\Omega_1} = \frac{2^7\pi\alpha^3}{(m\omega)^3} p_1 B(\mathbf{e}\cdot\mathbf{n}_1)^2(\mathbf{n}\cdot\mathbf{n}_1)^2, \quad \mathbf{n} = \frac{\mathbf{k}}{\omega}, \quad (41)$$

$$B = \int \frac{d\mathbf{q}}{(2\pi)^3}|S(\mathbf{q})|^2 = \int d\mathbf{r}\Psi_i^2(\mathbf{r},\mathbf{r}). \quad (42)$$

The integration of (41) in the angles is simple. As the two final electrons are identical, the result should be divided by 2,

$$\sigma_{\text{c.r.}}^{++} = \frac{2^8\pi^2\alpha^3}{15m^5}\left(\frac{m}{\omega}\right)^{5/2} B. \quad (43)$$

4. THE SINGLE IONIZATION CROSS SECTION

Let us determine the single ionization cross section σ^+ of the helium atom in the nonrelativistic approximation for a high-frequency photon and an arbitrary symmetrical function of the initial state. The amplitude of the single photoeffect is[4]

$$M^+ = 2\langle\Psi_f|\hat{\gamma}|\Psi_i\rangle = \frac{\sqrt{2}}{m}\int d\mathbf{r}d\mathbf{r}' \times [e^{-i\mathbf{p}\cdot\mathbf{r}}\varphi_{1s}(\mathbf{r}') + e^{-i\mathbf{p}\cdot\mathbf{r}'}\varphi_{1s}(\mathbf{r})] \times (-i\mathbf{e}\nabla)e^{i\mathbf{k}\cdot\mathbf{r}}\Psi_i(\mathbf{r},\mathbf{r}'). \quad (44)$$

Here, $p = \sqrt{2m\omega}$ and the Coulomb parameter $\xi = \eta/p = \sqrt{I/\omega} \ll 1$. For this reason, we use a plane wave as the wave function of the free electron and the bound electron in the final state is described by the Coulomb wave function φ_{1s} (we consider the ionization of the ground state of the helium atom). The two-electron wave function of the initial state $\Psi_i(\mathbf{r},\mathbf{r}')$ can be an arbitrary analytic function or given numerically.

[3] If $\Psi_i(\mathbf{r},\mathbf{r}')$ is given by the product of Coulomb functions $\varphi_{1s}(\mathbf{r})\varphi_{1s}(\mathbf{r}')$, then $S(\mathbf{q}) = [\mu^2/(q^2+\mu^2)]^2$, where $\mu = 2\eta$.

[4] The $\sqrt{2\pi\alpha/\omega}$ multiplier of the $\hat{\gamma}$ operator is included into the equation for the cross section.

The integration by parts transforms (44) into

$$M^+ = \frac{\sqrt{2}}{m}\int d\mathbf{r}d\mathbf{r}' \times\left\{\mathbf{e}\cdot\mathbf{p}e^{-i(\mathbf{p}-\mathbf{k})\cdot\mathbf{r}}\varphi_{1s}(\mathbf{r}') - i\eta\frac{\mathbf{e}\cdot\mathbf{r}}{r}e^{i\mathbf{k}\cdot\mathbf{r}}\varphi_{1s}(\mathbf{r})e^{-i\mathbf{p}\cdot\mathbf{r}'}\right\} \times\Psi_i(\mathbf{r},\mathbf{r}') \approx \frac{\sqrt{2}}{m}\mathbf{e}\mathbf{p} \times\int d\mathbf{r}d\mathbf{r}'e^{-i\mathbf{p}\cdot\mathbf{r}}\varphi_{1s}(\mathbf{r}')\Psi_i(\mathbf{r},\mathbf{r}'). \quad (45)$$

In the derivation of (45), it was taken into account that $\omega/p = \sqrt{\omega/2m} \ll 1$ and only the principal term was left in the expansion in this parameter; that is, the dipole approximation was used.

Next, consider the integral

$$J(\mathbf{p},\mathbf{r}') = \int d\mathbf{r}e^{-i\mathbf{p}\cdot\mathbf{r}}\Psi_i(\mathbf{r},\mathbf{r}'). \quad (46)$$

Using the Green theorem and the smallness of $r \sim p^{-1}$, we obtain

$$J(\mathbf{p},\mathbf{r}') = -\frac{1}{p^2}\int d\mathbf{r}\Psi_i\Delta e^{-i\mathbf{p}\cdot\mathbf{r}} = -\frac{1}{p^2}\int d\mathbf{r}e^{-i\mathbf{p}\cdot\mathbf{r}}\Delta\Psi_i \approx -\frac{1}{p^2}\int d\mathbf{r}e^{-i\mathbf{p}\cdot\mathbf{r}}(\Delta\Psi_i)|_{r\to 0}. \quad (47)$$

The Schrödinger equation for the helium atom in the variables $\mathbf{r}$ and $\mathbf{r}'$ has the form

$$\left(-\frac{\Delta}{2m} - \frac{\Delta'}{2m} - \frac{\alpha Z}{r} - \frac{\alpha Z}{r'} + \frac{\alpha}{|\mathbf{r}-\mathbf{r}'|} - \varepsilon_i\right) \times\Psi_i(\mathbf{r},\mathbf{r}') = 0, \quad (48)$$

where $\Delta(\Delta')$ is the Laplace operator in the variable $\mathbf{r}$ ($\mathbf{r}'$). It follows from this equation that

$$\Delta\Psi_i(\mathbf{r},\mathbf{r}')|_{r\to 0} = -\frac{2\eta}{r}\Psi_i(0,\mathbf{r}'). \quad (49)$$

Substituting (49) into (47) yields

$$J(\mathbf{p},\mathbf{r}') = \frac{8\pi\eta}{p^4}\Psi_i(0,\mathbf{r}') \quad (50)$$

and the amplitude [Eq. (45)] takes the form

$$M^+ = \frac{\sqrt{2}}{m}\mathbf{e}\cdot\mathbf{p}\int d\mathbf{r}'\varphi_{1s}(\mathbf{r}')J(\mathbf{p},\mathbf{r}') = \sqrt{2}\frac{\mathbf{e}\cdot\mathbf{p}}{m}\frac{8\pi\eta}{p^4}N, \quad (51)$$

$$N = \int d\mathbf{r}\varphi_{1s}(\mathbf{r})\Psi_i(0,\mathbf{r}). \quad (52)$$

In the one-particle approximation with Coulomb functions, we have

$$N = N_{1s} = (\eta^3/\pi)^{1/2}.$$

The differential and total cross sections for the single ionization of the K shell of the atom are given by the equations

$$d\sigma^+ = \frac{2\pi\alpha}{\omega}|M^+|^2\frac{d\mathbf{p}}{(2\pi)^3}2\pi\delta(E+I-\omega) = \frac{\alpha}{2\pi\omega}|M^+|^2 mpd\Omega_p, \quad (53)$$

$$\sigma^+ = \frac{32\sqrt{2}\pi^2\alpha^3Z^2}{3m^5}\left(\frac{m}{\omega}\right)^{7/2}N^2. \quad (54)$$

5. THE CONTRIBUTION OF THE CENTRAL REGION TO THE RATIO OF THE CROSS SECTIONS

The ratio between the nonrelativistic cross sections of single and double ionization in the high-frequency limit can be represented as the sum of R_0 (the contribution of the edge spectrum region) and R' (the contribution of the central spectrum region),

$$R = \frac{\sigma^{++}}{\sigma^+} = \frac{\sigma^{++}_{\text{e.r.}}}{\sigma^+} + \frac{\sigma^{++}_{\text{c.r.}}}{\sigma^+} = R_0 + R'. \quad (55)$$

The R_0 value is independent of the photon energy, and its value has been well established, $R_0 = 1.67\%$ [5, 10]. Conversely, the R' contribution linearly depends on the photon energy, as is easy to see when we divide (43) by (54),

$$R' = \frac{4\sqrt{2}}{5Z^2}C\frac{\omega}{m}, \quad (56)$$

$$C = \frac{B}{N^2} = \frac{\int d\mathbf{r}|\Psi_i(\mathbf{r},\mathbf{r})|^2}{\left|\int d\mathbf{r}\varphi_{1s}(\mathbf{r})\Psi_i(0,\mathbf{r})\right|^2}. \quad (57)$$

Equation (56) only differs from the one obtained by Drukarev [2] by the 1/2 multiplier. Unfortunately, [2] does not contain details of calculations, and we cannot therefore identify the reason for the discrepancy. The more recent paper by Kornberg and Miraglia [5] also contains errors, because it ignores interelectronic interaction in the final state, which results in the appearance of terms proportional to ω^{-1} in the equation for R'. Nevertheless, the term proportional to ω was calculated in [5] correctly, however, only for the $\Psi_i(\mathbf{r}, \mathbf{r}')$ function of a special form. Using the same function, we obtain $R' = 3.49\times10^{-5}\omega$ [keV] (cf. $4.02\times10^{-5}\omega$ [keV] in [5]). A 15% difference arises because we use (54) for σ^+, whereas Kornberg and Miraglia use the equation from [11] with the ground state wave function from [12]. The advantage of our approach is the use of the same initial state wave function for double and single ionizations, whereas Kornberg and Miraglia use different initial functions for σ^{++} and σ^+.

6. CONCLUSIONS

We derived equations that described the contribution of the central region of the electron energy spectrum to the total cross section of double ionization and to the ratio between the double and single ionization cross sections for the initial wave functions of the most general form. We assert that the contribution of the central region to the cross section ratio obtained in [2] is exaggerated twofold. The result for R' obtained in [5] is erroneous. It contains terms proportional to ω and ω^{-1}. The term proportional to ω correctly describes the contribution of the central region to R, whereas the term proportional to ω^{-1} disappears if interelectronic interaction in the final state is included.

ACKNOWLEDGMENTS

This work was financially supported by the Russian Foundation for Basic Research (project nos. 01-02-17246 and 00-15-96610).

REFERENCES

1. M. Ya. Amusia, E. G. Drukarev, V. G. Gorshkov, and M. P. Kazachkov, J. Phys. B **8**, 1248 (1975).
2. E. G. Drukarev, Phys. Rev. A **51**, R2684 (1995).
3. E. G. Drukarev and F. F. Karpeshin, J. Phys. B **9**, 399 (1976).
4. A. I. Mikhaĭlov and I. A. Mikhaĭlov, Zh. Éksp. Teor. Fiz. **114**, 1537 (1998) [JETP **87**, 833 (1998)].
5. M. A. Kornberg and J. E. Miraglia, Phys. Rev. A **60**, R1743 (1999).
6. R. Krivec, M. Ya. Amusia, and V. B. Mandelzweig, Phys. Rev. A **64**, 012713 (2001).
7. M. Haftel, R. Krivec, and V. B. Mandelzweig, J. Comput. Phys. **123**, 149 (1996).
8. R. Krivec, V. B. Mandelzweig, and K. Varga, Phys. Rev. A **61**, 062503 (2000).
9. E. G. Drukarev, N. B. Avdonina, and R. H. Pratt, J. Phys. B **34**, 1 (2001).
10. A. S. Kheifets and I. Bray, Phys. Rev. A **58**, 4501 (1998).
11. A. L. Stewart and T. G. Webb, Proc. Phys. Soc. London **82**, 532 (1963).
12. N. Sabelli and J. Hinze, J. Chem. Phys. **50**, 648 (1969).

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 98, No. 2, 2004, pp. 254–271.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 2, 2004, pp. 288–306.

ATOMS, SPECTRA, RADIATION

Dynamic Hyperpolarizabilities of Excited States of Hydrogen

N. L. Manakov, S. I. Marmo, and E. A. Pronin
Voronezh State University, Voronezh, 394006 Russia
e-mail: manakov@thp.vsu.ru
Received August 4, 2003

Abstract—On the basis of the generalized Sturm expansion of the radial part of the Coulomb Green function, a computational method is proposed and numerical results are presented for the dynamic hyperpolarizability γ and the corrections $E^{(4)}$ (quadratic in the light intensity) to the quasi-energy of the ground and excited states of hydrogen with principal quantum numbers $n \le 5$ in a monochromatic light field. In this approach, the problem is reduced to the summation of well-convergent double series of the hypergeometric kind, which ensures reliable numerical results both for states with a large n, and in a wide range of field frequencies ω, including the above-threshold frequency range of $\hbar\omega \gg |E_n|$ ($|E_n|$ is the ionization potential of the state $|nlm\rangle$ under investigation). We consider the frequency dependence of γ and $E^{(4)}$, their differences for the cases of linear and circular polarizations of the field, and the relation between their real and imaginary parts, which determine the laser field-induced corrections to the position and width of energy levels. For $n = 5$, the significant role of mixing the $|nlm\rangle$ states with different values of l by a laser field in the region of resonances on intermediate bound states is demonstrated. The linear (in intensity) corrections to the photoionization cross section for excited states are analyzed and the threshold intensity corresponding to the onset of atomic level stabilization is estimated for a number of states with $n = 3$ and $n = 5$. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The analysis of interaction of an atomic system with high-intensity laser radiation is simplified considerably in the case of a periodic (in particular, harmonic) time dependence of the external field. Spectral characteristics of an atom interacting with a monochromatic light wave,

$$\mathbf{F}(t) = F\mathrm{Re}\{\mathbf{e}e^{-i\omega t}\} \tag{1}$$

(F and ω are the amplitude and frequency of the wave and $\mathbf{e}$ is the unit vector of polarization, $\mathbf{e}\cdot\mathbf{e}^* = 1$), are determined by the spectrum of complex quasi-energies $E_n = \mathrm{Re}E_n - (i/2)\Gamma_n$, which correspond to the unperturbed atomic spectrum $E_n^{(0)}$ and determine the shift ($\mathrm{Re}E_n - E_n^{(0)}$) and ionization broadening (Γ_n) of energy levels $E_n^{(0)}$ by the field of the light wave [1]. If the field amplitude (1) is small as compared to the characteristic intraatomic field F_n ($F_n = F_{\mathrm{at}}/n^3$, $F_{\mathrm{at}} = 5.142 \times 10^9$ V/cm), the quasi-energy can be determined in perturbation theory by calculating the corrections to unperturbed energy level $E^{(0)}$. For states that do not possess an intrinsic dipole moment, the corrections contain only the terms with even powers of F:

$$E = E^{(0)} + E^{(2)} + E^{(4)} + \ldots, \quad E^{(k)} \sim F^k. \tag{2}$$

Excited states of the hydrogen atom may possess a dipole moment, but the effect of this quantity on the quasi-energy spectrum is not pronounced in the frequency range $\hbar\omega \gg F|\langle nlm|\mathbf{e}\cdot\mathbf{r}|nl'm'\rangle|$ considered here [1]; consequently, the expansion of E has the form (2) in this case also.

Perturbation of the spectrum is usually analyzed in the language of dynamic polarizability (DP) α and hyperpolarizability (DHP) γ, which determine the second-order corrections

$$E^{(2)} = -\frac{1}{4}\alpha F^2, \tag{3}$$

as well as the fourth-order corrections

$$E^{(4)} = -\frac{1}{8}\gamma F^4, \tag{4}$$

in F to the unperturbed energy level in relation (2).

Parameters α and γ introduced via relations (3) and (4) also determine the first- and third-order terms in F in the expansion of the atomic dipole moment $\mathbf{P}(t)$ induced by an external field at frequency ω:

$$P_j(t) = \mathrm{Re}\{P_j^{(\omega)}e^{-i\omega t} + P_j^{(3\omega)}e^{-3i\omega t} + \ldots\},$$
$$P_j^{(\omega)} = \alpha_{jk}e_kF + \gamma_{jklm}e_k^*e_le_mF^3, \tag{5}$$
$$\alpha = \alpha_{jk}e_j^*e_k, \quad \gamma = \gamma_{jklm}e_j^*e_k^*e_le_m.$$

Consequently, these parameters determine the cross section of Rayleigh scattering of light by an atom and the corrections to this cross section, which are linear in the wave intensity (for a gaseous medium, these parameters determine the refractive index including the corrections on the order of F^2). It should be borne in mind that tensors α_{jk} and γ_{jklm} are usually determined in terms of the mean value of the dipole moment operator on wave functions calculated in perturbation theory up to terms on the order of F^3. However, such a definition of γ becomes meaningless for frequencies exceeding the single-photon ionization threshold ($\hbar\omega > |E_n^{(0)}|$) since the components of tensor γ_{jklm} go to infinity, while relation (4) gives, as before, the shift of an atomic level [2]. Formally, the divergence is due to the fact that the fourth-order matrix elements, which determine γ_{jklm}, contain Green functions $G^{(+)}(E_n^{(0)} + \hbar\omega)$ and $G^{(-)}(E_n^{(0)} + \hbar\omega)$ for an atomic electron with an asymptotic form of diverging and converging waves with the same energy; this leads to divergence of the corresponding matrix elements for $E_n^{(0)} + \hbar\omega > 0$ [2]. At the same time, quantity γ in relation (4) contains only the Green functions with the asymptotic form of diverging waves ($G^{(+)}(\mathscr{E})$) and, hence, remains finite for $E_n^{(0)} + \hbar\omega > 0$.[1] According to the discussion in [4], the divergences mentioned above are due to the fact that, in a strong laser field (when effects on the order of F^4 become significant), it is necessary to take into account the quasi-stationary form of atomic levels and to calculate the mean value of the dipole moment operator using the appropriately normalized "dual" wave functions of the quasi-stationary state. The "dual" dipole moment defined in this way remains finite for $E_n^{(0)} + \hbar\omega > 0$, coincides with the conventional definition for $E_n^{(0)} + \hbar\omega < 0$, and leads to the same definition of DHP γ as in formula (4).

The polarizabilities of hydrogen-like states, which determine the main (in F) correction to energy and level width (for $\hbar\omega > |E_n^{(0)}|$), have been studied in numerous publications starting from the middle of the 1960s and have been investigated in detail (see, for example, [1, 5–7] and the literature cited in [7]). In particular, the closed analytic expressions obtained recently in [7] for the DP of arbitrary states $|nlm\rangle$ make it possible to carry out rather simple numerical calculations and to analyze the asymptotic behavior of DP (including that in the Rydberg spectral region). The DHP and the quadratic (in wave intensity) correction to energy in relation (2) have been studied less thoroughly since computations in this case cannot be carried out analytically even for the ground state[2] and the results obtained from numerical calculations are scarce. Most publications pertain to the ground state and below-threshold frequency values [1, 5]. Although, for a Coulomb potential, numerical methods exist for calculating directly the $E^{(k)}$ terms of expansion (2) even for values of k much greater than 4 (see, for example, [10, 11]), the application of these methods for highly excited states and/or above-threshold frequencies involves considerable computational difficulties. The dispersion dependence of γ for below-threshold and above-threshold frequencies $\hbar\omega \sim |E_n|$ for the ground state and first excited states of hydrogen was analyzed in [2], where it was proved that typical values of γ in different frequency intervals may differ by many orders of magnitude.

In the present study, a new method is proposed on the basis of the special representation for the radial Coulomb Green function obtained in [12] (see also [7]). This method can be used for calculating the DHPs for excited states of hydrogen with the principal quantum number $n \sim 10$, including those for above-threshold frequencies exceeding the ionization threshold $|E_n|$ for the state under study by more than an order of magnitude. The atomic parameters describing multiphoton processes in the above-threshold frequency range are interesting in view of the use of high (ultraviolet) harmonics of radiation from optical-range lasers in recent experiments, in view of the development of the nonlinear laser spectroscopy methods of highly excited atomic levels (for which even the frequencies of optical-range lasers correspond to the far above-threshold region), and also in connection with the possibility of using hard UV radiation from free-electron lasers in atomic experiments. Above-threshold multiphoton transitions via virtual states of the continuum are of special interest in analyzing the stabilization (retardation) effect in the decay of an atom in a high-frequency field upon an increase in the field intensity, which was discovered experimentally in 1993 [13]. Experiments on stabilization of hydrogen-like $5g$ states of neon [14, 15] show that the onset of stabilization corresponds to the threshold range of intensities $I_{thr} \sim 10^{13}$–10^{14} W/cm^2, which are smaller than the intraatomic intensity ($\sim 3.5 \times 10^{16}$ W/cm^2) so that I_{thr} can be estimated by analyzing the imaginary part of sequential terms of expansion (2) (see Section 3.2).

The general formulas required for calculating the DHPs of excited states are given in Section 2, where the method for calculating the compound radial matrix elements is described (the calculation of these elements presents the main difficulty in analysis of the amplitudes of multiphoton transitions in the above-threshold frequency range). Section 3.1 contains the results of numerical calculations of DHP as well as the shift and

[1] The formula for α contains only one Green's function; for this reason, the definitions of DP in terms of tensor α_{jk} and by formula (3) give the same result; see Section 59 in [3].

[2] Except for the low-frequency region in which γ can be expanded into a converging series in ω^2 with rational coefficients [8, 9].

width of excited Coulomb states, while linear (in wave intensity) corrections to the probability of the photoeffect are considered in Section 3.2, where the critical fields corresponding to the onset of stabilization of atomic levels are estimated. The algorithm for computing compound matrix elements for above-threshold frequencies, which can be used for analyzing arbitrary fourth-order transitions with exact account for virtual states of the continuum, is described in detail in the Appendix.

2. GENERAL FORMULAS FOR DYNAMIC HYPERPOLARIZABILITY AND COMPUTATIONAL TECHNIQUE

2.1. Isolated Levels

In spite of the fact that unperturbed energy levels for an electron in a central field are degenerate in the orbital angular momentum component m, quantity m also remains a conserved quantum number for quasi-energy states with a special choice of the quantization axis for linear and circular polarizations of field $\mathbf{F}(t)$ (in this paper, we confine our analysis to such polarizations only). In the case of linear polarization, the quantization axis is directed along the polarization vector, while for a circularly polarized wave, this axis coincides with the direction of wave propagation. In the Coulomb field, the energy levels are also degenerate in the orbital quantum number l, the nondiagonal elements of the quasi-energy matrix $Q_{l,\,l'=l\pm 2}(F)$ differing from zero even in the first nonvanishing ($\sim F^2$) order of perturbation theory. Consequently, sublevels of the shell n with different values of l and with a fixed m are mixed by the field even for linear or circular polarization of the wave [1]. Only the states $|nlm\rangle$ with $|m| = l = n-1$, $|m| = n-2$, $l = n-1, n-2$ and $|m| = n-3$, $l = n-2$ can be regarded as isolated. Such states include, in particular, the 1*s* and 2*s* states as well as 2*p* and 3*p* states with any values of *m*. For such states, the expressions for DP and DHP can be derived using the standard formula of perturbation theory for quasi-energy states [16] disregarding degeneracy. Choosing the operator V describing the interaction between an electron and a wave in the "form of velocity," i.e., using the notation with the momentum operator $\mathbf{p} = -i\nabla$ (here and below, atomic units are used except for specially stipulated cases),

$$V(\mathbf{r}, t) = -\frac{1}{c}\mathbf{A}(t)\cdot\mathbf{p} + \frac{1}{2c^2}\mathbf{A}^2(t),$$
$$\mathbf{A}(t) = \frac{c}{\omega}F\mathrm{Im}\{\mathbf{e}e^{-i\omega t}\}, \tag{6}$$

we can write the DHP as a combination of fourth-order matrix elements of perturbation theory,

$$\begin{aligned}\gamma_{nlm} = -\frac{F^4}{2\omega^4}e_p^* e_q^* e_r e_s[&(T_{rp}^{(1)}(l, l; -\omega) + T_{pr}^{(1)}(l, l; \omega))\\ &\times (T_{sq}^{(2)}(l, l; -\omega) + T_{qs}^{(2)}(l, l; \omega))\\ &- T_{pqrs}(l, l; \omega, 2\omega, \omega) - T_{rspq}(l, l; -\omega, -2\omega, -\omega)\\ &- T_{prqs}^R(l, l; \omega, 0, \omega)\\ &- T_{rpsq}^R(l, l; -\omega, 0, -\omega) - T_{prsq}^R(l, l; \omega, 0, -\omega)\\ &- T_{rpqs}^R(l, l; -\omega, 0, \omega)],\end{aligned} \tag{7}$$

where

$$\begin{aligned}T_{pq}^{(N)}(l, l'; \omega) &= \langle nl'm|\nabla_p[G_{E_n^{(0)}+\omega}]^N\nabla_q|nlm\rangle,\\ T_{pqrs}(l, l'; \omega_1, \omega_2, \omega_3) &= \langle nl'm|\nabla_p G_{E_n^{(0)}+\omega_1}\\ &\times\nabla_q G_{E_n^{(0)}+\omega_2}\nabla_r G_{E_n^{(0)}+\omega_3}\nabla_s|nlm\rangle,\\ T_{pqrs}^R(l, l'; \omega_1, 0, \omega_3) &= \langle nl'm|\nabla_p G_{E_n^{(0)}+\omega_1}\\ &\times\nabla_q G_{E_n^{(0)}}^R\nabla_r G_{E_n^{(0)}+\omega_3}\nabla_s|nlm\rangle,\end{aligned} \tag{8}$$

G_E is the Coulomb Green function and $G_{E_n^{(0)}}^R$ is the reduced Coulomb Green function (in the spectral decomposition of the latter function, all states corresponding to the unperturbed value of energy $E_n^{(0)}$ are omitted). It should also be noted that, in our notation, the DP has the form

$$\alpha_{nlm} = \frac{1}{\omega_2}\{-1 + e_p^* e_r[T_{rp}^{(1)}(l, l; -\omega) + T_{pr}^{(1)}(l, l; \omega)]\}, \tag{9}$$

and $E^{(2)} = -(1/4)\alpha_{nlm}F^2$ in accordance with formula (3).

It can be seen from expression (7) that the term with $\mathbf{A}^2$ in operator (6) makes zero contribution to the DHP of an isolated energy level. The expression for γ_{nlm} using the interaction operator in the "form of length",

$$V = \mathbf{F}(t)\cdot\mathbf{r}, \tag{10}$$

can be derived from formulas (7) and (8) using the simple substitution $\nabla/\omega \longrightarrow \mathbf{r}$. The choice of the interaction operator in form (6) makes it possible to separate the factor ω^{-4} in the expression for γ_{nlm} (see Eq. (7)); this reduces the mutual compensation of matrix elements at high frequencies and, as a result, leads to a higher accuracy in computations. Consequently, the

"velocity form" of operator V turns out to be more convenient when the DHP is calculated in the range of above-threshold frequencies $\omega > |E_n|$. In addition, operator (6) leads to more compact expressions for matrix elements in calculations with the generalized Sturm expansion of the Coulomb Green function [7] (see Eq. (33) below), which will be used in this study. At the same time, the operator of interaction in the "form of length" makes it possible to reduce the compensation of matrix elements in the low-frequency range and leads to more compact expressions for γ_{nlm} when the standard Sturm expansion (31) of the Coulomb Green function is used.

2.2. Degenerate Levels

It was noted above that, in view of the Coulomb degeneracy of energy levels in the orbital angular momentum l, the perturbation theory for degenerate levels should be used for calculating the perturbation of the spectrum of excited states (except for the states stipulated in the previous section) even in the case of a linearly or circularly polarized wave. In this case, the computational procedure for quasi-energy E is as follows [16]. Suppose that we have N states corresponding to an unperturbed energy level $E_n^{(0)}$. The proper wavefunction of the zeroth approximation can be represented in the form of a linear combination of N unperturbed functions $|\Phi_i^{(0)}\rangle$. Coefficients a_i of this combination and quasi-energy E satisfy the system of equations

$$\sum_{j=1}^{N} a_j(Q_{ij} - \Delta E\delta_{ij}) = 0, \tag{11}$$

where $\Delta E = E - E_n^{(0)}$. Matrix elements $Q_{ij} = Q_{ij}(F)$ can be expanded into a power series in F:

$$Q_{ij} = Q_{ij}^{(2)} + Q_{ij}^{(4)} + \ldots, \quad Q_{ij}^{(k)} \sim F^k. \tag{12}$$

In the first nonvanishing order in F, we have

$$\sum_{j=1}^{N} a_j^{(0)}(Q_{ij}^{(2)} - E^{(2)}\delta_{ij}) = 0, \tag{13}$$

where matrix elements $Q_{ij}^{(2)}$ written in terms of operator ∇ have the form

$$Q_{ij}^{(2)} = -\frac{F^2}{4\omega^2}\delta_{ij} + \frac{F^2}{4\omega^2}(\langle njm|(\mathbf{e}^*\cdot\nabla)G_{E_n^{(0)}+\omega}(\mathbf{e}\cdot\nabla)|nim\rangle + \langle njm|(\mathbf{e}\cdot\nabla)G_{E_n^{(0)}-\omega}(\mathbf{e}^*\cdot\nabla)|nim\rangle) \tag{14}$$

and exhibit the symmetry property, $Q_{ij}^{(2)} = Q_{ji}^{(2)}$ (which becomes obvious after integrating the matrix elements in Eq. (14) with respect to angular variables; see [6, 7]). Solving Eq. (13), we obtain a set of second-order corrections $E_\alpha^{(2)}$ to energy $E_n^{(0)}$ and the corresponding coefficients $a_{i,\alpha}^{(0)}$ defining the regular wave function $|\Phi_\alpha\rangle$ in the zeroth approximation. For above-threshold frequencies ($\omega > |E_n^{(0)}|$), the Green function $G_{E_n^{(0)}+\omega}$ in Eq. (14) has an imaginary part so that $E_\alpha^{(2)}$ and $a_{i,\alpha}^{(0)}$ are complex-valued and $|\Phi_\alpha\rangle$ describes a quasi-stationary quasi-energy state decaying as a result of single-photon ionization of degenerate states mixed by the field. We assume that coefficients $a_{i,\alpha}^{(0)}$ satisfy the relation

$$\sum_{i=1}^{N}(a_{i,\alpha}^{(0)})^2 = 1, \tag{15}$$

which corresponds to the normalization condition imposed on the wave functions of quasi-stationary quasi-energy states [4, 17]. Condition (15) corresponds to the fact that wave function $\langle\Phi_\alpha|$ in the first nonvanishing order of perturbation theory has the form

$$\langle\Phi_\alpha| = \sum_{i=1}^{N} a_{i,\alpha}^{(0)}\langle\Phi_i^{(0)}|, \tag{16}$$

although coefficients $a_{i,\alpha}^{(0)}$ may be complex-valued (for frequencies higher the single-photon ionization threshold); i.e., the bra function is obtained not as a simple complex conjugation of the ket function $|\Phi_\alpha\rangle$, which is due to the non-Hermitian form of symmetric matrix $Q_{ij}^{(2)}$.

In order to determine the fourth-order correction $E_\alpha^{(4)}$ to quasi-energy $E_n^{(0)} + E_\alpha^{(2)}$, we extend the method proposed in [18] for stationary perturbation theory in the presence of closely spaced levels to quasi-stationary quasi-energy states. We write Eq. (11) to within the fourth-order terms in F:

$$\sum_{j=1}^{N}[a_{j,\alpha}^{(2)}(Q_{ij}^{(2)} - E_\alpha^{(2)}\delta_{ij}) + a_{j,\alpha}^{(0)}(Q_{ij}^{(4)} - E_\alpha^{(4)}\delta_{ij})] = 0. \tag{17}$$

Multiplying this equation by $a_{i,\alpha}^{(0)}$, summing over i with due regard for the symmetry of $Q_{ij}^{(2)}$, and taking into account Eq. (13), we find that the first term in Eq. (17) vanishes, while the second term, on account of normal-

ization condition (15), gives the following expression for correction $E_\alpha^{(4)}$:

$$E_\alpha^{(4)} = \sum_{i,j=1}^{N} a_{i,\alpha}^{(0)} a_{j,\alpha}^{(0)} Q_{ij}^{(4)}. \tag{18}$$

The explicit expression for matrix elements $Q_{ij}^{(4)}$ in terms of operator ∇ has the form

$$\begin{aligned} Q_{ij}^{(4)} &= \frac{F^2}{4\omega^2} e_p e_r^* (T_{pr}^{(2)}(i,j;-\omega) + T_{rp}^{(2)}(i,j;\omega)) \\ &\times \left(E_\alpha^{(2)} + \frac{F^2}{4\omega^2} \right) - \frac{F^4}{16\omega^4} e_p e_q e_r^* e_s^* \\ &\times (T_{pqrs}(i,j;-\omega,-2\omega,-\omega) \\ &+ T_{rspq}(i,j;\omega,2\omega,\omega) \\ &+ T_{rpsq}^R(i,j;\omega,0,\omega) + T_{prqs}^R(i,j;-\omega,0,-\omega) \\ &+ T_{rpqs}^R(i,j;\omega,0,-\omega) + T_{prsq}^R(i,j;-\omega,0,\omega)), \end{aligned} \tag{19}$$

where we have retained notation (8), while indices and arguments *i* and *j* correspond to the values of the orbital angular momentum for degenerate states. It should be noted that, in calculating correction $E^{(4)}$ for degenerate states, the term with $\mathbf{A}^2$ in Eq. (6) makes a nonzero contribution since the expression for $E_\alpha^{(2)}$ in Eq. (19) cannot be written any longer in a simple analytic form following from formula (9) (the first term in formula (9) eliminates the dependence on $\mathbf{A}^2$ in the expression for γ in the case of an isolated energy level).

2.3. Separation of Angular Variables

Expressions (7) and (19) can be simplified by integrating with respect to angular variables in the matrix elements and summing over the components of polarization vector $\mathbf{e}$ for the linear ($\mathbf{e} = \mathbf{e}_z$) and circular ($\mathbf{e} = (\mathbf{e}_x + i\xi\mathbf{e}_y)/\sqrt{2}$, $\xi = \pm 1$) polarizations of the light field. Writing the Green functions in the form of a multipole expansion,

$$G_E(\mathbf{r},\mathbf{r}') = \sum_{lm} g_l(E;r,r') Y_{lm}(\hat{\mathbf{r}}) Y_{lm}^*(\hat{\mathbf{r}}'), \tag{20}$$

where g_l is the radial component of the Green function, $Y_{lm}(\hat{\mathbf{r}})$ are spherical functions, and $\hat{\mathbf{r}} = \mathbf{r}/r$, and carrying out the standard computations using the irreducible tensor operator technique [19], we obtain the following expression for the fourth-order matrix element of the general form:

$$\begin{aligned} & e_{1p}^{(4)} e_{1q}^{(3)} e_{1r}^{(2)} e_{1s}^{(1)} T_{pqrs}(l,l';\mathscr{E}_1,\mathscr{E}_2,\mathscr{E}_3) \\ &= \langle nl'm|(\mathbf{e}_4\cdot\nabla) G_{\mathscr{E}_3}(\mathbf{e}_3\cdot\nabla) \\ &\times G_{\mathscr{E}_2}(\mathbf{e}_2\cdot\nabla) G_{\mathscr{E}_1}(\mathbf{e}_1\cdot\nabla)|nlm\rangle \\ &= \sum_p C_{l'mp0}^{lm} \sum_{p_1 p_2} \{\{\mathbf{e}_3\otimes\mathbf{e}_4\}_{p_1}\otimes\{\mathbf{e}_2\otimes\mathbf{e}_1\}_{p_2}\}_{p_0} \\ &\times \left(\frac{\Pi_{p_1 p_2 p}}{\Pi_l}\right)^{1/2} \sum_{l_1 l_2 l_3} \begin{Bmatrix} 1 & 1 & p_1 \\ l' & l_2 & l_3 \end{Bmatrix} \\ &\times \begin{Bmatrix} 1 & 1 & p_2 \\ l & l_2 & l_1 \end{Bmatrix} \begin{Bmatrix} p_1 & p_2 & p \\ l & l' & l_2 \end{Bmatrix} \\ &\times \langle nl'|\nabla g_{l_3}(\mathscr{E}_3)\nabla g_{l_2}(\mathscr{E}_2)\nabla g_{l_1}(\mathscr{E}_1)\nabla|nl\rangle; \end{aligned} \tag{21}$$

here, $\mathbf{e}_i = \mathbf{e}$ or $\mathbf{e}^*$ in the case of photon emission or absorption, respectively; $\Pi_{l_i} = \sqrt{2l_i+1}$; and the standard notation [19] is used for the Clebsch–Gordon coefficients and 6*j* symbols. Passing from the reduced to radial matrix elements in accordance with the well-known formula [19]

$$\begin{aligned} \langle n'l'||\nabla||nl\rangle &= \sqrt{l+1}\langle n'l'|\frac{d}{dr} - \frac{l}{r}|nl\rangle\delta_{l',l+1} \\ &- \sqrt{l}\langle n'l'|\frac{d}{dr} + \frac{l+1}{r}|nl\rangle\delta_{l',l-1}, \end{aligned} \tag{22}$$

we can write expression (21) in terms of the combination of fourth-order radial matrix elements,

$$\begin{aligned} R_{ll'}^{l_1 l_2 l_3}(\mathscr{E}_1,\mathscr{E}_2,\mathscr{E}_3) &= \langle nl'|\hat{D}(l',l_3) g_{l_3}(\mathscr{E}_3)\hat{D}(l_3,l_2) \\ &\times g_{l_2}(\mathscr{E}_2)\hat{D}(l_2,l_1) g_{l_1}(\mathscr{E}_1)\hat{D}(l_1,l)|nl\rangle, \end{aligned} \tag{23}$$

where

$$\hat{D}(l_1,l_2) = \frac{d}{dr} + \frac{\mathrm{sgn}(l_2-l_1)\max(l_1,l_2)+1}{r}. \tag{24}$$

Relation (22) implies that, for a fixed *l*, the quantity *l'* in Eq. (21) may assume values of l, $l\pm 2$, and $l\pm 4$; in other words, only the *l*th sublevels of the *n*th shell with identical parities are mixed. Analogously, the first term in Eq. (19) can be reduced to a linear combination of the products of the second- and third-order radial matrix elements:

$$\begin{aligned} R_{ll'}^{L(N)}(\mathscr{E}) &= \langle nl'|\hat{D}(l',L)[g_L(\mathscr{E})]^N\hat{D}(L,l)|nl\rangle, \\ N &= 1,2. \end{aligned} \tag{25}$$

Ultimately, the dependence on the magnetic quantum number is separated in a correction on the order of

F^4 for isolated levels, and quantity $E^{(4)}$ in linear (L) and circular (c) fields can be expressed in terms of parameters $\gamma_p^{L,c}(\omega)$, which depend on the frequency and polarization type [5, 20],

$$E^{(4)} = -\frac{F^4}{8}\sum_{p=0}^{4} C_{lmp0}^{lm}\left[\frac{(2l+2)_p}{(2l+1-p)_p}\right]^{1/2}\gamma_p^{L,c}(\omega), \quad (26)$$

where $(a)_p$ is Pochhammer's symbol. Expression (26) can be written in elementary form. In a linear field, only three parameters γ_p^L with even p differ from zero:

$$E_{nlm}^{(4)L} = -\frac{F^4}{8}\left[\gamma_0^L + \gamma_2^L\frac{3m^2 - l(l+1)}{l(2l-l)}\right.$$
$$\left. + \gamma_4^L\frac{35m^4 + (25-2l-2l^2)m^2 + 3l(l^2-1)(l+2)}{2l(l-1)(2l-1)(2l-3)}\right]. \quad (27)$$

In a circular field, the fourth-order correction to quasi-energy is determined by five invariant parameters; in this case, $\gamma_p^L(\omega)$ in formula (27) is replaced by $\gamma_p^c(\omega)$, and the terms with odd powers of m are added:

$$\xi\gamma_1^c\frac{m}{l} + \xi\gamma_3^c\frac{m[5m^2+1-3l(l+1)]}{l(l-1)(2l-1)}. \quad (28)$$

Rather cumbersome expressions for irreducible components of $\gamma_p^{L,c}(\omega)$ in terms of radial matrix elements for an arbitrary l are of no interest and are not given here (the results for the s, p, and d states can be found, for example, in [20]). It should only be noted that, in view of the difference in the dipole selection rules, parameters $\gamma_p^L(\omega)$ and $\gamma_p^c(\omega)$ contain different combinations of radial matrix elements so that their frequency dependences differ significantly. In particular, parameter $\gamma_p^L(\omega)$ for s states exhibits two-photon resonant singularities on intermediate $n's$ states, while no such resonances are observed in the case of circular polarization.

Parameters $\gamma_p^{L,c}$ in Eq. (26) are similar to irreducible DP components of state $|nlm\rangle$, which determine the quadratic (in F) correction to quasi-energy,

$$E^{(2)} = -\frac{F^2}{4}\sum_{p=0}^{2} C_{lmp0}^{lm}\left[\frac{(2l+2)_p}{(2l+1-p)_p}\right]^{1/2}\alpha_p^{L,c}(\omega); \quad (29)$$

the only difference is that the value of $E^{(2)}$ in linear and circular fields is determined by the same parameters $\alpha_0^L = \alpha_0^c$ and $\alpha_2^L = -2\alpha_2^c$, while the DHP tensor components γ_p^L and γ_p^c with $p = 0, 2, 4$ are linearly independent.

For degenerate energy levels, corrections $E^{(2)}$ and $E^{(4)}$ cannot be parameterized explicitly analogously to Eqs. (29) and (26), although the dependence on the magnetic quantum number in matrix elements $Q_{ij}^{(2)}$ and $Q_{ij}^{(4)}$ can be separated by using the Clebsch–Gordon coefficients in this case also. After this, the evaluation of the corrections is reduced to the calculation of radial matrix elements, the standard procedure of diagonalization of matrix $Q_{ll'}^{(2)}$ in the second order in F (see Eq. (13)), and the determination of $E^{(4)}$ in accordance with formula (18).

2.4. Calculation of Compound Radial Matrix Elements for Virtual Transition to the Continuum

The main difficulty in the calculation of DHP for above-threshold values of frequency is associated with the evaluation of fourth-order radial matrix elements (23). The main condition for evaluating the compound matrix elements successfully is the existence of an appropriate representation for the radial component of the Coulomb Green function. The use of the spectral decomposition

$$g_l(E; r, r') = \sum_{n=l+1}^{\infty}\frac{R_{nl}(r)R_{nl}(r')}{E_n - E}$$
$$+ \int_0^{\infty} d\varepsilon\frac{R_{\varepsilon l}(r)R_{\varepsilon l}(r')}{\varepsilon - E - i0} \quad (30)$$

is helpful in a general analysis of radial matrix elements (the investigation of the resonance structure and high-frequency asymptotic form as well as the separation of the real and imaginary parts of transition amplitudes). However, this spectral decomposition is absolutely ineffective for numerical calculations because of the necessity to integrate over the continuous spectrum, the slow convergence of the sum over the discrete spectrum, and the compensation of contributions from these spectra to matrix elements. It is most convenient to evaluate matrix element (23) using the expansion of the Coulomb Green function in the Sturm functions $S_{kl}(2r/\nu)$ [21],

$$g_l(E; r, r') = \nu\sum_{k=0}^{\infty}\frac{k!S_{kl}(2r/\nu)S_{kl}(2r'/\nu)}{(k+2l+1)!(k+l+1-\eta)}, \quad (31)$$

where

$$S_{kl}\left(\frac{2r}{\nu}\right) = \frac{2}{\nu}\left(\frac{2r}{\nu}\right)^l\exp\left(-\frac{r}{\nu}\right)L_k^{2l+1}\left(\frac{2r}{\nu}\right), \quad (32)$$

L_n^α is the generalized Laguerre polynomial, $\nu = (-2E - i0)^{-1/2}$, and $\eta = Z\nu$, Z being the nuclear charge. Using relation (31), it is possible to represent matrix elements (23) and (25) in the form of series of hypergeometrical polynomials, which converge, however, only for negative energy parameters of the Green function ($\mathcal{E}_i < 0$). For above-threshold frequency values, when at least some of the energies in the Green functions become positive, the convergence of the Sturm series for matrix elements is violated; in this case, the evaluation of the matrix elements is a more complicated problem. In order to solve this problem, a generalized Sturm expansion of the Coulomb Green function was proposed in [22]. This expansion contained a free parameter α, the appropriate choice of which can ensure the convergence of the series for matrix elements, including the case of positive energies ($\mathcal{E}_i > 0$) in the Coulomb Green function. Such a representation of the Coulomb Green function was used for calculating the nonlinear susceptibilities of the hydrogen atom at frequencies exceeding the ionization threshold (in particular, DHP for the 1*s*, 2*s*, and 2*p* states [2]). However, the effectiveness of this method decreases rapidly with increasing frequency and principal quantum number *n* of the initial atomic state.

A convenient version of the expansion of the Coulomb Green function in the Sturm functions with free parameters was proposed in [7, 12], where $g_l(E)$ was represented in the form of a double series in functions $S_{kl}(2r/\alpha)$ and $S_{k'l}(2r'/\alpha')$:

$$g_l(E; r, r') = \sum_{k, k' = 0}^{\infty} g_{kk'}^l(\nu; \alpha, \alpha') S_{kl}\left(\frac{2r}{\alpha}\right) S_{k'l}\left(\frac{2r'}{\alpha'}\right). \quad (33)$$

Coefficients $g_{kk'}^l$ of this expansion can be expressed in terms of the hypergeometric Gaussian function ${}_2F_1$ and Appell function F_1:

$$g_{kk'}^l(\nu; \alpha, \alpha') = f(\alpha, \alpha') \times \left[\frac{{}_2F_1(-k, l + 1 - \eta; 2l + 2; z)}{l + 1 - \eta} \right.$$

$$\times F_1(l + 1 - \eta; -k', k' + 2l + 2; l + 2 - \eta; y, y') \quad (34)$$

$$\left. + \sum_{p = 1}^{k} C_k^p (-z)^p \times \frac{{}_2F_1(-k + p, l + 1 - \eta + p; 2l + 2 + p; z)}{(2l + 2)_p} \Phi_p^{lk'} \right].$$

Here,

$$z = -\frac{4\alpha\nu}{(\alpha - \nu)^2},$$

$$\Phi_p^{lk'} = -\frac{(l + 2 + \eta - p)_{p-1} (1 - y)^{k'}}{(1 - y')^{k' + 2l + 2}}$$

$$\times F_1\left(-p + 1; -k', k' + 2l + 2; l + 2 + \eta - p; \frac{1}{1 - y}, \frac{1}{1 - y'}\right),$$

$$y = \frac{\alpha - \nu}{\alpha + \nu} \frac{\alpha' + \nu}{\alpha' - \nu}, \quad y' = \frac{\alpha - \nu}{\alpha + \nu} \frac{\alpha' - \nu}{\alpha' + \nu},$$

and C_k^p is the binomial coefficient. Factor f has the form

$$f(\alpha, \alpha') = \nu \frac{(4\nu\sqrt{\alpha\alpha'})^{2l + 2}}{\Gamma(2l + 2)} \frac{(\alpha - \nu)^k}{(\alpha + \nu)^{k + 2l + 2}} \times \frac{(\alpha' - \nu)^{k'}}{(\alpha' + \nu)^{k' + 2l + 2}}.$$

For identical parameters $\alpha = \alpha'$, coefficients $g_{kk'}^l$ can be simplified,

$$g_{kk'}^l(\nu; \alpha, \alpha) = \frac{\nu}{\Gamma(2l + 2)} \left(\frac{\alpha - \nu}{\alpha + \nu}\right)^{k_<} \times {}_2F_1(-k_<, l + 1 - \eta; 2l + 2; z) \left(\frac{\alpha^2 - \nu^2}{4\alpha\nu}\right)^{k_>} \quad (35)$$

$$\times \frac{k_>! {}_2F_1(k_> + 1, k_> + 2l + 2; k_> + l + 2 - \eta; z^{-1})}{(l + 1 - \eta)_{k_> + 1}}$$

(where $k_< = \min\{k, k'\}$, $k_> = \max\{k, k'\}$, and expression (33) coincides with the result obtained in [23]. For $\alpha = \alpha' = \nu$, expansion (33) is transformed into (31).

An important circumstance ensuring considerable versatility of application of expansion (33) for practical purposes is the factorized dependence of the terms in the series on r, r', and energy parameter ν. The entire energy dependence is contained in kernel $g_{kk'}^l(\nu; \alpha, \alpha')$, which is independent of radial variables. A rational choice of (generally complex-valued) parameters α and α' in accordance with the specific features of a concrete problem makes it possible in some cases to radically simplify the procedure for calculating matrix elements. Representation (33) turned out to be helpful for analytic calculations and made it possible to obtain universal closed expressions for the matrix elements of two-photon bound-bound and bound-free transitions from an arbitrary $|nl\rangle$ state in the form of a linear combination of four quantities $g_{kk'}^l$ in the general case [7]. It will be shown in the Appendix that the application of expansion (33) is

also quite effective for calculating numerically the amplitudes of higher-order multiphoton transitions.

3. NUMERICAL RESULTS AND DISCUSSION

3.1. Dispersion Dependence of Dynamic Hyperpolarizability and Level Shift in a Strong Light Field

The algorithm described in the Appendix was used for systematic computation of the DHP for states with $n = 1$–5 and in individual test computations for higher energy levels. For 1*s*, 2*s*, and 2*p* states, the results of computations are in complete agreement with available results in the frequency range below the two-photon ionization threshold [2, 24]; the results obtained in the above-threshold range considerably extend the results described in [2]. The accuracy of computations was also controlled by independently calculating the value of γ with the interaction operator in the **A**–**p** (Eq. (6)) and **F**–**r** (Eq. (10)) representations; the results were found to be completely identical.

It follows from the general formulas in Section 2 that DHP γ (or correction $E^{(4)}$ to the quasi-energy, which is quadratic in intensity) exhibits a complex dependence on the quantum numbers of the atomic state in question as well as on the frequency and polarization of the light wave. Among general properties of the DHP, the pole and threshold singularities in the dispersion dependence are worth noting; these are one-photon (third-order) resonances and two-photon (first-order) resonances, as well as threshold anomalies [25] typical of potentials with the Coulomb asymptotic form, which are observed when one- and two-photon ionization channels are open.

In the case of nondegenerate $|nlm\rangle$ states, DHP $\gamma = \gamma_{nlm}^{L,c}$ in the linear (L) and circular (c) fields can be written in terms of irreducible components, which are proportional to quantities $\gamma_p^{L,c}$ in formula (26). For spherically symmetric *s* states, the DHP is determined by only one parameter (scalar DHP), which, in contrast to the DP, is different for linear and circular fields:

$$\gamma_{n00}^{L} = \gamma_L, \quad \gamma_{n00}^{c} = \gamma_c.$$

In the case of *p* states, after the separation of the dependence on the magnetic quantum number, the quantity γ_{nlm} is determined by the scalar ($\gamma_{L,c}^{(s)}$), vector ($\gamma_c^{(v)}$), and tensor ($\gamma_{L,c}^{(t)}$) DHPs, which are analogous to the corresponding irreducible polarizability components of the $|nlm\rangle$ state [2],

$$\begin{aligned}\gamma_{n1m}^{L} &= \gamma_L^{(s)} + \gamma_L^{(t)}(3m^2 - 2),\\ \gamma_{n1m}^{c} &= \gamma_c^{(s)} - \gamma_c^{(v)}\xi m + \gamma_c^{(t)}(3m^2 - 2),\end{aligned} \tag{36}$$

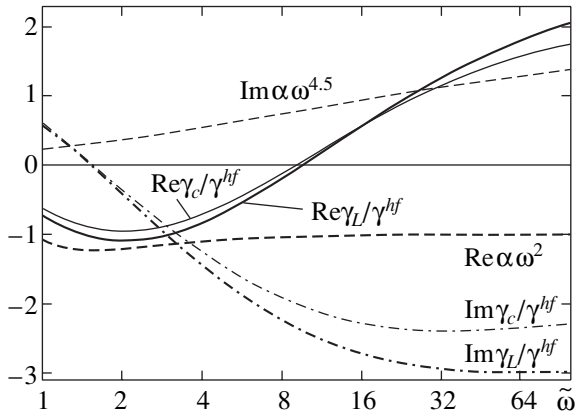


Fig. 1. Real and imaginary parts of the DHP for the 1*s* state of hydrogen above the single-photon ionization threshold in a linearly (bold curves) and circularly (fine curves) polarized fields. The solid and dot-and-dash curves correspond to the real and imaginary parts of the DHP, respectively. The curves were plotted by separating the high-frequency asymptotic factor $\gamma^{\mathrm{hf}} = A\omega^{-7.5}$ (see formula (37)). The real part of the DP (bold dashed curve) and its imaginary part (fine dashed line) are shown for comparison. The field frequency is measured in units of the ionization potential: $\tilde{\omega} = \omega/|E_n|$, $n = 1$.

where $\xi = \pm 1$ is the degree of circular polarization of the light wave.

Figure 1 shows the DHPs of the 1*s* state for linearly and circularly polarized fields at frequencies above the single-photon ionization threshold up to values of $\omega = 100|E_{1s}| = 50$ at. units. (At higher frequencies, the applicability of the dipole approximation becomes questionable since the order of magnitude of nondipole corrections is determined by parameter $(\alpha\omega)^2$, $\alpha = 1/137$.) The results shown in the figure demonstrate the computational potentialities of the method and singularities in the DHP approaching the asymptotic form in the high-frequency range. The curves were plotted by separating the asymptotic form of the DHP [2]:

$$\begin{aligned}\gamma_{L,c} &= a_{L,c}A\frac{1-i}{\omega^{7.5}}, \quad a_L = 3,\\ a_c &= 2, \quad A = 0.26.\end{aligned} \tag{37}$$

The general form of this relation was established using the method described in [26], and constant A was determined numerically in the region of $\omega \sim 200|E_{1s}|$. It can be seen from Fig. 1 that the asymptotic value of DHP is attained at much higher frequencies as compared to the case of DP $\alpha_{1s}(\omega)$ (especially for the real part that reverses its sign for $\omega \approx 8|E_{1s}|$), where $\mathrm{Re}\,\alpha_{1s}(\omega)$ is very close to its asymptotic value of $-1/\omega^2$. This apparently complicates approximation of the DHP by simple analytic formulas even in the range of above-threshold frequencies. It can be mentioned for comparison that the asymptotic ratio obtained analytically is $a_L/a_c = 1.5$,

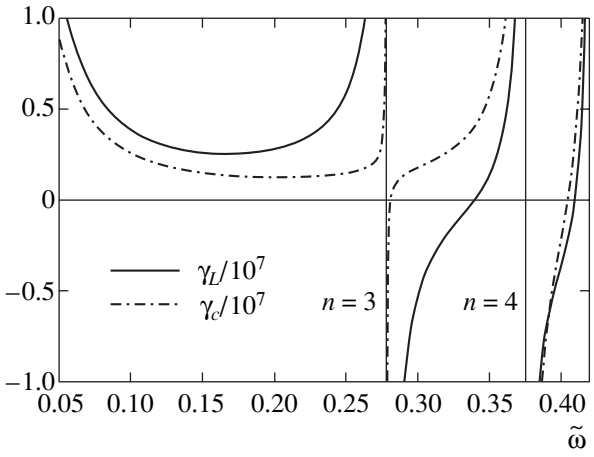


Fig. 2. DHP for the 2*s* state of hydrogen in the below-threshold frequency range in a linearly (solid curves) and circularly (dot-and-dash curves) polarized fields. Fine vertical lines indicate the position of two-photon resonances on states with $n = 3$ (3*s*, 3*d* for linearly and 3*d* for circularly polarized fields) and $n = 4$ (4*s* and 4*d* for linearly and 4*d* for circularly polarized fields); $\tilde{\omega} = \omega/|E_n|$, $n = 2$.

while the numerical ratio of the real and imaginary parts of the DHP for linear and circular polarizations at frequencies $\omega = 100|E_{1s}|$, $150|E_{1s}|$, $250|E_{1s}|$ is $\mathrm{Re}\gamma_L/\mathrm{Re}\gamma_c = 1.17$, 1.19, and 1.22, while $\mathrm{Im}\gamma_L/\mathrm{Im}\gamma_c = 1.30$, 1.33, and 1.36, respectively.

Figures 2–4 show the frequency dependence of γ for the 2*s* state in the below-threshold and above-threshold regions. It should be noted that the real part of the DHP for excited states has "down" resonances in the above-threshold frequency range (corresponding to one- and two-photon resonant transitions of an electron to low-lying levels, which are allowed by the dipole selection rules). Figure 4 illustrates the dependence of the resonance structure on the wave polarization: some resonances that are observed in the linearly polarized field are absent in the circularly polarized field. In the given case, a two-photon down resonance on the 1*s* state is observed in the linear field at $\omega = 1.5|E_n|$, while in the circular field this resonance is forbidden by the selection rules in magnetic quantum number *m* (since $\Delta m = \pm 2$ in this case, and the two-photon 2*s*–1*s* transition is ruled out). Obviously, $\mathrm{Im}\gamma$ shows no down resonances.

Figures 5 and 6 show the irreducible components of $\gamma^{s,\,v,\,t}$ for the DHP of the 2*p* state in the above-threshold frequency region. In this case, only a single-photon resonance on the lower-lying 1*s* state is obviously possible at $\omega = 3|E_n|$. It should be noted that a comparison of Figs. 5 and 6 with Fig. 4 shows that the value of γ_{2p} at high frequencies decreases slightly more rapidly than γ_{2s}.

The behavior of hyperpolarizability of highly excited states is illustrated in Figs. 7–11, which represent some of the results obtained for states with $n = 5$. Figures 7 and 8 show the dispersion dependence of the real and imaginary parts of the fourth-order correction $E^{(4)}$ (determined by DHP γ_{nm}) for states $|n = 5, m = 0\rangle$. It should be recalled that these states are degenerate in the orbital quantum number (the states with identical parity being mixed); consequently, the position of energy levels in a light field should be calculated by diagonalizing the quasi-energy matrix even in the sec-

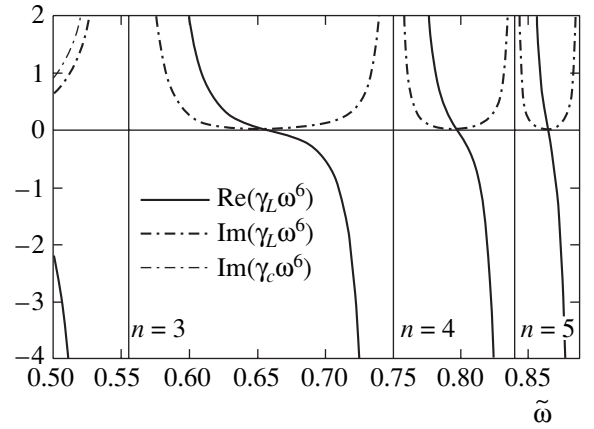


Fig. 3. Real and imaginary parts of the DHP for the 2*s* state of hydrogen in the frequency range between the two- and one-photon ionization thresholds for a linearly polarized field. Solid and dot-and-dash curves correspond to the real and imaginary parts of the DHP, respectively. Fine vertical lines indicate the position of one-photon resonances on states 3*p*, 4*p*, and 5*p*. In this frequency region, the DHPs for circularly and linearly polarized fields virtually coincide; slight difference is observed only for imaginary parts $\mathrm{Im}\gamma_{L,\,c}$ between the one-photon ionization threshold and the first resonance, where $\mathrm{Im}\gamma_c$ is shown by the fine dot-and-dash curve.

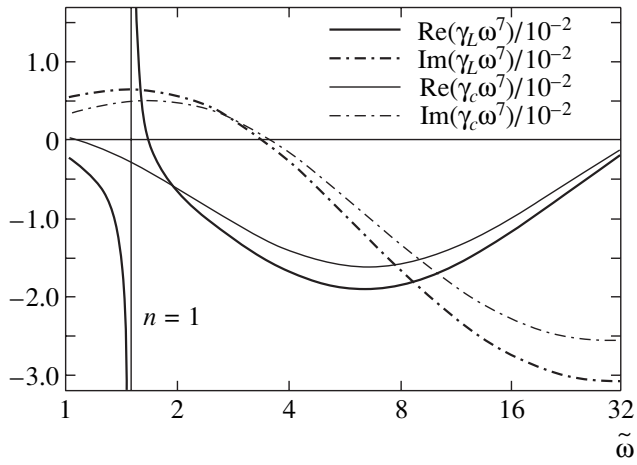


Fig. 4. Real and imaginary parts of the DHP for the 2*s* state of hydrogen behind the single-photon ionization threshold for linearly (bold curves) and circularly (fine curves) polarized fields. Solid curves correspond to the real parts, and dot-and-dash curves, to the imaginary part of the DHP; $\tilde{\omega} = \omega/|E_n|$, $n = 2$. The fine vertical line corresponds to the two-photon "down" resonance on the 1*s* state, which is absent in the case of circular polarization of the wave.

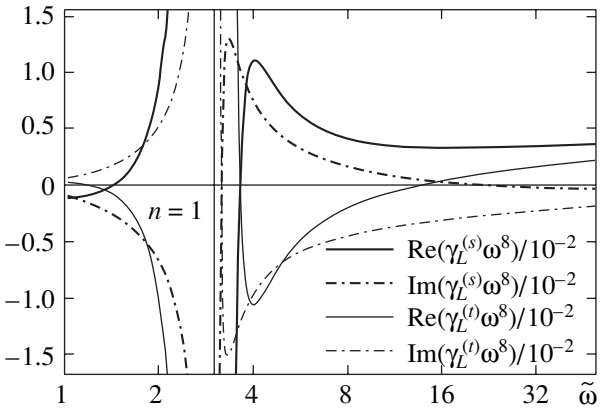


Fig. 5. Real and imaginary parts of the scalar and tensor DHPs for the 2*p* state of hydrogen at frequencies above the single-photon ionization threshold in a linearly polarized field. Solid and dot-and-dash curves correspond to the real and imaginary parts of the DHP, respectively; $\tilde{\omega} = \omega/|E_n|$, $n = 2$. The fine vertical line corresponds to the one-photon "down" resonance on the 1*s* state.

ond order in the field amplitude. Calculations show that degeneracy is especially significant for low frequencies ($\omega \lesssim |E_n|$) as well as in the vicinity of resonances. At frequencies much higher than the ionization threshold, the effect of degeneracy is virtually absent in view of a more rapid decrease in nondiagonal (in l, l') elements of the quasi-energy matrix for $\omega \longrightarrow \infty$ [26]. In addition, as can be seen from the figures, the state with $l = 4$ practically does not mix with the states corresponding to $l = 0$ and $l = 2$ at any frequency. The curves show that two from the three even states obtained as a result of mixing do not exhibit a one-photon "down" resonance on the 3*p* and 2*s* states, while this resonance is absent only for one of the three states with $l = 4$ if degeneracy is neglected. This fact of resonance "vanishing" becomes obvious if we write the quasi-energy matrix in the resonance approximation in the second order of perturbation theory and then calculate $E^{(4)}$. In this case, the terms containing resonance singularities are mutually compensated for two out of the three levels.

The frequency region in the vicinity of resonances requires special treatment since perturbation theory becomes inapplicable when ω approaches the resonance transition frequency. A peculiar situation arises when the energy levels of the initial n shell are mixed in orbital angular momentum l. It can be seen from Fig. 7 that an "extra" two-photon down resonance, which is not associated with physical factors, appears in the 2*p* state in the imaginary part of $E^{(4)}$. This is due to peculiarities in the normalization of quasi-stationary quasi-energy states (15), for which the dual bra function can be obtained using formula (16). In this case, the resonance existing only in real-valued fourth-order matrix elements appears in the imaginary part of $E^{(4)}$ (which cannot have resonance singularities in the given situation) due also to the complex-valuedness of coefficients $a^{(0)}_{i,\alpha}$ of the mixture. A correct analysis of the frequency range in the vicinity of the given resonances requires taking into account the fact that quasi-resonance levels for which $E^{(0)}_n - E^{(0)}_{n'} \approx k\omega$ are also regarded as close levels in the quasi-energy formalism. In our case of a two-photon resonance on the 2*p* state, we have $E^{(0)}_5 - E^{(0)}_2 \approx 2\omega$; i.e., $n = 5$, $n' = 2$, $k = 2$. Thus, a proper wave function of the zeroth approximation with a definite value of angular momentum projection m must contain a linear combination of both the quasi-energy states $|nlm\rangle$ of the initial shell n, which are degenerate in l, and (quasi-resonance) states $|n'l'm\rangle\exp(ik\omega t)$. In this case, in the spectral decomposition of Green function $G_{E^{(0)}_n - k\omega}$, states $|n'l'm\rangle$ that have already been included in the wave function in the zeroth approximation should be omitted. The results of calculations with an admixture of quasi-resonance states to the regular wave function

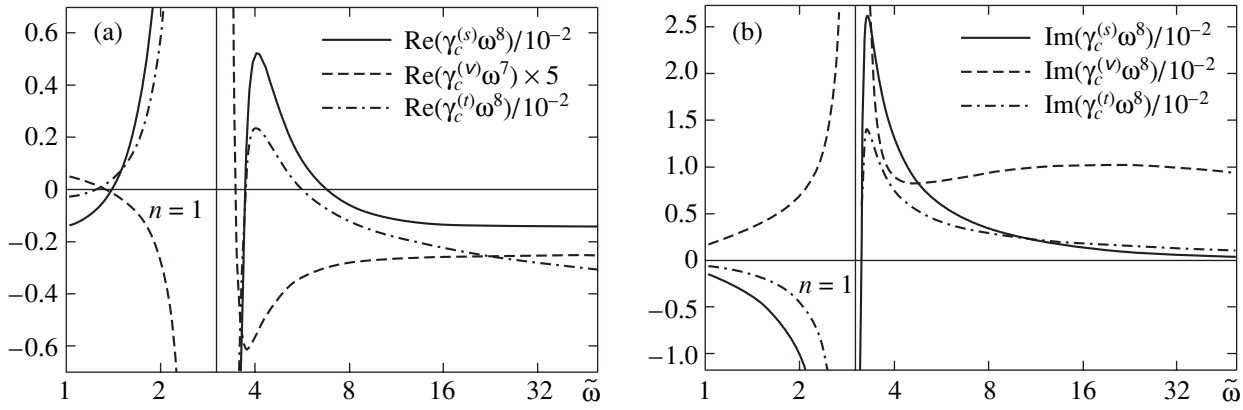


Fig. 6. Real (a) and imaginary (b) parts of the scalar, vector, and tensor hyperpolarizabilities for the 2*p* state of hydrogen in the above-threshold region for circularly polarized field; $\tilde{\omega} = \omega/|E_n|$, $n = 2$. The fine vertical line corresponds to the one-photon down resonance on the 1*s* state.

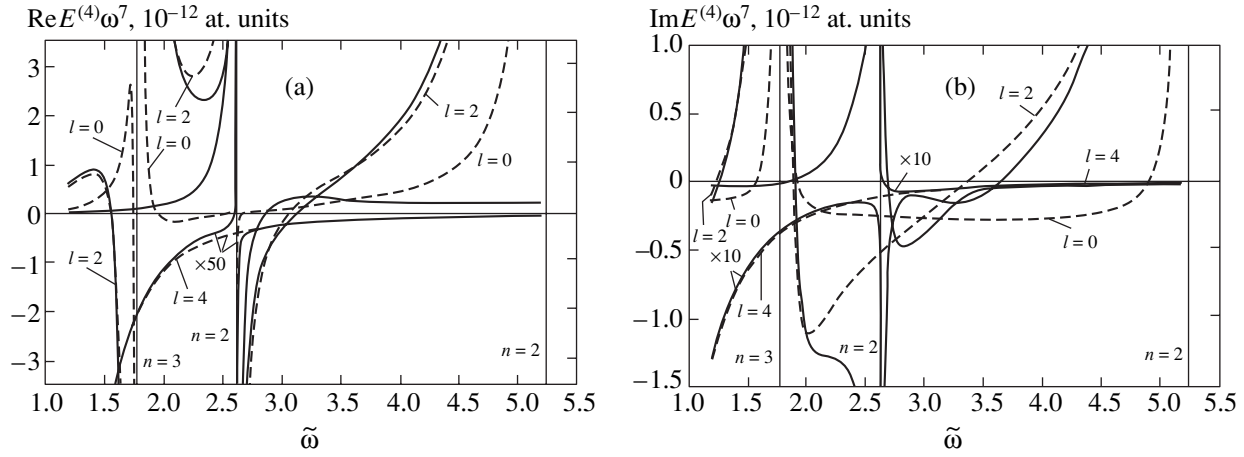


Fig. 7. Real (a) and imaginary (b) parts of the fourth-order correction $E^{(4)}$ for even states with $n = 5$, $m = 0$ in the above-threshold region for linearly polarized field; $F = 0.01$ at. unit, $\tilde{\omega} = \omega/|E_n|$. Solid and dashed curves are plotted taking into account and disregarding the degeneracy in l, respectively.

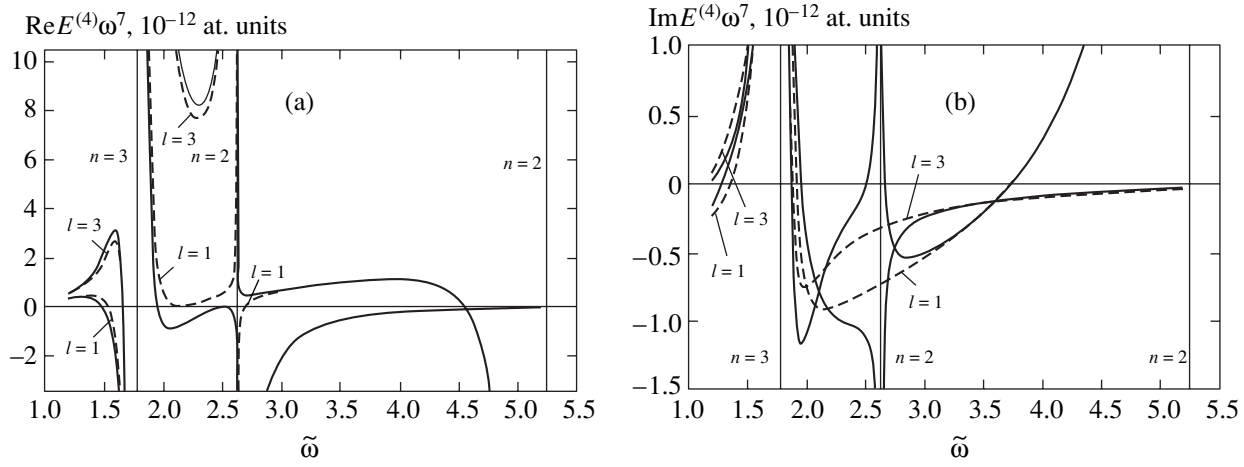


Fig. 8. Real (a) and imaginary (b) parts of the fourth-order correction $E^{(4)}$ for odd states with $n = 5$, $m = 0$ in the above-threshold region for linearly polarized field; $F = 0.01$ at. units, $\tilde{\omega} = \omega/|E_n|$. Solid and dashed curves are plotted taking into account and disregarding the degeneracy in l, respectively.

of the zeroth approximation are shown in Fig. 9 illustrating the dispersion dependence of $E^{(4)}$ for levels $|n = 5, m = 0\rangle$ in the vicinity of the two-photon "down" resonance on the $2p$ state for odd states. It can be seen from the figures that correction $E^{(4)}$ in the resonance region becomes a smooth function of frequency; as the frequency deviates farther and farther from the resonance, the effect of admixed energy levels from the shell with a different quantum number n' ($n' = 2$ in our case) becomes insignificant.

Figure 10 shows the dependence of the level shift on field amplitude F taking into account the terms on the orders of F^2 and F^4 for states with $n = 5$, $m = 0$. The calculations (both including and disregarding the degeneracy in l) were carried out for a field with linear polarization and frequencies $\omega = 1.5|E_n|$ and $\omega = 4|E_n|$. The values of F were limited by the condition of applicability of perturbation theory; i.e., the calculations were performed up to values of field intensity such that the contributions from terms on the order of F^4 were several times smaller than the contribution from terms on the order of F^2. Figure 11 shows for comparison the dispersion dependence of the DHP for isolated state $|n = 5, l = 4, m = 4\rangle$ in a linearly polarized field.

Numerical calculations for excited states show, on the whole, that the behavior of γ with increasing n considerably depends on the frequency range: for small values of ω, the DHP increases rapidly with n analogously to the static case; consequently, the contribution of $E^{(4)}$ to ΔE becomes significant. The role of fourth-order corrections is most significant in the vicinity of the ω values for which polarizability α vanishes and in the vicinity of two-photon resonances, which are observed for $E^{(4)}$ and are absent for $E^{(2)}$. Behind the sin-

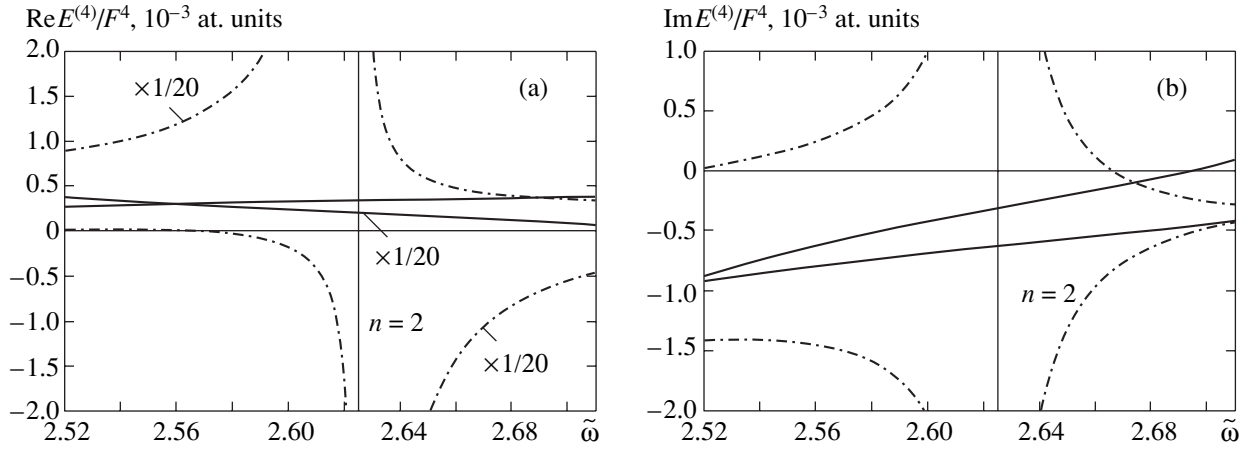


Fig. 9. Real (a) and imaginary (b) parts of the fourth-order correction $E^{(4)}/F^4$ for odd states with $n = 5$, $m = 0$ in the above-threshold frequency region for linearly polarized field taking into account (solid curves) and disregarding (dot-and-dash curves) the admixture of the $|2s\rangle$ resonance state; $\tilde{\omega} = \omega/|E_n|$.

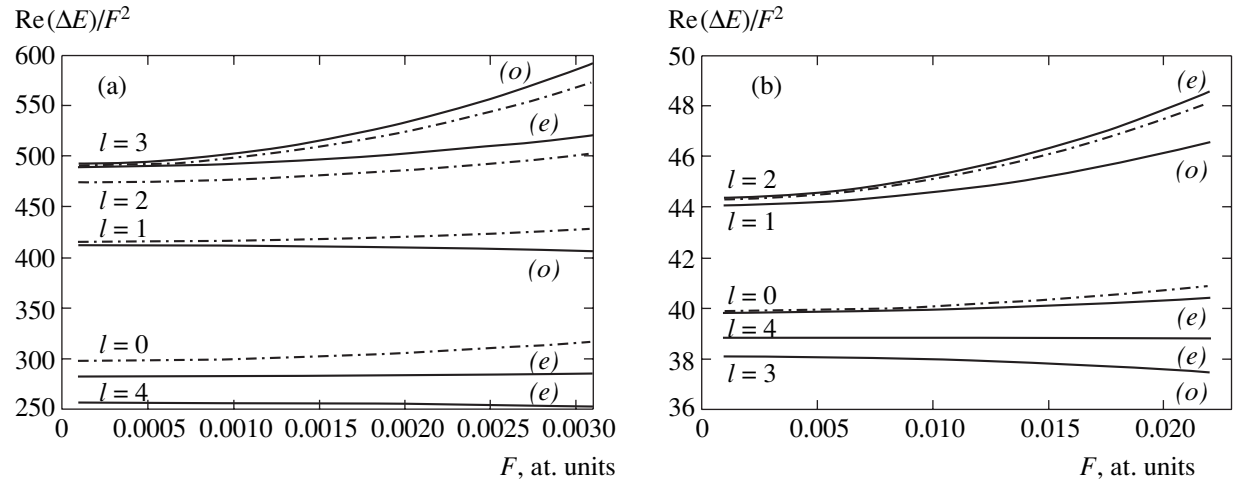


Fig. 10. Dependence of the level shift on the amplitude F of a linearly polarized field for states with $n = 5$, $m = 0$ taking into account (solid curves) and disregarding (dot-and-dash curves) the degeneracy in l. Field frequency $\omega = 1.5|E_n|$ (a) and $4|E_n|$ (b). Symbols (e) and (o) mark even and odd energy levels.

gle-photon ionization and outside the down resonance region, the value of γ decreases monotonically but rapidly with increasing n, and the role of corrections on the order of F^4 to the Stark shift and to the level splitting is insignificant. However, for $\omega > |E_n|$, the imaginary parts of α and γ are most interesting.

3.2. Corrections to the Probability of Photoeffect in a Strong Light Field

For above-threshold frequencies, correction ΔE to quasi-energy acquires an imaginary part determining the ionization broadening Γ of an isolated or degenerate level,

$$\Gamma = -2\mathrm{Im}\Delta E.$$

In the frequency interval $|E_n|/2 < \omega < |E_n|$, the polarizability is real-valued and the imaginary part of correction $\Delta E = E^{(2)} + E^{(4)}$ is determined by the imaginary part of the DHP. Direct calculation of Imγ is based on the application of the familiar relation

$$\frac{1}{x - i0} = \mathrm{V.p.}\left(\frac{1}{x}\right) + i\pi\delta(x) \qquad (38)$$

(symbol V.p. stands for the principal value of integral) for the transformation of the integral term in the spectral decomposition of Green function (30) with a positive energy of $E_n + 2\omega$. In the case of isolated states, evaluating $\mathrm{Im}\,R_{ll}^{l_1 l_2 l_3}(E_n + \omega, E_n + 2\omega, E_n + \omega)$ with the help of relation (38), we can easily obtain an explicit expression for Imγ in the form of a combination of the

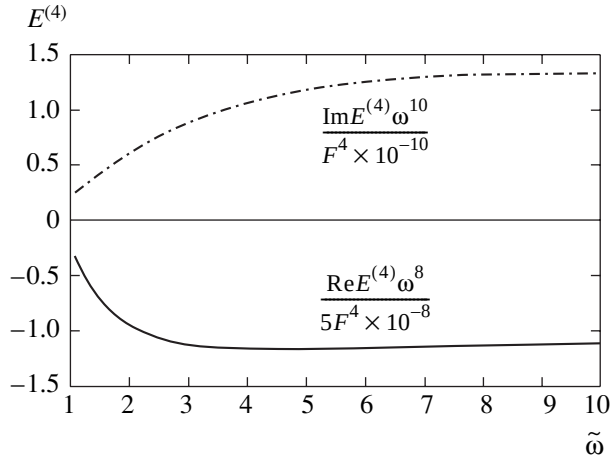


Fig. 11. Dispersion dependence of the fourth-order correction $E^{(4)}$ for state $|n = 5, l = 4, m = 4\rangle$ in a linearly polarized field; $\tilde{\omega} = \omega/|E_n|$.

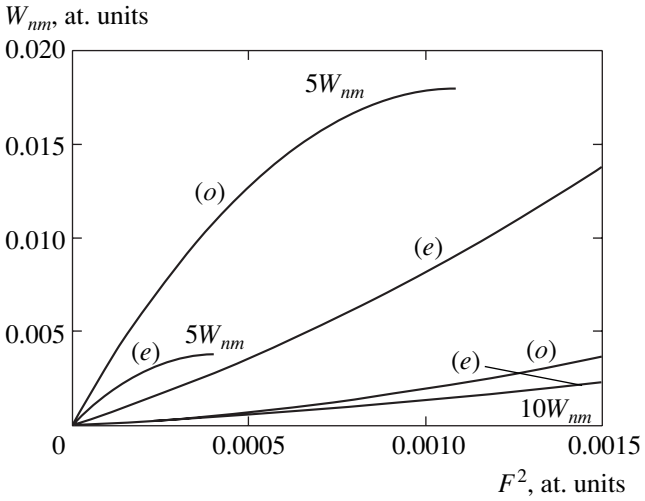


Fig. 12. Dependence of the total ionization probability $W_{nm} = -2\mathrm{Im}\{E^{(2)} + E^{(4)}\} = W^{(1)} + W^{(1-3)} + W^{(2)}$ on F^2 for states $|n = 5, m = 0\rangle$ mixed in the orbital angular momentum l in a linearly polarized field with frequency $\omega = 4|E_{n=5}| = 2.168$ eV. Symbols (*e*) and (*o*) mark even and odd states.

products of two-photon radial matrix elements of transitions to the continuous spectrum,

$$\left|A^{(2)}_{l_1 l_2}\right|^2, \quad 2\mathrm{Re}(A^{(2)*}_{l+1,\,l}A^{(2)}_{l-1,\,l}),$$

$$A^{(2)}_{l_1 l_2} = \langle E_n + 2\omega, l_2|\hat{D}(l_2, l_1) \times g_{l_1}(E_n + \omega)\hat{D}(l_1, l)|nl\rangle; \qquad (39)$$

$$l_1 = l \pm 1, \quad l_2 = l_1 \pm 1.$$

As a result, as follows from general considerations, the imaginary part of $E^{(4)}$ can be expressed in terms of two-photon ionization probability $W^{(2)}$:

$$W^{(2)} = -2\mathrm{Im}E^{(4)} = \frac{1}{4}\mathrm{Im}\gamma F^4. \qquad (40)$$

For $\omega > |E_n|$, the one-photon ionization channel is opened; the ionization probability $W^{(1)}$ in the lowest order of perturbation theory is determined by the imaginary part of the DP:

$$W^{(1)} = -2\mathrm{Im}E^{(2)} = \frac{1}{2}\mathrm{Im}\alpha F^2. \qquad (41)$$

In this frequency range, $\mathrm{Im}\gamma$ is due to the presence of the imaginary part not only in the Green function $g_L(E_n + 2\omega)$, but also in $g_L(E_n + \omega)$. Accordingly, $\mathrm{Im}E^{(4)}$ acquires additional terms from matrix elements $R^{ll}_{l_1 l_2 l_3}(\mathscr{E}_1, \mathscr{E}_2, \mathscr{E}_3)$ with energy $\mathscr{E}_1 = E_n + \omega$ or/and $\mathscr{E}_3 = E_n + \omega$:

$$\text{(a)}\ \ 2\mathrm{Re}(A^{(1)*}_{l_1}A^{(3)}_{l_1 l_2 l_3}(\omega, 2\omega)),$$

$$\text{(b)}\ \ 2\mathrm{Re}(A^{(1)*}_{l_1}A^{(3)}_{l_1 l_2 l_3}(\omega, 0)), \qquad (42)$$

$$\text{(c)}\ \ 2\mathrm{Re}(A^{(1)*}_{l_1}A^{(3)}_{l_1 l_2 l_3}(-\omega, 0)),$$

where

$$A^{(1)}_{l_1} = \langle E_n + \omega, l_1|\hat{D}(l_1, l)|nl\rangle,$$

$$A^{(3)}_{l_1 l_2 l_3}(\omega_1, \omega_2) = \langle E_n + \omega, l_1|\hat{D}(l_1, l_2)g_{l_2}(E_n + \omega_2) \times \hat{D}(l_2, l_3)g_{l_3}(E_n + \omega_1)\hat{D}(l_3, l)|nl\rangle,$$

$$l_1 = l \pm 1, \quad l_2 = l_1 \pm 1, \quad l_3 = l_2 \pm 1 = l + 1.$$

These matrix elements describe the interference of the amplitudes of conventional photoeffect and a three-photon process with reemission of photons (according to the scheme $E_n + 2\omega - \omega = E_n + \omega$. Thus, in this frequency range, $\mathrm{Im}E^{(4)}$ is the sum of two components:

$$-2\mathrm{Im}E^{(4)} = W^{(2)} + W^{(1-3)}. \qquad (43)$$

In addition to terms (42), component $W^{(1-3)}$ also includes the terms originating from the imaginary part of products $R^{(1)}R^{(2)}$ (so-called "nondiagrammatic" terms). The probability of single-photon ionization (resulting in the production of photoelectrons with energy $E_n + \omega$), calculated to within terms on the order of F^4, is defined by the sum

$$W_i = W^{(1)} + W^{(1-3)}. \qquad (44)$$

Direct calculation of correction $W^{(1-3)}$ by the methods of nonstationary perturbation theory involves considerable difficulties associated, in particular, with the necessity of computing matrix elements $A^{(3)}_{l_1 l_2 l_3}(\omega, 2\omega)$ and $A^{(3)}_{l_1 l_2 l_3}(\omega, 0)$ of three-photon transitions to the con-

Probabilities of one-photon ($W^{(1)}$) and two-photon ($W^{(2)}$) ionization and linear (in intensity) correction $W^{(1)}$ ($W^{(1-3)}$) for a number of states with principal quantum numbers $n = 3$ and 5; I_{thr} is the intensity corresponding to the stabilization threshold

n	m	l	w, eV	$W^{(1)}/F^2$, at. units	$W^{(2)}/F^4$, at. units	$W^{(1-3)}F^4$, at. units	I_{thr}, W/cm^2
3	0	1	5	1.4254	658.410	–100.10	2.570×10^{14}
3	1	1	5	0.9130	27.58	–60.811	2.710×10^{14}
3	1	2	5	0.2624	3.3032	14.305	–
3	2	2	5	0.1577	1.6677	–11.966	2.379×10^{14}
5	2	3	2	1.0034	205.10	220.13	–
5	3	3	2	0.5740	95.591	–605.95	1.710×10^{13}
5	3	4	2	0.1019	8.0334	–32.508	5.659×10^{13}
5	4	4	2	0.0568	3.7133	–47.731	2.148×10^{13}

tinuum. In these matrix elements, the energy of the first Green function coincides with the electron energy in the final state, which leads to their divergence. It was noted in [16, 27] that such divergences cannot be eliminated by regularizing the integrals and are compensated only in the final expression for $W^{(1-3)}$. Correct calculations according to the above scheme (which would lead to corrections not only to the total probability of photoeffect, but also to the angular distribution of photoelectrons) have not been carried out as yet. However, in our approach, the value of $W^{(1-3)}$ can be determined from Eq. (43) by subtracting the two-photon ionization probability $W^{(2)}$ (which can easily be calculated; see, for example, [28]) from the imaginary part of the DHP. It should be noted that, in contrast to probability $W^{(2)}$, which is essentially positive, correction $W^{(1-3)}$ (associated with reemission of photons) and, accordingly, the entire quantity $\mathrm{Im}E^{(4)}$ can be either positive or negative depending on frequency ω. Numerical calculations show (see Figs. 1, 4, 12, 13, and table) that both possibilities are indeed realized.

The interest in the correction term $W^{(1-3)}$ that determines the deviation from the linear dependence of photoeffect probability $W^{(1)}$ on the intensity upon an increase in F is due to the effect of stabilization of an atom in a strong field, which has been studied intensely in recent years (see reviews [29]). The stabilization effect consists in a slowing down (or even decrease) of the ionization probability with increasing field strength F. For a negative value of $W^{(1-3)}$ (for a given frequency), such slowing down must obviously be observed upon a transition from small to moderate field strengths, which can be regarded as the onset of stabilization. Consequently, the intensity range corresponding to the onset of stabilization can be estimated by calculating higher order corrections in perturbation theory in F to the conventional cross section of photoeffect (naturally, such estimates are valid only for fields $F < F_r$, where F_r determines the convergence radius of perturbation theory for the complex quasi-energy). The effect of stabilization in a linearly polarized laser field was observed experimentally in [14, 15] for the circular state $|n = 5, l = 4, m = 4\rangle$ of the Ne atom (with the maximal possible values of m and l for the given n), which is close to a hydrogen-like state in view of the smallness of quantum defects for states with $l = 4$. It was noted in Section 2.1 that states with $m = l = n - 1$ are isolated; consequently, the ionization probability for such states can be calculated using the standard formula of perturbation theory, disregarding degeneracy. Considering that transitions to states with an orbital angular momentum smaller than l are forbidden for such states in a linearly polarized field, we arrive at the following expressions:

$$W^{(1)} = \frac{\pi F^2}{2\omega^2}\frac{1}{2l+3}\left|A^{(1)}_{l+1}\right|^2, \tag{45}$$

$$W^{(1-3)} = \frac{F^4}{4\omega^4}\frac{\pi}{(2l+3)^2}$$

$$\times\left\{\mathrm{Re}\left[A^{(1)*}_{l+1}A^{(3)}_{l+1,l,l+1}(\omega,0) + A^{(1)*}_{l+1}A^{(3)}_{l+1,l,l+1}(-\omega,0)\right.\right.$$

$$+ A^{(1)*}_{l+1}A^{(3)}_{l+1,l,l+1}(\omega,2\omega) + \frac{4(l+1)}{2l+5}$$

$$\times\left(A^{(1)*}_{l+1}A^{(3)}_{l+1,l+2,l+1}(\omega,0)\right.$$

$$+ A^{(1)*}_{l+1}A^{(3)}_{l+1,l+2,l+1}(-\omega,0)$$

$$\left.\left.\left. + A^{(1)*}_{l+1}A^{(3)}_{l+1,l+2,l+1}(\omega,2\omega)\right)\right] \tag{46}$$

$$\left.\left. - \frac{1}{2\pi}\mathrm{Im}\left[\left(R^{l+1(1)}_{ll}(-\omega) + R^{l+1(1)}_{ll}(\omega)\right)\right.\right.\right.$$

$$\left.\left.\left.\times\left(R^{l+1(2)}_{ll}(-\omega) + R^{l+1(2)}_{ll}(\omega)\right)\right)\right]\right\}.$$

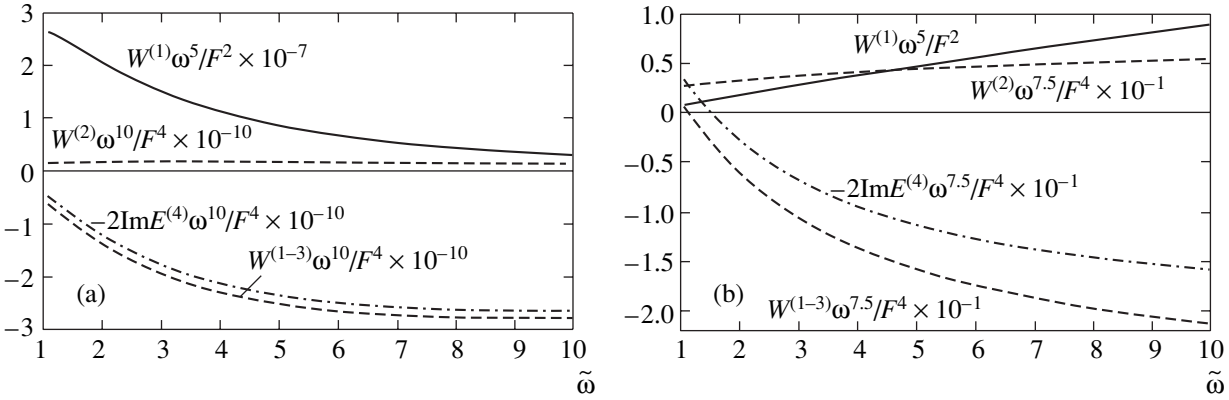


Fig. 13. Probabilities of one-photon ($W^{(1)}$) and two-photon ($W^{(2)}$) ionization, correction $W^{(1-3)}$ to the one-photon ionization probability, and the imaginary part of the fourth-order correction ($-2\mathrm{Im}E^{(4)} = W^{(2)} + W^{(1-3)}$) for state $|n = 5, l = 4, m = 4\rangle$ (a) and for the 1*s* state (b). All quantities are measured in atomic units; $\tilde{\omega} = \omega/|E_n|$.

It should be observed that these formulas are valid not only for states with $m = l = n - 1$, but also for states with $m = l = n - 2$.

Tikhonova *et al.* [30] analyzed experiment [15] in terms of continuum-interference stabilization; in the framework of perturbation theory, this can be interpreted as the inclusion of diagrammatic terms of form (a) from formula (42) in correction $W^{(1-3)}$ to the photoeffect probability. Matrix element $A^{(3)}_{l_1 l_2 l_3}(\omega, 2\omega)$ in this case was estimated with the help of the "pole" approximation [31], in which only the imaginary part corresponding to the term with the δ function on the right-hand side of Eq. (38) is taken into account in both Green functions in amplitude $A^{(3)}_{l_1 l_2 l_3}(\omega, 2\omega)$. In the pole approximation, the amplitude is factorized into single-photon factors, which simplifies the computation of this amplitude considerably (in particular, with such an approach, the above-mentioned divergences do not arise). An additional simplification used in [30] lies in the inclusion of only one matrix element of form (a) in Eq. (42) with $l_1 = l + 1$, $l_2 = l + 2$ from the two matrix elements allowed by the selection rules for linear polarization. This is substantiated by the Bethe rule, according to which it is this matrix element that makes the largest contribution to the total amplitude. As a result, we obtain the following approximate expression for probability W_i:

$$W_i = W^{(1)}\left[1 - \left(\frac{\pi F}{\omega}\right)^2 \frac{2(l+1)}{(2l+5)(2l+3)} \times \left|\langle E_n + 2\omega, l+1 | D(l+1, l+2) | E_n + \omega, l+2\rangle\right|^2\right]. \quad (47)$$

This estimate is attractive due to its simplicity; however, its correctness can be established only from a comparison with exact calculations. The considerable difference between estimate (47) and the exact relation given by the sum of expressions (45) and (46) lies in its fixed sign: in accordance with formula (47), correction $W^{(1-3)}$ is always negative (i.e., it reduces the ionization probability), while general formula (46) does lead to this result.

Formula (44) is valid only for values of F at which the correction term is small; however, this formula can be extrapolated, for obtaining estimates, to the region of values of F for which the first and second terms in Eq. (44) are on the same order of magnitude. Since $W^{(1)} \sim F^2$ and $W^{(1-3)} \sim F^4$, probability W_i (44) as a function of the wave intensity is a parabola with a finite maximum at $W^{(1-3)} < 0$. The value of F^2 for which W_i attains its maximum value is treated in [30] as the stabilization threshold I_{thr}:

$$I_{\mathrm{thr}} = \frac{W^{(1-3)}}{2W^{(1)}}.$$

In contrast to [30], we disregard in our analysis the shape of the laser pulse; for this reason, our value of I_{thr} is smaller by a factor of $\sqrt{2}$ than the value given by formula (8) in [30]. The results of exact numerical calculations are represented in Figs. 12, 13 and the table. Figure 12 shows the dependence of ionization probability W_{nm} on F^2 at frequency $\omega = 4|E_{n=5}| = 2.168$ eV for each from five possible states with $n = 5$, $m = 0$. It can be seen that two from the five $|n = 5, m = 0\rangle$ states exhibit a tendency to stabilization since the total ionization probability for these states increases with F at a lower rate as compared to F^2. Figure 13 shows the frequency dependences of photoionization probability $W^{(1)}$, two-photon ionization probability $W^{(2)}$, and correction $W^{(1-3)}$ associated with the reemission of photons as well as total correction (43) in a linearly polarized wave. It can be seen from the figures that $W^{(2)}$ and $W^{(1-3)}$ have the same order

of magnitude in the near-threshold region, while the contribution from $W^{(1-3)}$ to $\mathrm{Im}E^{(4)}$ becomes predominant upon an increase in ω. Pay attention to different modes of behavior of $W^{(1-3)}$ for the ground 1*s* state and excited circular states with $m = l = n - 1$ and with $m = l = n - 2$ (in particular, $|n = 5, l = 4, m = 4\rangle$). In the latter case, the value of $W^{(1-3)}$ is negative everywhere, indicating a tendency to stabilization at any above-threshold frequency, while $W^{(1-3)} > 0$ for the ground state in the near-threshold frequency range.

The table contains the values of $F^{-2}W^{(1)}$, $F^{-4}W^{(1-3)}$, and I_{thr} for frequencies $\omega = 0.0735 = 2$ eV and $\omega = 0.1838 = 5$ eV for some isolated states. A comparison of the value of $I_{\mathrm{thr}} = 2.1 \times 10^{13}$ W/cm^2 for state $|n = 5, l = 4, m = 4\rangle$ at frequency $\omega = 3.675|E_5| = 2$ eV from the table with the value of $I_{\mathrm{thr}} = 2.8 \times 10^{14}$ W/cm^2 obtained in [30] shows that the results differ by an order of magnitude. A similar situation takes place for state $|n = 3, l = 2, m = 2\rangle$ at a frequency of $\omega = 3.308|E_3| = 5$ eV: $I_{\mathrm{thr}} = 2.4 \times 10^{14}$ W/cm^2 according to our calculations and $I_{\mathrm{thr}} = 1.4 \times 10^{15}$ W/cm^2 according to [30]. In the case of the 1*s* state for which estimate (47) could be applied, this estimate does not even provide a correct qualitative dependence of the single-photon ionization probability W_{1s} on the field strength at frequencies near the ionization threshold (see above).

Thus, in numerical analysis of the amplitudes of multiphoton processes described by higher order matrix elements in perturbation theory, which correspond to virtual multiphoton transitions to the continuum (i.e., containing two or more Green functions with positive energies $\mathscr{E}_i$), various estimates based on the inclusion of a finite number of intermediate states in the discrete spectrum or in a narrow region of states in the continuum should be treated with care. As a matter of fact, the numerical values of these amplitudes depend significantly both on the interference (mutual compensation) of partial radial matrix elements and on the quantum numbers of the initial and final states as well as on energies $\mathscr{E}_i$ of intermediate states. The application of the generalized Sturm expansion of the Coulomb Green function makes it possible to avoid integration over virtual states in the continuum and to reduce the problem to computing well-converging series of the hypergeometrical type. Thus, the technique described above is equivalent to analytic continuation of the Sturm series for compound radial matrix elements to the region of positive energies of the Green functions, in which the standard Sturm series diverge. The results described above show that the method proposed here makes it possible to calculate correctly the higher order radial matrix elements in a wide frequency range up to frequencies two or three orders of magnitude higher than ionization potential $|E_n|$ as well as for highly excited states.

ACKNOWLEDGMENTS

This study was partly financed by the CRDF and the Ministry of Education of the Russian Federation (grant no. VZ-010-0) as well as by the Competition Center of the Ministry of Education of the Russian Federation (grant no. E00-3.2-515).

APPENDIX

We will describe the computational algorithm for the fourth-order matrix elements (23) using the generalized Sturm expansion (33) of the Coulomb Green function. Direct substitution of expression (33) into (23) leads to the sextuple series with six free parameters α_{1-3}, α'_{1-3}:

$$R_{ll'}^{l_1 l_2 l_3}(\mathscr{E}_1, \mathscr{E}_2, \mathscr{E}_3) = \langle nl'|\hat{D}(l, l_3) \times g_{l_3}(\mathscr{E}_3, \alpha_3, \alpha'_3)\hat{D}(l_3, l_2)g_{l_2}(\mathscr{E}_2, \alpha_2, \alpha'_2) \times \hat{D}(l_2, l_1)g_{l_1}(\mathscr{E}_1, \alpha_1, \alpha'_1)\hat{D}(l_1, l)|nl\rangle. \quad (A.1)$$

For the values

$$\alpha_1 = \alpha_2 = \alpha'_2 = \alpha'_3, \qquad \alpha_3 = \alpha'_1 = n/Z, \quad (A.2)$$

of the free parameters, four series from six are terminated due to the emerging orthogonality of the Laguerre polynomials [32]:

$$\int_0^\infty e^{-\rho}\rho^{-\beta}L_n^\beta(\rho)L_m^\beta(\rho)d\rho = \frac{\Gamma(\beta + n + 1)}{n!}\delta_{mn}. \quad (A.3)$$

Indeed, let us write the result of action of operators $\hat{D}$ on radial functions R_{nl} in the special form

$$\hat{D}(l+1, l)R_{nl}(r) = \frac{2Z^{5/2}}{n^3[(d)_{2l+1}]^{1/2}} \times \rho^l e^{-\rho/2}(L_{d-3}^{2l+3}(\rho) - L_{d-1}^{2l+3}(\rho)),$$

$$\hat{D}(l+1, l+2)R_{nl+2}(r) = \frac{2Z^{5/2}}{n^3[(d-3)_{2l+5}]^{1/2}} \quad (A.4)$$

$$\times \rho^l e^{-\rho/2}((s+1)(s+2)L_{d-3}^{2l+3}(\rho) - (d-1)(d-2)L_{d-1}^{2l+3}(\rho)),$$

where

$$d = n - l, \quad s = n + l, \quad \rho = 2Zr/n.$$

Since the Sturm function S_{kl} depends on coordinate r in the same way as $R_{nl}(r)$, we can write $\hat{D}(l', l)S_{kl}(2Zr/n)$ in a form analogous to (A.4). After this, all integrals in expression (A.1) are reduced to (A.3). As a result,

matrix element $R_{ll'}^{l_1 l_2 l_3}$ can be expressed via a linear combination of 16 double series of the form

$$\sum_{k,k'=0}^{\infty} g_{kk'}^{l_2}(\nu_2;\alpha,\alpha)A_{k'}(n,\nu_1,\nu_2,\alpha)B_k(n,\nu_2,\nu_3,\alpha) \equiv \sum_{k,k'=0}^{\infty} a_{kk'}. \tag{A.5}$$

It should be noted that a fourth-order matrix element can also be reduced to a combination of series of the same form if we use the **F**–**r** representation (10) for operator *V*. In this case, the required transformations can be carried out using the recurrent relation for Laguerre polynomials (see [32]):

$$L_n^{\alpha-1}(x) = L_n^{\alpha}(x) - L_{n-1}^{\alpha}(x).$$

However, the corresponding expression for $R_{l_1 l_2 l_3}$ turns out to be much more cumbersome and contains 125 double series of type (A.5) in the general case.

An appropriate choice of the remaining free parameter $\alpha_1 \equiv \alpha$ can ensure the convergence of series $\sum a_{kk'}$ in expression (A.5) for positive energies of the Green function and accelerate the convergence of these series for negative energies. For illustration, we will consider the asymptotic forms of the general term of series (A.5) (these forms can be determined using the results obtained in [22]) at the edges and on the diagonal of matrix $\alpha_{kk'}$:

(1) for a fixed k and $k' \longrightarrow \infty$:

$$a_{kk'} \sim \left(\frac{\alpha-\nu_2}{\alpha+\nu_2}\right)^{k'} \times \left[C_1(k')^n\left(\frac{\alpha-n}{\alpha+n}\right)^{k'} + C_2(k')^{\nu_1}\left(\frac{\alpha-\nu_1}{\alpha+\nu_1}\right)^{k'}\right]; \tag{A.6}$$

(2) for a fixed k' and $k \longrightarrow \infty$:

$$\alpha_{kk'} \sim \left(\frac{\alpha-\nu_2}{\alpha+\nu_2}\right)^{k} \times \left[C_3 k^n\left(\frac{\alpha-n}{\alpha+n}\right)^{k} + C_4 k^{\nu_3}\left(\frac{\alpha-\nu_3}{\alpha+\nu_3}\right)^{k}\right]; \tag{A.7}$$

(3) for $k' = k \longrightarrow \infty$:

$$a_{kk'} \sim \left[C_1 k^n\left(\frac{\alpha-n}{\alpha+n}\right)^{k} + C_2 k^{\nu_1}\left(\frac{\alpha-\nu_1}{\alpha+\nu_1}\right)^{k}\right] \times \left[C_3 k^n\left(\frac{\alpha-n}{\alpha+n}\right)^{k} + C_4 k^{\nu_3}\left(\frac{\alpha-\nu_3}{\alpha+\nu_3}\right)^{k}\right] \times \left[C_5 k^{2\nu_2}\left(\frac{\alpha-\nu_2}{\alpha+\nu_2}\right)^{2k} + C_6\right]. \tag{A.8}$$

Here, C_i are constants independent of k and k'. It follows from relations (A.6)–(A.8) that, by choosing parameter α in complex form with $\mathrm{Re}\alpha > 0$ and $\mathrm{Im}\alpha > 0$, we ensure the convergence of the series in the below-threshold ($\nu_i > 0$) and above-threshold ($\mathrm{Im}\nu_i > 0$) cases since all radixes with powers k, k' are modulo smaller than unity. At the same time, the application of the standard Sturm expansion of $g_{l_2}(\alpha' = \alpha = \nu_2)$ in formula (A.1) leads to a diverging series.

As regards the numerical calculation of the kernel $g_{kk'}^{l}$ of expansion (33) of the Coulomb Green function, it should be observed that the Appell function in the first term of expression (34) contains an integer negative parameter $-k'$ and is equivalent to the linear combination of $k' + 1$ hypergeometric functions ${}_2F_1$ that cannot be reduced to polynomials, while the Appell function in the expression for $\Phi_p^{lk'}$ is a finite polynomial in both arguments. Consequently, $g_{kk'}^{l}$ contains two essentially different groups of terms: complete (nonpolynomial) hypergeometric functions ${}_2F_1$ and the product of hypergeometric polynomials of one (${}_2F_1$) and two variables (F_1). However, the application of such a representation of $g_{kk'}^{l}$ for practical computations is often complicated in view of considerable compensation of the terms in linear combinations for function F_1. In most cases, it is more convenient to calculate the Appell function via its representation in the form of a single integral [33].

REFERENCES

1. L. P. Rapoport, B. A. Zon, and N. L. Manakov, *Theory of Multiphoton Processes in Atoms* (Atomizdat, Moscow, 1978).
2. N. L. Manakov, S. I. Marmo, and A. G. Faĭnshteĭn, Zh. Éksp. Teor. Fiz. **91**, 51 (1986) [Sov. Phys. JETP **64**, 29 (1986)].
3. V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ, *Quantum Electrodynamics*, 3rd ed. (Nauka, Moscow, 1989; Pergamon Press, Oxford, 1982).
4. N. L. Manakov, M. V. Frolov, A. F. Starace, and I. I. Fabrikant, J. Phys. B **33**, R141 (2000).
5. N. L. Manakov, V. D. Ovsiannikov, and L. P. Rapoport, Phys. Rep. **141**, 319 (1986).
6. A. A. Krylovetsky, N. L. Manakov, and S. I. Marmo, Laser Phys. **7**, 817 (1997).
7. A. A. Krylovetskiĭ, N. L. Manakov, and S. I. Marmo, Zh. Éksp. Teor. Fiz. **119**, 45 (2001) [JETP **92**, 37 (2001)].
8. N. L. Manakov, M. A. Preobrazhenskiĭ, and L. P. Rapoport, Opt. Spektrosk. **35**, 24 (1973).
9. M. Pont, R. M. Potvliege, R. Shakeshaft, and Z.-J. Tang, Phys. Rev. A **45**, 8235 (1992).
10. L. Pan, K. T. Teilor, and C. W. Clark, Phys. Rev. Lett. **61**, 2673 (1988).
11. R. Shakeshaft, R. M. Potvliege, M. Dörr, and W. E. Cooke, Phys. Rev. A **42**, 1656 (1990).

12. N. L. Manakov, A. Maquet, S. I. Marmo, and C. Szymanowski, Phys. Lett. A **237**, 234 (1998).
13. P. R. Jones, D. Schumacher, and P. Bucksbaum, Phys. Rev. A **47**, R49 (1993); M. P. de Boer, J. H. Hoogenraad, R. B. Vrijen, *et al.*, Phys. Rev. Lett. **71**, 3263 (1993).
14. M. P. de Boer, J. H. Hoogenraad, R. B. Vrijen, *et al.*, Phys. Rev. A **50**, 4085 (1994).
15. N. J. van Drutten, R. C. Constantinescu, J. M. Schins, *et al.*, Phys. Rev. A **55**, 622 (1997).
16. N. L. Manakov, M. A. Preobrazhenskiĭ, L. P. Rapoport, and A. G. Faĭnshteĭn, Zh. Éksp. Teor. Fiz. **75**, 1243 (1978) [Sov. Phys. JETP **48**, 626 (1978)].
17. N. Moiseyev, Phys. Rep. **302**, 211 (1998).
18. V. D. Ovsiannikov and S. V. Goosev, Phys. Scr. **57**, 506 (1998).
19. D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskiĭ, *Quantum Theory of Angular Momentum* (Nauka, Leningrad, 1975; World Sci., Singapore, 1988).
20. V. A. Davydkin and V. D. Ovsiannikov, J. Phys. B **19**, 2071 (1986).
21. L. H. Hostler, J. Math. Phys. **11**, 2966 (1970).
22. N. L. Manakov, S. I. Marmo, and A. G. Faĭnshteĭn, Teor. Mat. Fiz. **59**, 49 (1984).
23. E. J. Heller, Phys. Rev. A **12**, 1222 (1975).
24. D. P. Shelton, Phys. Rev. A **36**, 3032 (1987).
25. A. I. Baz', Ya. B. Zel'dovich, and A. M. Perelomov, *Scattering, Reactions and Decays in Nonrelativistic Quantum Mechanics*, 2nd ed. (Nauka, Moscow, 1971; Israel Program for Scientific Translations, Jerusalem, 1966).
26. N. L. Manakov, V. A. Sviridov, and A. G. Faĭnshteĭn, Zh. Éksp. Teor. Fiz. **95**, 790 (1989) [Sov. Phys. JETP **68**, 451 (1989)].
27. M. Edvards and R. Shakeshaft, Z. Phys. D **8**, 51 (1988).
28. A. A. Krylovetskiĭ, N. L. Manakov, S. I. Marmo, and A. F. Staras, Zh. Éksp. Teor. Fiz. **122**, 1168 (2002) [JETP **95**, 1006 (2002)].
29. M. Gavrila, J. Phys. B **35**, R147 (2002); A. M. Popov, O. V. Tikhonova, and E. A. Volkova, J. Phys. B **36**, R125 (2003).
30. O. V. Tikhonova, A. M. Popov, and M. V. Fedorov, Phys. Rev. A **65**, 053404 (2002).
31. Z. Deng and J. H. Eberly, Phys. Rev. Lett. **53**, 1810 (1984); J. Opt. Soc. Am. B **2**, 486 (1985).
32. *Higher Transcendental Functions (Bateman Manuscript Project)*, Ed. by A. Erdelyi (McGraw-Hill, New York, 1953; Nauka, Moscow, 1974), Vol. 2.
33. *Higher Transcendental Functions (Bateman Manuscript Project)*, Ed. by A. Erdelyi (McGraw-Hill, New York, 1953; Nauka, Moscow, 1973), Vol. 1.

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 98, No. 2, 2004, pp. 272–286.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 2, 2004, pp. 307–323.

ATOMS, SPECTRA, RADIATION

Parametric Frequency Conversion in Layered Nonlinear Media

G. Kh. Kitaeva and A. N. Penin
Moscow State University, Vorob'evy gory, Moscow, 119992 Russia
e-mail: kit@qopt.phys.msu.su
Received May 22, 2003

Abstract—Frequency-angular distributions of signal wave intensity are calculated for spontaneous parametric down-conversion and parametric frequency conversion in spatially nonuniform nonlinear media. Wave reflection from interfaces is taken into account, and both regular and irregular nonuniform distributions of second-order nonlinear susceptibility are considered. A unified approach using a scattering matrix and a generalized Kirchhoff law is applied in calculations of spontaneous and stimulated processes in dissipative nonlinear media. Interference of electromagnetic zero-point fluctuations of the vacuum, nonlinear interference, and nonlinear diffraction are examined for media with various absorptive properties. Theoretical foundations are developed for diagnostics of nonuniform distributions of the second-order susceptibility, based on measurement of the line profiles of nonlinear signals. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Media with spatially varying linear and nonlinear optical properties are currently attracting much attention [1, 2]. They are widely used in nonlinear optics and laser physics [3–9]. Characteristics of parametric processes in crystals with deep modulation of the second-order nonlinear susceptibility are the subject of ongoing research [3, 4, 6–8]. In our view, progress in this area is important for laser spectroscopy of solids and polymers, because parametric processes offer novel opportunities for analyzing the structure of inhomogeneous multidomain crystals and polymer materials and their transformations [9].

Spatial nonuniformity (periodic or random, depending on a particular specimen) strongly affects the nonlinear optical processes in which phase matching is a necessary condition. Energy conservation is a strict requirement. Under steady-state conditions, it reduces to the zero algebraic sum of the frequencies of the interacting waves. By contrast, momentum conservation may hold up to a certain mismatch when part of the momentum carried by light waves is transferred to the medium. In the general case, parametric processes in inhomogeneous media depend on changes in linear and nonlinear optical parameters across interfaces, geometry of individual regions, characteristics of boundaries, scattering and absorption coefficients, the coefficients characterizing nonlinear optical conversion, etc. Analysis of the integral effect of the combination of all factors on a two- or three-dimensional scattering spectrum is a difficult task. The problem is substantially simplified by invoking the generalized Kirchhoff law (a nonlinear analog of the Kirchhoff law) formulated by Klyshko [10, 11] in the framework of a unified phenomenological approach to spontaneous and stimulated parametric interactions in weakly nonlinear dissipative media. The generalized Kirchhoff law can be used to determine second-order correlation functions for the output field by using the corresponding functions prescribed at the input end and to calculate both frequency-angular distribution of the output radiation intensity and its statistical characteristics. In calculations of this kind, a linear relationship between the Heisenberg operators of the input and output fields is postulated. (The only exception is the pump field, which is treated as a classical wave of prescribed intensity.) This relationship is formulated in terms of a scattering matrix. The matrix can be calculated within the framework of classical nonlinear optics, and its elements determine the relations between the classical field strengths for all input and output modes.

In this paper, we present the results obtained by applying the generalized Kirchhoff law to parametric processes in media with nonuniform distributions of optical parameters. We consider spontaneous parametric down-conversion (SPDC) and stimulated parametric frequency conversion (PC) in media without an inversion center [10]. SPDC is the scattering or decay of pump photons of frequency ω_0 caused by quantum field fluctuations in a medium with nonzero second-order nonlinear susceptibility $\chi^{(2)}$.[1] As a result, photon pairs correlated with respect to the time and location of their origin are created in the medium. The frequencies of these photons, ω_1 and ω_2, are related as follows:

$$\hbar\omega_0 = \hbar\omega_1 + \hbar\omega_2. \quad (1)$$

[1] SPDC [12–14] was predicted by Klyshko in a paper presented in 1966 at the All-Union Conference on Nonlinear Media in Chernogolovka.

Their propagation directions are determined by the phase matching condition

$$\mathbf{k}_0 = \mathbf{k}_1 + \mathbf{k}_2 - \Delta\mathbf{k}, \quad (2)$$

where $\mathbf{k}_j$ is the jth wave vector, $|\mathbf{k}_j| = n_j\omega_j/c$, and n_j is the refractive index at ω_j (j = 0, 1, 2). The frequencies ω_1 and ω_2 may vary from zero to ω_0. By convention, the wave with the frequency between $\omega_0/2$ and ω_0 is the called signal wave (ω_1) and the other one, with the frequency below $\omega_0/2$, is called the idler wave (ω_2). Normally, the frequency of the signal wave generated in an experiment lies in a range well suited for measurement. The phase matching condition for the wave vectors $\mathbf{k}_j$ is satisfied up to a mismatch $\Delta\mathbf{k}$ that depends on the absorption, geometry of the illuminated specimen, and dispersion of the refractive index.

The emergence of photons at ω_1 and ω_2 via SPDC (without any input at these frequencies) can be explained only in the framework of a consistent quantum-theoretical treatment. Outside the absorption band, scattered radiation leaves the medium, consisting of correlated photon pairs (biphotons) [15, 16]. SPDC can be interpreted as the result of scattering of the pump wave by electromagnetic fluctuations of the vacuum in a nonlinear medium. If one of the output frequencies lies in an absorption band, e.g., a photon resonance absorption band, then a photon–polariton pair is created. In this case, the pump is scattered both by vacuum fluctuations and by thermal fluctuations of the scattering medium [17, 18]. When the input contains not an only pump wave, but also a signal- or idler-frequency wave, then spontaneous light scattering is accompanied by parametric conversion of pump photons into biphotons. In contrast to SPDC, stimulated parametric frequency conversion can be described by both quantum and classical models [19].

Here, we use the generalized Kirchhoff law in analyzing some specific effects of spatial inhomogeneity on SPDC and PC spectra. In Section 2, we apply the generalized Kirchhoff law to calculate the signal line profile in the simplest case of a slab of a nonlinear medium. This model was examined previously in analyses based on different approaches [10]. The present study includes calculation of the scattering matrix and provides a methodological basis for further analysis. In Section 3, we calculate the frequency-angular distributions for SPDC and PC signal waves in a slab with reflecting boundaries [20]. We analyze the fine interference structure of line profiles for various cases of signal and idler absorption. Next, we consider the effects of electromagnetic zero-point fluctuations of the vacuum (Section 4) and spatial distribution of the effective second-order susceptibility (Section 5) on the SPDC spectrum.

The present study is focused on the characteristics of three-wave SPDC and parametric difference-frequency generation in a spatially nonuniform medium. However, the results obtained for difference-frequency generation can be modified to describe optical harmonic generation and sum-frequency generation, as well as cascade processes of higher orders, in the constant-pump approximation by assuming linear amplification of the remaining waves.

2. SCATTERING MATRIX AND SIGNAL LINE PROFILE IN A SLAB

For three-wave parametric interactions, the average number of signal and idler output photons, $N_{1'} \equiv \langle a_{1'}^\dagger a_{1'}\rangle$ and $N_{2'} \equiv \langle a_{2'}^\dagger a_{2'}\rangle$, dictated by the generalized Kirchhoff law is

$$N_{1'} = \hat{U}_{1'1}(N_1 + N_0 + I)\hat{U}_{1'1}^\dagger + \hat{U}_{1'2}(\tilde{N}_2 - N_0)\hat{U}_{1'2}^\dagger - N_0 - I, \quad (3)$$

$$N_{2'} = \hat{U}_{2'1}(\tilde{N}_1 + N_0 + I)\hat{U}_{2'1}^\dagger + \hat{U}_{2'2}(N_2 - N_0)\hat{U}_{2'2}^\dagger + N_0, \quad (4)$$

where I is the identity matrix and $\hat{U}_{i'j}$ is an entry of the scattering matrix

$$\hat{U} = \begin{pmatrix} \hat{U}_{1'1} & \hat{U}_{1'2} \\ \hat{U}_{2'1} & \hat{U}_{2'2} \end{pmatrix}.$$

The subscripts 1' and 2' run through all signal and idler output modes; the subscripts 1 and 2, through all corresponding input modes;

$$(N_1)_{ij} \equiv \langle a_{1i}^\dagger a_{1j}\rangle, \quad (N_2)_{ij} \equiv \langle a_{2i}^\dagger a_{2j}\rangle$$

are the matrices representing the second-order moments of the input field; $\tilde{N}_1$ and $\tilde{N}_2$ are transposed matrices; $a_k^\dagger$ and a_k are the creation and annihilation operators for kth-mode photons;

$$N_0 \equiv 1/[\exp(\hbar\omega_p/kT) - 1]$$

is the temperature factor, where ω_p is the phonon resonance frequency. Normally, this factor can be neglected since it is much less than unity at room temperature. The difference between expressions (3) and (4) is due to the assumption that only idler waves are absorbed. This assumption is more likely to hold for real processes, as compared to substantial signal or pump absorption.

The scattering matrix $\hat{U}$ defines the relationship between the creation and annihilation operators for the

input and output fields and their first-order moments, $(\langle a_1^\dagger\rangle, \langle a_2\rangle)$ and $(\langle a_{1'}^\dagger\rangle, \langle a_{2'}\rangle)$:

$$\begin{aligned}\langle a_{1'}^\dagger\rangle &= \hat{U}_{1'1}\langle a_1^\dagger\rangle + \hat{U}_{1'2}\langle a_2\rangle,\\ \langle a_{2'}\rangle &= \hat{U}_{2'1}\langle a_1^\dagger\rangle + \hat{U}_{2'2}\langle a_2\rangle.\end{aligned} \tag{5}$$

Each mode satisfies the following relations outside the nonlinear medium up to general normalization factors:

$$\langle a_j^\dagger\rangle \sim \langle a_j\rangle \sim A_j \equiv E_j\sqrt{\frac{\cos\theta_j}{\omega_j}},$$

where E_j is the electric field amplitude and θ_j is the angle of wave incidence (the angle between the normal vector to the surface and the wave vector). In the absence of scattering, the scattering matrix is unitary and

$$\hat{U}\hat{\sigma}\hat{U}^\dagger = \hat{\sigma},$$

where $\sigma_{mn} = (-1)^{m+1}\delta_{mn}$ (δ_{mn} is Kronecker's delta). Each element of the matrix can be determined experimentally. However, it is frequently sufficient to calculate these elements theoretically by solving equations for slowly varying field amplitudes. Substituting the resulting elements of $\hat{U}$ into (3) and finding the average number of photons for each output mode, one can calculate the output intensity $P_{\omega\Omega}$ as a function of frequency and scattering (observation) angle, i.e., the line profile or form factor. (Intensity is defined here as the energy emitted by a unit surface area into unit solid angle per unit spectral interval per unit time.) The output signal intensity is

$$P_{\omega_1\Omega_1} = \frac{\hbar\omega_1^3\cos\theta_1}{c^2 v}N_{1'}, \tag{6}$$

where v is the volume per mode in the wave-vector space. Accordingly, the signal received with a quantum detector efficiency η_1 in the far-field region is

$$P_1 = \int\limits_{\Delta\omega_{\text{det}}} d\omega_1 \int\limits_{\Delta\Omega_{\text{det}}} d\Omega_1 \xi_1(\omega_1, \Omega_1) P_{\omega_1\Omega_1}, \tag{7}$$

where $\xi_1(\omega_1, \Omega_1) \equiv \eta_1/\hbar\omega_1 c\cos\theta_1$. The integral is calculated over the bandwidth and aperture of the detector, $(\Delta\omega_{\text{det}})$ and $(\Delta\Omega_{\text{det}})$.

The correlation moment $K' \equiv \langle a_{1'}a_{2'}\rangle^*$, which determines the statistical properties of the output biphoton field, is expressed as

$$\tilde{K} = \hat{U}_{2'1}(\tilde{N}_1 + N_0 + I)\tilde{\hat{U}}_{1'1} + \tilde{\hat{U}}_{2'2}(N_2 - N_0)\tilde{\hat{U}}_{1'2}. \tag{8}$$

The moments $N_{1'}$ and $N_{2'}$ and the second-order correlator K' determine the fourth-order moment of the scattered field,

$$\langle a_{1_i'}^\dagger a_{1_j'}^\dagger a_{2_m'}^\dagger a_{2_n'}\rangle = N_{1_i'1_j'}N_{2_m'2_n'} + K_{1_i'2_m'}K_{1_j'2_n'}^*,$$

which characterizes the correlation between signal and idler photons. It can be found by measuring the cross correlation of the signal and idler detector outputs [15, 16].

Consider a strong pump wave $E_0\exp(i\mathbf{k}_0\cdot\mathbf{r} - i\omega_0 t)$ and much weaker signal and idler waves $E_{10}\exp(i\mathbf{k}_1\cdot\mathbf{r} - i\omega_1 t)$ and $E_{20}\exp(i\mathbf{k}_2\cdot\mathbf{r} - i\omega_2 t)$ incident on a slab of thickness l. In the constant-pump approximation, the reduced wave equations describing the process are written as follows [21]:

$$\begin{aligned}\frac{dE_1(z)}{dz} + i\sigma_1 E_0 E_2^*(z)\exp\left(-\frac{iz\Delta}{l}\right) + \frac{y_1}{l}E_1(z) = 0,\\ \frac{dE_2(z)}{dz} + i\sigma_2 E_0 E_1^*(z)\exp\left(-\frac{iz\Delta}{l}\right) + \frac{y_2}{l}E_2(z) = 0.\end{aligned} \tag{9}$$

Here,

$$\Delta \equiv l(|k_{1z}| + |k_{2z}| - |k_{0z}|) \equiv \delta_1 + \delta_2 - \delta_0$$

is the dimensionless wave-vector mismatch, $\sigma_j \equiv 2\pi\chi\omega_j/cn_j\cos\vartheta_j$, and $y_j \equiv \alpha_j l/2\cos\vartheta_j$. The convolution of second-order nonlinear susceptibility tensor $\hat{\chi}^{(2)}$ with signal, idler, and pump unit polarization vectors is denoted by χ; α_j denotes the absorption coefficient at the frequency ω_j; and ϑ_j is the angle between the normal vector to the surface and a particular wave vector inside the slab. In the case of weak absorption, α_j is related to the imaginary part of permittivity ε_j'' as follows: $\alpha_j = \omega_j\varepsilon_j''/cn_j$.

Interaction between $E_1(z)$, $E_2(z)$, and $E_0\exp(i\mathbf{k}_0\cdot\mathbf{r} - i\omega_0 t)$ can be described by the parametric interaction matrix $\hat{w}$ relating plane waves satisfying conditions (1) and (2):

$$\begin{pmatrix} A_1(l/2)\\ A_2(l/2)\end{pmatrix} = \begin{pmatrix} w_{11} & w_{12}\\ w_{21} & w_{12}\end{pmatrix}\begin{pmatrix} A_1(-l/2)\\ A_2(-l/2)\end{pmatrix}, \tag{10}$$

where

$$A_{1,2}\left(\pm\frac{l}{2}\right) \equiv E_{1,2}\left(\pm\frac{l}{2}\right)\sqrt{\frac{\cos\theta_{1,2}}{\omega_{1,2}}},$$

with $E_{1,2}(-l/2)$ and $E_{1,2}(l/2)$ denoting input and output wave amplitudes. Since the matrix $\hat{w}$ is identical to $\hat{U}$ in the case under analysis, its elements are expressed as follows by virtue of (9):

$$\begin{aligned} w_{11} &= e^{-\mu}\left(\cosh\gamma + \frac{\eta\sinh\gamma}{\gamma}\right),\\ w_{22} &= e^{-\mu^*}\left(\cosh\gamma - \frac{\mu\sinh\gamma}{\gamma}\right),\\ w_{12} &= -i\beta e^{-y/2}\frac{\sinh\gamma}{\gamma}, \quad w_{21} = i\beta^* e^{-y/2}\frac{\sinh\gamma}{\gamma},\end{aligned} \tag{11}$$

where

$$y \equiv y_1 + y_2, \quad \mu \equiv (y_1 + y_2 + i\Delta)/2,$$
$$\eta \equiv (-y_1 + y_2 + i\Delta)/2,$$
$$\gamma \equiv \sqrt{|\beta|^2 + |\eta|^2}, \tag{12}$$
$$\beta \equiv \frac{2\pi\omega_1\omega_2\chi l E_0}{c^2\sqrt{k_{1z}k_{2z}}}\sqrt{\frac{\cos\theta_0}{n_0\cos\vartheta_0}}.$$

In cases of experimental interest, only the lowest order terms can be retained in the power series expansions of $\hat{w}$. Then,

$$w_{11} \approx e^{-y_1} + |\beta|^2 f_1(\eta),$$
$$w_{22} \approx e^{-y_2} + |\beta|^2 f_2(\eta), \tag{13}$$
$$w_{12} \approx i\beta f(\eta), \quad w_{21} \approx -i\beta^* f(\eta),$$

where

$$f_1(\eta) \equiv \frac{1}{(2\eta)^2}[e^{-y_2 - i\Delta} - (1-2\eta)e^{-y_1}],$$
$$f_2(\eta) \equiv \frac{1}{(2\eta)^2}[e^{-y_1 + i\Delta} - (1+2\eta)e^{-y_2}], \tag{14}$$
$$f(\eta) \equiv \frac{1}{2\eta}[e^{-y_2 - i\Delta/2} - e^{-y_1 + i\Delta/2}].$$

When only idler waves are absorbed, the generalized Kirchhoff law entails the following energy relation between SPDC signal characteristics:

$$N_{1'}^{\mathrm{SPDC}} = |w_{11}|^2 - 1.$$

By virtue of (6) and (13), this leads to the following expression for SPDC intensity in the linear approximation with respect to pump:

$$P^{\mathrm{SPDC}}_{\omega_1\Omega_1}\Big|_{\substack{y_1 = 0 \\ y_2 \neq 0}} = C_0 g(\Delta, y_2), \tag{15}$$

where

$$g(\Delta, y_2) \equiv \frac{2}{(\Delta^2 + y_2^2)^2}[(\Delta^2 - y_2^2)(1 - e^{-y_2}\cos\Delta) \tag{15'}$$
$$- 2y_2\Delta e^{-y_2}\sin\Delta + y_2(\Delta^2 + y_2^2)]$$

is the form factor that determines signal intensity as a function of phase mismatch. Coefficient C_0 depends on pump intensity P_0, specular transmissivity t_0^2 of the input surface, the layer thickness, and the effective second-order susceptibility χ:

$$C_0 \equiv \frac{\hbar\omega_1^4\omega_2}{c^5 n_0 n_1 n_2} P_0 t_0^2 \chi^2 l^2 \frac{\cos\theta_0\cos\theta_1}{\cos\vartheta_0\cos\vartheta_1\cos\vartheta_2}.$$

For parametric conversion of an idler input of radiance N_2, we have $N_{1'} = N_{1'}^{\mathrm{SPDC}} + N_{1'}^{\mathrm{PC}}$, where N^{SPDC} is spontaneous "noise" and $N_{1'}^{\mathrm{PC}} = |w_{12}|^2 N_2$ is the PC signal radiance measured in units of number of photons per mode. In the linear (spontaneous–stimulated) regime of conversion, the corresponding intensity is

$$P^{\mathrm{PC}}_{\omega_1\Omega_1}\Big|_{\substack{y_1 \neq 0 \\ y_2 \neq 0}} = C_0 F(\Delta, y_1, y_2) N_2, \tag{16}$$

where

$$F(\Delta, y_1, y_2) \equiv \frac{\exp(-2y_1) - 2\exp[-(y_1 + y_2)]\cos\Delta + \exp(-2y_2)}{(y_1 - y_2)^2 + \Delta^2}$$

is the form factor describing the observed spectrum corrected for absorption at both signal and idler frequencies. In particular, if only idler waves are absorbed, then the line profile due to parametric conversion characterized by an isotropic input intensity distribution (when all modes in the idler channel are uniformly populated) is

$$F(\Delta, 0, y_2) \equiv \frac{1 - 2\exp(-y_2)\cos\Delta + \exp(-2y_2)}{y_2^2 + \Delta^2}. \tag{16'}$$

Expressions (15') and (16') for the SPDC and PC form factors imply that the line profiles of isotropic SPDC and PC waves are different in the case of nonzero, but weak, absorption at the idler frequency ($y_2 \sim 1$). It should be noted here that absorption at a signal frequency does not entail any distinction of this kind. However, the SPDC and PC line profiles are similar in the case of strong absorption at the idler frequency ($y_2 \gg 1$). Then, both waves have Lorentzian profiles as functions of the wave-vector mismatch:

$$P^{\mathrm{SPDC}}_{\omega_1\Omega_1}\Big|_{\substack{y_1 = 0 \\ y_2 \gg 1}} = C_0\left(\frac{2y_2}{y_2^2 + \Delta^2}\right),$$
$$P^{\mathrm{PC}}_{\omega_1\Omega_1}\Big|_{\substack{y_1 = 0 \\ y_2 \gg 1}} = C_0 N_2\left(\frac{1}{y_2^2 + \Delta^2}\right). \tag{17}$$

When the medium is transparent at all frequencies, we obtain well-known expressions for spontaneous [10] and

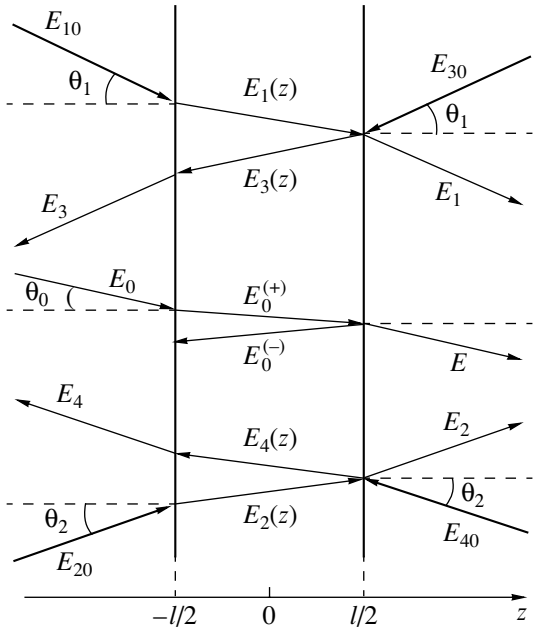


Fig. 1. Diagrams of waves propagating outside and inside a reflecting slab.

stimulated [19] waves as functions of the mismatch:

$$P^{\mathrm{SPDC}}_{\omega_1\Omega_1}\Big|_{\substack{y_1=0\\ y_2=0}} = C_0\mathrm{sinc}^2\frac{\Delta}{2},$$
$$P^{\mathrm{PC}}_{\omega_1\Omega_1}\Big|_{\substack{y_1=0\\ y_2=0}} = C_0N_2\mathrm{sinc}^2\frac{\Delta}{2}, \qquad (18)$$

where $\mathrm{sinc}x = \sin x/x$.

When the slab is transparent, the angular frequency bandwidths of these waves depend only on its thickness. In the case of substantial absorption (condition $y_j \ll 1$ is violated for $j = 1$ and/or $j = 2$), the widths are determined by both thickness and absorption coefficients α_j for the slab material. As an example, consider the line shape characteristics for signal with a wavelength of 625 nm when parametric interaction of *ooe* type takes place inside a slab of KDP crystal parallel to the z axis. The pump has extraordinary polarization and a wavelength of 488 nm and propagates along the normal to the slab. KDP crystal is virtually transparent to pump and signal in the visible range. The corresponding idler wavelength is 2.23 μm. The idler absorption coefficient is $\alpha_2 \approx 0.5$ cm^{-1}. Expressions (18) apply to slabs of thickness between 0.1 and 2 cm (the line profile scales with $\mathrm{sinc}^2(\Delta/2)$). The spectral width for the principal maximum in the signal profile emitted by a 1-cm-thick slab is 0.5 nm, and the corresponding angular width is 0.03°. When the slab thickness is ten times smaller (1 mm), these widths are 5.4 nm and 0.35°, respectively. With decreasing signal wavelength, the idler wavelength increases and the absorption coefficient α_2 grows even faster. For example, if a signal is detected at 572 nm, then the idler wavelength is 3.32 μm. When $\alpha_2 \approx 100$ cm^{-1} expressions (17) apply to slabs 0.5 cm or thicker. The FWHM of signal emitted by a 1 cm thick slab is 2 nm, and the corresponding angular width is 0.2°.

3. SCATTERING MATRIX AND SIGNAL LINE PROFILE FOR A SLAB WITH REFLECTING BOUNDARIES

The analysis developed above does not take into account reflection from the boundaries of nonlinear medium. It is clear that reflection gives rise to effects due to interference, which affects the scattered-wave line profile.

The pump, signal, and idler modes inside and outside a reflecting slab can be divided into forward and backward ones (propagating in the positive and negative directions along the z axis) (see Fig. 1). The forward and backward wave vectors associated with the same frequency differ only by the signs of their z-components. Wave vectors having equal magnitudes and opposite signs arise when signal and idler waves are incident upon the slab from both directions. Denote the forward ($k_{1z} > 0$) and backward ($k_{1z} < 0$) signal waves by subscripts 1 and 3, respectively, and the forward and backward idler waves by subscripts 2 and 4, respectively. The amplitude of each wave varies as the wave propagates across the slab because of parametric interaction. Additional variation of the amplitude is caused by absorption inside the slab and reflection from its boundaries. Linear scattering is neglected here. We denote the amplitudes of the forward and backward signal and idler waves leaving the slab by E_j ($j = 1, 2, 3, 4$), those of the incident waves by E_{j0}, and those propagating across the slab by $E_j(z)$.

The boundary conditions relating the amplitudes outside and inside the slab (at $z = \pm l/2$) can be represented as follows [20]:

$$\mathbf{A}' + \hat{\rho}\mathbf{A}'' = \hat{\tau}\mathbf{A}_0, \quad \mathbf{A} = \hat{\tau}\mathbf{A}'' + \hat{\rho}^*\mathbf{A}_0, \qquad (19)$$

where

$$\mathbf{A}_0 \equiv \{A_{10}, A_{20}^*, A_{30}, A_{40}^*\}, \quad \mathbf{A} \equiv \{A_1, A_2^*, A_3, A_4^*\},$$
$$\mathbf{A}' \equiv \{A_1(-l/2), A_2^*(-l/2), A_3(l/2), A_4^*(l/2)\},$$
$$\mathbf{A}'' \equiv \{A_1(l/2), A_2^*(l/2), A_3(-l/2), A_4^*(-l/2)\}.$$

The elements of the diagonal matrix $\hat{\tau}$ are

$$\tau_1 = \tau_3 = t_1\sqrt{\frac{n_1\cos\vartheta_1}{\cos\theta_1}}, \quad \tau_2 = \tau_4 = t_2\sqrt{\frac{n_2\cos\vartheta_2}{\cos\theta_2}},$$

where t_1 and t_2 are the amplitude transmissivities characterizing the slab boundaries. The nonzero compo-

nents of the matrix $\hat{\rho}$ are

$$\rho_{13} = \rho_{31} \equiv \rho_1 \equiv r_1 e^{i\delta_1}, \quad \rho_{24} = \rho_{42} \equiv \rho_2 \equiv r_2 e^{-i\delta_2},$$

$$\delta_j \equiv |k_{jz}| l = \frac{l n_j \omega_j}{c} \cos\vartheta_j,$$

where r_j denotes the amplitude reflectivities of the slab boundaries. Both r_j and t_j are calculated by using the Fresnel formulas [22]. If the optical density of the medium outside the slab is lower, then $r_j < 0$. When anisotropy is allowed for, the coefficients r_j and t_j depend both on the propagation direction of the incident waves and on their respective angles of incidence. Moreover, we may consider the effects due to birefringence, in which case each incident plane wave is associated with two forward and backward waves propagating across the slab. To simplify further analysis, we assume that the parametric interaction involves only signal and idler waves of definite polarization. We also assume that the orientation of the optical axes of the anisotropic layer is such that birefringence effects vanish and the polarizations of linearly polarized incident waves remain invariant after reflection and refraction at the slab boundary.

Reduced equations (9) relate both the forward waves $E_1(z)$ and $E_2(z)$ and the backward waves $E_3(z)$ and $E_4(z)$ pairwise. The only difference between the pairs is that the forward pump amplitude $E_0^{(+)}$ must be substituted into (9) to calculate $E_1(z)$ and $E_2(z)$, whereas the backward pump amplitude $E_0^{(-)}$ must be substituted to calculate $E_3(z)$ and $E_4(z)$. The incident pump amplitude is related to the amplitudes $E_0^{(+)}$ and $E_0^{(-)}$ inside the slab by the Airy formulas

$$E_0^{(+)} = \frac{t_0}{1 - r_0^2 \exp(2i\delta_0)} E_0,$$

$$E_0^{(-)} = \frac{t_0 r_0 \exp(i\delta_0)}{1 - r_0^2 \exp(2i\delta_0)} E_0 = r_0 \exp(i\delta_0) E_0^{(+)}.$$

Interaction between $E_1(z)$, $E_2(z)$, $E_3(z)$, and $E_4(z)$ in the presence of pumps $E_0^{(+)}$ and $E_0^{(-)}$ can be represented in terms of a general parametric interaction matrix as $\mathbf{A}'' = \hat{w}\mathbf{A}'$, where

$$\hat{w} = \begin{pmatrix} w_{11} & w_{12} & 0 & 0 \\ w_{21} & w_{22} & 0 & 0 \\ 0 & 0 & w_{11}^{(-)} & w_{12}^{(-)} \\ 0 & 0 & w_{21}^{(-)} & w_{22}^{(-)} \end{pmatrix}. \tag{20}$$

The elements w_{ij} ($i = 1, 2$) are identical to those of the parametric interaction matrix defined by (11) in the absence of reflection (after the change $E_0 \longrightarrow E_0^{(+)}$). The components $w_{ij}^{(-)}$ characterize interaction between the backward waves and are similar to w_{ij}, with the exception that E_0 is replaced by $E_0^{(-)}$ in expression (12) for β. The relation $\hat{w}^{(-1)} = \hat{w}(-l)$ means that the inverse matrix characterizes conversion of waves propagating in the negative direction along the z axis.

Combining (19) with (20), we find the scattering matrix $\hat{U}$ for a slab:

$$\mathbf{A} = \hat{U}\mathbf{A}_0, \quad \hat{U} = \hat{t}\hat{w}(\hat{I} + \hat{\rho}\hat{w})^{-1}\hat{t} + \hat{\rho}^*. \tag{21}$$

Expressions for the elements of $\hat{U}$ were presented in [20].

If neither signal nor idler wave is absorbed, then we have the condition

$$|U_{11}|^2 + |U_{13}|^2 + (-|U_{12}|^2 - |U_{14}|^2) = 1. \tag{22}$$

By the generalized Kirchhoff law (represented by (3)), the SPDC signal radiance in a slab that partially absorbs idler waves and partially reflects all waves is expressed as

$$N_{1'}^{\mathrm{SPDC}} = |U_{11}|^2 + |U_{13}|^2 - 1. \tag{23}$$

For PC of an idler input of radiance N_2, we have

$$N_{1'} = N_{1'}^{\mathrm{SPDC}} + N_{1'}^{\mathrm{PC}}, \quad N_{1'}^{\mathrm{PC}} = |U_{12}|^2 N_2. \tag{24}$$

Instead of (15), we obtain the following expression for the SPDC signal intensity:

$$P_{\omega_1\Omega_1}^{\mathrm{SPDC}}\Big|_{\substack{y_1 = 0 \\ y_2 \neq 0}} = C_0 \frac{1 - R_1}{D_0 D_1 \tilde{D}_2} \{\tilde{D}_2(1 + R_0R_1) g(\Delta, y_2)$$

$$+ [\tilde{R}_2(1 + R_0R_1)(\cos 2\delta_2 - \tilde{R}_2) - 2r_0r_1\tilde{r}_2(1 - \tilde{R}_2)$$

$$\times \cos(\delta_0 + \delta_1)\cos\delta_2] g'(\Delta, y_2) \tag{25}$$

$$- [\tilde{R}_2(1 + R_0R_1)\sin 2\delta_2 - 2r_0r_1\tilde{r}_2(1 + \tilde{R}_2)$$

$$\times \cos(\delta_0 + \delta_1)\sin\delta_2] g''(\Delta, y_2) \}.$$

Here, the supplementary functions determining the

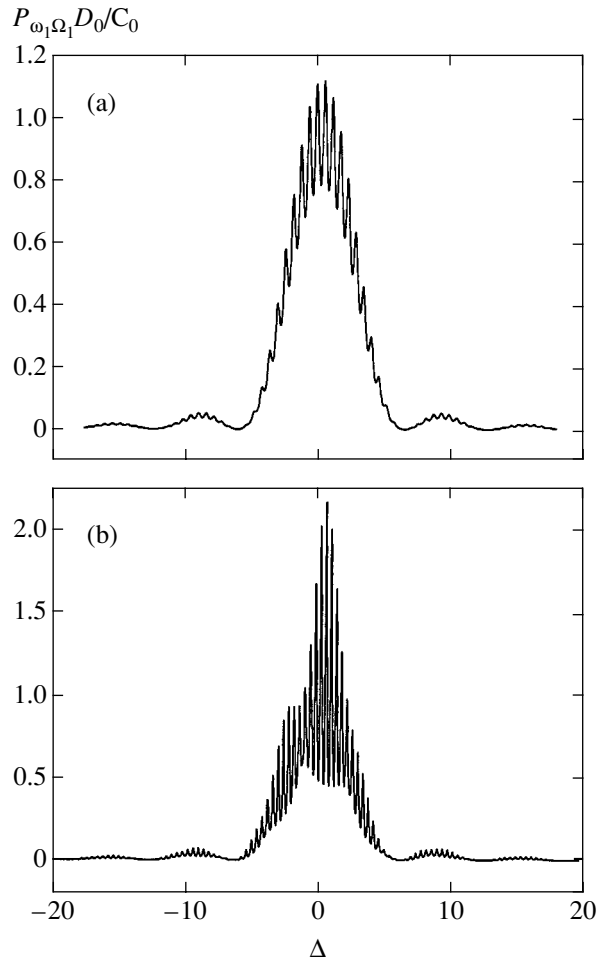


Fig. 2. SPDC line shape in a transparent slab that reflects signal waves ($r_1 = -0.3$): (a) other waves are not reflected; (b) idler and pump waves are reflected ($r_2 = -0.5$ and $r_0 = -0.5$).

form factor for scattered waves are

$$g'(\Delta, y_2) \equiv \frac{4}{(\Delta^2 + y_2^2)^2}[(\Delta^2 - y_2^2) \times (1 - \cosh y_2 \cos\Delta) - 2y_2\Delta \sinh y_2 \sin\Delta],$$
$$g''(\Delta, y_2) \equiv \frac{4}{(\Delta^2 + y_2^2)^2}[(\Delta^2 - y_2^2)\sinh y_2 \sin\Delta - 2y_2\Delta(1 - \cosh y_2 \cos\Delta)], \quad (26)$$
$$D_j = 1 - 2R_j \cos 2\delta_j + R_j^2,$$
$$\tilde{D}_j = 1 - 2\tilde{R}_j \cos 2\delta_j + \tilde{R}_j^2,$$

$R_j \equiv r_j^2$ is the reflection coefficient defined as the corresponding reflected-to-incident intensity ratio for a wave of frequency ω_j, $\tilde{R}_j \equiv \tilde{r}_j^2$, and $\tilde{r}_j \equiv r_j \exp(-y_i)$.

For a medium that is transparent at idler frequencies, but partially absorbs signal waves, an expression for SPDC intensity is derived from the generalized Kirchhoff law written represented by (4). Instead of (25), we obtain

$$P^{\mathrm{SPDC}}_{\omega_1\Omega_1}\Big|_{\substack{y_1 \neq 0 \\ y_2 = 0}} = C_0 \frac{(1-R_1)(1-R_2)}{D_0 \tilde{D}_1 D_2} F(\Delta, y_1, 0) \times \{1 + R_2 + R_0\tilde{R}_1 + R_0\tilde{R}_1R_2 - 2r_0\tilde{r}_1r_2[\cos(\Delta + 2\delta_0) + \cos\delta]\}. \quad (27)$$

Here, the value of $\delta \equiv \delta_2 - \delta_1 - \delta_0$ differs from that of $\Delta \equiv \delta_2 + \delta_1 - \delta_0$.

In the case of arbitrary laws of wave absorption and reflection in the slab, the PC signal intensity is expressed as

$$P^{\mathrm{PC}}_{\omega_1\Omega_1}\Big|_{\substack{y_1 \neq 0 \\ y_2 \neq 0}} = C_0 N_2 \frac{(1-R_1)(1-R_2)}{D_0 \tilde{D}_1 \tilde{D}_2} F(\Delta, y_1, y_2) \times (1 - 2r_0\tilde{r}_1\tilde{r}_2\cos\delta + R_0\tilde{R}_1\tilde{R}_2). \quad (28)$$

It is obvious that the PC line profile in the presence of reflections differs from the SPDC line profile in the absence of both absorption at signal frequencies (cf. (25) and (28)) and appreciable absorption at idler frequencies (cf. (27) and (28)).

Reflections in a slab give rise to fine interference structure in SPDC and PC line profiles. When the crystal is transparent, the modulation due to frequency- or angle-dependent variation of $\mathrm{sinc}^2(\Delta/2)$ combines with additional modulation of two types. One of these is associated with the behavior of terms containing $2\delta_1$ and $2\delta_2$. Moreover, these factors characterize the Airy linear interference of signal and idler waves in a slab. Interference of this kind occurs when the superposition of incident and reflected waves of each particular frequency is independent of the presence of other waves. This interference also manifests itself in spectra associated with nonlinear interaction. It leads to double periodic modulation of spectra with periods determined by the conditions $\delta_j = \pi n$, where n is an integer. The corresponding modulation depth is determined by the reflection coefficients R_1 and R_2.

Figure 2a shows an example of a frequency-dependent SPDC line profile in the case of Airy linear interference between reflected signal waves when $R_2 = 0$. Whether the pump is reflected or not, the SPDC line profile remains invariant. The relation between the Airy modulation period and the modulation controlled by the function $\mathrm{sinc}^2(\Delta/2)$ is determined by specific parameters of the medium. The SPDC signal intensity mainly concentrates within the principal maximum of

$\mathrm{sinc}^2(\Delta/2)$, i.e., within the frequency bandwidth $\Delta\omega_{1s}$ given, as in a nonreflecting slab, by the expression

$$\Delta\omega_{1s} \equiv \frac{\int P_{\omega_1\Omega_1} d\omega_1}{P_{\omega_{1s}\Omega_1}} = \frac{2\pi\cos(\vartheta_2 - \vartheta_0)}{\left|u_2^{-1} - u_1^{-1}\cos(\vartheta_1 - \vartheta_2)\right| l}, \quad (29)$$

where u_m denotes group velocities in the crystal. To observe extra peaks in the frequency profile of a scattered-wave line, we must have the condition $\Delta\omega_{ml} < \Delta\omega_{1s}$, where $\Delta\omega_{ml} = \pi u_m/l\cos\vartheta_m$ is the frequency spacing between the peaks due to linear interference of signal or idler waves. Analogous relations for angular intervals determine the observability of Airy interference in the angular SPDC line profile. For example, when parametric interaction of the type considered at the end of the preceding section takes place inside a 1-cm-thick slab of KDP crystal, the modulation period associated with wave interference varies between 0.013 and 0.014 nm, whereas the spectral width of the principal maximum is substantially larger (0.5 nm). The corresponding period of interference modulation in the angular line profile is similar to the angular width of the principal maximum (0.02°).

The other type of SPDC line profile modulation, due to wave reflection in the slab, is of greater interest. Modulation of this type is associated with change in the combination of parameters δ_j:

$$\Delta + 2\delta_0, \quad \delta, \quad 2(\delta_2 - \delta_1) \equiv 2(\delta + \delta_0),$$
$$2(\delta_1 + \delta_2) \equiv 2(\Delta + \delta_0).$$

It can be observed only when both signal and idler waves are reflected by the boundaries. Note that the reflected pump wave somewhat changes the overall intensity distribution, but it is not required to observe interference. The peaks associated with changes in the mismatch δ by multiples of π (which is equivalent to changes in $\delta_2 - \delta_1$ by multiples of π) are spaced approximately as the Airy interference maxima. For example, the inverse frequency widths of the Airy maxima, $\Delta\omega_{ml}$ and $\Delta\omega_1'$, satisfy the relation

$$(\Delta\omega_1')^{-1} = (\Delta\omega_{1l})^{-1} + (\Delta\omega_{2l})^{-1}. \quad (30)$$

However, the occurrence of the peaks associated with changes by multiples of π in the mismatch $\Delta + 2\delta_0$ and, therefore, in $\delta_2 + \delta_1$ (if δ_0 is constant) is strictly related to the initial mismatch. Interference of this type can be classified as nonlinear, because the locations of the peaks depend on the phases of all waves involved in a parametric interaction. The frequency bandwidth due to the nonlinear interference associated with the phase

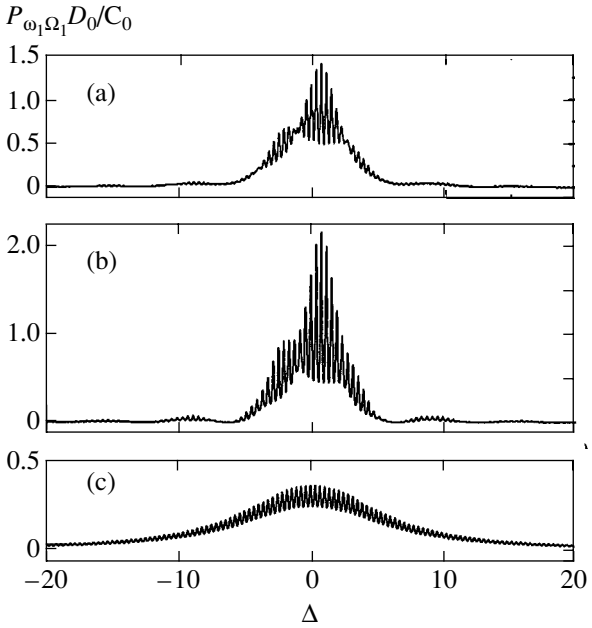


Fig. 3. SPDC line shape in a reflecting slab ($r_0 = -0.5$, $r_1 = -0.3$, $r_2 = -0.5$) for several values of absorption coefficients at signal and idler frequencies: (a) $\alpha_1 l = 0$, $\alpha_2 l = 1$; (b) $\alpha_1 l = 1$, $\alpha_2 l = 0$; (c) $\alpha_1 l = 0$, $\alpha_2 l = 10$.

incursion $2(\delta_1 + \delta_2) \equiv 2(\Delta + \delta_0)$ is determined by the relation

$$(\Delta\omega_{1nl})^{-1} = \left|(\Delta\omega_{1l})^{-1} - (\Delta\omega_{2l})^{-1}\right|. \quad (31)$$

The corresponding modulation period is smaller than the frequency width of the central maximum of $\mathrm{sinc}^2(\Delta/2)$ by a factor of four. The period of the nonlinear interference associated with the phase incursion $\Delta + 2\delta_0$ is smaller than the frequency width of the same central maximum by a factor of two. Figure 2b shows an example of frequency-domain scattered-wave line profile observed in the case when the signal, idler, and pump waves are reflected from the boundaries. It demonstrates that nonlinear interference leads to periodic variation in the amplitude of high-frequency modulation of the spectrum.

Nonlinear interference is caused by parametric interaction between forward and backward waves and linear interference between signal waves of frequency ω_1. Similar effects were observed, for example, in studies of frequency–angular intensity distribution and correlation between SPDC signal and idler channel intensities in slabs of a nonlinear crystal separated by layers of an optically linear medium [23–26].

Figure 3 illustrates the influence of the signal and idler waves on the interference pattern observed at the signal frequency. With increasing absorption, the visibility of the interference pattern, which is determined by linear and nonlinear interference of the absorbed wave, decreases. The variation of visibility concurs

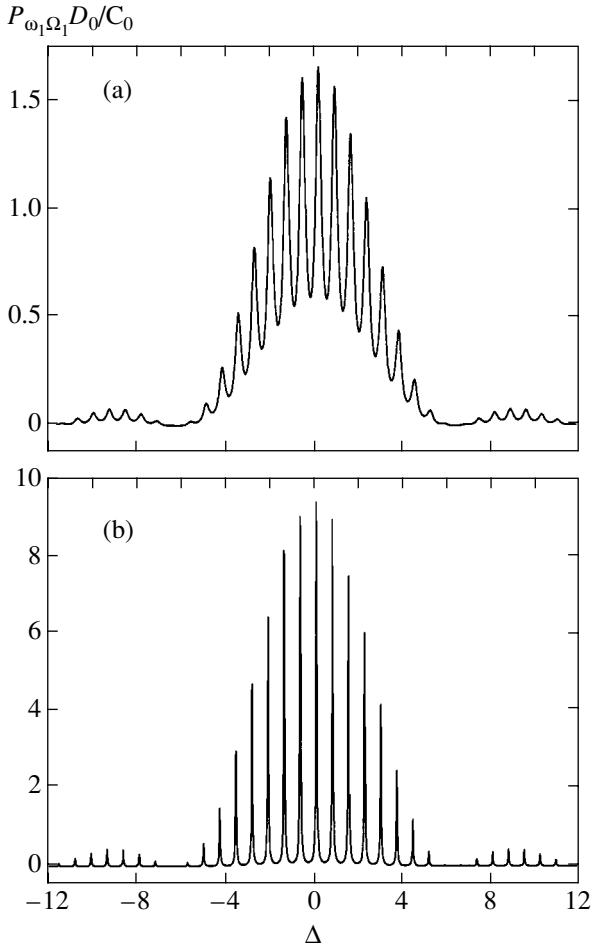


Fig. 4. SPDC line shape for a slab that reflects only idler waves ($r_0 = r_1 = 0$): (a) $r_2 = -0.5$, (b) $r_2 = -0.9$.

with analogous variations of the basic line profile, which is characteristic of the slab in the absence of reflection. The increase in absorption at both frequencies leads to gradual decrease in contrast, and the interference structure of the line profile is eventually blurred out. The effects of reflection on the SPDC line profile can be observed experimentally only under certain conditions. These conditions are generally violated when SPDC is used in spectroscopy, but this does not preclude determination of volume-averaged values of refractive index and absorption coefficient at idler frequencies from smoothed scattered-wave line profiles. Reflection effects must be taken into account in measurements of integral scattered-wave intensity. For example, this is required when SPDC is used in photometry [20, 27–29].

4. INTERFERENCE OF ELECTROMAGNETIC ZERO-POINT FLUCTUATIONS OF THE VACUUM

Consider SPDC in the case when the idler wave is reflected by the slab boundaries, whereas the signal and pump are not. Suppose that the idler wave can be absorbed. Figure 4 shows the frequency-domain line profiles corresponding to several values of the absorption coefficient r_2. It is clear that the overall signal line profile is modulated. The modulation frequency and depth are described by the Airy linear interference for idler (rather than signal) waves.

Generally, the efficiency of each act of parametric conversion depends, among other factors, on the average values of field operators in the input idler and signal modes [10]. In SPDC observations, only the pump wave is incident on the crystal. The thermal population of the modes is too low to ensure the required scattering efficiency. In this case, the role of seed is played only by the so-called electromagnetic zero-point fluctuations of the vacuum at the idler frequency when a signal wave is detected or at the signal frequency when an idler wave is detected. Recall that the effective radiance of zero-point fluctuations of the vacuum in PC processes is one photon per mode [30]. If the seed is an external free wave "separated" from its source, then its intensity inside the crystal will depend on the absorption and reflection coefficients of the medium at the idler frequency and therefore affect the signal spectrum. In SPDC, idler waves are generated only via decay of the pump. The "free" idler wave is then reflected by the slab boundaries and absorbed as it propagates through the medium. In the linear-amplification approximation, the emerging idler wave weakly interacts with the pump and feedback is negligible. Therefore, the idler wave does not affect the signal spectrum, and the line profile shown in Fig. 4 must not be sensitive to the intensity distribution of the idler wave generated in the medium. Thus, the signal line profile shown here is due only to the field fluctuations at the idler frequency, which participate initially in each elementary act of scattering and subsequently in the interference of the signal waves originating from different points in a nonlinear medium. The interference structure of the signal line with parameters determined by the idler-frequency properties of the crystal can be attributed to interference of zero-point fluctuations of the vacuum [31].

The interference pattern in the line profile or the signal–idler intensity cross-correlation function reflects the spatial distribution of the vacuum states of the field in the slab. Spatial nonuniformity must be taken into account in a comprehensive quantum-theoretic analysis of the problem even at the stage of field quantization. To perform quantization, the field strength operator in an individual mode is represented as follows [32]:

$$E_j(\mathbf{r}, t) = \nu_j(\mathbf{r})a_j(t) + \nu_j^*(\mathbf{r})a_j^\dagger(t), \qquad (32)$$

where $a_j^\dagger(t) = a_j^\dagger \exp(i\omega_j t)$ and $a_j(t) = a_j \exp(-i\omega_j t)$ are the jth-mode photon creation and annihilation operators, respectively, and $\nu_j(\mathbf{r})$ is an eigenfunction of the

corresponding boundary value problem for the spatial field distribution described by Maxwell's equations and medium-specific boundary conditions. It is obvious that every jth-mode field has the same spatial distribution $\nu_j(\mathbf{r})$ irrespective of its quantum state, which may be a Fock-space state with a constant number of photons, a mixed or vacuum one, etc. When the slab has reflecting boundaries, an interference pattern corresponding to $\nu_j(\mathbf{r})$ is observed, and similar patterns are characteristic of both a single-mode equilibrium thermal field and a vacuum-state field. Since both signal and idler input fields are in the vacuum state in SPDC, the signal spectrum exhibits an interference structure associated with the vacuum states of the field at the idler frequency. A different quantum state of the field added to the input in the idler channel can change the frequency-domain line profile of the scattered wave. However, the change will occur only when all modes are populated nonuniformly, i.e., when the mode occupation number (mean number of photons in a mode) varies from mode to mode because of the additional field. Otherwise, the signal line profile will be similar to that due to the interference of zero-point fluctuations of the vacuum described here.

If R_0, R_1, and y_2 in (25) are assumed to be small, this expression describes light scattering by interference polaritons [33] in those spectral regions where the contribution of the Raman tensor to the scattered-wave intensity is negligible. The contribution to light scattering by polaritons due to parametric processes is the main source of signal in spectral regions far from phonon resonances when the difference of idler and phonon wavenumbers (measured in cm^{-1}) is greater than the phonon decay constant. The curves connecting the interference peaks of the same order $m = \delta_2/\pi = k_{2z}l/\pi$ can be used to determine the dispersion of interference polaritons. As $R_2 \longrightarrow 1$, the conditions for total reflection of the polariton waves trapped in the slab are satisfied. In this case, peaks in the fine signal-profile structure correspond to scattering by waveguide polaritons [33–35]. For example, this is characteristic of parametric interaction in lithium niobate crystals, when the idler wave corresponds to the upper polariton branch with wavenumbers about 2000 cm^{-1}. In this case, observation of linear interference of zero-point fluctuations of the vacuum requires either a receiver with a sufficiently high resolution or a sufficiently thin crystal film, because the spacing between the lines is 0.2 cm^{-1} and 4 cm^{-1} for 1 cm and 500 μm thick slabs.

5. SCATTERING MATRIX AND SIGNAL LINE PROFILE IN MEDIA WITH SPATIALLY MODULATED SECOND-ORDER SUSCEPTIBILITY

We consider three-wave parametric interaction in a specimen of a layered medium with second-order susceptibility varying in one direction. First, we suppose that the value of $\chi^{(2)}$ varies periodically. We assume that the spatial modulation of linear susceptibility is so weak that the effects due to variation of refractive index and absorption coefficient can be neglected. This approximation is valid, for example, in analysis of parametric processes in ferroelectric crystals with regular domain structure [1, 3, 4, 6–9]. In such crystals, a change in the sign of spontaneous polarization between domains is associated with a change in all even-order susceptibilities in the expansion of polarizability in powers of field strength, while the odd-order susceptibilities remain invariant [36].

Periodically varying second-order susceptibility can be represented as the Fourier series

$$\chi^{(2)}(\mathbf{r}) = \sum_{m=-\infty}^{\infty} \chi_m \exp(i\mathbf{q}_m \cdot \mathbf{r}) \tag{33}$$

with spatial-harmonic amplitudes

$$\chi_m = \frac{1}{d}\int_{-d/2}^{d/2} \chi^{(2)}(\mathbf{r})\exp(-i\mathbf{q}_m \cdot \mathbf{r})d\mathbf{r}, \tag{34}$$

where d is the period of a one-dimensional "nonlinear superlattice" (modulation period) and $\mathbf{q}_m = \mathbf{q}m$ denotes vectors of the reciprocal superlattice, with an integer m and $\mathbf{q} \equiv (2\pi/d)\mathbf{n}$ ($\mathbf{n}$ is the unit normal vector to the layer boundaries). In the approximation considered here, the amplitudes of forward waves are continuous across the input boundary of the medium, and the scattering matrix is identical to the matrix of the parametric interaction of forward waves. The nonlinear polarization at the frequency ω_1 is expressed as

$$P_1(\omega_1 = \omega_0 - \omega_2) = \sum_{m=-\infty}^{\infty} \chi_m E_0 E_2^* \exp[i(\mathbf{k}_0 - \mathbf{k}_2 + \mathbf{q}_m)\cdot\mathbf{r}]. \tag{35}$$

The nonlinear polarization at the frequency ω_2 is given by a similar expression. The corresponding reduced equations are

$$\frac{dE_1(z)}{dz} + i\sum_m \sigma_{1m}E_0E_2^*(z)\exp\left(-\frac{i\Delta_m z}{l}\right) + \frac{y_1}{l}E_1(z) = 0,$$
$$\frac{dE_2(z)}{dz} + i\sum_m \sigma_{2m}E_0E_1^*(z)\exp\left(-\frac{i\Delta_m z}{l}\right) + \frac{y_2}{l}E_2(z) = 0. \tag{36}$$

Here, the z axis is aligned with the normal to the nonlinear layers,

$$\Delta_m \equiv l(k_{1z} + k_{2z} - k_{0z} - q_m) \equiv \delta_1 + \delta_2 - \delta_0 - q_m l$$

is the dimensionless wave-vector mismatch for the mth harmonic, Δ_m/l differs from the mismatch Δ/l in a spatially uniform medium by the magnitude of the reciprocal-superlattice vector q_m, and

$$\sigma_{jm} \equiv 2\pi\chi_m\omega_j/cn_j\cos\vartheta_j \quad (j = 1, 2).$$

When the PC coefficient is small, the absolute values of the parameters $\beta_{1m} \equiv i\sigma_{1m}E_0l$ and $\beta_{2m} \equiv -i\sigma^*_{2m}E_0l$ characterizing the efficiency of wave interaction in the slab are small: $|\beta_m| = \sqrt{\beta_{1m}\beta_{2m}} \ll 1$. The parameter β_m is defined analogously to the parameter β for a homogeneous medium (see (12)), with $\chi \longrightarrow \chi_m$.

The solution of Eq. (36) determines the elements of the scattering matrix, which is identical to the parametric interaction matrix for waves in the layered medium considered here. In the linear-amplification approximation (up to terms of order β_m^2),

$$\begin{aligned} w_{11} &\approx \exp(-y_1) + \sum_{m=-\infty}^{\infty} |\beta_m|^2 f_1(\eta_m) \\ &+ \sum_{m=-\infty}^{\infty} \sum_{\substack{m'=-\infty \\ m'\neq m}} \beta_{1m}\beta_{2m'} \frac{\exp(\eta_m - \eta_{m'} - y_1)}{2\eta_{m'}} \\ &\times [\varphi(\eta_m) - \varphi(\eta_m - \eta_{m'})], \\ w_{12} &\approx \sum_{m=-\infty}^{\infty} i\beta_m f(\eta_m), \end{aligned} \tag{37}$$

where

$$\varphi(x) \equiv \frac{e^{-2x} - 1}{2x}.$$

The functions $f(\eta_m)$ and $f_1(\eta_m)$ are defined by (14) with the replacements $\Delta \longrightarrow \Delta_m$ and $\eta \longrightarrow \eta_m \equiv (-y_1 + y_2 + i\Delta_m)/2 = \eta - imql$:

$$\begin{aligned} f_1(\eta_m) &\equiv \frac{1}{(2\eta_m)^2}[\exp(-y_2 - i\Delta_m) \\ &- (1 - 2\eta_m)\exp(-y_1)] = \frac{\exp(-y_1)[\varphi(\eta_m) + 1]}{2\eta_m}, \\ f(\eta_m) &\equiv \frac{1}{2\eta_m}\left[\exp\left(-y_2 - \frac{i\Delta_m}{2}\right)\right. \\ &\left. - \exp\left(-y_1 + \frac{i\Delta_m}{2}\right)\right] = \exp\left(-y_1 + \frac{i\Delta_m}{2}\right)\varphi(\eta_m). \end{aligned} \tag{38}$$

In (37), the element w_{12} contains contributions of all spatial harmonics χ_m, which are similar to the expression for a homogeneous specimen (cf. (37) and (13)), but with $\Delta \longrightarrow \Delta_m$ and $\chi \longrightarrow \chi_m$. The expression for w_{11} contains additional summands corresponding to interference of the contributions due to different harmonics χ_m and $\chi_{m'}$ ($m \neq m'$).

When the idler wave is absorbed, the generalized Kirchhoff law dictates the following expression for the signal intensity $P_{\omega\Omega}$ in SPDC:

$$\begin{aligned} P_{\omega_1\Omega_1} &= C\sum_{m=-\infty}^{\infty} |\chi_m|^2 g(\Delta_m, y_2) \\ &+ C\sum_{m=-\infty}^{\infty} \sum_{\substack{m'=-\infty \\ m'\neq m}} (-1)^{n(m'-m)} \chi^*_m \chi_{m'} \\ &\times \left[\frac{\exp(-i\Delta - y_2) - 1}{(y_2 + i\Delta - imql)(y_2 + i\Delta - im'ql)}\right. \\ &\left. + \frac{\exp(i\Delta - y_2) - 1}{(y_2 - i\Delta - imql)(y_2 - i\Delta - im'ql)}\right]. \end{aligned} \tag{39}$$

Here, the coefficient C is equal to C_0/χ^2 for a homogeneous medium (see (15), (16), (25)–(28)) and n is the number of periods spanned by the thickness l ($n \equiv l/d$). The scattered-wave line profile in a periodically modulated medium is characterized by two terms. One of them is the sum of contributions of individual harmonics. In each summand, the intensity distribution relative to the maximum is similar to that in a homogeneous medium (function $g(\Delta, y_2)$ in (15)). For each component, the new maxima are shifted according to the quasi-matching condition $\Delta(w_1, \theta_1) = -q_ml \equiv -2\pi mn$. The value of each summand is proportional to the squared amplitude of the corresponding harmonic. The other term in (39) contains products of different harmonics χ_m and describes their interference.

When neither signal nor idler wave is absorbed, the SPDC intensity is

$$P_{\omega_1\Omega_1}\Big|_{\substack{y_1=0 \\ y_2=0}} = C\left|\sum_{m=-\infty}^{\infty} \chi_m \mathrm{sinc}\left(\frac{\Delta_m}{2}\right)\right|^2. \tag{40}$$

Figure 5 shows examples of SPDC intensity versus phase mismatch $\Delta k/q \equiv \Delta/2\pi n$ calculated by using (40) in the simplest case of meander-type periodic distribution of $\chi^{(2)}(z)$. This distribution is characteristic of multidomain crystals in which the sign of second-order nonlinear susceptibility sharply reverses across the

domain boundaries and the transition-layer thickness is negligible as compared to the thickness of a domain of either sign. In this case, (40) yields the following expression for the SPDC line profile:

$$P_{\omega_1\Omega_1}\Big|_{\substack{y_1=0\\ y_2=0}} = C\bar{\chi}^2\frac{\mathrm{sinc}^2(\Delta/2)}{\sin^2(\Delta/2n)}$$

$$\times\left(1-2\cos\frac{\rho\Delta}{2n}\cos\frac{\Delta}{2n}+\cos^2\frac{\Delta}{2n}\right),$$

where $\rho \equiv (l_1 - l_2)/d$ is a parameter characterizing the asymmetry of the domain system ($d = l_1 + l_2$ is the superlattice period). Both spectral and angular spacings between the peaks due to nonlinear diffraction increase with decreasing superlattice period. Unlike the widths of the peaks, they are independent of the total thickness of a specimen. For example, consider collinear parametric interaction of *ooe* type in a lithium niobate crystal with regular domain structure and domain walls parallel to the *xz* crystallographic plane. For this configuration, the difference between the curves corresponding to zeroth- and first-order quasi-matching is about 100 cm^{-1} if the domain layer thickness is 5 μm. Accordingly, the difference is smaller by an order of magnitude if the thickness is 50 μm. These values are obtained for a 488-nm pump propagating in the crystallographic planes *yz* of domains at an angle of 57°, crossing the domain walls. In this case, collinear interaction is observed when the signal wavelength is about 510 nm.

Expression (40) for signal intensity contains the amplitudes $|\chi_m|$ and phases φ_m of the spatial harmonics $\chi_m \equiv |\chi_m|\exp(i\varphi_m)$ (which are complex-valued in the general case). When the imaginary part of $\chi^{(2)}(\mathbf{r})$ vanishes (far from the resonances of the medium), it holds that

$$\chi_m = (\chi_{-m})^*, \quad |\chi_m| = |\chi_{-m}|, \quad \varphi_m = -\varphi_{-m}.$$

The direct relation between SPDC spectrum and spatial distribution of second-order susceptibility can be used as a basis for measuring $\chi^{(2)}(\mathbf{r})$ in layered spatially nonuniform crystals and nonlinear structures. Measuring the ratio of the peak intensities associated with nonlinear diffraction in different orders of quasi-matching, for which $|\Delta| = 2\pi mn$, one can find the ratio of different harmonic amplitudes $|\chi_m|$. Since the interference terms directly depend on the phases φ_m, the signal intensity is redistributed among the quasi-matching maxima when $\Delta \neq 2\pi mn$. The interference is most pronounced in scattering by specimens spanning just a few superlattice periods.

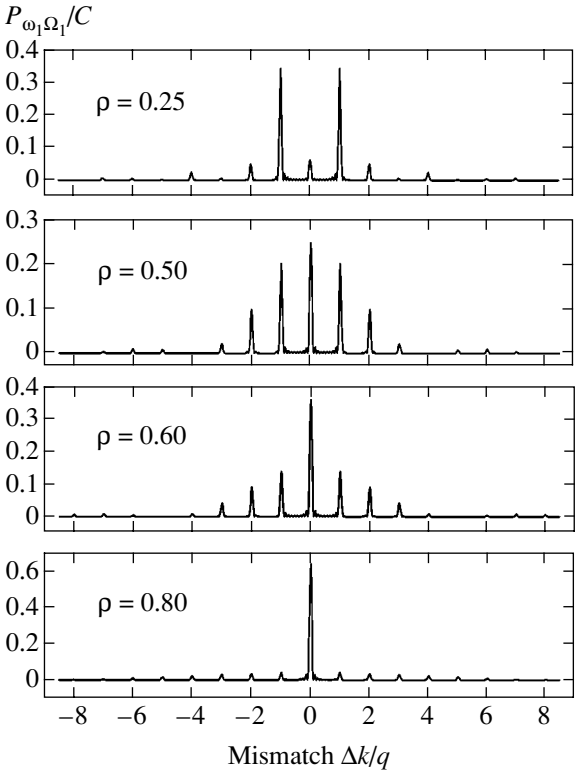


Fig. 5. SPDC line shape in a transparent ferroelectric with regular domain structure for different domain thickness ratios.

When the idler wave is absorbed, each peak associated with nonlinear diffraction is broadened, but not shifted (see Fig. 6). If $y_2 \gg 1$ and the number of domains is so large that $n \gg y_2/n$, then (39) reduces to

$$P_{\omega_1\Omega_1} = 2C_0 y_2 \sum_{m=-\infty}^{\infty}\frac{|\chi_m|^2}{y_2^2 + (\Delta - 2\pi mn)^2}. \qquad (41)$$

The width of each nonlinear-diffraction peak on the scale of Δ is determined by the integral absorption y_2 over the entire crystal thickness *nd*, whereas information about the phases φ_m is lost. Nevertheless, the nonlinear-diffraction bands can be observed if the absorption over a superlattice period, y_2/n, is relatively weak. For example, if the domain size measured along the idler-wave propagation direction does not exceed 5 μm, then the diffraction pattern can be observed in the range of idler frequencies such that $\alpha_2 \leq 2000$ cm^{-1}. Conversely, if absorption is so strong that $y_2/n \gg 1$, then the nonlinear-diffraction bands are blurred out and the line of a parametrically scattered wave is a single broad Lorentzian curve centered at Δ = 0. In the latter case, even the values of $|\chi_m|$ cannot be measured.

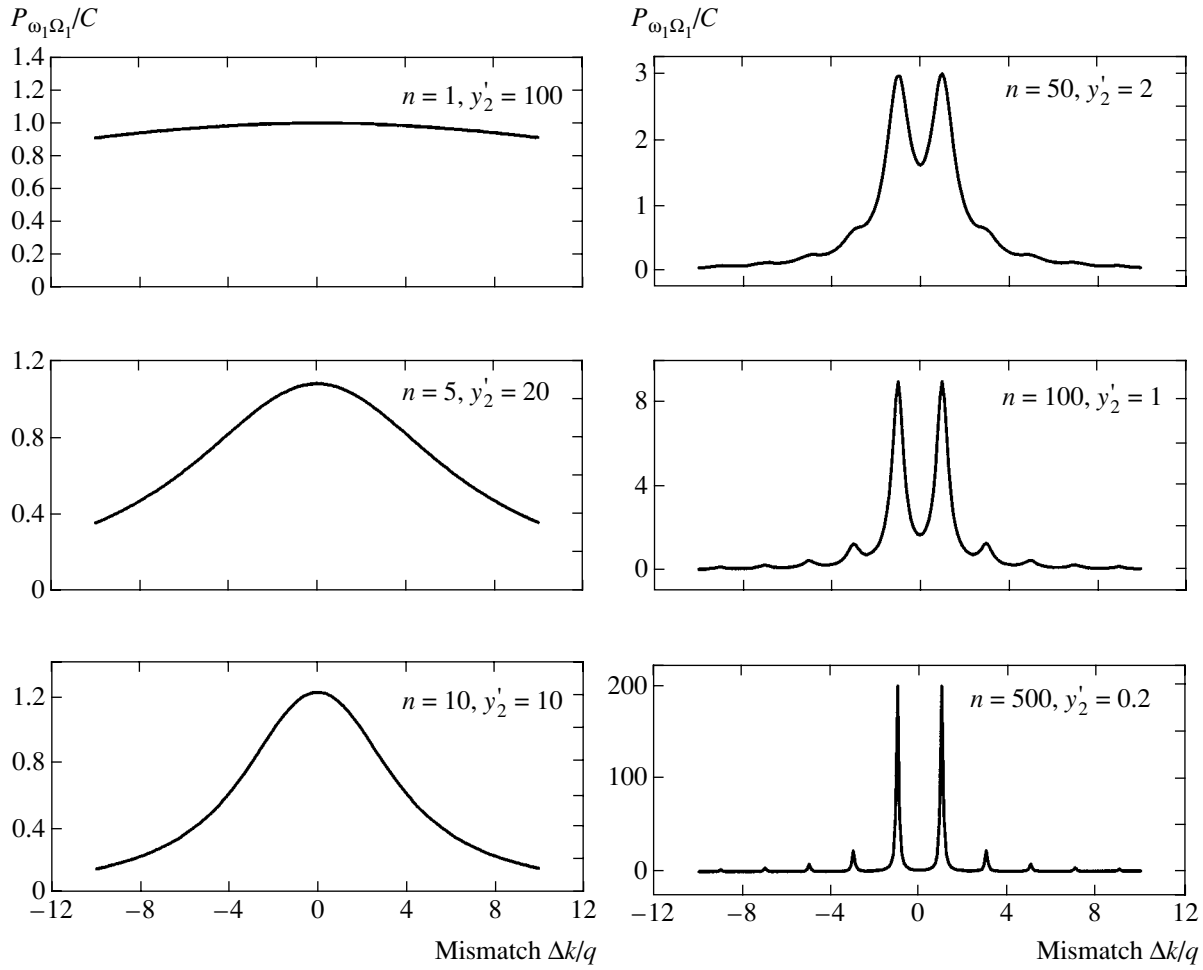


Fig. 6. SPDC line shape in a periodically poled ferroelectric with regular domain structure for different n and $y_2/n = y_2'$. Both crystal thickness and integral absorption y_2 are the same in all cases.

SPDC spectra can also be used to measure the complete profile $\chi^{(2)}(z)$ in a weakly absorbing specimen of thickness l with an arbitrary distribution of second-order susceptibility. Consider the situation when the specimen is a part of a periodic medium in which the period of $\chi^{(2)}(z)$ is equal to l. If the pump occupies only a slab $\{-l/2, l/2\}$ in the medium, then the parametrically scattered signal is identical to that generated by a specimen of thickness l. The corresponding SPDC intensity distribution is described by (39) and (40) with $n = 1$ and

$$\chi_m = \frac{1}{l}\int_{-l/2}^{l/2}\chi^{(2)}(z)\exp\left(-i\frac{2\pi m}{l}z\right)dz. \qquad (42)$$

In principle, the inverse problem of reconstruction of a complete profile $\chi^{(2)}(z)$ can be solved by using any three-wave parametric process, including stimulated up-conversion, second-harmonic generation, and cascade processes [3, 4, 6–9]. However, when using a stimulated process, one always has to deal with insufficient or nonuniform population of the converter modes induced by input radiation. For example, the line profile of parametric conversion determined by the scattering matrix elements w_{12} is generally expressed as

$$\begin{aligned} P^{\mathrm{PC}}(\omega_1, \theta_1) &= P(\omega_2, \theta_2)\frac{\omega_1^3\cos\theta_1}{\omega_2^3\cos\theta_2}|w_{12}|^2 \\ &= CN(\omega_2, \theta_2)\left|\sum_{m=-\infty}^{\infty}\chi_m f(\eta_m)\right|^2, \end{aligned} \qquad (43)$$

where $N(\omega_2, \theta_2)$ is the distribution of the photon number in the input idler mode, whereas strictly uniform population of input idler modes is guaranteed (by virtue of zero-point fluctuations of the vacuum) when spontaneous SPDC is used in the linear-amplification approximation. Information about domain structure can also be obtained by measuring a line profile even under the simplest conditions, when the effects due to interference are negligible. The number and relative intensity

of shifted curves can be used to find the number of spatial harmonics and the relative values of $|\chi_m|$; the shift direction and spacing between the parametric curves, to determine the direction and period of variation of second-order nonlinear susceptibility [9, 37–40].

6. CONCLUSIONS

In this study, we describe frequency-angular distributions for spontaneous parametric down-conversion and stimulated parametric conversion by applying a general approach based on the scattering matrix and the generalized Kirchhoff law. This approach makes it possible to solve a problem in quantum optics without invoking the formal procedure of field quantization for a spatially nonuniform nonlinear medium consisting of a nonlinear crystal and the surrounding space. The eigenfunctions of the corresponding wave problem describing a spatial distribution governed by Maxwell's equations subject to boundary conditions are not required. The elements of the scattering matrix are calculated for classical waves and then used as coefficients relating the quantum operators of input and output waves. The approach relies on the similarity between Maxwell's equations for classical fields and the Heisenberg equations for field operators in free space. Thus, intensity distributions can be obtained both for classical down-converted signals and for signals of essentially quantum nature generated via scattering of a pump by electromagnetic zero-point fluctuations of the vacuum.

Even though the generalized Kirchhoff law allows one to ignore the spatial distribution of modes in calculations, it can be used to obtain correct spatial and spectral distributions of scattered-wave intensity. In particular, we have obtained expressions for SPDC involving interference and waveguide polaritons without using the spatial dependence of the wavefunctions describing the corresponding states of the field in a slab. The calculated SPDC line profiles reflect the behavior of $\nu_m(\mathbf{r})$ for eigenmodes in spatially nonuniform media, including crystalline slabs characterized by absorption, semitransparent boundaries, and nonuniformly distributed second-order susceptibility. In addition to its convenience, an important advantage of the approach based on the generalized Kirchhoff law is the possibility to allow for absorption of signal or idler waves. As in formulations of fluctuation-dissipation theorems, absorption is taken into account on a phenomenological level. The distinction is that the generalized Kirchhoff law deals with directly measurable characteristics: correlation functions of input and output fields.

It should be noted that the use of the generalized Kirchhoff law in the calculation of the line profile for a stimulated difference-frequency signal reduces to a standard classical calculation of the input field intensity based on the scattering-matrix elements relating the input and output fields. Expressions (16), (28), and (43) for the line profile for a stimulated difference-frequency signal generated in a layered inhomogeneous medium are obtained directly by solving the reduced equations for slowly varying field amplitudes. Analogous equations describe both sum-frequency (in particular, second-harmonic) generation and each three-wave stage of stimulated cascade processes of higher orders. The only difference lies in the definitions of wave-vector mismatch Δ in terms of δ_j for sum-frequency generation: the signal and idler wave vectors must be taken with opposite signs.

Analysis of the SPDC intensity spectrum can serve as a basis for measuring spatial variations of linear and nonlinear optical susceptibilities. One advantage is that SPDC has a very broad spectrum determined by the spectrum of zero-point fluctuations of the vacuum, whereas the signal spectrum in any stimulated frequency-conversion process is restricted to the spectrum of incident radiation. Note also that the method based on SPDC makes it possible to examine domain structures lying in the bulk of a crystal, man-made layered structures, and photonic crystals. The results presented here can be applied to measure one-dimensional distributions of $\chi^{(2)}$ in media with vanishingly small variations of linear susceptibility. The approach based on the generalized Kirchhoff law can be extended to media with nonuniform nonlinearities characterized by arbitrary irregular distributions of refractive indices and absorption coefficients.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project nos. 02-02-16843 and 03-02-16364, and by grant NSh-166.2003.02 under the program of State Support of Leading Scientific Schools.

REFERENCES

1. A. V. Golenishchev-Kutuzov, V. A. Golenishchev-Kutuzov, and R. I. Kalimullin, Usp. Fiz. Nauk **170**, 697 (2000) [Phys. Usp. **43**, 647 (2000)].
2. G. D'Aguanno, M. Centini, M. Scalora, *et al.*, Phys. Rev. E **64**, 016609 (2001).
3. R. L. Byer, J. Nonlinear Opt. Phys. Mater. **6**, 549 (1997).
4. D. K. Serkland, M. M. Fejer, R. L. Byer, and Y. Yamamoto, Opt. Lett. **20**, 1649 (1995).
5. M. Scalora, M. J. Bloemer, A. S. Manka, *et al.*, Phys. Rev. A **56**, 3166 (1997).
6. N. V. Kravtsov, G. D. Laptev, I. I. Naumova, *et al.*, Kvantovaya Élektron. (Moscow) **32**, 923 (2002).
7. A. S. Chirkin, V. V. Volkov, G. D. Laptev, and E. Yu. Morozov, Kvantovaya Élektron. (Moscow) **30**, 847 (2000).
8. S. G. Grechin and V. G. Dmitriev, Kvantovaya Élektron. (Moscow) **26**, 151 (1999).
9. G. Kh. Kitaeva, A. A. Mikhaĭlovskiĭ, and A. N. Penin, Zh. Éksp. Teor. Fiz. **112**, 2001 (1997) [JETP **85**, 1094 (1997)].

10. D. N. Klyshko, *Photons and Nonlinear Optics* (Nauka, Moscow, 1980).
11. D. N. Klyshko, Izv. Akad. Nauk SSSR, Ser. Fiz. **46**, 1478 (1982).
12. D. N. Klyshko, Zh. Éksp. Teor. Fiz. **55**, 1006 (1968) [Sov. Phys. JETP **28**, 522 (1968)].
13. D. N. Klyshko, Pis'ma Zh. Éksp. Teor. Fiz. **6**, 490 (1967) [JETP Lett. **6**, 23 (1967)].
14. D. N. Klyshko, *Physical Fundamentals of Quantum Electronics* (Nauka, Moscow, 1986).
15. D. N. Klyshko, Zh. Éksp. Teor. Fiz. **83**, 1313 (1982) [Sov. Phys. JETP **56**, 753 (1982)].
16. D. N. Klyshko, Usp. Fiz. Nauk **166**, 613 (1996) [Phys. Usp. **39**, 573 (1996)].
17. D. N. Klyshko, A. N. Penin, and B. F. Polkovnikov, Pis'ma Zh. Éksp. Teor. Fiz. **11**, 11 (1970) [JETP Lett. **11**, 5 (1970)].
18. D. N. Klyshko, V. F. Kutsov, A. N. Penin, and B. F. Polkovnikov, Zh. Éksp. Teor. Fiz. **62**, 1291 (1972) [Sov. Phys. JETP **35**, 682 (1972)].
19. Y. R. Shen, *The Principles of Nonlinear Optics* (Wiley, New York, 1984; Nauka, Moscow, 1989).
20. G. Kh. Kitaeva, D. N. Klyshko, and I. V. Taubin, Kvantovaya Élektron. (Moscow) **9**, 560 (1982).
21. S. A. Akhmanov and R. V. Khokhlov, *Problems in Nonlinear Optics* (Akad. Nauk SSSR, Moscow, 1965; Gordon and Breach, New York, 1972).
22. M. Born and E. Wolf, *Principles of Optics*, 4th ed. (Pergamon Press, Oxford, 1969; Nauka, Moscow, 1970).
23. A. V. Burlakov, S. P. Kulik, A. N. Penin, and M. V. Chekhova, Zh. Éksp. Teor. Fiz. **113**, 1991 (1998) [JETP **86**, 1090 (1998)].
24. D. Yu. Korystov, S. P. Kulik, and A. N. Penin, Kvantovaya Élektron. (Moscow) **30**, 921 (2000).
25. A. V. Burlakov, Yu. B. Mamaeva, A. N. Penin, and M. V. Chekhova, Zh. Éksp. Teor. Fiz. **120**, 67 (2001) [JETP **93**, 55 (2001)].
26. D. Yu. Korystov, S. P. Kulik, and A. N. Penin, Pis'ma Zh. Éksp. Teor. Fiz. **73**, 248 (2001) [JETP Lett. **73**, 214 (2001)].
27. D. N. Klyshko, Kvantovaya Élektron. (Moscow) **4**, 1056 (1977).
28. D. N. Klyshko, Kvantovaya Élektron. (Moscow) **7**, 1932 (1980).
29. D. N. Klyshko and A. N. Penin, Usp. Fiz. Nauk **152**, 653 (1987) [Sov. Phys. Usp. **30**, 716 (1987)].
30. O. A. Abroskina, G. Kh. Kitaeva, and A. N. Penin, Dokl. Akad. Nauk SSSR **280**, 584 (1985) [Sov. Phys. Dokl. **30**, 67 (1985)].
31. G. Kh. Kitaeva, A. V. Sergienko, and A. N. Penin, Dokl. Akad. Nauk SSSR **293**, 848 (1987) [Sov. Phys. Dokl. **32**, 293 (1987)].
32. R. Glauber, in *Quantum Optics and Electronics*, Ed. by C. DeWitt, A. Blandin, and C. Cohen-Tannoudji (Gordon and Breach, New York, 1965; Mir, Moscow, 1966).
33. *Surface Polaritons*, Ed. by V. M. Agranovich and D. L. Mills (North-Holland, Amsterdam, 1982; Nauka, Moscow, 1985).
34. V. N. Denisov, T. A. Leskova, B. N. Mavrin, and V. B. Podobedov, Zh. Éksp. Teor. Fiz. **94** (5), 261 (1988) [Sov. Phys. JETP **67**, 1013 (1988)].
35. Y. Sasaki and S. Ushioda, Phys. Rev. B **27**, 1122 (1983).
36. A. L. Aleksandrovskiĭ, O. A. Gliko, I. I. Naumova, and V. I. Pryalkin, Kvantovaya Élektron. (Moscow) **23**, 657 (1996).
37. A. L. Aleksandrovskiĭ, G. Kh. Kitaeva, S. P. Kulik, and A. N. Penin, Zh. Éksp. Teor. Fiz. **90**, 1051 (1986) [Sov. Phys. JETP **63**, 613 (1986)].
38. G. Kh. Kitaeva, S. P. Kulik, and A. N. Penin, Fiz. Tverd. Tela (St. Petersburg) **34**, 3440 (1992) [Sov. Phys. Solid State **34**, 1841 (1992)].
39. G. Kh. Kitaeva, S. P. Kulik, and A. N. Penin, Ferroelectrics **172**, 469 (1995).
40. G. Kh. Kitaeva, A. A. Mikhailovsky, I. I. Naumova, *et al.*, Appl. Phys. B **66**, 201 (1998).

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 98, No. 2, 2004, pp. 287–304.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 2, 2004, pp. 324–344.

PLASMA, GASES

Coagulation and Growth Mechanisms for Dust Particles in a Low-Temperature Plasma

M. A. Olevanov, Yu. A. Mankelevich, and T. V. Rakhimova

Skobel'tsyn Research Institute of Nuclear Physics, Moscow State University, Vorob'evy gory, Moscow, 119992 Russia

e-mail: olevanov_m@mtu-net.ru; ymankelevich@mics.msu.su

Received November 4, 2002

Abstract—We suggest a theoretical model of dust clustering in a low-temperature plasma that includes a description of all the main stages of this process, from the initial growth and coagulation of particles to the saturation phase. Based on the constructed theory, we have explained the experimentally observed threshold behavior of the coagulation process for the first time and estimated the critical microparticle size upon reaching which transition from the growth of particles through the deposition of material from the gas phase to their coagulation becomes possible. Using the derived analytical expressions for the coagulation rate constant, we numerically simulated the clustering process based on data taken for real experimental conditions and studied the evolution of the particle size distribution function during the entire process. A direct comparison of the numerical calculations with experimental data shows them to be in good agreement with the actually observed pattern of the phenomenon. © *2004 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION

In recent years, studies in the field of a dusty low-temperature plasma have aroused broad interest and have been of great practical importance in connection with rapid progress of microelectronics and with the transition of production to nanotechnology. Micron- and submicron-sized particles have long been known to nucleate and grow as a side effect in most commercial plants used in semiconductor production. Dust clusters have been found in plasma etching [1, 2], chemical vapor deposition [3, 4], and sputtering deposition [5, 6] reactors. Since macroparticles in a plasma generally acquire significant negative charges, they prove to be locked by an electrostatic field and can grow for a long period until the discharge is quenched or they are removed from the discharge glow region under self-gravity. State-of-the-art semiconductor technology allows one to deal with objects with sizes on the order of 0.1 μm even now. However, further sophistication and miniaturization of electronic devices require placing an increasingly large number of elements on a chip, which cannot be done without exploring the range of sizes of several tens of nanometers. Therefore, the appearance of dust during a technological cycle is a serious problem, because macroparticles with sizes of 20–100 nm falling on the substrate can give rise to a fatal defect, thereby sharply reducing the chip yield.

However, apart from the processes in which the presence of nanoparticles leads to undesirable effects, there is a wide range of problems related to the production of materials with special properties. Thus, investigating and using the nucleation and growth of particles in a plasma are of great interest, because their properties, such as size monodispersity and chemical composition, may be well controllable. Nanoparticles synthesized in a plasma are used in ceramic production [7], and powders of pure metals, alloys, and composite materials are produced in magnetron sputtering reactors [8].

In recent years, the need for understanding the growth of dust particles and controlling these processes has led to the appearance of many experimental works in this field. The growth of dust clusters in silane [9–19], germanium [20, 21], and methane [22, 23] plasmas, where argon and helium were commonly used as the buffer gas, has been studied extensively. The experimental parameters were typical of the conditions in chemical film deposition reactors: the SiH_4, GeH_4, and CH_4 contents were varied within the range 5–15%, the pressure in the reactor was 10–80 Pa, the discharge power was 10–100 W, and the frequency was 3.5–28 MHz. The growth dynamics of dust particles has also been investigated in experiments in which the initial material for clustering was produced via substrate sputtering. These experiments were carried out with various materials (copper, aluminum, graphite, tungsten, titanium, and others) [24–26]. Apart from an analysis of the clustering dynamics itself, the factors and parameters that had the strongest effect on the particle growth rate were sought. The influence of a gas flow [27, 28], temperature gradient, and frequency modulation [29] on dust formation was studied. These experiments showed that noticeable suppression of the clustering process could be achieved by a special choice of conditions.

A generalization of the experimental results yields the following pattern of dust particle formation and growth. In general, this process proceeds in three phases: the initial growth phase, the coagulation phase, and the saturation phase. In the first phase, the particles uniformly grow to sizes on the order of 10 nm. They have an almost regular spherical shape and a narrow size distribution. In the coagulation phase, the growth rate of dust particles rapidly increases, because the probability of collisions between the particles themselves increases sharply. The surfaces of the dust grains formed in this phase can have distinctly different shapes and topologies: both a spherical one [22, 30] and a fractal one formed from linked chains of initial crystallites [13, 21]. In the saturation phase, the coagulation essentially ceases and the particles continue to grow very slowly via the gas-phase deposition of material. In general, the characteristic clustering time can vary from fractions of a second to several hours, depending on the conditions of a specific experiment. However, the dust growth rate is, on average, much higher than the growth rate of films from the gas phase. For comparison, we can give data from [21], where the dust growth rate reached 100–800 nm s^{-1}, while the characteristic film deposition rate is only 0.064–0.12 nm s^{-1} [17].

In addition to experimental studies, extensive theoretical investigations aimed at offering a proper explanation for the mechanism of rapid cluster growth have also been carried out. The initial phase of dust particle nucleation from the gas phase in silane was studied in detail in [17, 31–35]. The authors considered the sequences of chemical reactions that resulted in $[Si_nH_x]$ clustering; the clusters with $n \geq 10$ were assumed to be formed particles that grew further via the gas-phase deposition of radicals [36, 37].

The coagulation phase of dust particles begins when they reach some critical size, typically 2–10 nm. Although the dynamics of the process has been well studied experimentally, the mechanism of this phenomenon still remains unclear in many respects. The principal question that many investigators have attempted to answer is how dust particles can actively grow in the presence of the significant like charges accumulated by them. According to the approximation of restricted orbits [38], the collision cross section for such particles must be exponentially small. Several models that solve this problem have been suggested in the literature. The first ballistic model assumes that dust particles coagulate into clusters through a collision between charged and uncharged dust grains [32, 39] or during a collision between weakly charged particles [40–42]. The fact that the dust number density in the initial coagulation phase generally exceeds the ion and electron number density in a plasma is an argument for this assumption. As a result, according to the quasineutrality condition, a large fraction of the macroparticles are uncharged. This model can account for the onset of saturation and the termination of coagulation. In this case, the dust number density falls below the ion number density in the volume. As a result, almost all the macroparticles are charged, which is why the process terminates. However, the existence of the initial critical macroparticle size below which no coagulation takes place remains incomprehensible in this model. In addition, as has been repeatedly pointed out by many authors [15, 21], the collision frequency that corresponds to the thermal cross section for the interaction between dust grains cannot account for such a high cluster growth rate observed experimentally. Thus, we conclude that additional attraction must exist between particles.

The anomalously high coagulation rate can be explained by assuming that some of the dust particles are positively charged and that collisions take place mostly between oppositely charged particles. The authors of this model point out that, since the mean dust particle charge in the initial coagulation phase is small and undergoes fluctuations about zero and since collisions of dust grains with high-energy electrons can lead to secondary electron emission from their surfaces, a positive charge is actually present on any macroparticle for some time [13, 43]. However, there is strong doubt that this hypothesis is fruitful, because the presence of a sufficient number of high-energy electrons in the volume that are capable of providing the necessary number density of positively charged macroparticles will most likely require creating special conditions. Meanwhile, practice shows that dust clusters actively grow in ordinary situations.

In addition, we should also take into account the fact that it would be more appropriate to consider the evolution of the charge on an individual macroparticle rather than the total content of the positively charged dust in the volume. In general, the characteristic fluctuation time of the dust grain charge is much shorter than the time it takes for the grain to collide with another particle. Since the charging time is mainly determined by collisions with electrons, we can write $\tau_e/\tau_d \sim n_d v_d/n_e v_e$ for estimation, where n_e, n_d are the electron and macroparticle number densities, and v_e, v_d are their thermal velocities. For silicon particles 10 nm in size at room temperature, $v_d \sim 10^2$ cm s^{-1}; the electron velocity at a temperature $T_e \sim 3$ eV characteristic of most experiments is $v_e \sim 10^8$ cm s^{-1}. Since $v_d/v_e \sim 10^{-5}$–10^{-6} under these conditions and the number density ratio n_d/n_e does not exceed 10^2–10^3 in most cases, we obtain $\tau_e/\tau_d \ll 1$. This implies that, when considering a pair particle–particle interaction, we must deal with the mean particle charge rather than the instantaneous one. Thus, it should be recognized that the currently existing theoretical description of the coagulation process is unsatisfactory and needs to be improved further.

Here, based on our previously suggested polarization model for the interaction between dust particles [44], we theoretically calculate the formation and growth of dust clusters. The main idea of the model is that, being

placed in the electric field produced by a neighboring macroparticle, a dust particle is polarized, which, in turn, leads to a redistribution of ion flows along its surface. This asymmetry gives rise to an additional force acting on the particle along the electric vector. The principal distinctive feature of the suggested model is that, in this case, the force from the plasma flows is proportional to the electric field strength, as is the force of Coulomb repulsion. Our expression for the force exerted on the macroparticles from a plasma flow allows the contribution of the plasma to be parametrically taken into account when calculating the pair interaction. Depending on conditions in the system, this contribution may prove to be dominant and can lead to effective attraction between the particles.

The goal of our study is to construct a theoretical model for the growth of macroparticles in a plasma that describes the three clustering phases mentioned above as well as the time evolution of the particle size distribution function and the time dependence of the dust particle number density and mean size. As the starting point for constructing our theory, we took the experimentally obtained pattern of the phenomenon from [13], which is the most typical behavior of the system. Here, we briefly give basic parameters of the experiment [13] and the main evolutionary stages of the system observed in this case. The experiment was carried out in an argon plasma with a 10% SiH_4 content under a pressure of 13 Pa. The discharge was maintained by an electric field with a frequency of 6.5 MHz; the applied power was 40 W. Within the first approximately 0.2 s after discharge initiation, the particles were outside the resolution range of the methods for determining the sizes used in the experiment (about 5 nm). Within 0.5 s after discharge initiation, the particle size increased to 15 nm and coagulation began. After 1.0 s, the appearance of two groups of particles with a narrow size distribution was clearly observed. The group of small particles subsequently remained within the same range from 10 to 20 nm over the entire period of the experiment from 0.5 until 4 s, while their density decreased from 1.5×10^9 cm^{-3} 0.5 s after discharge initiation to 3×10^8 cm^{-3} 1.0 s after the start. Photographs taken with an electron microscope show that the particles from the first group have a more or less regular spherical shape, while the particles from the second group are fractal structures consisting of coalesced particles 10–20 nm in size. Their growth rate in the initial phase was very high; the mean size reached 75 nm 1.2 s after the start of the experiment. Subsequently, saturation set in, and the particles continued to grow more slowly: up to approximately 110 nm by 4.0 s. Their number density remained virtually constant, 4×10^7 cm^{-3}.

Below, we perform all our calculations by using the above data. The initial growth of dust particles and their coagulation will be considered separately. The attainment of some critical size by the particles, which will also be estimated in our model, will be used as the criterion for the transition from one cluster growth regime to the other.

2. THE INITIAL GROWTH OF DUST PARTICLES

The main feature of the evolution of a system in the phase of initial growth is the formation of a fraction of dust particles in the volume with an approximately constant number density that have an almost regular spherical shape and a narrow size distribution. Below, we calculate the growth rate of clusters in this phase and estimate the dust particle size dispersion.

Simplistically, the initial growth of clusters can be subdivided into two concurrent subprocesses. The first describes the nucleation of future particles, and the second describes their subsequent growth via the deposition of material from a plasma. Assuming the particle shape to be spherical, we will describe the evolution of the dust subsystem by introducing a particle volume distribution function $f(V, t)$, where $V = (4/3)\pi a^3$ is the volume of the individual particle and a is its radius. We normalize the distribution function of the dust number density in the volume as follows:

$$n_d(t) = \int_0^\infty f(V, t)dV. \quad (1)$$

The equation for the distribution function that describes the particle growth via the gas-phase deposition of material can be written as

$$\frac{\partial f(V, t)}{\partial t} = \nu_0(V - u_0) f(V - u_0, t) - \nu_0(V) f(V, t). \quad (2)$$

In this equation, ν_0 is the frequency of collisions of a dust particle with radicals and gas atoms, and u_0 is the increase in the volume of the macroparticle through the attachment of one particle from the gas phase to it. Below, we will estimate this quantity as the volume of the molecule being attached.

We estimate the collision frequency as

$$\nu_0 \approx \pi a^2 n_a \bar{v}_a, \quad (3)$$

where $\bar{v}_a$ is the mean thermal velocity of the active radicals and n_a is their number density. In this case, the probability that an incident atom will stick to the surface of a dust particle when they collide is assumed to be equal to unity. As was shown in several papers [17, 19], for a silane plasma, the main particle responsible for the growth of clusters from the gas phase is the neutral SiH_2 radical.

Since the increase in the volume of the macroparticle as it attaches one molecule from the gas is much

smaller than its volume, Eq. (2) can be rewritten in differential form:

$$\frac{\partial f(V,t)}{\partial t} = -u_0 \frac{\partial \nu_0(V) f(V,t)}{\partial V}. \quad (4)$$

Supplementing this equation with the initial and boundary conditions

$$f(V, t=0) = 0,$$
$$\frac{\partial f(V_0,t)}{\partial t} = \frac{k(t)}{u_0} - \nu_0(V_0) f(V_0,t), \quad (5)$$

we obtain a completely definite problem for the distribution function to be found. In the last equation, $k(t)$ is the nucleation rate of future macroparticles with an initial volume $V_0 \gg u_0$.

Equation (4) can be easily solved by using the Laplace transform, and its solution can be written as

$$f(V,t) = \frac{\nu_0(V_0)}{\nu_0(V)} f_0(t-b)\eta(t-b), \quad (6)$$

where

$$f_0(t) = \frac{1}{u_0}\int_0^t k(t')\exp(\nu_0(V_0)(t'-t))dt'. \quad (7)$$

In Eq. (6), the Heaviside function

$$\eta(t) = \begin{cases} 1, & t \geq 0, \\ 0, & t < 0, \end{cases} \quad (8)$$

and b is the time before the appearance of particles with a given size:

$$b = \frac{1}{u_0}\int_{V_0}^{V} \frac{dV}{\nu_0(V)}. \quad (9)$$

To calculate the cluster growth rate, we need to pass from the macroparticle volume distribution function to the macroparticle size distribution function by assuming that all particles are spherical in shape and that $dV = 4\pi a^2 da$. Equation (6) then transforms to

$$f(a,t) = 4\pi a_0^2 f_0(t-b)\eta(t-b), \quad (10)$$

where the radius of the nucleation center is denoted by a_0. Substituting (3) into (9) and performing integration over the sizes yields the following expression for the delay time:

$$b(a) = \frac{4}{\bar{\nu}_a n_a u_0}(a - a_0). \quad (11)$$

As we see from Eq. (10) for the distribution function, the pattern of its evolution with time is the propagation of a wave in the space of macroparticle sizes. In this case, for any time, the shape of the curve in the space of macroparticle sizes does not change, while the curve itself drifts toward larger sizes at a constant rate c_f:

$$c_f = \frac{\bar{\nu}_a n_a u_0}{4}. \quad (12)$$

We see from Eq. (10) that the form of the macroparticle size distribution function for any time is completely determined by the time dependence of the nucleation rate. The nucleation process is a complex sequence that consists of several tens of chemical reactions [32, 33], and the determination of its temporal characteristics is an independent and complex problem. To find the form of the distribution function in our case and to make further estimates, we will consider this process in simplified form.

Let only one reaction that involves active radicals of one type and molecules of the initial gas underlie the macroparticle nucleation process. Denote, as previously, the number densities of the radicals and the initial gas molecules by n_a and n_0, respectively. If the electron impact dissociation of initial gas molecules is taken as the main production mechanism of radicals, then the production rate of radicals can be written as $c = k_e n_e n_0$. The destruction of the radicals that formed will be governed by two processes—the reaction with particles of the initial gas followed by the nucleation of future macroparticles and the sticking to the already existing dust particles in the volume. Assuming that the steady state comes fast, we write the equilibrium equation for the reactions as

$$k_0 n_a n_0 + k_d n_a n_d = k_e n_e n_0, \quad (13)$$

where we use the following notation: k_0—the rate constant for the nucleation reaction, k_d—the rate constant for the destruction on the surfaces of existing dust particles, and n_d—the dust number density.

We see from Eq. (13) that the dust nucleation and the deposition of material on the already existing nucleation centers compete with each other. In the initial phase of the process, when the dust content in the volume is small, the main removal mechanism of active radicals is nucleation. However, over the course of time, as the number density and surface area of the nucleation centers increase, this process must give way to the destruction of molecules on dust particles. In this case, the formation of new growth centers in the volume terminates almost completely. The satisfaction of the following condition can serve as the criterion for the transition from one regime to the other:

$$k_0 n_0 \sim k_d n_d. \quad (14)$$

Let us estimate the time in which the transition from the first regime to the second will take place. The nucleation rate is defined by the relation $k(t) = k_0 n_a n_0$. Expressing n_a from (13) and substituting it into the last

expression, we obtain

$$\frac{dn_d}{dt} = \frac{k_e n_e n_0}{1 + k_d n_d / k_0 n_0}. \quad (15)$$

Since $k_0 n_0 \gg k_d n_d$ in the nucleation phase, the dust particle size remains virtually unchanged, as does the rate constant for the destruction of radicals on macroparticles, k_d. In this case, the solution of Eq. (15) can be written as

$$n_d = \frac{k_0 n_0}{k_d}\left(\sqrt{1 + 2\frac{k_d k_e n_e t}{k_0}} - 1\right). \quad (16)$$

Let us estimate the time it takes for the nucleation to almost cease by using (16) and condition (14):

$$\tau \approx \frac{k_0}{k_d k_e n_e}. \quad (17)$$

For our estimates, we take the following parameters: $T_e = 2.5$ eV; $T_a = 0.025$ eV — the gas temperature, which is assumed to be the room one; $n_e \sim 1 \times 10^9$ cm^{-3}; and $a_a \sim 2.5$ Å—the characteristic size of the initial gas molecules and the active radicals. According to [33], the typical particle nucleation center is a $[Si_nH_x]$-type cluster, where $n \sim 10$. We estimate its size as $10^{1/3} a_a \approx 5.4$ Å. The thermal velocities $\bar{v}_e = 9.3 \times 10^7$ cm s^{-1} and $\bar{v}_a = 4 \times 10^4$ cm s^{-1} correspond to the above temperatures. Therefore, using the expressions for the interaction rate constants $k = \pi a^2 \bar{v}$, we obtain $k_0 \approx 7.9 \times 10^{-11}$ cm^3 s^{-1}, $k_d \approx 3.6 \times 10^{-10}$ cm^3 s^{-1}, and $k_e \approx 1.8 \times 10^{-7}$ cm^3 s^{-1}. Substituting all parameters into (17), we write the estimate for the nucleation time as $\tau \approx 1.2 \times 10^{-3}$ s.

Let us now estimate the rate of macroparticle growth via the gas-phase deposition of active radicals. This is nothing but the drift rate of the distribution function in the space of particle sizes, c_f, that we calculated above (see (12)). The number density of the radicals can be derived from Eq. (13). In the initial nucleation phase, the following estimate is valid for it:

$$n_a \sim \frac{k_e n_e}{k_0} = 2.3 \times 10^{12} \text{ cm}^{-3}.$$

We estimate the change in the volume of the macroparticle when a radical is attached to it as

$$u_0 \sim \frac{4}{3}\pi a_a^3 = 6.5 \times 10^{-23} \text{ cm}^3.$$

For the growth rate, we then obtain $c_f \approx 15$ nm s^{-1}. In fact, this estimate is an upper limit, because when calculating this quantity, we assumed that the number density of the radicals was constant. However, as the dust particle size and, accordingly, the sticking rate constant increase, the number density of the radicals decreases, which will eventually cause the material deposition rate to decrease and the further cluster growth to slow down.

Intense nucleation takes place only for a short time, following which the number density of the nucleation centers changes only slightly. Therefore, the width of the macroparticle size distribution function can be estimated as $\Delta a \sim c_f \tau \approx 0.02$ nm.

Thus, we arrive at the following pattern of the initial growth of clusters: in this phase, the nucleation centers of future particles are produced for a short time on the order of 10^{-3}, following which their number density ceases to change and their size begins to increase. This growth proceeds at an approximately constant rate of 10–15 nm s^{-1}. Since the nucleation centers are produced only for a very short time, the particle size distribution is virtually monodisperse. Our model is in close agreement with experimental data (see the Introduction) both qualitatively and quantitatively. In the next section, we consider the coagulation of dust particles.

3. THE COAGULATION RATE OF DUST PARTICLES IN THE APPROXIMATION OF POLARIZATION INTERACTION

We calculated the force exerted on a dust particle in an electric field with allowance made for its polarization previously [44]. Below, we write out only the resulting expression:

$$F_d = Z_d e E_0 \left(\frac{\frac{3}{2} + \frac{n_0 \sigma_0}{n_d \pi \bar{a}^2} \Phi(\xi)}{\frac{Z_i \tilde{n}_i}{Z_d n_d} \frac{\bar{a}^2}{\bar{a}^2} \frac{}{1 + \frac{n_0 \sigma_0}{n_d \pi \bar{a}^2}}} - 1 \right). \quad (18)$$

Here, Z_d is the dust grain charge, E_0 is the electric field, Z_i is the ion charge, $\tilde{n}_i$ is the ion number density near the macroparticle surface, n_d is the dust number density, n_0 is the number density of the neutral gas atoms, σ_0 is the cross section for the interaction of ions with neutral gas atoms, $\bar{a}^2$ is the mean square of the macroparticle size, a is the radius of the chosen particles, and

$$\Phi(\xi) = \xi^2 [1 - 2\xi^2 + 2\sqrt{\pi}\xi^3 \exp(\xi^2)\operatorname{erfc}(\xi)], \quad (19)$$

where

$$\xi^2 = \frac{Z_i e |\varphi_0|}{T_i} \frac{1}{1 + \frac{n_0 \sigma_0}{n_d \pi \bar{a}^2}}. \quad (20)$$

In the latter relation, T_i is the ion temperature and φ_0 is the potential on the dust particle surface. When the macroparticle is spherical in shape, the potential on its surface is related to the particle size by

$$e|\varphi_0| = \frac{Z_d e^2}{a}. \quad (21)$$

Next, let us consider two dust particles a_1 and a_2 that interact with each other with forces $\boldsymbol{f}_{12}$ and $\boldsymbol{f}_{21}$, where $\boldsymbol{f}_{12}$ is the force exerted on the first particle from the second one and $\boldsymbol{f}_{21}$ is the force exerted on the second particle from the first one. We take (18), in which the electric field is produced by the charge of the neighboring particle, as the expression for the force. As a result, we obtain a system of two particles with a pair central interaction in which the presence of plasma is taken into account parametrically. Next, we will solve the problem of the scattering of two particles by each other in this system.

However, it should be noted that, despite its similarity to the standard problem that considers the pair interaction between two particles, the problem of dust particle interaction in a plasma has an important distinction. Given the ion flows, the third Newton law will not hold for the interaction forces; i.e., in general, $\boldsymbol{f}_{12} \neq -\boldsymbol{f}_{21}$ for particles of different sizes. This conclusion is not surprising, because the presence of plasma makes the system unclosed. As a result, situations are possible where one of the particles runs away from the second particle (it has such a size that the electrostatic repulsion exceeds the force from the ion flow) and the second particle catches up with it. This process results in energy transfer from the plasma to the dust subsystem accompanied by its heating. The heating of the dust subsystem was actually observed experimentally [45, 46].

Denote the radius vector that connects the first and the second particles by $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$, where $\mathbf{r}_1$ and $\mathbf{r}_2$ are the radius vectors of the first and the second particles in a fixed frame of reference. Its change with time can then be described by the equation

$$\ddot{\mathbf{r}} = \frac{\boldsymbol{f}_{21}}{m_2} - \frac{\boldsymbol{f}_{12}}{m_1}. \tag{22}$$

Substituting the expressions for the forces into (22) yields

$$\ddot{\mathbf{r}} = -\frac{1}{\mu} \times \left\{ \frac{Z_i}{n_d \overline{a^2}} \left(\frac{\tilde{n}_{i1} a_1^2}{Z_{d1}} \frac{m_2}{m_1 + m_2} \frac{\frac{3}{2} + \frac{n_0 \sigma_0}{n_d \pi \overline{a^2}} \Phi(\xi_1)}{1 + \frac{n_0 \sigma_0}{n_d \pi \overline{a^2}}} + \frac{\tilde{n}_{i2} a_2^2}{Z_{d2}} \frac{m_1}{m_1 + m_2} \frac{\frac{3}{2} + \frac{n_0 \sigma_0}{n_d \pi \overline{a^2}} \Phi(\xi_2)}{1 + \frac{n_0 \sigma_0}{n_d \pi \overline{a^2}}} \right) - 1 \right\} f_{\text{e.s.}}, \tag{23}$$

where the subscripts denote the parameters pertaining to the first and the second particles, respectively,

$$\mu = \frac{m_1 m_2}{m_1 + m_2}$$

is the reduced mass, and $\boldsymbol{f}_{\text{e.s.}}$ is the force of electrostatic repulsion that would act between the two particles in the absence of plasma.

Thus, we find that the original problem is equivalent to the problem of the motion of a particle with mass μ in a centrally symmetric field with an interaction potential energy $U(r)$:

$$U(r) = -\left\{ \frac{Z_i}{n_d \overline{a^2}} \left(\frac{\tilde{n}_{i1} a_1^2}{Z_{d1}} \frac{m_2}{m_1 + m_2} \frac{\frac{3}{2} + \frac{n_0 \sigma_0}{n_d \pi \overline{a^2}} \Phi(\xi_1)}{1 + \frac{n_0 \sigma_0}{n_d \pi \overline{a^2}}} + \frac{\tilde{n}_{i2} a_2^2}{Z_{d2}} \frac{m_1}{m_1 + m_2} \frac{\frac{3}{2} + \frac{n_0 \sigma_0}{n_d \pi \overline{a^2}} \Phi(\xi_2)}{1 + \frac{n_0 \sigma_0}{n_d \pi \overline{a^2}}} \right) - 1 \right\} U_{\text{e.s.}}, \tag{24}$$

where $U_{\text{e.s.}}(r)$ is the energy of electrostatic repulsion between the dust grains.

The problem of the motion of a macroparticle in the field of a fixed center with a given centrally symmetric potential can be easily solved. In this case, the collision cross section is

$$\sigma_{dd} = \pi R_{\min}^2 \left(1 + \frac{2U(R_{\min})}{\mu v^2} \right), \tag{25}$$

where $R_{\min}$ is the minimum distance to which the particle can approach the center and upon reaching which a collision occurs and v is the relative velocity.

For spherical particles, $R_{\min} = a_1 + a_2$, $m_1, m_2 \sim a_1^3$, a_2^3, and $U_{\text{e.s.}}$ is given by

$$U_{\text{e.s.}}(R_{\min}) = \frac{Z_{d1} Z_{d2} e^2}{a_1 + a_2}.$$

In our subsequent analysis, we also assume that, on average, the potential on the surfaces of all particles is the same; $\varphi_0 = \text{const}$. This conclusion follows from the fact that the macroparticles are in a conducting medium and the potential difference would be rapidly offset by the currents produced in it. For $U_{\text{e.s.}}$, we can then write

$$U_{\text{e.s.}}(a_1 + a_2) = \varphi_0^2 \frac{a_1 a_2}{a_1 + a_2}. \tag{26}$$

The ion number densities near the macroparticles, $\tilde{n}_{i1}$ and $\tilde{n}_{i2}$, are determined by the potential on the dust

grain surfaces as well as by the particle sizes and the conditions in the plasma. The number density can vary greatly, depending on the situation being realized. The two main factors that affect the form of the ion distribution function are the absorption of ions on the macroparticle surface and the relationship of the dust grain size a to the ion mean free path λ_i in a plasma. The former factor is significant for large particles, because the lifetime of the ions near a macroparticle for $a \gtrsim \lambda_i$ decreases sharply, causing their number density to decrease appreciably compared to the case without absorption on the surface. This situation corresponds to the hydrodynamic limit. However, under typical conditions of the experiments in which the growth of particles from the gas phase is studied, the opposite case is typical. In this case, the Boltzmann exponential law can serve as a valid estimate of $\tilde{n}_i$:

$$\tilde{n}_i \approx n_i \exp\frac{Z_i e|\varphi_0|}{T_i}, \tag{27}$$

where n_i is the volume-averaged ion number density.

For the interaction cross section, we then finally obtain

$$\sigma_{a_1a_2} = \pi(a_1 + a_2)^2\left\{1 + \frac{2U_{\text{e.s.}}(a_1 + a_2)}{\mu v^2}\right.$$
$$\times\left(\frac{Z_i e n_i \exp(Z_i e|\varphi_0|/T_i)}{n_d\overline{a^2}|\varphi_0|}\frac{a_1a_2(a_1^2 + a_2^2)}{a_1^3 + a_2^3}\right. \tag{28}$$
$$\left.\left.\times\frac{\frac{3}{2} + \frac{n_0\sigma_0}{n_d\pi\overline{a^2}}\Phi(\xi)}{1 + \frac{n_0\sigma_0}{n_d\pi\overline{a^2}}} - 1\right)\right\}.$$

Next, let us introduce the parameter

$$\chi = \frac{Z_i e n_i \exp(Z_i e|\varphi_0|/T_i)}{n_d\overline{a^2}|\varphi_0|}$$
$$\times\frac{a_1a_2(a_1^2 + a_2^2)}{a_1^3 + a_2^3}\frac{\frac{3}{2} + \frac{n_0\sigma_0}{n_d\pi\overline{a^2}}\Phi(\xi)}{1 + \frac{n_0\sigma_0}{n_d\pi\overline{a^2}}} - 1, \tag{29}$$

which determines the sign of the interaction between the two particles. The particles are attracted for $\chi > 0$ and repelled for $\chi < 0$. Denote also the cross section for the interaction between uncharged particles by $\sigma_T = \pi(\alpha_1 + \alpha_2)^2$.

Let us now calculate the rate constant for the interaction between macroparticles of different sizes:

$$k_{a_1a_2} = \langle\sigma_{a_1a_2}(v)v\rangle. \tag{30}$$

In this expression, the averaging is over the relative particle velocities. Assume that the macroparticle velocity distribution is Maxwellian with a temperature T_0 equal to the gas temperature and consider separately two cases: $\chi \geq 0$ and $\chi < 0$.

For $\chi \geq 0$, which corresponds to the attraction between the particles,

$$k_{a_1a_2} = \left(\frac{\mu}{2\pi T_0}\right)^{3/2}\sigma_T\int_0^\infty v\left(1 + \frac{2U_{\text{e.s.}}}{\mu v^2}\chi\right)$$
$$\times\exp\left(-\frac{\mu v^2}{2T_0}\right)4\pi v^2 dv. \tag{31}$$

Integrating (31) yields

$$k_{a_1a_2} = 2\bar{v}\sigma_T\left(1 + \frac{2U_{\text{e.s.}}}{T_0}\chi\right), \tag{32}$$

where $\bar{v}$ is the mean relative velocity of the dust particles.

For the repulsion between the particles ($\chi < 0$), we write

$$k_{a_1a_2} = \left(\frac{\mu}{2\pi T_0}\right)^{3/2}\sigma_T\int_{v_{\min}}^\infty v\left(1 - \frac{2U_{\text{e.s.}}}{\mu v^2}|\chi|\right)$$
$$\times\exp\left(-\frac{\mu v^2}{2T_0}\right)4\pi v^2 dv, \tag{33}$$

where $v_{\min} = \sqrt{2U_{\text{e.s.}}|\chi|/\mu}$ is the minimum relative velocity of the dust grains at which a collision is possible. Performing integration as in the previous case yields

$$k_{a_1a_2} = 2\bar{v}\sigma_T\exp\left(-\frac{U_{\text{e.s.}}}{T_0}|\chi|\right). \tag{34}$$

Let us now analyze the results obtained and first consider the expression for χ. For our estimates, we take the neutral gas number density $n_0 = 3.25 \times 10^{15}$ cm^{-3}, which corresponds to the pressure $P = 13$ Pa, and the dust number density $n_d = 1.5 \times 10^9$ cm^{-3}; we estimate the cross section for the interaction of ions with neutral gas atoms from the following considerations. The main

ions in a mixture of Ar and SiH_4 are SiH_x^+ because of the lower ionization potential of silane and the nonresonant charge exchange of argon ions on SiH_x. SiH_x^+ interact with neutral gas particles through polarization, and the interaction cross section was estimated to be $\sigma_0 \approx 0.5 \times 10^{-14}$ cm^2. We use this estimate in our subsequent calculations. In this case, the relation

$$\frac{n_0\sigma_0}{n_d\pi\overline{a^2}} \gg 1$$

holds good up to root-mean-square macroparticle sizes $a \sim 0.8$ μm. We then obtain

$$\xi^2 \approx \frac{Z_i e|\varphi_0|}{T_i}\frac{n_d\pi\overline{a^2}}{n_0\sigma_0}. \tag{35}$$

Let us now consider the ratio $Z_i e|\varphi_0|/T_i$. The initial model that we constructed to describe the interaction between dust particles assumes a Maxwellian velocity distribution with an equilibrium temperature T_i for the ions. However, in the case discussed here, the dust particle sizes are much smaller than the characteristic ion mean free path in a plasma, $a \ll \lambda_i$. As a result, the ions between collisions with neutral gas atoms spend much of their time by moving in finite trajectories in the macroparticle field. Thus, the ion velocity distribution cannot be considered an equilibrium one, and the mean kinetic energy of the ions near the macroparticles is much higher than that in the surrounding gas. In our model, we take into account the latter circumstance by applying a correction for the ion temperature parameter and by assuming that, on average, the increase in the kinetic energy of the ions is proportional to the energy gained by them in the macroparticle field. This assumption is confirmed by our previous calculations using the method of molecular dynamics [47], although, undoubtedly, it is a forced assumption and stems from the fact that there is probably no exact analytical expression for the ion distribution function in the range of parameters under consideration. Thus, we assume that

$$T_i \approx \zeta Z_i e|\varphi_0|, \tag{36}$$

where ζ is a dimensionless parameter, $\zeta \lesssim 1$. Returning to expression (35) and taking the neutral gas temperature to be equal to the room temperature $T_0 = 0.025$ eV, we find that, in general, the inequality $\xi^2 \ll 1$ always holds in the range of particle sizes concerned, 10–100 nm, with number densities $n_d \sim 1 \times 10^9$ cm^{-3}.

For low values of ξ, the asymptotic behavior of the function $\Phi(\xi)$ is

$$\Phi(\xi) \sim \xi^2. \tag{37}$$

Substituting (37) into (29) and using (35) and (36), as well as the condition $n_0\sigma_0/n_d\pi\overline{a^2} \gg 1$, then yields

$$\chi = \frac{Z_i e\pi n_i}{n_0\sigma_0|\varphi_0|}\frac{a_1a_2(a_1^2+a_2^2)}{a_1^3+a_2^3}\frac{\exp(1/\zeta)}{\zeta} - 1. \tag{38}$$

Consider the dependence of χ on the sizes of the interacting particles. Let, for definiteness, a_1 and a_2 be the sizes of the smaller and larger particles, respectively. Denote their ratio by α: $\alpha = a_2/a_1 \geq 1$. The size-dependent part of expression (38) can then be rewritten as

$$\frac{a_1a_2(a_1^2+a_2^2)}{a_1^3+a_2^3} = a_1 f(\alpha) = a_1\frac{\alpha(1+\alpha^2)}{1+\alpha^3}. \tag{39}$$

The values of the function $f(\alpha)$ in the last expression lie within the range $1 \leq f(\alpha) \lesssim 1.1$. Thus, we find that the sign of the interaction between two particles of different sizes is determined almost completely by the size of the smaller particle.

Let us now consider the growth phase of particles that precedes the onset of their coagulation. As we showed in the previous section, the dust particle size distribution in this evolutionary stage of the system is in the form of a narrow peak and their shape is a regular sphere. The number density of the macroparticles in the volume is virtually constant, while their size uniformly increases through the gas-phase deposition of radicals.

Let us determine how the particle interaction cross section behaves as the particle size increases. To calculate the potential on the macroparticle surface, we use expression (21) and the quasineutrality relation:

$$\overline{Z}_d n_d + n_e = Z_i n_i, \tag{40}$$

where $\overline{Z}_d$ is the mean charge accumulated by the macroparticle. Since the particle size distribution is narrow, we may conclude that all of the particles carry almost identical charges equal to $\overline{Z}_d$. Denote also $P = n_e/n_i$. After all transformations, we then obtain

$$\chi\frac{U_{\text{e.s.}}}{T_0} = \left(\frac{Z_i n_i(1-P)}{n_d}\right)^2\frac{e^2/2a}{T_0}\left(\frac{a}{a_*} - 1\right), \tag{41}$$

where we also denoted

$$a_* \approx \sqrt{\frac{n_0\sigma_0(1-P)}{\pi n_d}}\sqrt{\zeta}\exp\left(-\frac{1}{2\zeta}\right). \tag{42}$$

When the dust content in the volume is large and the dust number density is comparable to the carrier number densities, which is almost always true for the range

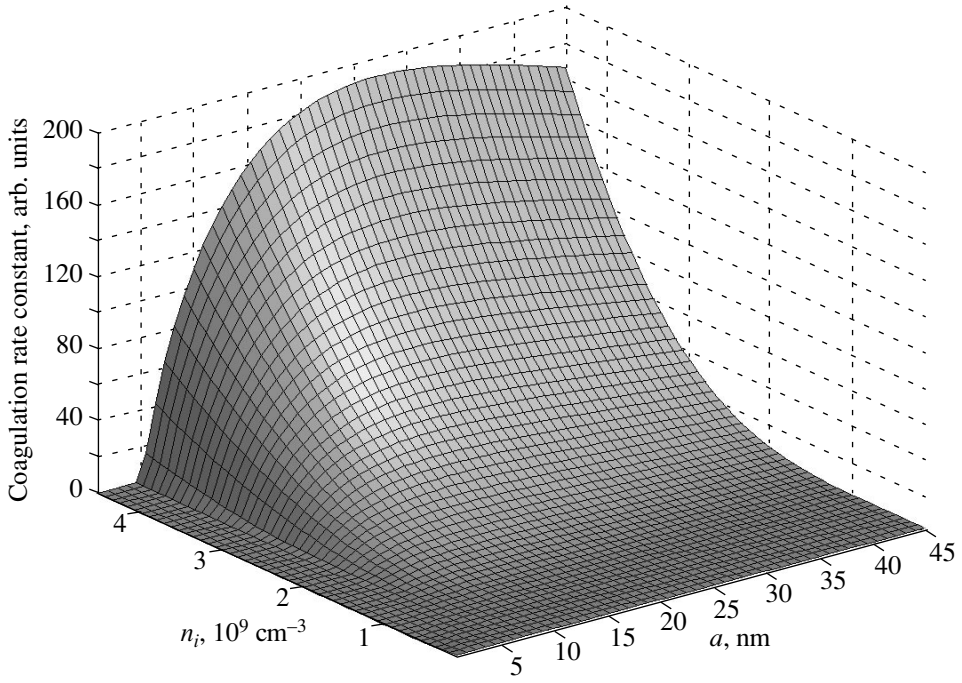


Fig. 1. Coagulation rate constant for spherical particles versus their radius a and ion number density n_i. The coagulation rate constant is given in units of the thermal collision rate constant $k_T = 2\bar{v}\sigma_T$, where $\bar{v}$ is the mean relative particle velocity, and $\sigma_T = 2\pi a^2$ is the thermal collision cross section.

of problems under consideration, the parameter $P \ll 1$ and it may be omitted in subsequent calculations.

Relation (41) is the main factor that governs the dust particle coagulation rate. If the dust particle radius is smaller than some critical size, which we denoted above by a_*, then χ will be negative, and, according to (34), the interaction rate constant will exponentially decrease with increasing $U_{\text{e.s.}}/T_0$. However, as soon as the particle size exceeds the critical value, the coagulation rate increases sharply, because the energy of the electrostatic interaction between dust grains significantly exceeds the energy of their thermal motion. In Fig. 1, the coagulation rate constant $k_{a_1 a_2}$ between dust particles divided by the rate constant of the collisions caused by the thermal motion of the dust grains alone is plotted against particle radius and ion number density. The particle radius is along one of the axes, and the ion number density is along the other axis. In constructing this plot, we assumed that the particle size distribution was monodisperse, which is characteristic of the initial coagulation phase. For our calculations, we took the above number densities of the macroparticles and the neutral gas atoms and assumed the parameter ζ to be equal to 0.13. In this case, we obtain an estimate of $a_* \approx 4.8$ nm for the critical macroparticle radius. We see from the figure that when the particles reach this size, the cross section for their interaction increases rapidly.

Thus, we may conclude that the transition from the initial growth of dust particles to their coagulation is a distinct threshold process, and this result is in close agreement with a large number of experimentally obtained descriptions of the process.

We would also like to note yet another feature of our model. As many investigators have pointed out, the coagulation process takes place under conditions when the dust number density is either higher than or approximately equal to the ion number density. In this case, the accumulation of large negative charges by dust particles, which, in turn, leads to a strengthening of the electrostatic repulsion between them, has so far been assumed to be mainly responsible for the termination of the coagulation. However, in this approach, explaining the high growth rates of the particles as they coagulate involves serious difficulties. Meanwhile, in our model, the role of the electrostatic interaction between macroparticles proves to be opposite. This is because the forces exerted on the particles from the ion flows, as well as the forces of electrostatic repulsion, are proportional to the electric field strength. We clearly see from Fig. 1 that, as the ion number density increases, which is equivalent to an increase in the charge accumulated by macroparticles, the coagulation rate constant rapidly increases. This conclusion is confirmed experimentally: the coagulation rate of dust particles increases with discharge power, which, in turn, governs the ionization rate in a plasma [29].

In the next section, we investigate the growth dynamics of dust particles in the coagulation phase by using the interaction rate constants derived above and discuss the causes of the termination of coagulation at some stage.

4. THE DYNAMICS OF A DUST SYSTEM IN THE COAGULATION PHASE

Thus, we have found that the onset of coagulation between dust particles is a threshold process, and the coagulation sets in when they reach a critical size. To describe the coagulation process, we introduce, as above, a particle volume distribution function, $f(V, t)$. The coagulation equation with the solution $f(V, t)$ is generally written as

$$\frac{\partial f(V,t)}{\partial t} = -f(V,t)\int_0^\infty k_{VV'}(t) f(V',t)dV' + \frac{1}{2}\int_0^V k_{V'V-V'}(t) f(V',t) f(V-V',t)dV', \quad (43)$$

where the first term on the right-hand side reflects the departure of particles from a given size range and the second term corresponds to their arrival. The coagulation rate constant $k(t)$ in the integrand is given by relations (32) and (34), in which the initial parameters are defined by expressions (26), (29), and (36). This constant is a function of time due to the time dependence of the integrated quantities that characterize the properties of the dust system. These include the potential on the macroparticle surface φ_0 and the total macroparticle surface area $n_d\pi\overline{a^2}$, which is a measure of the intensity of interaction between the dust and the ion subsystem. In general, the ion number density in real conditions can also change appreciably as the dust clusters grow and can, in turn, affect the dynamics of the process under consideration. However, in constructing our model, we assume it to be constant, because otherwise we would have to solve the full system of plasma–chemical equations, which would require knowledge of many additional details and specific experimental conditions.

Equation (43), which describes the coagulation dynamics in integral form, assumes that the dust particle size distribution function is continuous, or, more specifically, that the minimum possible change in particle size during a collision is much smaller than the characteristic sizes of the particles themselves. Since the initial particle distribution in our problem is very narrow, the step of change in their size during collisions is comparable to the size itself; as a result, the above approach is inapplicable. To properly describe the process, we will then need to rewrite Eq. (43) in discrete form.

During coagulation, particles with an approximately equal volume $V_0 = (4/3)\pi a_0^3$, where $a_0 \sim a_*$, are the building material for clustering. Therefore, instead of the continuous distribution function $f(V, t)$, we may introduce a discrete sequence of the number densities of clusters with different sizes, f_n, where $n \geq 1$ is the number of initial particles involved in clustering. The discrete analog of Eq. (43) can then be written as

$$\frac{df_n(t)}{dt} = -f_n(t)\sum_{i=1}^{\infty} k_{in}(t) f_i(t) + \frac{1}{2}\sum_{i=1}^{i<n} k_{in-i}(t) f_i(t) f_{n-i}(t). \quad (44)$$

As the initial condition, we assume that all particles had the same size and that their number density was n_{d0} at the start of the coagulation process:

$$f_n(t=0) = \begin{cases} n_{d0}, & n = 1, \\ 0, & n > 1. \end{cases} \quad (45)$$

The formulas for the coagulation rate constant derived in the previous section describe the interaction between two spherical particles of arbitrary sizes. However, in calculating the interaction between the clusters that are formed when macroparticle coalesce, these formulas become inapplicable in the form in which they were originally written and need to be generalized.

Let the cluster consist of n spherical microparticles with volume V_0. As practice shows, the surface topology of this structure can be completely arbitrary—both nearly spherical and strongly fractal, consisting of chains of coalesced dust microparticles. Let us write out the basic cluster parameters in the limiting situations. In both cases, the mass of the cluster is defined as the sum of the masses of its constituent particles, $m_n = nm_0$. For a spherical topology, dense packing, and large n, the cluster surface area is $S_n \approx n^{2/3}S_0$ and the cluster size is defined as $a_n \approx n^{1/3}a_0$, where S_0 and a_0 are the surface area and size of the initial crystallites, respectively. An expression for $k_{n_1n_2}$ can then be derived by substituting $a_1 = n_1^{1/3}a_0$ and $a_2 = n_2^{1/3}a_0$ into the above formulas.

The case of a fractal topology is more complicated and should be considered separately. As the model of such a macroparticle, we will consider a combination of n_c chains of initial particles coalescing at one point that is the cluster center. The total number of particles in the cluster is equal to n as before. The surface area of the cluster in this case is close to the total area of its constituent particles, $S_n \approx nS_0$. The additional conclu-

sion that the integrated parameter $n_d\pi\overline{a^2}$ remains constant during the entire process when modeling the growth of fractal structures also follows from this assertion. The question as to what should be meant by the size of a fractal particle is much more difficult to answer unambiguously. Nevertheless, this parameter is of great importance in performing calculations, because the electric capacitance of the dust cluster is related to it. In general, the representation of a fractal structure as a point charge when modeling the electrostatic interactions is a rough approximation. However, for lack of a better approximation, in our model we will consider clusters in the shape of a sphere with an effective radius a_{eff}, which is chosen in such a way that the area of the corresponding spherical surface is equal to the real surface area of the cluster: $a_{\text{eff}} = n^{1/2}a_0$.

The next feature of fractal structures that should be taken into account when modeling the interaction is the sharp reduction in the efficiency of attraction between them compared to spherical particles of the same size. This is because the dragging force from the ion flows that we calculated in the spherical geometry is proportional to the particle surface area. For fractal particles, however, provided that they may be considered to be conductive, their accumulated charges will concentrate at the ends of the chains. Accordingly, ion flows will have a directed action only on a limited region of the cluster surface. If the cluster consists of n_c chains, with the total number of constituent particles being n, then its geometrical size can be estimated as $a \approx n/n_c a_0$. The surface area of the sphere that corresponds to this size is $S = 4\pi(n/n_c)^2 a_0^2$, while the effective surface area that contributes to the interaction is much smaller: $S_{\text{eff}} \approx 4\pi a_0^2 n_c$. In this case, the efficiency of the attractive force exerted on the cluster changes by a factor of γ:

$$\gamma = \frac{S_{\text{eff}}}{S} = \frac{n_c^3}{n^2}. \quad (46)$$

For large fractal structures, the number of chains is much smaller than the number of their constituent microparticles. As a result, $\gamma \ll 1$ and the efficiency of the attraction between large particles sharply decreases. In this case, the cross section for the interaction between them is determined by electrostatic repulsion. Thus, we conclude that fractal clusters can grow only via the attachment of the smallest particles, i.e., those that serve as the initial building material in the coagulation phase, to them. Given that the force of microparticle attraction to the cluster is proportional to the electric field strength and provided that the charges in the cluster concentrate at the ends of the chains, we find that these ends will be the most likely locations of the attachment of new particles. This is the explanation for the typical formation pattern of such fractal structures, more specifically, the fact that the number of chains remains virtually constant during the growth of a fractal cluster, while their length increases rapidly.

Thus, the properties of the microparticle material from which clusters are formed play a key role in determining the macroparticle formation mechanism in the coagulation phase; the higher the conductivity of the material, the higher the fractal dimension of the growing structures. In general, this tendency was observed by the authors of several experimental works [24–26] who studied the growth of dust structures with various substances served as their material.

When the charges on the macroparticle surface are distributed uniformly and are not accumulated on a small area, which is characteristic of nonconducting materials, new particles are attached to the cluster at an arbitrary location. In this case, the number of chains steadily increases, and in the limit of a spherical topology, this number has the same order of magnitude as the number of microcrystallites on the cluster surface, $n_c \sim n^{2/3}$. Then, as would be expected, $\gamma \approx 1$.

In this study, we numerically simulate the coagulation process in which fractal structures grow, and we assume in our subsequent calculations that the number of chains in the cluster increases as n^κ, where $\kappa = 2/3$ for $n \le 6$ and smoothly tends to zero for large n. Thus, we take into account the fact that the efficiency of the attractive forces exerted on small clusters is comparable to the efficiency of the interaction between spherical particles.

Given the aforesaid, the following substitution should be made in expression (38) for χ for fractal particles to modify it:

$$\frac{a_1 a_2 (a_1^2 + a_2^2)}{a_1^3 + a_2^3} \longrightarrow a_0 \frac{n_{c1} n_1^{-1/2} n_2 + n_{c2} n_2^{-1/2} n_1}{n_1 + n_2}, \quad (47)$$

where a_0 is the radius of the initial microparticles from which clusters grow; n_1 and n_2 are the numbers of crystallites that constitute the interacting macroparticles; and n_{c1} and n_{c2} are the numbers of chains in the first and the second clusters. We should also make changes to the expression for $U_{\text{e.s.}}$:

$$U_{\text{e.s.}}(n_1 + n_2) = \varphi_0^2 a_0 \frac{n_1^{1/2} n_2^{1/2}}{n_1^{1/2} + n_2^{1/2}}. \quad (48)$$

The coagulation rate constant $k_{a_1 a_2}$ is plotted against cluster size at the initial time in Fig. 2. For our calculations, we took the initial particle size to be $a_0 = 7.5$ nm, which slightly exceeds the threshold value, and the dust particle number density to be $n_i \approx 1.5 \times 10^9$ cm^{-3}. We see from the plot that the most likely collisions are those between the particles with the largest size difference. As a result, the dust cloud separates into two groups of particles: the first group consists of small

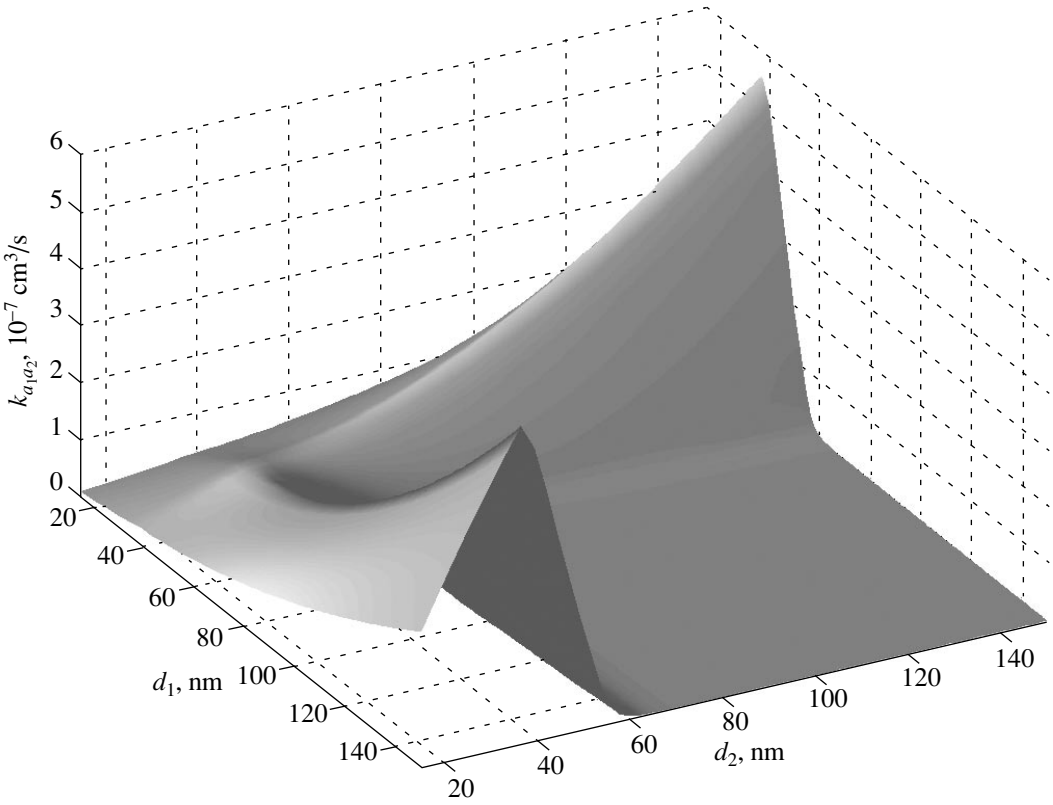


Fig. 2. Coagulation rate constant $k_{a_1a_2}$ for fractal particles versus their sizes d_1 and d_2. By the particle size, we mean twice its effective radius, $d = 2a_{\text{eff}}$. The plot corresponds to the start of coagulation: $t = 0$.

crystallites with a roughly regular shape, while the second group consists of branched structures growing through the attachment of particles from the first group. This characteristic feature of the distribution function was observed in many experiments [9–21]. Here, it should be noted that, the higher the selectivity of the interaction rate constant to the selection of particles with differing sizes, the narrower the particle size dispersion in both groups. The latter assertion is general and should be taken into account when constructing models of the interaction in dust clouds different from our model.

Next, we numerically solved Eq. (44) by using a simple explicit scheme:

$$f_n(t+\tau) = \frac{f_n(t) + \frac{1}{2}\tau\sum_{i=1}^{i<n} k_{in-i}(t) f_i(t) f_{n-i}(t)}{1 + \tau\sum_{i=1}^{\infty} k_{in}(t) f_i(t)} \quad (49)$$

with an adaptive time step τ. All of the mean quantities and the interaction rate constants that depend on them were recalculated at each time step. For fractal particles, the main quantity that governs the cluster growth dynamics is the potential on the macroparticle surface. Assuming that this potential is the same for all particles and taking into account the quasineutrality relation and the smallness of P, we can express it in terms of the dust number density and the mean effective particle radius:

$$e|\varphi_0| = \frac{n_i e^2}{n_d \bar{a}_{\text{eff}}}. \quad (50)$$

Let us now discuss the results obtained. The coagulation dynamics is shown in Figs. 4–7. First, note that the saturation phase begins in the course of time and the process virtually terminates. This closely corresponds to the experimentally observed behavior of the dust system in the coagulation phase, but the mechanism of this phenomenon requires a special analysis.

In Fig. 3, the coagulation rate constant is plotted against cluster size at the time the computation finishes, which corresponds to 2.3 s after the start of the process. A comparison of this plot with the analogous plot at the initial time (Fig. 2) indicates that the pattern of particle interaction changes radically in the time during which clusters rapidly grow. The dependence of the interaction rate constant on such integrated parameters as the macroparticle potential and number density and the total surface area is responsible for this change.

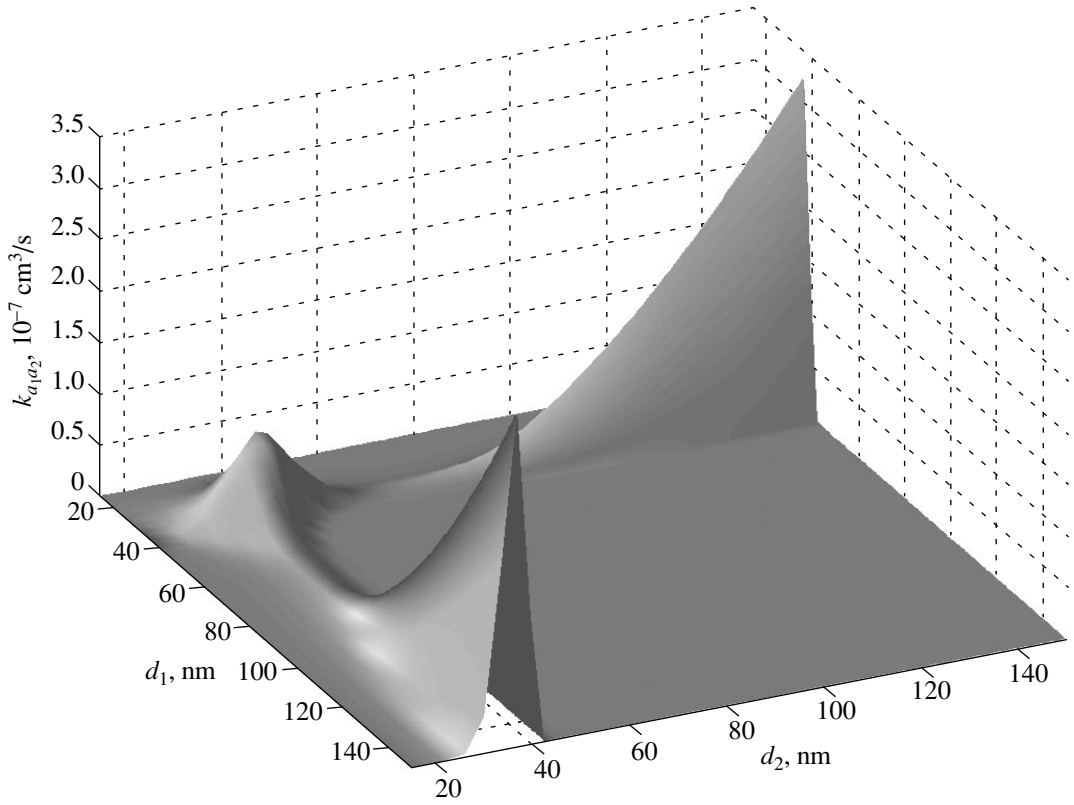


Fig. 3. Coagulation rate constant $k_{a_1a_2}$ for fractal particles versus their sizes d_1 and d_2. The plot was constructed for t = 2.3 s after the onset of coagulation and corresponds to the saturation phase.

As we noted above, when the interaction between fractal particles is considered, their potential is the main parameter that affects the pattern of particle interaction. Its change with time is shown in Fig. 4. We see from the figure that this parameter monotonically increases during the entire experiment, although its absolute change is moderately large, a factor of 2.5 compared to its initial value. Nevertheless, this change has a crucial effect on the subsequent development of the entire process. The point is that the onset of coagulation itself is possible only after the initial microparticles reach some threshold size. Subsequently, they serve as the source of material for the growth of clusters, but their sizes still remain in the threshold range. The latter, in turn, is very sensitive to the integrated parameters of the system and can shift in one or another direction as they change. As the particle charges grow, the behavior of the interaction rate constant in the threshold range also changes in pattern. It becomes sharper, causing the sensitivity of the process to a threshold shift to increase further.

As the macroparticle potential increases, the coagulation transition region shifts toward larger sizes. As a result, the main reservoir of the crystallites involved in clustering proves to be in the range of parameters for which collisions between particles become unlikely. This is clearly seen from Fig. 3 and is consistent with the above estimate of the critical crystallite size (42), which suggests that the critical size must increase with decreasing dust number density.

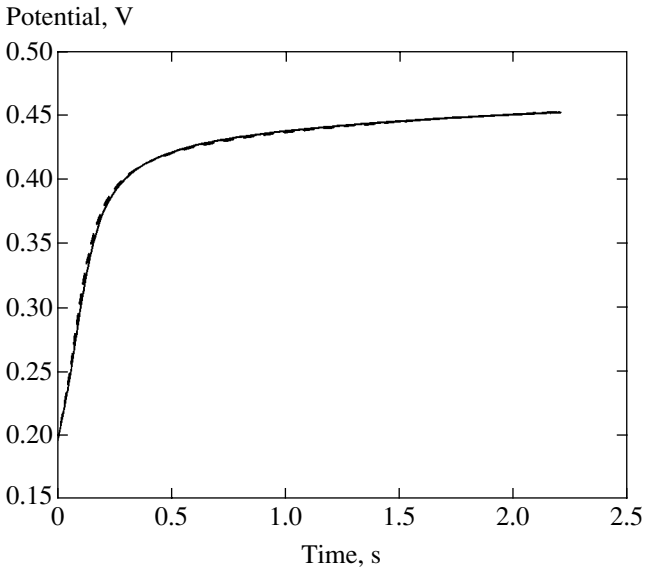


Fig. 4. Potential on the macroparticle surface φ_0 versus time. The solid curve corresponds to the direct calculation, and the dashed curve was obtained by using relation (52).

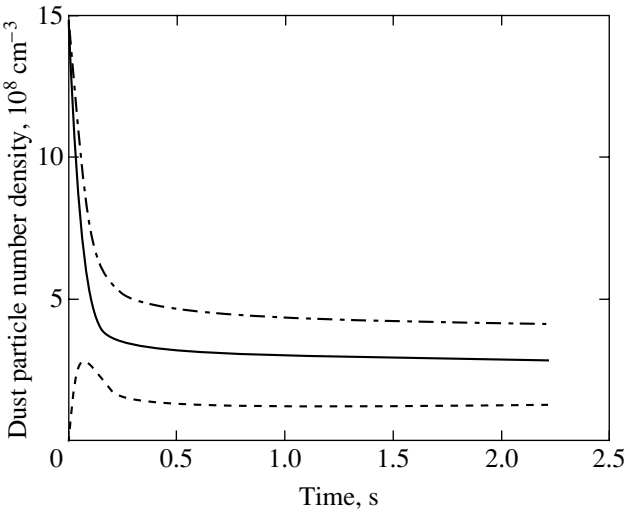


Fig. 5. Dust particle number density versus time. The solid curve represents the evolution of the number density of the initial crystallites that serve as the building material in clustering; the dashed curve indicates the change in cluster number density with time; and the dash–dotted curve indicates the change in total dust number density in the volume, including the macroparticles of both groups.

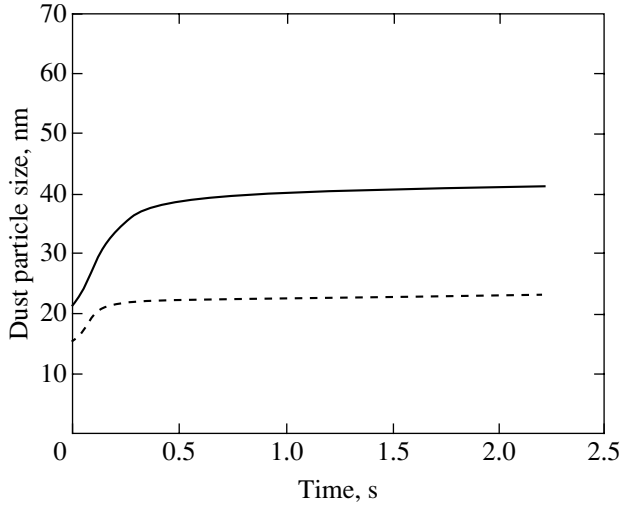


Fig. 6. Mean macroparticle size versus time. By the particle size, we mean twice its effective radius, $d = 2a_{\text{eff}}$. The solid line corresponds to the size of the particles from the larger group formed during the coalescence of initial crystallites; the dashed line corresponds to the mean particle size in the system as a whole.

Let us now trace the dynamics of the process. We will consider two groups of particles: the first group consists of initial spherical particles with size a_0 and serves as the major source for the growth of clusters, and the second group includes clusters composed of two or more crystallites. The dust particle number density is plotted against time in Fig. 5. The solid and dashed curves correspond to the first and the second groups, respectively. Also shown here is the change in total dust number density (dash–dotted line). Within the first 0.2 s after the start of coagulation, the number density of the initial particles rapidly decreases by about an order of magnitude, while the number density of the forming clusters increases synchronously. Subsequently, there comes a time when the initial particle group under the changed conditions ceases to be involved in the growth of dust structures, and the decrease in their number density slows down sharply. At the same time, the cluster number density reaches its maximum and begins to decrease. This is because new material ceases to be supplied for the building and the particles in the second group grow further via collisions between them; this growth gradually slows down as the coagulation threshold shifts and as the number of small clusters decreases. As a result, the saturation phase accompanied by the attainment of steady-state values by all system parameters starts about 0.5 s after the onset of coagulation.

Figure 6 shows the dynamics of the change in mean particle size. By the particle size, we mean their diameter, because most of the experimental works deal precisely with this parameter. The calculated values of $\bar{a}_{\text{eff}}$ served as the input data for constructing the plots. The solid line corresponds to the mean cluster size, while the dashed line corresponds to the mean particle size in the system as a whole; the latter size, in particular, takes into account the presence of seed microparticles in the volume. We see from the figure that the mean cluster size is much larger than the sizes of the initial microparticles, while the total mean particle size in the system is much smaller because of the high percentage of the crystallites from the first group remaining in the volume.

The detailed dynamics of the change in the macroparticle size distribution function or, more precisely, in the partial cluster number densities f_n is shown in Fig. 7 for various times. The particle diameter (in nm) is along the x axis, the time is along the y axis, and the decimal logarithm of the number density is along the z axis. We clearly see from the figure that two distinct groups with narrow dispersion appear in the particle size distribution in the course of time. The time of distinct separation of the larger group comes shortly after the departure of initial microparticles into the range of parameters inaccessible to interaction about 0.3 s after the start of the process. This is because the size of macroparticles increases further via small clusters composed of several crystallites.

Apart from the dynamics of the change in system parameters with time, the relations between the integrated quantities themselves, in particular, the relation between the dust number density n_d, the mean particle size $\bar{a}_{\text{eff}}$, and the potential on the particle surface, are of great interest. For the particle sizes and number densi-

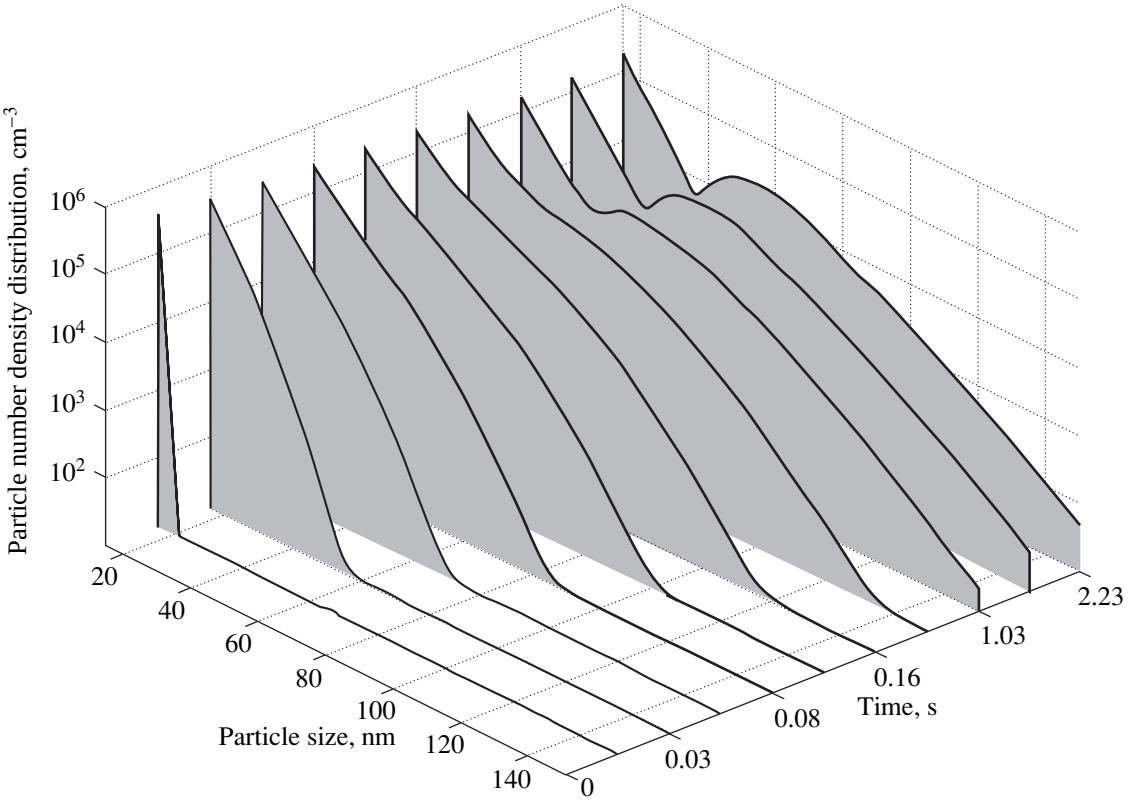


Fig. 7. Time evolution of the macroparticle size distribution function.

ties taken at arbitrary times, the following relation holds with a high accuracy:

$$n_d(t)\bar{a}_{\text{eff}}^3(t) = \text{const}. \quad (51)$$

This is illustrated by Fig. 8. Accordingly, given (50), we find that the potential is related to the particle size by

$$\varphi_0(\bar{a}_{\text{eff}}) \approx \varphi_0(a_0)\left(\frac{\bar{a}_{\text{eff}}}{a_0}\right)^2. \quad (52)$$

The accuracy with which this relation holds can be estimated from Fig. 4, which shows two curves for the potential. The solid curve corresponds to the direct calculation, while the dashed curve was constructed by using (52).

The result obtained is somewhat unexpected. The point is that expression (51) resembles the mass conservation condition for spherical particles of equal sizes with radius a_{eff}. At the same time, it is not related in any way to the conservation of material, because the mass conservation condition for fractal clusters is completely different:

$$n_d(t)\bar{n}(t) = n_{d0}. \quad (53)$$

Here, we denoted the dust number density at the onset of coagulation by n_{d0} and the mean number of crystallites in the clusters by $\bar{n}$. During our calculations, this relation served as the stability criterion for the numerical scheme and held with an accuracy as high as 0.1%. The result is also unusual in that the criterion for choos-

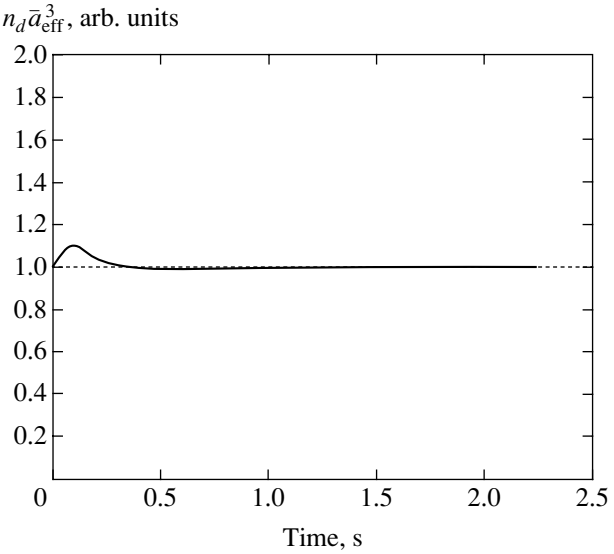


Fig. 8. Relationship between the dust number density and the mean particle size for an arbitrary time. The ratio of $n_d\bar{a}_{\text{eff}}^3$ to its value taken at $t = 0$ is plotted along the vertical axis.

Comparison of the numerical simulations with experimental data [13]

Parameter	Experiment	Numerical calculation
Rate of increase in microparticle radius during initial growth, c_f	15 nm s^{-1}	15 nm s^{-1}
Crystallite radius corresponding to coagulation threshold, a^*	7.5–10 nm	$\gtrsim$4.8 nm
Dispersion of particle size distribution during initial growth, Δa	–	~0.02 nm
Separation time of two particle groups in coagulation phase (from start)	0.5 s	0.3 s
Time at which coagulation is saturated (from start)	0.7 s	0.5 s
Number density of particles in first group in saturation phase	3×10^8 cm^{-3}	2.9×10^8 cm^{-3}
Number density of particles in second group in saturation phase	4×10^7 cm^{-3}	1.2×10^8 cm^{-3}
Cluster size in saturation phase	75 nm	40 nm

ing $\bar{a}_{\text{eff}}$ that we used here was largely arbitrary, and we did not make any special estimates on this score. Thus, the high accuracy with which relations (51) and (52) hold when numerically simulating the dynamics of the process and the fact that an interaction model can generally be chosen in many ways suggest that our relations are universal, although the reason why they are valid is not quite clear.

DISCUSSION

To summarize, let us formulate our main results and make several additional remarks regarding their possible applications and the ways of furthering the studies in this field.

We constructed a theoretical model that includes a description of all the main stages of the growth of dust structures in a low-temperature plasma, from the initial growth of particles and their coagulation to the saturation phase. We also completely traced the evolution of the particle size distribution function over the entire process, and the main features of its change with time closely corresponded to the actually observed pattern of the phenomenon. These include the high monodispersity of the particle size distribution in the phase of initial growth and its separation into two narrow groups of particles in the coagulation phase. Based on the constructed model, we explained the threshold behavior of the coagulation process and estimated the critical microparticle size upon reaching which a transition from the growth of particles via the gas-phase deposition of material to their coagulation becomes possible.

Using the analytically derived expression for the coagulation rate constant, we numerically simulated the clustering process based on data from real experiments. The results of these experiments were briefly outlined in the Introduction, and they can now be directly compared with the results of our numerical simulations. The main parameters to be compared include the temporal characteristics of the process that describe the duration of its various phases; the sizes of the forming particles; and the particle number densities in the saturation phase. The results of our comparison are given in the table.

Looking at these data, we may conclude that, in general, our model yields parameters close to their experimental values. There is a difference only in the number density and sizes of the forming clusters. However, this difference is attributable to the peculiarities of separating the groups in [13] and in our study. In our study, the first group includes only initial crystallites, while structures composed of two or three microparticles contribute noticeably to the number density of the clusters included in the second group. These agglomerates are also responsible for the slightly smaller mean size of the particles in the second group. We see from Fig. 7 that the peak in the size distribution of clusters in the second group lies well above the value given in the table and corresponds to approximately 60 nm. In [13], the separation of particles into groups was not so distinct; in addition, a different method of estimating the cluster size based on calculations of the area of their projection onto the plane was used by the authors of [13]. The numerically calculated and experimentally obtained temporal characteristics of the particle growth proved to be similar. However, it should be noted that the ion number density strongly affects them. In our calculations, we assumed it to be equal to the initial dust particle number density, because it is mentioned in the experimental work that these number densities have similar values, but no accurate n_i is given there. Our additional calculations indicate that increasing the ion number density causes an appreciable acceleration of the process, although the characteristic cluster number densities and sizes in the saturation phase do not change much.

The above comparison indicates that our model for the growth of dust particles may well serve for estimations when describing the actually observed processes. However, it should be noted that our theory have several points that need to be seriously improved in the future. The first consists in an arbitrary definition of the dependence of the number of chains in a cluster on the total number of its constituent crystallites. The choice

of this dependence mainly affects the final particle size in the saturation phase and the dispersion of their size distribution, but the temporal characteristics of this process do not change appreciably.

The main difficulty in studying the model under discussion is that there are currently no well-developed analytical methods for calculating the distribution function of ions and electrons near nanometer-size dust particles. To perform our calculations, we had to use an equilibrium ion distribution function in which T_i, the ion temperature near a microparticle, took into account the heating of the ion subsystem in the dust grain field and was taken in the form $T_i \approx \zeta Z_i e|\varphi_0|$ as the basis. This approximation was taken on the basis of our previous numerical calculations of the ion distribution near a macroparticle [47]. The choice of ζ for simulating the coagulation process in terms of our model proved to be of crucial importance, because the critical crystallite size after reaching which the coagulation process begins strongly depends on this parameter. Thus, we conclude that knowledge of the specific ion distribution function in the immediate vicinity of dust clusters is the key to describing their growth, and particular attention should be given to this question in further studies.

In general, our main results pertain to the simulation of the coagulation and growth of dust structures with a fractal topology. However, the growth of spherical clusters is also a widespread phenomenon and is commonly observed experimentally [22, 30]. As has been noted above, the material from which a cluster is formed must play a major role in determining the shapes of the growing structures and a fractal topology of the cluster surface in our model corresponds to conducting materials. Since the expression for the forces exerted on macroparticles from ion flows was initially derived for this case, the pattern of interaction between nonconducting particles should be discussed separately.

In the latter case, the polarization of particles in the electric field of its neighbors is low, and the associated additional ion flows cannot compensate for the forces of electrostatic repulsion between dust grains. Nevertheless, the ion flows can also become anisotropic here. The point is that when the dust number density is of the same order of magnitude as the ion number density or larger than the latter, the mean charges accumulated by the macroparticles do not exceed several electron charges. In this case, a sharp anisotropy is observed in the charge distribution along the cluster surface; as a result, the charge acquires a large dipole moment. The electric field of neighboring particles must lead to a turn of the cluster in such a way that its dipole moment is directed along the external field. We then arrive at the situation that existed for conducting particles with the only difference that the cluster dipole moment there was produced by a redistribution of moving charges along its surface. As in the case of conducting dust grains, we obtain a redistribution of ion flows on the cluster surface, and this will produce an effective force of attraction between particles under certain conditions.

An important feature of this description is the dependence of the pattern of interaction between microparticles on their size. First, increasing the size of the cluster causes its mass to increase, which will eventually make the time of its turn in an external field longer than the time of the change in its dipole moment caused by the absorption of ions and electrons from a plasma. As a result, the directed action of ion flows will be lost. Second, the charge of the cluster will increase with its size, and the charge distribution along the surface will become more uniform. In this case, the dipole moment will decrease, as will the anisotropy of the ion flows on the particle surface. Thus, we may conclude that the forces of attraction between nonconducting particles will act only on small crystallites; as a result, as in the case of a fractal topology, dust structures will grow via particles of smaller diameters, with the only difference that it will now be uniform, because the charge on the surfaces of nonconducting macroparticles is distributed uniformly. As in the case of fractal cluster growth considered above, the selectivity of the interaction rate constant with respect to the particle size will give rise to two groups of particles: the first will serve as the source of material, and the second will uniformly increase in size through particles from the initial group.

ACKNOWLEDGMENTS

We thank A.M. Dykhne, who took an active part in the discussion of questions related to the range of problems under consideration and whose comments played an invaluable role in preparing this paper. This work was supported in part by the Russian Foundation for Basic Research (project nos. 01-02-17726 and 00-15-96554).

REFERENCES

1. G. Selwyn, J. Singh, and R. Bennet, J. Vac. Sci. Technol. A **7**, 2758 (1989).
2. M. P. Garrity, T. W. Peterson, and J. F. O'Hanlon, J. Vac. Sci. Technol. A **14**, 550 (1996).
3. L. Boufendi, J. Hermann, A. Bouchoule, *et al.*, J. Appl. Phys. **76**, 148 (1994).
4. K. Spears, T. Robinson, and R. Roth, IEEE Trans. Plasma Sci. **14**, 179 (1986).
5. G. Selwyn, J. McKillop, K. Haller, and J. Wu, J. Vac. Sci. Technol. A **8**, 1726 (1990).
6. G. Jellum and D. Graves, J. Appl. Phys. **67**, 6490 (1990).
7. R. Buss and J. Babu, J. Vac. Sci. Technol. A **14**, 577 (1996).
8. H. Hahn and R. Averback, J. Appl. Phys. **67**, 1113 (1990).
9. T. Fukuzava, H. Kawasaki, M. Shiratani, and Y. Watanabe, Jpn. J. Appl. Phys., Part 1 **33** (7B), 4212 (1994).

10. H. Kawasaki, T. Fukuzava, H. Tsuruoka, *et al.*, Jpn. J. Appl. Phys., Part 1 **33** (7B), 4198 (1994).
11. W. Böhme, W. E. Köhler, M. Römheld, *et al.*, IEEE Trans. Plasma Sci. **22**, 110 (1994).
12. M. Shiratani, T. Fukuzawa, and Y. Watanabe, IEEE Trans. Plasma Sci. **22**, 103 (1994).
13. Y. Watanabe, M. Shiratani, H. Kawasaki, *et al.*, J. Vac. Sci. Technol. A **14**, 540 (1996).
14. Y. Watanabe and M. Shiratani, Jpn. J. Appl. Phys., Part 1 **32** (6B), 3074 (1993).
15. A. Bouchoule, L. Boufendi, J. Hermann, *et al.*, Pure Appl. Chem. **68**, 1121 (1996).
16. H. Kawasaki, K. Sakamoto, S. Maeda, *et al.*, Jpn. J. Appl. Phys., Part 1 **37** (10), 5757 (1998).
17. T. Fukuzava, S. Kushima, Y. Matsuoka, *et al.*, J. Appl. Phys. **86**, 3543 (1999).
18. K. Koga, Y. Matsuoka, K. Tanaka, *et al.*, Appl. Phys. Lett. **77**, 196 (2000).
19. M. Shiratani, T. Fukuzawa, and Y. Watanabe, Jpn. J. Appl. Phys., Part 1 **38** (7B), 4542 (1999).
20. H. Kawasaki, K. Sakamoto, and S. Maeda, Jpn. J. Appl. Phys., Part 2 **37** (10B), L1264 (1998).
21. H. Kawasaki, J. Kida, K. Sakamoto, *et al.*, J. Appl. Phys. **83**, 5665 (1998).
22. Y. Hayashi and K. Tachibana, Jpn. J. Appl. Phys., Part 1 **33** (7B), 4208 (1994).
23. Y. Hayashi and K. Tachibana, Jpn. J. Appl. Phys., Part 2 **33** (6A), L804 (1994).
24. G. Praburam and J. Goree, Astrophys. J. **441**, 830 (1995).
25. D. Samsonov and J. Goree, J. Vac. Sci. Technol. A **17**, 2835 (1999).
26. D. Samsonov and J. Goree, Phys. Rev. E **59**, 1047 (1999).
27. P. Haaland, S. Ibrani, and H. Jiang, Appl. Phys. Lett. **64**, 1629 (1994).
28. Y. Matsuoka, M. Shiratani, T. Fukuzava, *et al.*, Jpn. J. Appl. Phys., Part 1 **38** (7B), 4556 (1999).
29. M. Shiratani, S. Maeda, K. Koga, and Y. Watanabe, Jpn. J. Appl. Phys., Part 1 **39** (1), 287 (2000).
30. A. Garscadden, Pure Appl. Chem. **66**, 1319 (1994).
31. M. T. Swihart, S. Nijhawan, M. R. Mahajan, *et al.*, J. Aerosol Sci. **29**, S79 (1998).
32. U. R. Kortshagen, U. V. Bhandarkar, M. T. Swihart, and M. T. Girshick, Pure Appl. Chem. **71**, 1871 (1999).
33. M. T. Swihart and S. L. Girshick, J. Phys. Chem. B **103**, 64 (1999).
34. U. V. Bhandarkar, M. T. Swihart, S. L. Girshick, and U. R. Kortshagen, J. Phys. D: Appl. Phys. **33**, 2731 (2000).
35. Kio-Seon Kim and M. Ikegawa, Plasma Sources Sci. Technol. **5**, 311 (1996).
36. P. Haaland, A. Garscadden, and B. Ganguly, Appl. Phys. Lett. **69**, 904 (1996).
37. S. J. Choi and M. J. Kushner, J. Appl. Phys. **74**, 853 (1993).
38. V. N. Tsytovich, Usp. Fiz. Nauk **167**, 57 (1997) [Phys. Usp. **40**, 53 (1997)].
39. V. A. Schweigert and I. V. Schweigert, J. Phys. D: Appl. Phys. **29**, 655 (1996).
40. H. U. Keller, J. Blum, B. Donn, *et al.*, Adv. Space Res. **13** (7), 73 (1993).
41. F. Y. Huang, H. H. Hwang, and M. J. Kushner, J. Vac. Sci. Technol. A **14**, 562 (1996).
42. F. Y. Huang and M. J. Kushner, J. Appl. Phys. **81**, 5960 (1997).
43. K. Watanabe, K. Nishimura, and T. Sato, in *Advances in Dusty Plasmas*, Ed. by P. K. Shukla, D. A. Mendis, and T. Desai (World Sci., Singapore, 1998), p. 394.
44. Yu. A. Mankelevich, M. A. Olevanov, and T. V. Rakhimova, Zh. Éksp. Teor. Fiz. **121**, 1288 (2002) [JETP **94**, 1106 (2002)].
45. R. A. Quinn and J. Goree, Phys. Rev. E **61**, 3033 (2000).
46. R. A. Quinn and J. Goree, Phys. Plasmas **7**, 3904 (2002).
47. M. A. Olevanov, Yu. A. Mankelevich, and T. V. Rakhimova, Zh. Éksp. Teor. Fiz. **123**, 503 (2003) [JETP **96**, 444 (2003)].

Translated by V. Astakhov

Journal of Experimental and Theoretical Physics, Vol. 98, No. 2, 2004, pp. 305–315.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 2, 2004, pp. 345–355.

PLASMA, GASES

The Permittivity of a Monatomic Gas with Spatial Dispersion

V. P. Makarov and A. A. Rukhadze

Prokhorov Institute of General Physics, Russian Academy of Sciences, ul. Vavilova 38, Moscow, 119991 Russia
e-mail: asam@ran.gpi.ru

Received July 2, 2003

Abstract—The longitudinal, $\varepsilon^{l}(\omega, k)$, and transverse, $\varepsilon^{\mathrm{tr}}(\omega, k)$, permittivities of a monatomic gas were calculated. The frequency ranges in which the permittivity $\varepsilon(\omega)$ and permeability $\mu(\omega)$ of a gas without spatial dispersion have a physical meaning were determined. The limiting magnetic susceptibility $\chi(\omega)$ at $\omega = 0$ and the static magnetic susceptibility were found. The question of whether an electromagnetic wave with antiparallel group and phase velocities can propagate through a monatomic gas is discussed. *© 2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Determining the permittivity tensor $\varepsilon_{ij}(\omega, \mathbf{k})$ [1–3] reduces to calculating the density of the electric current induced by an electromagnetic field $\mathbf{E}(\mathbf{r}, t)$, $\mathbf{B}(\mathbf{r}, t)$ in a medium:

$$\mathbf{j}(\mathbf{r}, t) = \mathrm{Sp}(\hat{W}\hat{\mathbf{j}}(\mathbf{r}, t))/\mathrm{Sp}\,\hat{W}. \tag{1.1}$$

Here, $\hat{\mathbf{j}}(\mathbf{r}, t)$ is the electric current density operator, and $\hat{W}$ is the density matrix that satisfies the equation [4, Section 6]

$$\hat{W} = -i[\hat{H}, \hat{W}], \tag{1.2}$$

where $\hat{H}$ is the total Hamiltonian[1] of the "medium + electromagnetic field" system, including the interaction energy $\hat{U}$ between them.

Denote the complete set of quantum numbers that characterize the stationary states of the medium in the absence of a field by α (or β) and the corresponding energy levels by E_α (or E_β). In this notation, the solution of Eq. (1.2) in the linear (in $\hat{U}$) approximation is

$$\begin{aligned} W_{\alpha\beta}(t) &= W_\alpha^{(0)}\delta_{\alpha\beta} \\ &+ (W_\alpha^{(0)} - W_\beta^{(0)})\int \frac{U_{\alpha\beta}(\omega)}{\omega_{\alpha\beta} - \omega} e^{-i\omega t} d\omega, \end{aligned} \tag{1.3}$$

where $\omega_{\alpha\beta} = E_\alpha - E_\beta$, $W_{\alpha\beta}^{(0)} = W_\alpha^{(0)}\delta_{\alpha\beta}$ is the density matrix of the medium in the absence of a field and $U_{\alpha\beta}(\omega)$ is the Fourier component of the matrix element $U_{\alpha\beta}(t)$. We assume that the medium in the absence of a field is in thermodynamic equilibrium at temperature T. In this case [4, Section 31],

$$W_\alpha^{(0)} = e^{(F - E_\alpha)/T}, \quad \mathrm{Sp}\,\hat{W}^{(0)} = \sum_\alpha W_\alpha^{(0)} = 1, \tag{1.4}$$

where F is the free energy of the medium.[2] We see from (1.3) and (1.4) that $\mathrm{Sp}\,\hat{W} = 1$.

The current density operator $\hat{\mathbf{j}}(\mathbf{r}, t)$ consists of two parts:

$$\hat{\mathbf{j}}(\mathbf{r}, t) = \hat{\mathbf{j}}^{(0)}(\mathbf{r}) + \delta\hat{\mathbf{j}}(\mathbf{r}, t), \tag{1.5}$$

where $\hat{\mathbf{j}}^{(0)}(\mathbf{r})$ is the (time-independent) current density operator in the absence of a field, and the operator $\delta\hat{\mathbf{j}}(\mathbf{r}, t)$ is proportional to the field. Substituting (1.3) and (1.5) into (1.1), we obtain the following expression for the Fourier components of the current density in the linear (in field) approximation:

$$\begin{aligned} \mathbf{j}(\omega, \mathbf{k}) = \sum_\alpha W_\alpha^{(0)} \Bigg[& \delta\mathbf{j}_{\alpha\alpha}(\omega, \mathbf{k}) \\ & - \sum_\beta \left(\frac{U_{\alpha\beta}(\omega)\mathbf{j}_{\beta\alpha}^{(0)}(\mathbf{k})}{\omega_{\beta\alpha} + \omega} + \frac{\mathbf{j}_{\alpha\beta}^{(0)}(\mathbf{k})U_{\beta\alpha}(\omega)}{\omega_{\beta\alpha} - \omega} \right) \Bigg]. \end{aligned} \tag{1.6}$$

The permittivity tensor of the medium can be derived

[1] We use a system of units in which $\hbar = 1$.

[2] The temperature is measured in energy units.

from (1.6) [1, Section 2; 2, Section 12; 3, Section 31]:[3]

$$\varepsilon_{ij}(\omega, \mathbf{k}) = \delta_{ij} + \frac{4\pi i}{\omega}\sigma_{ij}(\omega, \mathbf{k}), \qquad j_i(\omega, \mathbf{k}) = \sigma_{ij}(\omega, \mathbf{k})E_j(\omega, \mathbf{k}). \tag{1.7}$$

For reasons of symmetry, the permittivity tensor for an isotropic nongyrotropic medium can be written as [1, Section 2; 2, Section 1; 3, Section 28]

$$\varepsilon_{ij}(\omega, \mathbf{k}) = \varepsilon^{\mathrm{tr}}(\omega, k)\left(\delta_{ij} - \frac{k_i k_j}{k^2}\right) + \varepsilon^{l}(\omega, k)\frac{k_i k_j}{k^2}, \tag{1.8}$$

where $\varepsilon^l(\omega, k)$ and $\varepsilon^{\mathrm{tr}}(\omega, k)$ are the longitudinal and transverse permittivities.

The scalar potential of the electromagnetic field is assumed to be equal to zero. The Fourier components of the fields $\mathbf{E}(\omega, \mathbf{k})$ and $\mathbf{B}(\omega, \mathbf{k})$ and the vector potential $\mathbf{A}(\omega, \mathbf{k})$ are then related by [2, Section 12]

$$\mathbf{E}(\omega, \mathbf{k}) = \frac{i\omega}{c}\mathbf{A}(\omega, \mathbf{k}), \quad \mathbf{B}(\omega, \mathbf{k}) = i\mathbf{k} \times \mathbf{A}(\omega, \mathbf{k}). \tag{1.9}$$

2. THE LONGITUDINAL AND TRANSVERSE PERMITTIVITIES OF A MONATOMIC GAS

If the medium is an ideal gas, then to determine the current density $\mathbf{j}(\mathbf{r}, t)$, it will suffice to calculate the current produced by one molecule: multiplying it by the total number of molecules NV (N is the number of molecules per unit volume, and V is the volume of the gas) yields an expression for $\mathbf{j}(\mathbf{r}, t)$ in the gas.

Denote the coordinates of Z electrons and the atomic nucleus in the laboratory frame of reference in which the field $\mathbf{E}(\mathbf{r}, t)$, $\mathbf{B}(\mathbf{r}, t)$ is specified by $\mathbf{R}_a$ ($a = 1, 2, \ldots, Z$) and $\mathbf{R}_n$, respectively. The atom–field interaction energy in the linear (in field) approximation if we disregard the terms on the order of m/M_n, where m and M_n are the electron and nuclear masses, respectively, is [5, Section 113]

$$\hat{U} = \mu_B \sum_a [\hat{\mathbf{P}}_a \cdot \mathbf{A}(\mathbf{R}_a, t) + \mathbf{A}(\mathbf{R}_a, t) \cdot \hat{\mathbf{P}}_a + 2\hat{\mathbf{s}}_a \cdot \mathbf{B}(\mathbf{R}_a, t)], \tag{2.1}$$

where $\mu_B = e/2mc$ is the Bohr magneton, and $\hat{\mathbf{P}}_a = -i\partial/\partial\mathbf{R}_a$ and $\hat{\mathbf{s}}_a$ are the electron momentum and spin operators, respectively.

[3] Summation is everywhere meant to be over the doubly repetitive indices $i, j, \ldots = x, y, z$.

The expression for the current density operator is

$$\frac{\hat{\mathbf{j}}(\mathbf{r}, t)}{NV} = -\frac{e}{2m}\sum_a [\delta(\mathbf{R}_a - \mathbf{r})(\hat{\mathbf{P}}_a + 2i\hat{\mathbf{P}}_a \times \hat{\mathbf{s}}_a) + (\hat{\mathbf{P}}_a - 2i\hat{\mathbf{P}}_a \times \hat{\mathbf{s}}_a)\delta(\mathbf{R}_a - \mathbf{r})] - \frac{e^2}{2mc}\mathbf{A}(\mathbf{r}, t)\sum_a \delta(\mathbf{R}_a - \mathbf{r}). \tag{2.2}$$

It differs from the standard expression for the current density [1, Section 31; 2, Section 12; 6, Section 24] only by the inclusion of spins.

To calculate the matrix elements in (1.6), we should pass from $\mathbf{R}_a$ and $\mathbf{R}_n$ to the coordinates of the electrons relative to the nucleus, $\mathbf{r}_a$, and to the coordinates of the atomic center of mass, $\mathbf{R}$ [7, Sections 13, 40]:

$$\mathbf{r}_a = \mathbf{R}_a - \mathbf{R}_n, \qquad \mathbf{R} = \frac{1}{M}\left(m\sum_a \mathbf{R}_a + M_n\mathbf{R}_n\right), \tag{2.3}$$

where $M = Zm + M_n$ is the atomic mass.

The complete set of quantum numbers and the atomic energy can be represented as

$$\alpha = \{\mathbf{P}, J, M, n\}, \quad E_\alpha = \frac{1}{2M}P^2 + E_{Jn}. \tag{2.4}$$

Here, $\mathbf{P}$ is the momentum of the atomic center of mass, J is the angular momentum ($J = 0, 1/2, 1, 3/2, \ldots$), $M = J, J-1, \ldots, -J$ is the angular momentum component along the z axis, and E_{Jn} is the atomic energy in the frame of reference in which $\mathbf{P} = 0$; n numbers atomic states with equal J and M, but with different energies.

Simple calculations yield the following expressions for the matrix elements of the Fourier components of operators (2.1) and (2.2):

$$\left\langle \mathbf{P}'J'M'n'\middle|\hat{U}(\omega)\middle|\mathbf{P}JMn\right\rangle = -\frac{ie}{m\omega}\frac{(2\pi)^3}{V} \times \mathbf{E}(\omega, \mathbf{P}' - \mathbf{P})\left\langle J'M'n'\middle|\hat{\mathbf{P}}(\mathbf{P}' - \mathbf{P})\middle|JMn\right\rangle,$$

$$\left\langle \mathbf{P}'J'M'n'\middle|\hat{\mathbf{j}}^{(0)}(\mathbf{k})\middle|\mathbf{P}JMn\right\rangle = -\frac{eNV}{m(2\pi)^3}\delta_{\mathbf{P}', \mathbf{P}-\mathbf{k}}\left\langle J'M'n'\middle|\hat{\mathbf{P}}(-\mathbf{k})\middle|JMn\right\rangle, \tag{2.5}$$

$$\langle \mathbf{P}'J'M'n'|\delta\hat{\mathbf{j}}(\omega, \mathbf{k})|\mathbf{P}JMn\rangle = \frac{ie^2N}{m\omega}\mathbf{E}(\omega, \mathbf{k} - \mathbf{P} + \mathbf{P}') \times \langle J'M'n'|\sum_a \exp[i(\mathbf{P} - \mathbf{P}') \cdot \mathbf{r}_a]|JMn\rangle.$$

We introduced the operators[4] $\hat{\mathbf{p}}_a = -i\partial/\partial\mathbf{r}_a$ and

$$\hat{\mathbf{P}}(\mathbf{k}) = \sum_a \left\{ \frac{1}{2}[\hat{\mathbf{p}}_a \exp(i\mathbf{k}\cdot\mathbf{r}_a) + \exp(i\mathbf{k}\cdot\mathbf{r}_a)\hat{\mathbf{p}}_a] + i\hat{\mathbf{s}}_a \times \mathbf{k}\exp(i\mathbf{k}\cdot\mathbf{r}_a) \right\}. \quad (2.6)$$

Given (2.4), the normalization condition (1.4) after passing from summation to integration over $\mathbf{P}$ takes the form

$$V\left(\frac{MT}{2\pi}\right)^{3/2} \sum_{Jn} (2J+1)\exp\left(\frac{F - F_{Jn}}{T}\right) = 1. \quad (2.7)$$

Using formulas (2.4)–(2.7), we can determine the current density $\mathbf{j}(\omega, \mathbf{k})$ (see (1.6)) and then the permittivity tensor (see (1.7)):

$$\varepsilon_{ij}(\omega, \mathbf{k}) = \left(1 - \frac{\omega_p^2}{\omega^2}\right)\delta_{ij} - \frac{4\pi e^2 NV}{m^2\omega^2}\left(\frac{MT}{2\pi}\right)^{3/2} \times \sum_{Jn} \exp\left(\frac{F - E_{Jn}}{T}\right) \times \sum_{J'n'} A^{ij}_{J'n',\,Jn}(\mathbf{k})\Phi_{J'n',\,Jn}(\omega, k;\, T). \quad (2.8)$$

Here, $\omega_p = \sqrt{4\pi e^2 N_e/m}$ is the plasma frequency of the atomic electrons ($N_e = ZN$),

$$\Phi_{J'n',\,Jn}(\omega, k;\, T) = \frac{1}{\sqrt{2}k\bar{v}}\left[\frac{F(z^{(+)}_{J'n',\,Jn})}{z^{(+)}_{J'n',\,Jn}} - \frac{F(z^{(-)}_{J'n',\,Jn})}{z^{(-)}_{J'n',\,Jn}}\right], \quad (2.9)$$

$$z^{(\pm)}_{J'n',\,Jn} = \frac{\omega \pm (\omega_{J'n',\,Jn} + k^2/2M)}{\sqrt{2}k\bar{v}},$$

$\bar{v} = \sqrt{T/M}$ is the mean thermal velocity of the atoms,

$$F(z) = \frac{z}{\sqrt{\pi}}\int_{-\infty}^{\infty} \frac{\exp(-t^2)}{t - z}dt, \quad (2.10)$$

$$A^{(ij)}_{J'n',\,Jn}(\mathbf{k}) = \sum_{MM'} \langle JMn|\hat{P}_i(-\mathbf{k})|J'M'n'\rangle \times \langle J'M'n'|\hat{P}_j(\mathbf{k})|JMn\rangle = A^{ji}_{J'n',\,Jn}(-\mathbf{k}). \quad (2.11)$$

[4] Strictly speaking, the term related to the spin–orbit interaction [6, Sections 55, 59] should be added to the operator $\hat{\mathbf{p}}_a$; the minus sign on the right-hand side of formula (55.7) in [6] was omitted, and m should be substituted for m^2 in formula (59.12).

The last equality in (2.11) can be easily obtained by taking into account the fact that after time reversal [5, Section 60],

$$|JMn\rangle \longrightarrow (-1)^{J-M}|J, -M, n\rangle,$$

$$\mathbf{r}_a \longrightarrow \mathbf{r}_a, \quad \hat{\mathbf{p}}_a \longrightarrow -\hat{\mathbf{p}}_a, \quad \hat{\mathbf{s}}_a \longrightarrow -\hat{\mathbf{s}}_a.$$

Since the states $|JMn\rangle$ with different M are transformed via the irreducible representation $D^{(J)}$ of the rotation group and since all of the states $|JMn\rangle$ (irrespective of M) are either even or odd, it is easy to show that (for given J and n) $A^{ij}_{J'n',\,Jn}(\mathbf{k})$ are transformed through the representation $D^{(0)} + D^{(2)}$ of the rotation group. Consequently, they can be written as

$$A^{(ij)}_{J'n',\,Jn}(\mathbf{k}) = \left(\delta_{ij} - \frac{k_i k_j}{k^2}\right)A^{\mathrm{tr}}_{J'n',\,Jn}(k) + \frac{k_i k_j}{k^2}A^{l}_{J'n',\,Jn}(k). \quad (2.12)$$

Finally, it follows from the equality $\hat{\mathbf{P}}^{(+)}(\mathbf{k}) = \hat{\mathbf{P}}(-\mathbf{k})$ that $A^{l,\,\mathrm{tr}}_{J'n',\,Jn}(\mathbf{k})$ are real. They can be calculated, for example, by using the formulas

$$A^{l}_{J'n',\,Jn}(k) = A^{zz}_{J'n',\,Jn}(k\mathbf{e}_z), \quad A^{\mathrm{tr}}_{J'n',\,Jn}(k) = A^{zz}_{J'n',\,Jn}(k\mathbf{e}_x), \quad (2.13)$$

where $\mathbf{e}_i$ are the unit vectors along the coordinate axes.

As must be the case, substituting (2.12) reduces the permittivity tensor (2.8) to form (1.8), where the longitudinal and transverse permittivities are given by

$$\varepsilon^{l,\,\mathrm{tr}}(\omega, k) = 1 - \frac{\omega_p^2}{\omega^2} - \frac{4\pi e^2 NV}{m^2\omega^2}\left(\frac{MT}{2\pi}\right)^{3/2} \times \sum_{Jn} \exp\left(\frac{E - E_{Jn}}{T}\right) \times \sum_{J'n'} \Phi_{J'n',\,Jn}(\omega, k;\, T)A^{l,\,\mathrm{tr}}_{J'n',\,Jn}(k). \quad (2.14)$$

The integral in (2.10) for real z has no meaning. We assume that $\omega > 0$ in (2.14) (and below) and that the field proportional to $\exp(-i\omega t)$ adiabatically switches on for $t \longrightarrow -\infty$ and make the corresponding substitution $\omega \longrightarrow \omega + i\delta$, where $\delta \longrightarrow +0$ [1, Section 11; 3, Section 29; 5, Section 42]. In plasma theory, this rule for bypassing the pole in integrals of form (2.10) is called the Landau bypassing rule [3, Section 29].

Below, we will need the following limiting expressions for the function $F(z)$, $z = z' + iz''$, $z'' > 0$ [1, Section 12; 3, Section 31]:

$$F(z) \approx \begin{cases} i\sqrt{\pi}z - 2z^2, & |z| \ll 1, \\ -1 - \dfrac{1}{2z^2} + i\sqrt{\pi}ze^{-z^2}, & |z| \gg 1, |z'| \gg z''. \end{cases} \quad (2.15)$$

We assume that the field is a moderately short-wavelength one: if the Bohr radius $a_B = 1/me^2$ is on the order of the atomic radius, then $ka_B \ll 1$. The matrix elements in (2.11) can then be expanded in power series of the small parameter ka_B. Retaining the terms of the second order of smallness, we write operator (2.6) as

$$\hat{P}_i(\mathbf{k}) = \hat{p}_i + \frac{i}{2}[(\hat{\mathbf{J}} + \hat{\mathbf{S}}) \times \mathbf{k}]_i + ik_j\hat{B}_{ij} - k_jk_k\hat{C}_{i,jk}, \quad (2.16)$$

where

$$\hat{\mathbf{p}} = \sum_a \hat{\mathbf{p}}_a, \quad \hat{\mathbf{S}} = \sum_a \hat{\mathbf{s}}_a, \quad \hat{\mathbf{J}} = \hat{\mathbf{S}} + \hat{\mathbf{L}},$$

$$\hat{\mathbf{L}} = \sum_a \hat{\mathbf{l}}_a, \quad \hat{\mathbf{l}}_a = \mathbf{r}_a \times \hat{\mathbf{p}}_a,$$

$$\hat{B}_{ij} = \frac{1}{2}\sum_a (\hat{p}_{ai}r_{aj} + r_{ai}\hat{p}_{aj}) = \hat{B}_{ji},$$

$$\hat{C}_{i,jk} = \frac{1}{4}\sum_a (\hat{p}_{ai}r_{aj}r_{ak} + r_{aj}r_{ak}\hat{p}_{ai} - 4e_{ijl}r_{ak}\hat{s}_{al}) = \delta_{ik}\hat{C}_j^{(1)}$$

$$- \delta_{jk}[\hat{C}_i^{(1)} - \hat{C}_i^{(2)} + (1 - 3\delta_{jz}\delta_{kz})\hat{C}_i^{(3)}] + \hat{\tilde{C}}_{i,jk}, \quad (2.17)$$

$$\hat{\mathbf{C}}^{(1)} = \frac{1}{2}\sum_a \mathbf{r}_a \times \hat{\mathbf{s}}_a,$$

$$\hat{\mathbf{C}}^{(2)} = \frac{1}{12}\sum_a (\hat{\mathbf{p}}_a r_a^2 + r_a^2\hat{\mathbf{p}}_a),$$

$$\hat{\mathbf{C}}^{(3)} = \frac{1}{20}$$

$$\times \sum_a \left[\mathbf{r}_a(\mathbf{r}_a \cdot \hat{\mathbf{p}}_a) + (\hat{\mathbf{p}}_a \cdot \mathbf{r}_a)\mathbf{r}_a - \frac{1}{3}(r_a^2\hat{\mathbf{p}}_a + \hat{\mathbf{p}}_a r_a^2)\right],$$

and $\hat{\tilde{C}}_{i,jk}$ are transformed via the representation $D^{(3)} + 2D^{(2)} + D^{(0)}$ and give no contribution to $A^{l,\mathrm{tr}}_{J'n',Jn}(k)$ with the adopted accuracy (on the order of $(ka_B)^2$). Note that

$$\hat{\mathbf{p}} = m\frac{d\mathbf{r}}{dt}, \quad \mathbf{r} = \sum_a \mathbf{r}_a = -\frac{\mathbf{d}}{e},$$

$$\hat{B}_{ij} = \frac{m}{2}\frac{d}{dt}R_{ij},$$

$$R_{ij} = \sum_a r_{ai}r_{aj} = \frac{1}{3}\delta_{ij}R_{kk} - \frac{1}{3e}Q_{ij}, \quad (2.18)$$

where $\mathbf{d}$ and Q_{ij} are the dipole and quadrupole electric moments of the atom, respectively [5, Section 75].

It is convenient to calculate the matrix elements of operators (2.16)–(2.18) by introducing the corresponding spherical tensors and using the Wigner–Eckart theorem and the standard properties of the 3j symbols [5, Sections 106, 107]. As a result, we obtain

$$A^{l,\mathrm{tr}}_{J'n',Jn}(k) = A^{l,\mathrm{tr}(d)}_{J'n',Jn}(k) + A^{l,\mathrm{tr}(m)}_{J'n',Jn}(k) + A^{l,\mathrm{tr}(q)}_{J'n',Jn}(k) + A^{l,\mathrm{tr}(s)}_{J'n',Jn}(k),$$

$$A^{l,\mathrm{tr}(d)}_{J'n',Jn}(k) = \frac{m^2\omega^2_{J'n',Jn}}{3e^2}|\langle Jn\|\mathbf{P}\|J'n'\rangle|^2 + \frac{2}{3}k^2\mathrm{Re}\langle Jn\|\mathbf{P}\|J'n'\rangle\langle J'n'\|C^{l,\mathrm{tr}}\|Jn\rangle,$$

$$\hat{\mathbf{C}}^l = -(\hat{\mathbf{C}}^{(2)} + 2\hat{\mathbf{C}}^{(3)}),$$

$$\hat{\mathbf{C}}^{\mathrm{tr}} = \hat{\mathbf{C}}^{(1)} - \hat{\mathbf{C}}^{(2)} + \hat{\mathbf{C}}^{(3)},$$

$$A^{l(m)}_{J'n',Jn}(k) = 0,$$

$$A^{\mathrm{tr}(m)}_{J'n',Jn}(k) = \frac{1}{12}k^2|\langle Jn\|J + S\|J'n'\rangle|^2, \quad (2.19)$$

$$A^{l(q)}_{J'n',Jn}(k) = \frac{4}{3}A^{\mathrm{tr}(q)}_{J'n',Jn}(k) = \frac{1}{270e^2}m^2\omega^2_{J'n',Jn}k^2|\langle Jn\|Q\|J'n'\rangle|^2,$$

$$A^{l(s)}_{J'n',Jn}(k) = \frac{1}{36}m^2\omega^2_{J'n',Jn}k^2|\langle Jn\|R_0\|J'n'\rangle|^2,$$

$$A^{\mathrm{tr}(s)}_{J'n',Jn}(k) = 0,$$

where $\langle Jn\|\ldots\|J'n'\rangle$ are the reduced matrix elements; $R_{00} = R_{ii}$ [5, Section 107]. Since $\hat{\mathbf{J}}$ commutes with the

Hamiltonian of the atom [5, Section 29],

$$\langle Jn\|J\|J'n'\rangle = \delta_{J'J}\delta_{n'n}\sqrt{J(J+1)(2J+1)}. \quad (2.20)$$

In the zero approximation in spin–orbit interaction, we can calculate the matrix elements $\langle Jn\|S\|J'n'\rangle$ to the end [8, Section 50]. Denote the atomic term unsplit by the spin–orbit interaction by LSn, so that $J = L + S, L + S - 1, \ldots, |L - S|$. Then,

$$\langle J'L'S'n'\|S\|JLSn\rangle = \frac{1}{2}\delta_{L'L}\delta_{S'S}\delta_{n'n}\left\{\delta_{J'J}\sqrt{\frac{2J+1}{J(J+1)}}[J(J+1)+S(S+1)-L(L+1)]\right.$$
$$-\delta_{J',J+1}\sqrt{\frac{(J+2+L+S)(J+1+S-L)(J+1+L-S)(L+S-J)}{J+1}} \quad (2.21)$$
$$\left.-\delta_{J',J-1}\sqrt{\frac{(J+L+S+1)(J+S-L)(J+L-S)(L+S+1-J)}{J}}\right\};$$

i.e., the matrix elements of the spin $\hat{\mathbf{S}}$ are nonzero only for transitions inside the fine structure of the term.

Formulas (2.14) and (2.19) give the final expressions for $\varepsilon^l(\omega, k)$ and $\varepsilon^{tr}(\omega, k)$. Below, we restrict our analysis to moderately high temperatures: if ω_0 is the energy interval between the ground and the first excited atomic levels, then $T \ll \omega_0$. Given (2.7), formula (2.14) then takes the form

$$\varepsilon^{l,\,tr}(\omega, k) = 1 - \frac{\omega_p^2}{\omega^2} - \frac{4\pi N e^2}{m^2\omega^2(2J_0+1)} \times \sum_{Jn}\Phi_{Jn,\,J_0n_0}(\omega, k; T)A^{l,\,tr}_{Jn,\,J_0n_0}(k). \quad (2.22)$$

We mark the states of the ground atomic level by the subscript 0: $|J_0M_0n_0\rangle$ and $E_{J_0n_0}$. If $L_0 = 0$ or $S_0 = 0$, then the ground term has no fine structure, with $\omega_0 \sim \omega_R = 1/(ma_B)^2$. If L_0 and $S_0 \neq 0$, then [5, Section 72]

$$J_0 = |L_0 - S_0|, \quad \omega_0 = |A_{L_0S_0}|(J_0+1),$$

for the normal multiplet, and

$$J_0 = L_0 + S_0, \quad \omega_0 = |A_{L_0S_0}|J_0,$$

for the inverted multiplet; here, the constant $|A_{L_0S_0}| \sim \omega_R(Ze^2/c)^2$.

Below, we will have to use a quantity proportional to the difference $\varepsilon^{tr}(\omega, k) - \varepsilon^l(\omega, k)$. We specially denote it by

$$1 - \frac{1}{\mu(\omega, k)} = \frac{\omega^2}{c^2k^2}[\varepsilon^{tr}(\omega, k) - \varepsilon^l(\omega, k)] = -\frac{\pi e^2 N}{3(mc)^2(2J_0+1)}\sum_{Jn}\Phi_{Jn,\,J_0n_0}(\omega, k; T)M_{Jn,\,J_0n_0}, \quad (2.23)$$

where

$$M_{Jn,\,J_0n_0} = M^{(d)}_{Jn,\,J_0n_0} + M^{(m)}_{Jn,\,J_0n_0} + M^{(q)}_{Jn,\,J_0n_0} + M^{(s)}_{Jn,\,J_0n_0},$$

$$M^{(d)}_{Jn,\,J_0n_0} = 8\mathrm{Re}\langle J_0n_0\|p\|Jn\rangle\langle Jn\|C^\mu\|J_0n_0\rangle,$$

$$\hat{\mathbf{C}}^\mu = \hat{\mathbf{C}}^{(1)} + 3\hat{\mathbf{C}}^{(3)}, \quad (2.24)$$

$$M^{(m)}_{Jn,\,J_0n_0} = |\langle J_0n_0\|J+S\|Jn\rangle|^2,$$

$$M^{(q)}_{Jn,\,J_0n_0} = -\frac{m^2\omega^2_{Jn,\,J_0n_0}}{90e^2}|\langle J_0n_0\|Q\|Jn\rangle|^2,$$

$$M^{(s)}_{Jn,\,J_0n_0} = -\frac{m^2\omega^2_{Jn,\,J_0n_0}}{3}|\langle J_0n_0\|R_0\|Jn\rangle|^2.$$

3. THE LONGITUDINAL AND TRANSVERSE PERMITTIVITIES IN VARIOUS FREQUENCY RANGES

We will derive expressions for $\varepsilon^{tr,\,l}(\omega, k)$ in various frequency ranges from formula (2.22). Let us first consider the case where ω is close to the frequency of a particular atomic transition: the detuning $\delta_{Jn} = \omega - \omega_{Jn,\,J_0n_0}$ satisfies the condition

$$|\delta_{Jn}| \ll |\omega_{Jn,\,J'n'}|, \quad (Jn) \neq (J_0n_0), \quad (J'n') \neq (Jn). \quad (3.1)$$

Since we disregard the (natural and collisional) width ν_{Jn} of the excited level, the detuning in (3.1) cannot be very small: $|\delta_{Jn}| \gg \nu_{Jn}$. When condition (3.1) is satis-

fied, it will suffice to retain only the second term in function (2.9). Then, as follows from (2.22)

$$\varepsilon^{l,\mathrm{tr}}(\omega, k) = \tilde{\varepsilon}(\omega) + \frac{4\pi e^2 N}{m^2\omega^2(2J_0+1)(\delta_{Jn} - k^2/2M)} \times F\left(\frac{\delta_{Jn} - k^2/2M}{\sqrt{2}k\bar{v}}\right) A^{l,\mathrm{tr}}_{Jn,\,J_0n_0}(k), \tag{3.2}$$

where $\tilde{\varepsilon}(\omega)$ is a smooth function of the frequency in range (3.1). Note that $k\bar{v}$ is identical to the Doppler width of the spectral line that corresponds to the $J_0n_0 \longleftrightarrow Jn$ transition [9, Sections 20; 10, Section 103]. Expression (2.23) takes the form

$$1 - \frac{1}{\mu(\omega, k)} = \frac{\pi e^2 N}{3(mc)^2(2J_0+1)(\delta_{Jn} - k^2/2M)} \times F\left(\frac{\delta_{Jn} - k^2/M}{\sqrt{2}k\bar{v}}\right) M_{Jn,\,J_0n_0}. \tag{3.3}$$

Let us now consider the case where ω is far from the resonance frequencies:

$$|\delta_{Jn}| \gg k\bar{v}, k^2/M, \quad (Jn) \neq (J_0n_0). \tag{3.4}$$

In this case, according to (2.9) and (2.15), we have with the adopted accuracy

$$\Phi_{Jn,\,J_0n_0}(\omega, k; T) = \frac{2\omega_{Jn,\,J_0n_0}}{\omega^2 - \omega^2_{Jn,\,J_0n_0}} \times \left[1 + \frac{k^2}{M(\omega^2 - \omega^2_{Jn,\,J_0n_0})} \times \left(\frac{\omega^2 + \omega^2_{Jn,\,J_0n_0}}{2\omega_{Jn,\,J_0n_0}} + \frac{3\omega^2 + \omega^2_{Jn,\,J_0n_0}}{\omega^2 - \omega^2_{Jn,\,J_0n_0}} T\right)\right]. \tag{3.5}$$

With the same accuracy, we obtain from (2.22), (2.19), (2.23), and (2.24)

$$\varepsilon^{l,\mathrm{tr}}(\omega, k) = 1 - \frac{\omega_p^2}{\omega^2} + \frac{8\pi e^2 N}{m^2\omega^2(2J_0+1)} \times \sum_{Jn}{}' \frac{\omega_{Jn,\,J_0n_0}}{\omega^2_{Jn,\,J_0n_0} - \omega^2} \times \left[A^{l,\mathrm{tr}}_{Jn,\,J_0n_0}(k) + \frac{m^2k^2\omega^2_{Jn,\,J_0n_0}}{3Me^2(\omega^2 - \omega^2_{Jn,\,J_0n_0})} \times |\langle J_0n_0\|d\|Jn\rangle|^2 \left(\frac{\omega^2 + \omega^2_{Jn,\,J_0n_0}}{2\omega_{Jn,\,J_0n_0}} + \frac{T(3\omega^2 + \omega^2_{Jn,\,J_0n_0})}{\omega^2 - \omega^2_{Jn,\,J_0n_0}}\right)\right] - \frac{4\pi e^2 N}{m^2\omega^2(2J_0+1)} A^{l,\mathrm{tr}}_{J_0n_0,\,J_0n_0}(k)\Phi_0(\omega, k; T), \tag{3.6}$$

$$1 - \frac{1}{\mu(\omega, k)} = \frac{\pi e^2 N}{3(mc)^2}\left[\frac{2}{2J_0+1}\sum_{Jn}\frac{\omega_{Jn,\,J_0n_0}}{\omega^2_{Jn,\,J_0n_0} - \omega^2} \times M_{Jn,\,J_0n_0} - J_0(J_0+1)g_0^2\Phi_0(\omega, k; T)\right], \tag{3.7}$$

where

$$\sum_{Jn}{}'(\ldots) = \sum_{(Jn)\neq(J_0n_0)}(\ldots),$$

$$\Phi_0(\omega, k; T) = \Phi_{J_0n_0,\,J_0n_0}(\omega, k; T),$$

$$g_0 = g_{J_0L_0S_0},$$

and

$$g_{JLS} = 1 + \frac{J(J+1) + S(S+1) - L(L+1)}{2J(J+1)} \tag{3.8}$$

is the Lande factor of the *JLSn* atomic level [5, Section 113]. In (3.6) and (3.7), we take into account the fact that, as follows from (2.19)–(2.21),

$$A^{l}_{Jn,\,Jn}(k) = 0, \quad A^{\mathrm{tr}}_{Jn,\,Jn}(k) = \frac{1}{12}k^2J(J+1)(2J+1)g^2_{JLS}. \tag{3.9}$$

If, in addition to conditions (3.4), conditions (3.1) are also satisfied, then the expressions derived from (3.6) and (3.7) are identical to the expressions derived from (3.2) and (3.3) when it is considered that in this case, $\omega^2 - \omega^2_{Jn,\,J_0n_0} \approx 2\omega_{Jn,\,J_0n_0}\delta_{Jn}$ and (see (2.15))

$$\frac{1}{z^{(-)}_{Jn,\,J_0n_0}}F(z^{(-)}_{Jn,\,J_0n_0}) \approx -\frac{\sqrt{2}k\bar{v}}{\delta_{Jn}}\left[1 + \frac{k^2}{M\delta_{Jn}}\left(\frac{1}{2} + \frac{T}{\delta_{Jn}}\right)\right], \tag{3.10}$$

$$\varepsilon^{l,\mathrm{tr}}(\omega, k) = \tilde{\varepsilon}(\omega) - \frac{4\pi e^2 N}{m^2\omega^2(2J_0+1)\delta_{Jn}} \times \left[A^{l,\mathrm{tr}}_{Jn,\,J_0n_0}(k) + \frac{m^2k^2\omega^2_{Jn,\,J_0n_0}}{3e^2M\delta_{Jn}} \times |\langle J_0n_0\|d\|Jn\rangle|^2\left(\frac{1}{2} + \frac{T}{\delta_{Jn}}\right)\right], \tag{3.11}$$

$$1 - \frac{1}{\mu(\omega, k)} = -\frac{\pi e^2 N}{3(mc)^2(2J_0+1)\delta_{Jn}}M_{Jn,\,J_0n_0}. \tag{3.12}$$

The case of low frequencies, $\omega \ll \omega_0$, is contained in formulas (3.6) and (3.7). The terms proportional to k^2 in the expression for $A^{l,\mathrm{tr}}_{Jn,J_0n_0}(k)$ at $Jn \neq J_0n_0$ are on the order of $m\omega_R(ka_B)^2$, while the remaining terms in square brackets in (3.6) at $\omega \ll \omega_0$ are on the order of $m\omega_R(ka_B)^2m/M$. We disregard these terms (on the order of m/M) from the outset. Therefore, at $\omega \ll \omega_0$,

$$\varepsilon^{l,\mathrm{tr}}(\omega,k) = 1-\frac{\omega_p^2}{\omega^2}-\frac{8\pi e^2N}{m^2\omega^2(2J_0+1)} \times{\sum_{Jn}}'\frac{\omega_{Jn,J_0n_0}}{\omega^2-\omega^2_{Jn,J_0n_0}}A^{l,\mathrm{tr}}_{Jn,J_0n_0}(k) - \frac{4\pi e^2N}{m^2\omega^2(2J_0+1)}A^{l,\mathrm{tr}}_{J_0n_0,J_0n_0}(k)\Phi_0(\omega,k;T). \tag{3.13}$$

We know [2, Section 12] how the pole at $\omega = 0$ in the expression for $\varepsilon_{ij}(\omega, \mathbf{k})$ can be eliminated if the contribution of only dipole transitions is considered. A similar procedure can also be performed in (3.13). First, note that

$$\frac{1}{\omega^2(\omega^2-\omega^2_{Jn,J_0n_0})} = \frac{1}{\omega^2_{Jn,J_0n_0}}\left(\frac{1}{\omega^2-\omega^2_{Jn,J_0n_0}}-\frac{1}{\omega^2}\right). \tag{3.14}$$

The sums (proportional to $1/\omega^2$) that result from the substitution of (3.14) into (3.13) can be calculated by using formulas (2.17)–(2.19), the Wigner–Eckart theorem, and the standard properties of the 3*j* and 6*j* symbols [5, Sections 106–110]. As a result, we obtain

$${\sum_{Jn}}'\frac{1}{\omega_{Jn,J_0n_0}}A^{l}_{Jn,J_0n_0}(k) = \frac{Zm}{2}(2J_0+1),$$

$${\sum_{Jn}}'\frac{A^{\mathrm{tr}}_{Jn,J_0n_0}(k)}{\omega_{Jn,J_0n_0}} = \frac{Zm}{2}(2J_0+1) + \frac{k^2}{12}\left[{\sum_{Jn}}'\frac{|\langle J_0n_0\|S\|Jn\rangle|^2}{\omega_{Jn,J_0n_0}} - \frac{m(2J_0+1)}{\sqrt{2L_0+1}}\langle J_0n_0\|R_0\|J_0n_0\rangle\right]. \tag{3.15}$$

The final expressions can be derived from (3.13)–(3.15) and (3.9)

$$\varepsilon^{l}(\omega,k) = 1-\frac{8\pi e^2N}{m^2(2J_0+1)} \times{\sum_{Jn}}'\frac{1}{\omega_{Jn,J_0n_0}(\omega^2-\omega^2_{Jn,J_0n_0})}A^{l}_{Jn,J_0n_0}(k), \tag{3.16}$$

$$\varepsilon^{\mathrm{tr}}(\omega,k) = 1-\frac{8\pi e^2N}{m^2(2J_0+1)} \times{\sum_{Jn}}'\frac{1}{\omega_{Jn,J_0n_0}(\omega^2-\omega^2_{Jn,J_0n_0})}A^{\mathrm{tr}}_{Jn,J_0n_0}(k) + \frac{2\pi e^2Nk^2}{3m^2\omega^2}\left[-\frac{m\langle L_0n_0\|R_0\|L_0n_0\rangle}{\sqrt{2L_0+1}} + \frac{1}{2J_0+1}{\sum_{Jn}}'\frac{|\langle J_0n_0\|S\|Jn\rangle|^2}{\omega_{Jn,J_0n_0}} - \frac{J_0(J_0+1)}{2}g_0^2\Phi_0(\omega,k;T)\right], \tag{3.17}$$

$$1-\frac{1}{\mu(\omega,k)} = -\frac{2\pi e^2N\omega^2}{3(mc)^2(2J_0+1)} \times{\sum_{Jn}}'\frac{1}{\omega_{Jn,J_0n_0}(\omega^2-\omega^2_{Jn,J_0n_0})} \times M_{Jn,J_0n_0} + \frac{2\pi e^2N}{3(mc)^2} \times\left[-\frac{m\langle L_0n_0\|R_0\|L_0n_0\rangle}{\sqrt{2L_0+1}} + \frac{1}{2J_0+1} \times{\sum_{Jn}}'\frac{|\langle J_0n_0\|S\|Jn\rangle|^2}{\omega_{Jn,J_0n_0}} - \frac{J_0(J_0+1)}{2}g_0^2\Phi_0(\omega,k;T)\right]. \tag{3.18}$$

Formula (3.18) can also be derived from (3.7) by using a transformation similar to the transformation used when passing from (3.13) to (3.16) and (3.17).

The longitudinal permittivity $\varepsilon^l(\omega, k)$ is a regular function of ω. Using (2.19), we obtain from (3.16)

$$\varepsilon^l(\omega, 0) = \varepsilon^l(\omega, k)\Big|_{\frac{ka_B}{\omega/\omega_R}\to 0} = 1 + \frac{8\pi N}{3(2J_0+1)} \times \sum_{Jn}{}' \frac{\omega_{Jn, J_0n_0}|\langle J_0n_0\|d\|Jn\rangle|^2}{\omega^2_{Jn, J_0n_0} - \omega^2}, \tag{3.19}$$

$$\varepsilon^l(0, k) = \varepsilon^l(\omega, k)\Big|_{\frac{\omega/\omega_R}{ka_B}\to 0} = 1 + \frac{8\pi e^2 N}{m^2(2J_0+1)} \times \sum_{Jn}{}' \frac{A^l_{Jn, J_0n_0}(k)}{\omega^3_{Jn, J_0n_0}}. \tag{3.20}$$

Thus, the limiting value

$$\varepsilon^l_0 = \varepsilon^l(\omega, 0)\big|_{\omega/\omega_R\to 0} = \varepsilon^l(0, k)\big|_{ka_B\to 0} = 1 + \frac{8\pi N}{3(2J_0+1)}\sum_{Jn}{}' \frac{|\langle J_0n_0\|d\|Jn\rangle|^2}{\omega_{Jn, J_0n_0}} \tag{3.21}$$

does not depend on the order of the passage to the limit: first $ka_B \longrightarrow 0$ and then $\omega/\omega_R \longrightarrow 0$, or vice versa, first $\omega/\omega_R \longrightarrow 0$ and then $ka_B \longrightarrow 0$.

In contrast to $\varepsilon^l(\omega, k)$, the transverse permittivity $\varepsilon^{tr}(\omega, k)$ and the permeability $\mu(\omega, k)$ have singularities at $\omega = 0$. The function $\Phi_0(\omega, k; T)$ has the following limiting expressions (see (2.9) and (2.15)):

$$\Phi_0(\omega, k; T) \approx \frac{k^2}{M\omega^2} \quad \text{for} \quad \omega \gg \frac{k^2}{M}, k\nabla, \tag{3.22}$$

$$\Phi_0(\omega, k; T) \approx -\frac{4M}{k^2} \quad \text{for} \quad k \gg \sqrt{M\omega}, M\nabla, \tag{3.23}$$

$$\Phi_0(\omega, k; T) \approx -\frac{1}{T} \quad \text{for} \quad \sqrt{M\omega} \ll k \ll M\nabla. \tag{3.24}$$

The conditions in (3.22) are satisfied for $ka_B \ll \omega/\omega_R$. Therefore, using (2.19) and (3.19), we obtain from (3.17) and (3.18)

$$\varepsilon^{tr}(\omega, 0) = \varepsilon^{tr}(\omega, k)\Big|_{\frac{ka_B}{\omega/\omega_R}\to 0} = \varepsilon^l(\omega, 0), \qquad \varepsilon^{tr}(\omega, 0)\Big|_{\frac{\omega}{\omega_R}\to 0} = \varepsilon^l_0, \tag{3.25}$$

$$1 - \frac{1}{\mu(\omega, 0)} = 1 - \frac{1}{\mu(\omega, k)}\Big|_{\frac{ka_B}{\omega/\omega_R}\to 0} = \frac{2\pi N}{3(2J_0+1)}\left(\frac{e\omega}{mc}\right)^2 \times \sum_{Jn}{}' \frac{M_{Jn, J_0n_0}}{\omega_{Jn, J_0n_0}(\omega^2_{Jn, J_0n_0} - \omega^2)} + \frac{2\pi e^2 N}{3(mc)^2} \times \left(-\frac{m}{\sqrt{2L_0+1}}\langle L_0n_0\|R_0\|L_0n_0\rangle + \frac{1}{2J_0+1}\sum_{Jn}{}' \frac{|\langle J_0n_0\|S\|Jn\rangle|^2}{\omega_{Jn, J_0n_0}}\right), \tag{3.26}$$

$$1 - \frac{1}{\mu(\omega, 0)}\Big|_{\frac{\omega}{\omega_R}\to 0} = \frac{2\pi e^2 N}{3(mc)^2} \times \left(-\frac{m}{\sqrt{2L_0+1}}\langle L_0n_0\|R_0\|L_0n_0\rangle + \frac{1}{2J_0+1}\sum_{Jn}{}' \frac{|\langle J_0n_0\|S\|Jn\rangle|^2}{\omega_{Jn, J_0n_0}}\right). \tag{3.27}$$

If the condition $k \gg \sqrt{M\omega}$ in (3.23) and (3.24) is satisfied, then the condition $ka_B \gg \omega/\omega_R$ is also satisfied. In this case, we obtain from (3.17) and (3.18)

$$\varepsilon^{tr}(\omega, k)\Big|_{\frac{\omega}{k^2/M}\to 0} = 1 + \frac{2\pi e^2 N k^2}{3m^2\omega^2} \times \left[-\frac{m}{\sqrt{2L_0+1}}\langle L_0n_0\|R_0\|L_0n_0\rangle + \frac{1}{2}J_0(J_0+1)g_0^2\Phi_0 + \frac{1}{2J_0+1}\sum_{Jn}{}' \frac{|\langle J_0n_0\|S\|Jn\rangle|^2}{\omega_{Jn, J_0n_0}}\right], \tag{3.28}$$

$$1 - \frac{1}{\mu(0, k)} = 1 - \frac{1}{\mu(\omega, k)}\Big|_{\frac{\omega}{k^2/M}\to 0} = \frac{2\pi e^2 N}{3(mc)^2} \times \left[-\frac{m}{\sqrt{2L_0+1}}\langle L_0n_0\|R_0\|L_0n_0\rangle + \frac{1}{2}J_0(J_0+1)g_0^2\Phi_0 + \frac{1}{2J_0+1}\sum_{Jn}{}' \frac{|\langle J_0n_0\|S\|Jn\rangle|^2}{\omega_{Jn, J_0n_0}}\right]. \tag{3.29}$$

The function $\Phi_0 = \Phi_0(\omega, k; T) = -4M/k^2$ if $k \gg M\nabla$ and $\Phi_0 = \Phi_0(\omega, k; T) = -1/T$ if $k \ll M\nabla$ (see (3.23) and (3.24)).

4. PERMITTIVITY $\varepsilon(\omega)$ AND PERMEABILITY $\mu(\omega)$

In the approach to the electrodynamics of material media in which, apart from the vectors **E** and **B**, the vectors **D** and **H** are also introduced, the properties of an isotropic medium are characterized by the permittivity $\varepsilon(\omega)$ and the permeability $\mu(\omega)$ [1, Section 2; 10, Section 77]:

$$\mathbf{D}(\omega) = \varepsilon(\omega)\mathbf{E}(\omega), \quad \mathbf{B}(\omega) = \mu(\omega)\mathbf{H}(\omega). \tag{4.1}$$

They are related to the longitudinal, $\varepsilon^l(\omega, k)$, and transverse, $\varepsilon^{tr}(\omega, k)$, permittivities considered here by [1, Section 2; 10, Section 103]

$$\varepsilon(\omega) = \varepsilon^l(\omega, k), \quad \mu(\omega) = \mu(\omega, k), \tag{4.2}$$

where $\mu(\omega, k)$ is the function introduced in (2.23). It is clear from these relations that $\varepsilon(\omega)$ and $\mu(\omega)$ have physical meaning only when the k dependences of $\varepsilon^l(\omega, k)$ and $\mu(\omega, k)$ may be ignored. Thus, as follows from the results of Section 3, $\varepsilon(\omega)$ loses its physical meaning (as the coefficient that relates $\mathbf{D}(\omega)$ and $\mathbf{E}(\omega)$) not only at ω close to the frequency of any permitted transition [10, Section 103], but also in all of the ω ranges in which electric quadrupole and completely symmetric transitions contribute appreciably to $\varepsilon^l(\omega, k)$.

According to (4.2), the magnetic permeability $\mu(\omega)$ at low ($\omega \ll \omega_R$) frequencies is defined by formulas (3.26) and (3.27). If we introduce the magnetic susceptibility $\chi(\omega) = [\mu(\omega) - 1]/4\pi$, then, according to (3.27), its low-frequency limit will be defined by the formula

$$\chi(0) = \frac{Ne^2}{6(mc)^2}\left(-\frac{m}{\sqrt{2L_0+1}}\langle L_0 n_0 \| R_0 \| L_0 n_0\rangle + \frac{1}{2J_0+1}\sum_{Jn}{}'\frac{|\langle J_0 n_0 \| S \| Jn\rangle|^2}{\omega_{Jn,\,J_0 n_0}}\right). \tag{4.3}$$

In order of magnitude, $|\chi(0)| \approx Na_B^3\omega_R/mc^2 \ll 1$, because $Na_B^3 < 10^{-4}$ and $\omega_R/mc^2 \approx 10^{-4}$. The static susceptibility in a uniform field can be obtained from (3.29) for $\Phi_0 = -1/T$,

$$\chi_{st}(0) = \chi(0) + \frac{1}{3T}N\mu_B^2 J_0(J_0+1)g_0^2, \tag{4.4}$$

where $\chi(0)$ is given by (4.3). If the g factor of the ground atomic level is nonzero, then the last term in (4.4) is much larger than $\chi(0)$; it is on the order of $\chi(0)\omega_R/T$; however, $\chi_{st}(0) \ll 1$ in this case as well. As must be the case, expression (4.4) for $\chi_{st}(0)$ matches the Van Vleck standard formula [4, Section 52]. The T-independent diamagnetic and paramagnetic terms are also retained at high frequencies (see (4.3) and (3.26)). Therefore, there is no reason to believe [10, Section 79] that, in contrast to $\varepsilon(\omega)$, $\mu(\omega)$ loses its physical meaning with increasing ω relatively early or that we should set $\mu(\omega) = 1$. However, at optical frequencies, the contribution of electric quadrupole transitions to $\mu(\omega)$ is comparable to the contribution of magnetic dipole and completely symmetric transitions (see (3.26) and (2.24)), so the vector

$$\mathbf{M} = \frac{1}{4\pi}(\mathbf{B} - \mathbf{H}) = \frac{1}{4\pi}[\mu(\omega) - 1]\mathbf{H}$$

at these frequencies loses the meaning of magnetic moment per unit volume of the gas in accordance with [10, Section 79].[5]

In conclusion, let us consider the question of whether a transverse electromagnetic wave with a group velocity antiparallel to its phase velocity can propagate through a monatomic gas. At present, this question is widely discussed in the literature (see, e.g., [11] and references therein). In [12], attention was given to interesting features of the propagation of an electromagnetic wave with antiparallel group and phase velocities through a transparent isotropic medium.[6] The dispersion relation for a transverse electromagnetic wave in an isotropic nongyrotropic medium is [1, Section 6]

$$\frac{c^2k^2}{\omega^2} = \varepsilon^{tr}(\omega, k). \tag{4.5}$$

For the wave damping to be ignored (i.e., for the wave vector k to be real), we must require for $\varepsilon^{tr} = \varepsilon^{tr'} + i\varepsilon^{tr''}$ that the following conditions be satisfied [1, Section 6]:

$$\varepsilon^{tr''}(\omega, k) \approx 0, \quad \varepsilon^{tr'}(\omega, k) > 0. \tag{4.6}$$

For the phase, $\mathbf{u}_f = \omega\mathbf{k}/k^2$, and group, $\mathbf{u}_{gr} = (\mathbf{k}/k)d\omega/dk$, velocities of the wave, we find from Eq. (4.5) that

$$\text{if } \frac{\omega^2}{c^2}\frac{\partial\varepsilon^{tr'}}{\partial k^2} < 1, \text{ then } \mathbf{u}_{gr} \parallel \mathbf{u}_f;$$
$$\text{if } \frac{\omega^2}{c^2}\frac{\partial\varepsilon^{tr'}}{\partial k^2} > 1, \text{ then } \mathbf{u}_{gr} \text{ and } \mathbf{u}_f \text{ are antiparallel.} \tag{4.7}$$

Thus, the group and phase velocities can be antiparallel only if the spatial dispersion is taken into account [2, Sections 7, 10]. In the frequency range where $\varepsilon(\omega)$ and $\mu(\omega)$ (see (4.2) and (2.23)) have physical meaning, Eq. (4.5) is equivalent to the equation [10, Section 83]

$$\frac{c^2k^2}{\omega^2} = \varepsilon(\omega)\mu(\omega), \tag{4.8}$$

[5] The reasoning given in [10] proves this result alone.

[6] It is clear from symmetry considerations that the group and phase velocities of any wave in an isotropic medium are either parallel or antiparallel.

while conditions (4.6) and (4.7) are equivalent to the following conditions [13]:

$$\varepsilon'' \approx 0, \quad \mu'' \approx 0, \quad \varepsilon'\mu' > 0, \tag{4.9}$$

$$\begin{aligned} &\text{if } \ \varepsilon', \mu' > 0, \ \text{ then } \ \mathbf{u}_{gr} \parallel \mathbf{u}_f, \\ &\text{if } \ \varepsilon', \mu' < 0, \ \text{ then } \ \mathbf{u}_{gr} \text{ and } \mathbf{u}_f \text{ are antiparallel.} \end{aligned} \tag{4.10}$$

The first condition in (4.6) is satisfied only if ω is not too close to any of the eigenfrequencies of the atom: the detuning $|\delta_{Jn}| \gg \tilde{\nu}_{Jn}$, where $\tilde{\nu}_{Jn}$ is the total width of the Jn level, including the natural and collisional width ν_{Jn} and the Doppler width $k\bar{v}$. In this case (see (3.6), (3.9), and (2.19)),

$$\begin{aligned} \frac{\omega^2}{c^2}\frac{\partial\varepsilon^{\mathrm{tr}}(\omega, k)}{\partial k^2} &= \frac{8\pi e^2 N}{3(mc)^2(2J_0+1)} \\ &\times\left\{-\frac{1}{8}J_0(J_0+1)g_0^2\Phi_0\right. \\ &+\sum_{Jn}{}'\frac{\omega_{Jn,J_0n_0}}{\omega_{Jn,J_0n_0}^2-\omega^2}\left[\frac{1}{4}|\langle J_0n_0\|S\|Jn\rangle|^2\right. \\ &+\frac{m^2\omega_{Jn,J_0n_0}^2}{120e^2}|\langle J_0n_0\|Q\|Jn\rangle|^2 \\ &+\frac{m^2\omega_{Jn,J_0n_0}^2}{Me^2(\omega^2-\omega_{Jn,J_0n_0}^2)}|\langle J_0n_0\|d\|Jn\rangle|^2 \\ &\left.\left.\times\left(\frac{\omega^2+\omega_{Jn,J_0n_0}^2}{2\omega_{Jn,J_0n_0}}+T\frac{3\omega^2+\omega_{Jn,J_0n_0}^2}{\omega^2-\omega_{Jn,J_0n_0}^2}\right)\right]\right\}. \end{aligned} \tag{4.11}$$

Assuming that ω and $|\delta_{Jn}| \sim \omega_R$ in (4.11) and using (3.22), we find that

$$\frac{\omega^2}{c^2}\left|\frac{\partial\varepsilon^{\mathrm{tr}}}{\partial k^2}\right| \sim Na_B^3\frac{\omega_R}{mc^2} \ll 1. \tag{4.12}$$

It follows from this relation and from (4.7) that $\mathbf{u}_{gr} \parallel \mathbf{u}_f$ for waves with frequencies far from the atomic transition frequencies.

At low frequencies ($\omega \ll T$), equality (4.11) reduces to (see (3.28))

$$\begin{aligned} \frac{\omega^2}{c^2}\frac{\partial\varepsilon^{\mathrm{tr}}(\omega, k)}{\partial k^2} &= \frac{2\pi e^2 N}{3(mc)^2}\left[-\frac{m}{\sqrt{2L_0+1}}\right. \\ &\times\langle L_0n_0\|R_0\|L_0n_0\rangle+\frac{1}{2T}J_0(J_0+1)g_0^2 \\ &\left.+\frac{1}{2J_0+1}\sum_{Jn}{}'\frac{|\langle J_0n_0\|S\|Jn\rangle|^2}{\omega_{Jn,J_0n_0}}\right]. \end{aligned} \tag{4.13}$$

If $g_0 = 0$, then (4.12) remains valid; if, however, $g_0 \neq 0$, then

$$\begin{aligned} \frac{\omega^2}{c^2}\frac{\partial\varepsilon^{\mathrm{tr}}(\omega, k)}{\partial k^2} &= \frac{4\pi\mu_B^2 N}{3T}g_0^2J_0(J_0+1) \\ &\sim Na_B^3\frac{\omega_R}{mc^2}\frac{\omega_R}{T} \ll 1. \end{aligned} \tag{4.14}$$

Consequently, $\mathbf{u}_{gr} \parallel \mathbf{u}_f$ for low-frequency waves as well.

It remains to consider a wave with a frequency close to ω_{Jn,J_0n_0} of a particular atomic transition: condition (3.1) is satisfied, but $|\delta_{Jn}| \gg \tilde{\nu}_{Jn}$ as before. In this case, equality (4.11) reduces to (see (3.11) and (2.19))

$$\begin{aligned} \frac{\omega^2}{c^2}\frac{\partial\varepsilon^{\mathrm{tr}}(\omega, k)}{\partial k^2} &= -\frac{4\pi e^2 N}{3(mc)^2(2J_0+1)\delta_{Jn}} \\ &\times\left\{\frac{1}{4}|\langle J_0n_0\|S\|Jn\rangle|^2+2\mathrm{Re}\langle J_0n_0\|p\|Jn\rangle\right. \\ &\times\left\langle Jn\middle\|C^{\mathrm{tr}}\middle\|J_0n_0\right\rangle+\frac{m^2\omega_{Jn,J_0n_0}^2}{e^2} \\ &\times\left[\frac{1}{120}|\langle J_0n_0\|Q\|Jn\rangle|^2\right. \\ &\left.\left.\left.+\frac{1}{M\delta_{Jn}}|\langle J_0n_0\|d\|Jn\rangle|^2\left(\frac{1}{2}+\frac{T}{\delta_{Jn}}\right)\right]\right\}. \end{aligned} \tag{4.15}$$

For the Jn levels to which magnetic dipole and electric quadrupole transitions are permitted, we find that

$$\frac{\omega^2}{c^2}\left|\frac{\partial\varepsilon^{\mathrm{tr}}}{\partial k^2}\right| \sim Na_B^3\frac{\omega_R}{mc^2}\frac{\omega_R}{|\delta_{Jn}|} \ll 1, \tag{4.16}$$

because even $\omega_R/\nu_{Jn} < 10^7$ and $|\delta_{Jn}| \gg \tilde{\nu}_{Jn} > \nu_{Jn}$.

For the Jn levels to which an electric dipole transition is permitted, we obtain from (4.15)

$$\begin{aligned} \frac{\omega^2}{c^2}\frac{\partial\varepsilon^{\mathrm{tr}}(\omega, k)}{\partial k^2} &= -\frac{2\pi N}{3Mc^2(2J_0+1)} \\ &\times\frac{\omega_{Jn,J_0n_0}^2}{\delta_{Jn}^2}|\langle J_0n_0\|d\|Jn\rangle|^2\left(1+\frac{2T}{\delta_{Jn}}\right). \end{aligned} \tag{4.17}$$

The contribution of the first term on the right-hand side is on the order of $Na_B^3(\omega_R/mc^2)(m/M)(\omega_R/\delta_{Jn})^2$. Since $m/M < 10^{-3}$, the first term is on the order of unity only for $|\delta_{Jn}| < 10^{-6}\omega_R$. If, however, $|\delta_{Jn}| \gg \tilde{\nu}_{Jn}$, then the first term on the right-hand side of (4.17) is smaller than

unity in order of magnitude. The second term on the right-hand side of (4.17) differs in order of magnitude from the first term by a factor of $(T/\omega_R)(\omega_R/|\delta_{Jn}|)$, so it can become approximately equal to unity even for $|\delta_{Jn}| \le 10^{-4}\omega_R$. If, in this case, $\delta_{Jn} < 0$, then the second condition (4.6) is also satisfied, and, according to (4.7), the group and phase velocities of the wave can be antiparallel.

Thus, the group velocity of the transverse electromagnetic wave in a monatomic gas at all frequencies coincides in direction with its phase velocity, except for frequencies slightly detuned from the frequencies of electric dipole transitions toward longer wavelengths. In this case (see (3.11) and (3.12)),

$$\varepsilon^{l}(\omega, k) = \varepsilon^{\mathrm{tr}}(\omega, k), \quad \mu(\omega, k) = 1,$$

and conditions (4.9) and (4.10) are inapplicable.

REFERENCES

1. V. P. Silin and A. A. Rukhadze, *Electromagnetic Properties of Plasma and Plasmalike Media* (Gosatomizdat, Moscow, 1961).
2. V. M. Agranovich and V. L. Ginzburg, *Crystal Optics with Spatial Dispersion, and Excitons*, 2nd ed. (Nauka, Moscow, 1979; Springer, New York, 1984).
3. E. M. Lifshitz and L. P. Pitaevskiĭ, *Physical Kinetics* (Nauka, Moscow, 1979; Pergamon Press, Oxford, 1981).
4. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 5: *Statistical Physics*, 3rd ed. (Nauka, Moscow, 1976; Pergamon Press, Oxford, 1980), Part 1.
5. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory*, 3rd ed. (Nauka, Moscow, 1974; Pergamon, New York, 1977).
6. E. M. Lifshitz and L. P. Pitaevskiĭ, *Course of Theoretical Physics*, Vol. 5: *Statistical Physics* (Nauka, Moscow, 1978; Pergamon Press, Oxford, 1980), Part 2.
7. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 1: *Mechanics*, 3rd ed. (Nauka, Moscow, 1973; Pergamon Press, Oxford, 1976).
8. V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ, *Quantum Electrodynamics*, 2nd ed. (Nauka, Moscow, 1980; Pergamon Press, Oxford, 1982).
9. L. A. Vaĭnshteĭn, I. I. Sobel'man, and E. A. Yukov, *Excitation of Atoms and Broadening of Spectral Lines* (Nauka, Moscow, 1979; Springer, Berlin, 1981).
10. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media*, 2nd ed. (Nauka, Moscow, 1982; Pergamon Press, Oxford, 1984).
11. A. A. Houck, J. B. Brock, and I. L. Chung, Phys. Rev. Lett. **90**, 137401 (2003).
12. L. I. Mandel'shtam, Zh. Éksp. Teor. Fiz. **15**, 475 (1945); in *Complete Works* (Akad. Nauk SSSR, Moscow, 1947), Vol. 2, p. 334; in *Complete Works* (Akad. Nauk SSSR, Moscow, 1950), Vol. 5, p. 461.
13. V. G. Veselago, Usp. Fiz. Nauk **92**, 517 (1967) [Sov. Phys. Usp. **10**, 509 (1968)].

Translated by V. Astakhov

Journal of Experimental and Theoretical Physics, Vol. 98, No. 2, 2004, pp. 316–321.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 2, 2004, pp. 356–361.

SOLIDS
Structure

Effect of Elastic Deformations on the Critical Behavior of Disordered Systems with Long-Range Interactions

S. V. Belim
Omsk State University, pr. Mira 55, Omsk, 644077 Russia
e-mail: belim@univer.omsk.su
Received June 25, 2003

Abstract—A field-theoretic approach is applied to describe behavior of three-dimensional, weakly disordered, elastically isotropic, compressible systems with long-range interactions at various values of a long-range interaction parameter. Renormalization-group equations are analyzed in the two-loop approximation by using the Padé–Borel summation technique. The fixed points corresponding to critical and tricritical behavior of the systems are determined. Elastic deformations are shown to changes in critical and tricritical behavior of disordered compressible systems with long-range interactions. The critical exponents characterizing a system in the critical and tricritical regions are determined. © *2004 MAIK "Nauka/Interperiodica".*

The effect of long-range interaction described by the power law $1/r^{-D-a}$ was studied analytically in the framework of the ε expansion [1–3] and numerically by Monte Carlo methods [4–6] for two- and one-dimensional systems. It was shown that effects due to long-range interaction are essential for the critical behavior of Ising systems when $a < 2$. The two-loop approximation applied in [7] directly in the three-dimensional space corroborates the predictions of the ε expansion for homogeneous systems with long-range interactions.

In structural phase transitions without piezoelectric effect in the paraphase, elastic deformations play the role of a secondary order parameter whose fluctuations are not critical in most cases [8, 9]. Since the main contribution to striction effects in the critical region is due to the distance dependence of the exchange integral, only isotropic elastic systems are discussed below.

It was shown in [10, 11] that coupling between order-parameter fluctuations and elastic deformations can be responsible both for change in critical behavior and for emergence of tricritical and tetracritical points. The introduction of frozen point impurities into a system both changes the critical behavior and eliminates multicritical points [12]. The analysis presented in [13] showed that frozen structural defects introduced into spin systems with long-range interactions increased the value of a long-range parameter corresponding to the transition to mean-field critical behavior. The effect of elastic deformations on homogeneous systems with long-range interactions also changes the critical behavior [14]. Therefore, it would be interesting to analyze the combined effect of structural defects and elastic deformations on systems with long-range interactions.

In this work, the critical behavior of disordered compressible systems with long-range interactions in the three-dimensional space is described for various values of the long-range parameter *a*.

The Hamiltonian of the disordered Ising model including elastic deformations and long-range interactions can be written as

$$\begin{aligned} H_0 = {} & \frac{1}{2}\int d^D q(\tau_0 + q^a) S_q S_{-q} + \frac{1}{2}\int d^D q \Delta\tau_q S_q S_{-q} \\ & + u_0 \int d^D q S_{q1} S_{q2} S_{q3} S_{-q1-q2-q3} \\ & + a_3 \int d^D q y_{q1} S_{q2} S_{-q1-q2} \\ & + \frac{a_3^{(0)}}{\Omega} y_0 \int d^D q S_q S_{-q} + \frac{1}{2} a_1 \int d^D q y_q y_{-q} \\ & + \frac{1}{2}\frac{a_1^{(0)}}{\Omega} y_0^2 + \int d^D q h_q y_q + \frac{h_0}{\Omega} y_0, \end{aligned} \tag{1}$$

where S_q is an order parameter, u_0 is a positive constant, $\tau_0 \sim |T - T_c|/T_c$, T_c is the phase-transition temperature, *a* is the long-range parameter, $\Delta\tau_q$ is a random impurity field (e.g., random temperature), a_1 and a_2 are the elastic constants of a crystal, and a_3 is the quadratic striction constant. The coupling between impurities and the nonfluctuating order parameter

$$y(x) = \sum_{\alpha=1}^{3} u_{\alpha\alpha}(x),$$

where $u_{\alpha\beta}$ is the strain tensor, is specified by the random field h_q thermodynamically conjugate to $u_{\alpha\alpha}(x)$. In Eq. (1), integration is performed in the terms depending on nonfluctuating variables not coupled to the order

parameter S_q, and the y_0 terms (describing homogeneous deformations) are separated. It was shown in [8] that this separation is necessary because inhomogeneous deformations y_q are responsible for exchange of acoustic phonons and for long-range effects that are absent under homogeneous deformations.

When the impurity concentration is low, random fields $\Delta\tau_q$, h_q, and h_0 can be treated as Gaussian and specified by the function

$$P[\Delta\tau, h, h_0] = A\exp\left[-\frac{1}{8b_1}\int\Delta\tau_q^2 d^D q - \frac{1}{8b_2}\int h_q^2 d^D q - \frac{1}{8b_3}\int h_0 d^D q - \frac{1}{4b_4}\int\Delta\tau_q h_q d^D q - \frac{1}{4b_5}\int\Delta\tau_q h_0 d^D q\right], \quad (2)$$

where A is a normalization factor and b_i denotes positive constants proportional to the concentration of frozen structural defects.

Using the replica procedure to average over the random fields generated by frozen structural defects, we obtain an effective Hamiltonian of the system:

$$H_R = \frac{1}{2}\int d^D q(\tau_0 + q^a)\sum_{a=1}^{m} S_q^a S_{-q}^a - \frac{\delta_0}{2}\sum_{a,b=1}^{m}\int d^D q(S_{q1}^a S_{q2}^a)(S_{q3}^b S_{-q1-q2-q3}^b) + u_0\sum_{a=1}^{m}\int d^D q S_{q1}^a S_{q2}^a S_{q3}^a S_{-q1-q2-q3}^a + g_0\sum_{a=1}^{m}\int d^D q y_{q1}^a S_{q2}^a S_{-q1-q2}^a + \frac{g_0^{(0)}}{\Omega}\sum_{a=1}^{m} y_0^a\int d^D q S_q^a S_{-q}^a + \frac{1}{2}\lambda\int d^D q y_q y_{-q} + \frac{1}{2}\frac{\lambda_0}{\Omega}y_0^2, \quad (3)$$

where the positive constants δ_0, g_0, $g_0^{(0)}$, λ, and λ_0 can be expressed in terms of a_i and b_i. The properties of the original system can be obtained in the limit $m \longrightarrow 0$, where m is the number of replicas (images).

Define an effective Hamiltonian depending only on the strongly fluctuating order parameter S by the relation

$$\exp\{-H[S]\} = B\int\exp\{-H_R[S, y]\}\prod dy_q. \quad (4)$$

If an experiment is carried out at constant volume, then y_0 is a constant and integration in Eq. (4) is performed over inhomogeneous deformations, whereas homogeneous deformations do not contribute to the effective Hamiltonian. At constant pressure, the Hamiltonian is modified by adding the term $P\Omega$, volume is represented in terms of the strain tensor components as

$$\Omega = \Omega_0\left[1 + \sum_{\alpha=1} u_{\alpha\alpha} + \sum_{\alpha\neq\beta} u_{\alpha\alpha}u_{\beta\beta} + O(u^3)\right], \quad (5)$$

and integration over homogeneous deformations is also performed in Eq. (4). It was noted in [15] that the quadratic terms in Eq. (5) can be important at high pressures and for crystals with large striction effects. The result is

$$H = \frac{1}{2}\int d^D q(\tau_0 + q^a)\sum_{a=1}^{m} S_q^a S_{-q}^a + \left(u_0 - \frac{z_0}{2}\right)\sum_{a=1}^{m}\int d^D\{q_i\}S_{q1}^a S_{q2}^a S_{q3}^a S_{-q1-q2-q3}^a - \frac{\delta}{2}\sum_{a,b=1}^{m}\int d^D\{q_i\}(S_{q1}^a S_{q2}^a)(S_{q3}^b S_{-q1-q2-q3}^b) + \frac{1}{2\Omega}(z_0 - w_0)\sum_{a=1}^{m}\int d^D\{q_i\}(S_{q1}^a S_{-q1}^a)(S_{q2}^a S_{-q2}^a), \quad (6)$$

$$z_0 = g_0^2/\lambda, \quad w_0 = g_0^{(0)2}/\lambda_0.$$

The effective coupling constant $v_0 = u_0 - z_0/2$, which arises in the Hamiltonian owing to the striction effects determined by g_0, can have positive and negative values. Therefore, this Hamiltonian describes both first- and second-order phase transitions. When $v_0 = 0$, the system exhibits tricritical behavior. Furthermore, the effective interaction determined by the difference $z_0 - w_0$ in Eq. (6) can also change the order of the phase transition. This representation of the effective Hamiltonian entails the existence of a higher order critical point where tricritical curves intersect when $v_0 = 0$ and $z_0 = w_0$ [16]. Note that Hamiltonian (6) is isomorphic to the Hamiltonian of the disordered Ising model with long-range interactions under the tricritical condition $z_0 = w_0$.

Using the standard renormalization-group procedure based on the Feynman diagram technique [17, 18] with the propagator $G(\mathbf{k}) = 1/(\tau + |\mathbf{k}|^q)$, we derive the following expressions for the functions β_v, β_δ, β_z, β_w, γ_t, and γ_φ specifying the differential renormalization-group equation:

$$\beta_v = -(4-D)v\Big[1-36vJ_0+24\delta J_0+1728\Big(2J_1-J_0^2-\frac{2}{9}G\Big)v^2$$
$$-2304\Big(2J_1-J_0^2-\frac{1}{6}G\Big)v\delta+672\Big(2J_1-J_0^2-\frac{2}{3}G\Big)\delta^2\Big],$$
$$\beta_\delta = -(4-D)\delta\Big[1-24vJ_0+16\delta J_0+576\Big(2J_1-J_0^2-\frac{2}{3}G_1\Big)v^2$$
$$-1152\Big(2J_1-J_0^2-\frac{1}{3}G\Big)v\delta+352\Big(2J_1-J_0^2-\frac{1}{22}G\Big)\delta^2\Big],$$
$$\beta_z = -(4-D)z\Big[1-24vJ_0-2zJ_0+8\delta J_0+576\Big(2J_1-J_0^2-\frac{2}{3}G\Big)v^2$$
$$-120\Big(2J_1-J_0^2-\frac{8}{5}G\Big)v\delta+96\Big(2J_1-J_0^2-\frac{2}{3}G\Big)\delta^2\Big],$$
$$\beta_w = -(4-D)w\Big[1-24vJ_0+8\delta J_0-4zJ_0+2wJ_0+576\Big(2J_1-J_0^2-\frac{2}{3}G\Big)v^2 \qquad (7)$$
$$-120\Big(2J_1-J_0^2-\frac{8}{5}G\Big)v\delta+96\Big(2J_1-J_0^2-\frac{2}{3}G\Big)\delta^2\Big],$$
$$\gamma_t = (4-D)\Big[-12vJ_0+4\delta J_0-2zJ_0+2wJ_0+288\Big(2J_1-J_0^2-\frac{1}{3}G\Big)v^2$$
$$-288\Big(2J_1-J_0^2-\frac{2}{3}G\Big)v\delta+32\Big(2J_1-J_0^2-\frac{1}{2}G\Big)\delta^2\Big],$$
$$\gamma_\varphi = (4-D)64G(3v^2-3v\delta+\delta^2),$$
$$J_1 = \int\frac{d^Dqd^Dp}{(1+|\mathbf{q}|^a)^2(1+|\mathbf{p}|^a)(1+|q^2+p^2+2\mathbf{p}\cdot\mathbf{q}|^{a/2})},$$
$$J_0 = \int\frac{d^Dq}{(1+|\mathbf{q}|^a)^2},$$
$$G = -\frac{\partial}{\partial|\mathbf{k}|^a}\int\frac{d^Dqd^Dp}{(1+|q^2+k^2+2\mathbf{k}\cdot\mathbf{q}|^a)(1+|\mathbf{p}|^a)(1+|q^2+p^2+2\mathbf{p}\cdot\mathbf{q}|^{a/2})}.$$

In terms of the new effective interaction vertices

$$v_1 = \frac{v}{J_0}, \quad v_2 = \frac{\delta}{J_0}, \quad v_3 = \frac{z}{J_0}, \quad v_4 = \frac{w}{J_0}, \qquad (8)$$

the functions β_i, γ_t, and γ_φ are expressed as

$$\beta_1 = -(4-D)\Big[1-36v_1+24v_2+1728\Big(2\tilde{J}_1-1-\frac{2}{9}\tilde{G}\Big)v_1^2$$
$$-2304\Big(2\tilde{J}_1-1-\frac{1}{6}\tilde{G}\Big)v_1v_2+672\Big(2\tilde{J}_1-1-\frac{2}{3}\tilde{G}\Big)v_2^2\Big],$$
$$\beta_2 = -(4-D)\delta\Big[1-24v_1+8v_2+576\Big(2\tilde{J}_1-1-\frac{2}{3}\tilde{G}\Big)v_1^2$$

$$-1152\left(2\tilde{J}_1-1-\frac{1}{3}\tilde{G}\right)v_1v_2+352\left(2\tilde{J}_1-1-\frac{1}{22}\tilde{G}\right)v_2^2\Big],$$

$$\beta_3 = -(4-D)v_3\Big[1-24v_1+16v_2-2v_3+576\left(2\tilde{J}_1-1-\frac{2}{3}\tilde{G}\right)v_1^2$$

$$-120\left(2\tilde{J}_1-1-\frac{8}{5}G\right)v_1v_2+96\left(2\tilde{J}_1-1-\frac{2}{3}G\right)v_2^2\Big], \quad (9)$$

$$\beta_4 = -(4-D)v_4\Big[1-24v_1+8v_2-4v_3+2v_4+576\left(2\tilde{J}_1-1-\frac{2}{3}\tilde{G}\right)v_1^2$$

$$-120\left(2\tilde{J}_1-1-\frac{8}{5}G\right)v_1v_2+96\left(2\tilde{J}_1-1-\frac{2}{3}G\right)v_2^2\Big],$$

$$\gamma_t = (4-D)\Big[-12v_1+4v_2-2v_3+2v_4+288\left(2\tilde{J}_1-1-\frac{1}{3}\tilde{G}\right)v_1^2$$

$$-192\left(2\tilde{J}_1-1-\frac{2}{3}\tilde{G}\right)v_1v_2+32\left(2\tilde{J}_1-1-\frac{1}{2}\tilde{G}\right)v_2^2\Big],$$

$$\gamma_\varphi = (4-D)64\tilde{G}(3v_1^2-3v_1v_2+v_2^2).$$

This redefinition is meaningful for $a \le D/2$. In this case, J_0, J_1, and G are divergent functions. Introducing the cutoff parameter Λ, we obtain finite expressions for the ratios

$$\frac{J_1}{J_0^2} = \frac{\int_0^\Lambda\int_0^\Lambda d^Dq\,d^Dp/((1+|\mathbf{q}|^a)^2(1+|\mathbf{p}|^a)(1+|q^2+p^2+2\mathbf{p}\cdot\mathbf{q}|^a))}{\left[\int_0^\Lambda d^Dq/(1+|\mathbf{q}|^a)^2\right]^2},$$

$$\frac{G}{J_0^2} = \frac{-\partial/(\partial|\mathbf{k}|^a)\int_0^\Lambda\int_0^\Lambda d^Dq\,d^Dp/((1+|q^2+k^2+2\mathbf{k}\cdot\mathbf{q}|^a)(1+|\mathbf{p}|^a)(1+|q^2+p^2+2\mathbf{p}\cdot\mathbf{q}|^a))}{\left[\int_0^\Lambda d^Dq/(1+|\mathbf{q}|^a)^2\right]^2}, \quad (10)$$

as $\Lambda \longrightarrow \infty$.

The integrals are performed numerically. For $a \le D/2$, a sequence of J_1/J_0^2 and G/J_0^2 corresponding to various values of Λ is calculated and extrapolated to infinity.

It is well known that perturbation-theory series are asymptotic and expressions (9) cannot be applied directly since the interaction vertices for fluctuations of order parameters in the fluctuation region are too large. For this reason, the required physical information was extracted from these expressions by applying the Padé–Borel method extended to the four-parameter case. The appropriate direct and inverse Borel transforms have the form

$$f(v,\delta,z,w) = \sum_{i_1,\ldots,i_4} c_{i_1,\ldots,i_4} v_1^{i_1}v_2^{i_2}v_3^{i_3}v_4^{i_4}$$

$$= \int_0^\infty e^{-t}F(v_1t, v_2t, v_3t, v_4t)\,dt, \quad (11)$$

$$F(v,\delta,z,w)$$

$$= \sum_{i_1,\ldots,i_4} \frac{c_{i_1,\ldots,i_4}}{(i_1+\ldots+i_4)!} v_1^{i_1}v_2^{i_2}v_3^{i_3}v_4^{i_4}.$$

Fixed points and eigenvalues of the stability matrix (for complex eigenvalues, only real parts are presented)

No.	v_1^*	v_2^*	v_3^*	v_4^*	b_1	b_2	b_3	b_4
				$a = 1.8$				
1	0.064189	0.046878	0	0	0.626*	0.626*	−0.123	−0.123
2	0.064189	0.046878	0.066101	0	0.626*	0.626*	0.124	0.125
3	0.064189	0.046878	0.066101	0.066101	0.626*	0.626*	0.124	−0.124
				$a = 1.9$				
4	0.066557	0.040818	0	0	0.559*	0.559*	−0.118	−0.118
5	0.066557	0.040818	0.065716	0	0.559*	0.559*	0.119	0.119
6	0.066557	0.040818	0.065716	0.065716	0.559*	0.559*	0.119	−0.119

To obtain an analytic continuation of the Borel transform of a function, a series in an auxiliary variable θ is introduced:

$$\tilde{F}(v, \delta, z, w, \theta) = \sum_{k=0}^{\infty} \theta^k \sum_{i_1, \ldots, i_4} \frac{c_{i_1, \ldots, i_4}}{k!} v_1^{i_1} v_2^{i_2} v_3^{i_3} v_4^{i_4} \delta_{i_1 + \ldots + i_4, k}. \quad (12)$$

The Padé approximant $[L/M]$ is applied to this series at $\theta = 1$. This procedure was proposed and tested in [19–22] for describing the critical behavior of a number of systems characterized by several interaction vertices for order-parameter fluctuations. It was found in [19–22] that Padé approximants in the variable θ preserve the symmetry of the system. This property is important for analysis of multivertex models. The approximants [2/1] are used to calculate the β functions in the two-loop approximation.

Critical behavior is completely determined by the stable fixed points of the renormalization group transformation. These points can be found from the condition

$$\beta_i(v_1^*, v_2^*, v_3^*, v_4^*) = 0 \quad (i = 1, 2, 3, 4). \quad (13)$$

The requirement of stability of a fixed point reduces to the condition that the eigenvalues b_i of the matrix

$$B_{i,j} = \frac{\partial \beta_i(v_1^*, v_2^*, v_3^*, v_4^*)}{\partial v_j} \quad (14)$$

be positive.

The exponent ν characterizing the growth of the correlation radius near the critical point ($R_c \propto |T - T_c|^{-\nu}$) is determined by the relation

$$\nu = \frac{1}{2}(1 + \gamma_t)^{-1}.$$

The Fisher exponent η describing the behavior of the correlation function near the critical point in the wave vector space ($G \propto k^{2+\eta}$) is determined by the scaling function γ_φ: $\eta = \gamma_\varphi(v_1^*, v_2^*, v_3^*, v_4^*)$. The remaining critical exponents can be determined from the scaling relations.

The table shows the stable fixed points of the renormalization group transformation and the fixed-point eigenvalues of the stability matrix for $a = 1.8$ and 1.9. It was shown in [13] that stable fixed points exist in the physical region ($v_i^* > 0$) for disordered systems only when $a \geq 1.8$. When $a < 1.8$, the stable points of any three-dimensional impurity system are characterized by negative values of the vertex v_1^*.

An analysis of the critical points and their stability shows that the fixed critical points of disordered systems with long-range interactions (nos. 1 and 4) are unstable under elastic deformations. The critical behavior of disordered compressible systems with long-range interactions is described by their respective fixed points (nos. 2 and 5). The phase diagram of the substance can contain tricritical points specified by fixed points (nos. 3 and 6).

Calculations of the critical exponents for the stable fixed points (nos. 2 and 5) yield

$$\begin{aligned} a &= 1.9, \quad \nu = 0.685, \quad \eta = 0.034, \\ a &= 1.8, \quad \nu = 0.682, \quad \eta = 0.051. \end{aligned} \quad (15)$$

For the tricritical points (nos. 3 and 6), the critical exponents are

$$\begin{aligned} a &= 1.9, \quad \nu = 0.652, \quad \eta = 0.034, \\ a &= 1.8, \quad \nu = 0.649, \quad \eta = 0.051. \end{aligned} \quad (16)$$

Thus, elastic deformations change both critical and tricritical behavior of three-dimensional disordered Ising systems with long-range interactions.

REFERENCES

1. M. E. Fisher, S.-K. Ma, and B. G. Nickel, Phys. Rev. Lett. **29**, 917 (1972).
2. J. Honkonen, J. Phys. A **23**, 825 (1990).
3. E. Luijten and H. Mebingfeld, Phys. Rev. Lett. **86**, 5305 (2001).
4. E. Bayong and H. T. Diep, Phys. Rev. B **59**, 11919 (1999).
5. E. Luijten, Phys. Rev. E **60**, 7558 (1999).
6. E. Luijten and H. W. J. Blooöte, Phys. Rev. B **56**, 8945 (1997).
7. S. V. Belim, Pis'ma Zh. Éksp. Teor. Fiz. **77**, 118 (2003) [JETP Lett. **77**, 112 (2003)].
8. A. I. Larkin and S. A. Pikin, Zh. Éksp. Teor. Fiz. **56**, 1664 (1969) [Sov. Phys. JETP **29**, 891 (1969)].
9. D. J. Bergman and B. I. Halperin, Phys. Rev. B **13**, 2145 (1976).
10. V. M. Laptev and Yu. N. Skryabin, Phys. Status Solidi B **91**, K143 (1979).
11. Y. N. Skryabin and A. V. Shchanov, Phys. Lett. A **234**, 147 (1997).
12. S. V. Belim and V. V. Prudnikov, Fiz. Tverd. Tela (St. Petersburg) **43**, 1299 (2001) [Phys. Solid State **43**, 1353 (2001)].
13. S. V. Belim, Pis'ma Zh. Éksp. Teor. Fiz. **77**, 509 (2003) [JETP Lett. **77**, 434 (2003)].
14. S. V. Belim, Pis'ma Zh. Éksp. Teor. Fiz. **77**, 659 (2003) [JETP Lett. **77**, 556 (2003)].
15. M. A. de Moura, T. C. Lubensky, Y. Imry, and A. Aharony, Phys. Rev. B **13**, 2176 (1976).
16. Y. Imry, Phys. Rev. Lett. **33**, 1304 (1974).
17. D. Amit, *Field Theory, the Renormalization Group, and Critical Phenomena* (McGraw-Hill, New York, 1976).
18. J. Zinn-Justin, *Quantum Field Theory and Critical Phenomena* (Clarendon Press, Oxford, 1989).
19. S. A. Antonenko and A. I. Sokolov, Phys. Rev. B **49**, 15901 (1994).
20. K. B. Varnashev and A. I. Sokolov, Fiz. Tverd. Tela (St. Petersburg) **38**, 3665 (1996) [Phys. Solid State **38**, 1996 (1996)].
21. A. I. Sokolov, K. B. Varnashev, and A. I. Mudrov, Int. J. Mod. Phys. B **12**, 1365 (1998).
22. A. I. Sokolov and K. B. Varnashev, Phys. Rev. B **59**, 8363 (1999).

Translated by R. Tyapaev

Journal of Experimental and Theoretical Physics, Vol. 98, No. 2, 2004, pp. 322–333.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 2, 2004, pp. 362–376.

SOLIDS
Structure

Cascades of Atomic Displacements in Solids: The Ballistic Stage

I. M. Dremin[a], O. V. Ivanov[a], V. A. Nechitailo[a], V. P. Shevel'ko[a], N. M. Sobolevskiĭ[b], and A. V. Subbotin[c]

[a]*Lebedev Institute of Physics, Russian Academy of Sciences, Moscow, 119991 Russia*
[b]*Institute for Nuclear Research, Russian Academy of Sciences, Moscow, 117312 Russia*
[c]*Dollezhal Research and Development Institute of Power Engineering, Moscow, 101000 Russia*
e-mail: ivanov@lpi.ru
Received June 9, 2003

Abstract—The evolution of cascades of atomic displacements in solids is analyzed. The charge-exchange, ionization, and elastic scattering cross sections are calculated for the atoms and ions involved in cascade evolution. The effects due to the material density are taken into account. These results are used to perform the first Monte Carlo computations of cascades based on the knowledge of microscopic processes without invoking phenomenological potentials. The proposed approach is unique in that detailed characteristics of atomic processes are obtained by *ab initio* calculation and applied to analyze cascades of atomic displacements. Subcascade development is described, and a relation between the number of Frenkel pairs and the energy of the primary knock-on atom is found for a wide energy range. This provides a basis for characterizing the dose dependence of the hardening of reactor pressure vessel steel and for comparing the effects of primary radiation damage for fission and fusion reactors. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Interactions of atomic particles passing through matter have been studied for a long time (e.g., see [1–3]). Understanding of the physical principles that underlie the key processes and calculation of their consequences are required primarily for evaluating the performance of structural materials used in nuclear power engineering. Current progress in computer simulation makes it possible to examine these processes in more detail. First, we can perform *ab initio* calculations of cross sections for various atomic collision processes and use the results to simulate cascades of atomic displacements in solids. Second, new numerical approaches based on the use of wavelet analysis in solving three-dimensional heat and mass transfer equations can be applied to investigate the quasi-thermodynamic stage of a cascade. In this paper, we consider the ballistic stage of the evolution of a cascade of atomic displacements in iron. The subsequent quasi-thermodynamic stage is analyzed in [4]. We consider metals irradiated to neutrons with energy sufficiently high to damage the crystalline structure. Further analysis shows that our approach applies to the case of irradiation by high-energy ions as well.

Irradiation-induced effects can be loosely defined as the sequence of events initiated in a solid by the neutron–nucleus interaction in which a primary knock-on atom (PKA) giving rise to a cascade of atom–atom interactions is created. This leads to the development of regions (with size varying from tens to hundreds of angstroms) characterized by relatively intense energy release (thermal spikes) and high concentration of interstitial–vacancy (Frenkel) pairs [5–8]. Their diffusive spread, driven by the elevated temperatures in thermal spikes and the gradients of the chemical potential of point defects, results in the formation of a supersaturated two-component solid solution of interstitials and vacancies. Its subsequent dissociation initiates the evolution of microstructural elements: nucleation, growth, and climb of various dislocation loops and dislocation elements; nucleation and growth of voids; "dissolution" and formation of ensembles of fine phase particles; diffusion of alloying and impurity atoms giving rise to dislocation atmospheres and segregations at grain boundaries; etc. The evolution of a microstructure, in turn, leads to irradiation-induced macroscale effects, such as radiation growth, swelling, and changes in electrical and thermal conductivity and in mechanical properties.

At high temperatures ($T \geq 0.3T_m$, where T_m is the melting point), the irradiation-induced effects are due solely to the dissociation of the supersaturated solid solution. It is well known that these effects can be described by reaction rate equations [9–13]. In addition to constants determined by the structure and properties of the solid and the point defects, one must also define "sources," i.e., calculate the rates of generation of point defects and their total number, which are determined by the flux and energy spectrum of neutrons.

Cascade evolution in solids has been discussed in extensive literature [14–19]. Recent publications were mainly focused on the number of Frenkel pairs that survive after the ballistic stage and on spontaneous recombination of such pairs and point-defect clusters, which can substantially contribute to the formation of a supersaturated solid solution. Cascades with PKA energies up to 60 keV in solids have been simulated by using molecular-dynamics models that take into account crystalline structure, but ignore ionization of atoms and electron capture by ions [18, 20].

However, irradiation-induced effects of different nature manifest themselves at temperatures of about $0.3T_m$, including the irradiation-induced hardening of VVER reactor pressure vessel steel. This effect is caused by the development of an ensemble of nanoscale precipitates initiated by subcascades of certain size. Therefore, description of subcascading (i.e., splitting of cascades into smaller cascades having well-defined parameters [4]) and evolution of particular subcascades at the quasi-thermodynamic stage is required to understand both precipitate formation and hardening kinetics as determined by the neutron energy spectrum and dose rate.

The present study of cascade evolution is focused on characterization of subcascades in a wide range of PKA energy, which can be done by treating the material as amorphous. Since the starting point of our analysis is PKA production, the methods developed here can also be applied to describe the effects induced by ion irradiation.

2. CASCADE EVOLUTION: AN OVERVIEW

Cascades of atomic displacements develop in a material irradiated by particles whose energy is sufficiently high to knock atoms out of their sites. This effect is of interest both for studies of physical processes and for practical applications. The PKA produced by an initial neutron has a high energy. Moving through the material, it collides with another atom and knocks it out of its site. When a relatively small amount of energy is transferred, the secondary atom either returns to the ground state after being excited or decelerates because of energy loss. In either case, the material is heated by phonon emission and absorption. When the transferred energy is relatively high, the secondary atom can initiate a new branch of the cascade. As the process repeats, the cascade branches out like a tree and Frenkel pairs are created. In the course of the cascade evolution, the charges of the incident and target atoms can change via ionization or electron capture.

All of these processes are described by the electromagnetic interaction Lagrangian. In any theory with a dimensionless coupling constant (including electrodynamics), a dominant role is played by processes characterized by relatively low energy transfer (recall the infrared singularity associated with massless photons, which is cut off by form factors). The relative significance of each process outlined above is determined by similar factors. Therefore, one can expect the following scenario of cascade evolution (confirmed by calculations) to occur. Initially, the PKA moves along an almost straight path, losing small portions of its energy and occasionally deviating as it displaces another high-energy atom, which initiates a new branch of the cascade. The degree of cascade branching depends on the initial energy and on the relative intensity of various processes in the material. The ballistic stage of the cascade evolution lasts until the initial energy is completely converted into heat in the material. Phase transitions can occur in regions of intense heat release.

Thus, to describe the evolution of a cascade of atomic displacements and possible ensuing effects, the following information is required:

(i) at the initial stage, the characteristics of the knock-on processes initiated by particles with different energies;

(ii) at the ballistic stage, the characteristics of interactions between atoms and ions for different energies of the primary atom (or ion) passing through the material (interaction potentials, charge-exchange cross sections, density effects, etc.);

(iii) at the quasi-thermodynamic stage, the three-dimensional temperature and structure evolution and the initiation of phase-particle formation depending on certain conditions.

These three stages are distinguished by their duration. A PKA is created by a neutron in a very short time, from 10^{-23} to 10^{-20} s, determined by the finite nuclear force range. Atomic collisions and recombination of Frenkel pairs at the ballistic stage last for 10^{-16} to 10^{-13} s. The thermal effects determine the duration of the entire process (about 10^{-10} s in the case of a phase transition characterized by a nonzero latent heat).

Here, we consider the first two stages as a unified ballistic stage, because they determine the dynamics of cascades of atomic displacements in iron irradiated by neutrons. We analyze the displacement of a PKA by an initial neutron and calculate cross sections for atom–atom and atom–ion collisions. These characteristics are used to perform Monte Carlo simulations of cascades. As a result, the key parameters of the regions of maximum energy release are determined.

This approach is unique in that characteristics of atom–atom and atom–ion interactions are calculated *ab initio* and applied to simulate and analyze cascades of atomic displacements.

3. SIMULATION OF CASCADES OF ATOMIC DISPLACEMENTS

To substantiate our method for simulating cascades of atomic displacements in solids, we describe the bal-

listic stage of cascade evolution outlined above in more detail.

1. First, we consider the knock-on process in which a primary iron atom is knocked out by a neutron. Since the atom–nucleus interaction is controlled by nuclear forces, it can be calculated as a sudden perturbation.

In the nonrelativistic case, the kinetic energy transferred from a neutron having a kinetic energy E_n reaches a maximum value $E_{\max}$ when the PKA moves in the same direction:

$$E_{\max} = \frac{4}{A+2}E_n, \quad (1)$$

where A is the mass number ($A = 56$ for iron). Thus, the kinetic energy of the knocked-on iron atom is at least 14.5 times lower than the neutron energy. Note that the angular distribution of scattered neutrons is isotropic in the center-of-mass system and the emission angle of the atom in the laboratory system θ is less than 90° when the energies are relatively low. In this case, the energy of the atom is found by multiplying the right-hand side of (1) by $\cos^2\theta$. The resulting average energy of the atom is lower than the neutron energy by a factor of 29. This implies that the iron energy spectrum is much softer than the neutron spectrum.

The energy of almost every neutron in nuclear power reactors does not exceed 10 MeV. An iron PKA cannot be ionized under this energy constraint. Indeed, the energy of the outermost ($4s^2$) electron in an iron atom is 7.8 eV. One can easily show that it cannot be knocked out when the atom is hit by a neutron with an energy below 11.6 MeV. This estimate is based on the equality of the bound-electron and PKA velocities, which follows from an analysis of sudden perturbations (see [21–23]). The squared velocity of an electron with energy 7.8 eV is about $3 \times 10^{-5}c^2$. The energy of an atom having this velocity is about 800 keV (electron energy multiplied by the atom-to-electron mass ratio). The estimate for the minimum neutron energy follows from Eq. (1). By virtue of the factor $\cos^{-2}\theta$, an even higher energy is required to create a PKA with a nonzero lateral velocity. The energy of electrons in the next-to-outermost ($3d^6$) shell is 9.0 eV. The corresponding lower bound for the neutron energy required to knock an electron out of this shell is 13.6 MeV. However, the probability of ionization is low even for higher energies. For a low nucleus velocity v_A, it is roughly estimated as follows [21]:

$$W \approx v_A^2/c^2 \ll 1. \quad (2)$$

Using the Thomas–Fermi model, one can calculate the total ionization energy, i.e., the energy required to remove all electrons from a neutral atom [24]. The most accurate estimate based on empirical data is $E_e = 16Z^{7/3} = 32$ keV, i.e., much higher than the electron energies in outer shells. These estimates demonstrate that primary ions with $Z = 26$ (completely ionized iron atoms) cannot play any significant role in the problem analyzed here. (The scattering cross section for these particles was used in a recent study of cascades [25].)

We conclude that the development of a cascade in iron exposed to neutrons with energy below 10–20 MeV must be initiated by a neutral PKA.

2. The ballistic stage of cascade evolution is essentially determined by four processes: elastic scattering of atoms (or ions), which preserves the charges of the colliding particles (while some energy is required to knock an atom out of its lattice site); ionization of the incident atom (or ion) by the target atom, which produces an ion with a higher charge; electron capture by the projectile ion, which reduces its charge; and ionization of the target atom. The first and last processes do not change the charge of the projectile, but reduce the projectile energy. These processes must be taken into account. The projectile charge is changed by the second and third processes.

The cross sections for these processes are calculated for the first time in this study by using results borrowed from [26–32]. The essential novelty of the present approach is that we use density functional theory in *ab initio* computations of cross sections without invoking phenomenological adjustable potentials. Moreover, both elastic scattering and charge-changing processes are taken into account. They are extremely important for our conclusions about the cascade evolution. The computations performed in this study are briefly described in Appendices 1 and 2. This description is sufficient to develop a Monte Carlo scheme and elucidate the relative roles played by various processes in cascade evolution according to the calculated cross sections.

Cascades are simulated by using the calculated cross sections in the optical-thickness approximation for an amorphous material (e.g., see [25, 33]). The optical thickness of a material with respect to the processes mentioned above for a projectile atom (or ion with charge Z) of energy E is defined as

$$O_n(Z, E) = \int_0^E \rho\sigma_n(Z, E')\left(\frac{dE'}{dR}\right)^{-1} dE', \quad (3)$$

where ρ is the material density and σ_n is the cross section for elastic scattering (n = el), projectile ionization (n = ip), or electron capture (n = ep) by the projectile ion. The stopping power of the material is expressed as

$$\frac{dE}{dR} = 2I\rho\sigma_{\text{it}}(Z, E) + \rho E\int_0^{\theta_c} \sin^2\frac{\theta}{2}\sigma_{\text{el}}(Z, E, \theta)d\theta, \quad (4)$$

where I is the effective ionization potential for a target atom, σ_{it} is the ionization cross section for the target atom, and $\sigma_{el}(Z, E, \theta)$ is the differential elastic scattering cross section. The integral is calculated to the angle θ_c defined by the minimum energy required to form a Frenkel pair. This energy is set equal to the standard average threshold displacement energy $E_d \approx 40$ eV [34], which is related to the angle θ_c by the kinematic relation $E_d = E\sin^2(\theta_c/2)$ [2].

The first term in Eq. (4) is associated with ionization and electronic excitation of the medium, whose contributions are equal (hence the factor 2 in this term); the second term is associated with phononic excitations. The total elastic scattering cross section, which is contained in (3) and is responsible for cascading, is related to the differential cross section as follows:

$$\sigma_{el}(Z, E) = \int_{\theta_c}^{\pi} \sigma_{el}(Z, E, \theta) d\theta. \quad (5)$$

As a first step in a cascade simulation, the optical thicknesses corresponding to the three possible processes involving ions (or two processes involving atoms) with initial energy E are calculated according to the formulas written out above. The cross sections as functions of energy are obtained in the calculations described in the Appendices.

Figure 1a shows the optical thickness with respect to elastic scattering versus energy for projectile Fe atoms ($Z = 0$) and ions ($Z = 1$, 2, and 8). The largest optical thickness obviously corresponds to scattering of neutral atoms. It decreases with increasing ion charge and has a much lower value for the Fe^{8+} ion (which has no valence electrons).[1]

Figure 1b shows the optical thickness with respect to ionization versus energy for projectile Fe atoms and ions with $Z = 1$ and 2. Note that the thickness is substantially lower for the ions as compared with the atom. Comparing Figs. 1a and 1b, one finds that elastic scattering dominates at energies below about 2 keV; i.e., charge-changing processes are less important at low energies, even though they should still be taken into account.

Figure 1c shows the optical thickness with respect to electron capture versus energy for projectile ions with $Z = 1$, 2, and 3. It is an essentially nonmonotonic function of the ion charge. Figures 1b and 1c demonstrate that the Fe^{2+} ion is characterized by a large mean free path.

[1] Note that mean free path is an exponential of optical thickness. Therefore, even a small change in optical thickness results in a drastic change in mean free path.

The functions $O_n(E)$ are used in the Monte Carlo method involving inverse functions to generate random free paths with respect to elastic scattering or ionization of an atom or ion and with respect to electron capture by an ion [25, 33]. To simulate a particular cascade, the energy lost in each channel is calculated and the process corresponding to the highest (i.e., closest to E) post-collision energy E' of an atom or ion is identified as an event. The coordinates of the event are found with the use of Eq. (4). In the case of a charge-changing event, the simulation is continued iteratively according to the same scheme (with the use of the optical thickness that corresponds to the resulting ion). In the case of elastic scattering, the scattering angle is determined and simulations are performed for both particles.

The relative probabilities of different processes are determined by their respective cross sections and kinematic characteristics. Elastic scattering cross sections are typically smaller than inelastic ones (see Appendices). Therefore, as an atom (or ion) moves through matter, its energy is dissipated mostly via excitation and charge exchange, whereas new cascade branches are initiated less often occasionally. This and other properties of cascades are demonstrated in the next section. Note that the development of a cascade involves recombination of interstitials and vacancies that are sufficiently close to one another. Its impact on cascade properties is also discussed in the next section.

In this paper, we consider the cascades initiated by primary knock-on atoms with energy having a certain value or uniformly distributed over a certain interval. In applications, one must analyze interactions of initial neutrons with target atoms to find the PKA energy distribution. If the neutron spectrum is known for a particular reactor, this can easily be done by using standard tables of experimental data concerning neutron–nucleus interactions in the energy range in question.

4. CHARACTERISTICS OF CASCADES OF ATOMIC DISPLACEMENTS IN SOLIDS

4.1. Structure of Cascades

As a cascade develops, atoms of the solid are knocked out of their lattice sites. Connecting the scattering sites (vacancies) successively with one another and with the corresponding interstitials, one obtains a graph. The number of vertices in a treelike cascade graph is the number of vacancies (or interstitials) minus one. The number of its edges (including the initial one) equals the total number of vacancies and interstitials, i.e., twice the number of Frenkel pairs.

The main objective is to find the locations of clusters of Frenkel pairs, their size, spatial distribution, and other characteristics of such clusters. Our analysis is focused on relatively large clusters (subcascades).

Figure 2a shows a simulated cascade corresponding to a PKA energy of 300 keV. (Since the two-dimen-

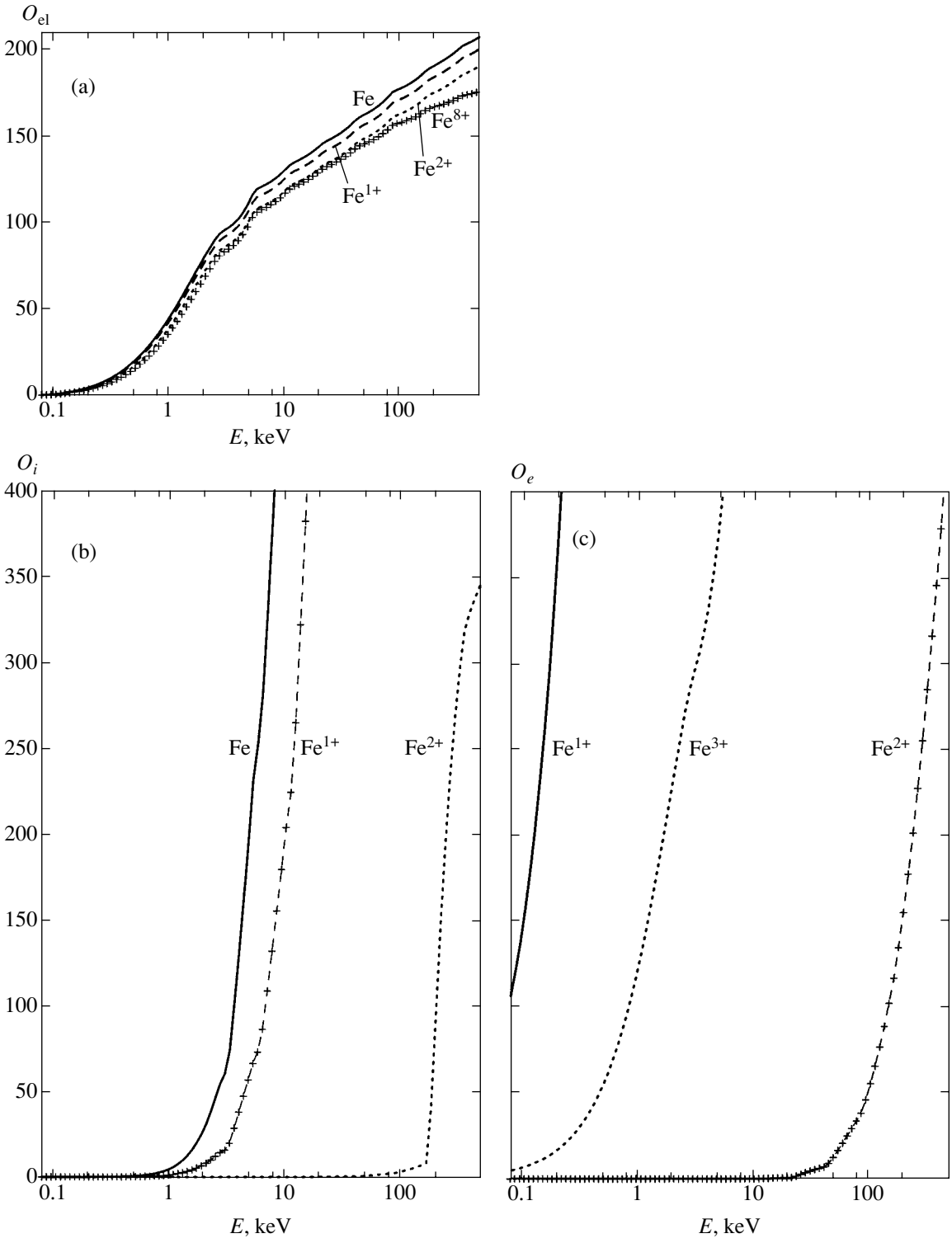


Fig. 1. Optical thickness vs. atom or ion energy: (a) elastic scattering of iron atom and ions with $Z = 1$, 2, and 8; (b) ionization of incident iron atom and ions with $Z = 1$ and 2; (c) electron capture by incident ions with $Z = 1$, 2, and 3.

sional projection of a three-dimensional cascade is plotted here, some elements that are separated in the direction perpendicular to the plane overlap in the figure.) Even a visual survey of the cascade reveals at least two types of processes associated with the vertices of a cascade tree. The most frequent process (in which a small amount of energy is lost) is either ionization or displacement of a relatively slow iron atom hit by a pro-

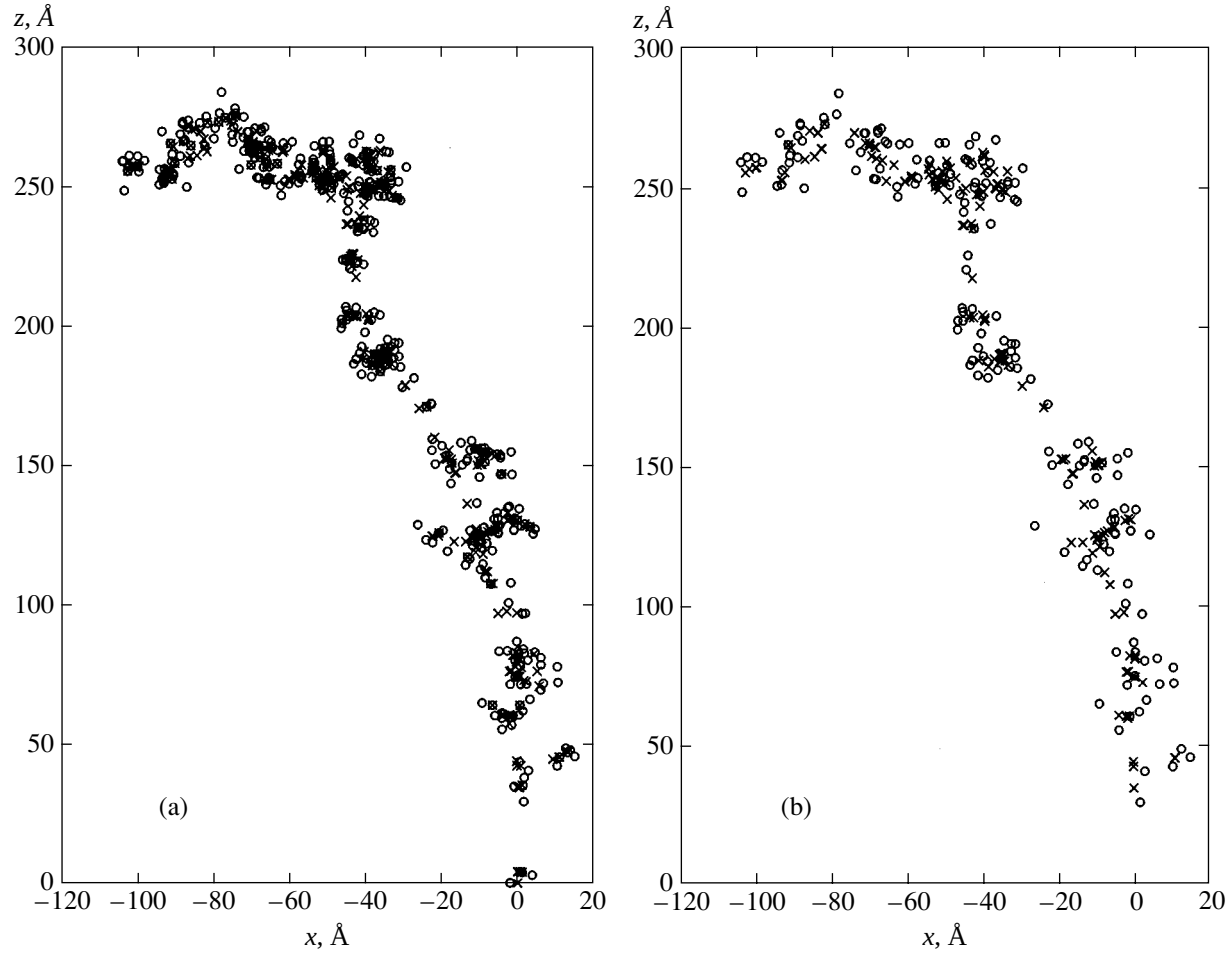


Fig. 2. Two-dimensional projection of a simulated cascade: (a) before recombination and (b) after spontaneous recombination. All pairs are assumed to recombine at distances less than 6 Å. Circles and crosses denote interstitials and vacancies, respectively.

jectile (atom or ion). The displaced atom either becomes an interstitial after traveling a short path or displaces several atoms, i.e., gives rise to a branch with few edges.

Much less frequently, the energy transferred to the displaced atom is sufficiently high to sustain the development of a branch that has many edges and involves dense groups of Frenkel pairs, which are called subcascades here. Computations show that the treelike structure on a subcascade level strongly depends on the simulated evolution of the charge states of atoms along their paths.

In the least frequent process, the energy of an atom is split into two almost equal parts and two similar branches are created. Both ensuing minor cascades involve subcascades.

This pattern of cascade evolution is explained by the fact that small amounts of energy are transferred more frequently.

The two-dimensional projection of a cascade shown in Fig. 2a demonstrates that interstitials tend to lie at its periphery, whereas vacancies group together around the cascade's centerline and subcascade centers.

Initially, each primary knock-on atom is neutral. However, if its energy is high (tens or hundreds of kiloelectronvolts), it loses its outermost shell in collisions and becomes a Fe^{2+} ion almost immediately. This is clear from comparison of the optical thicknesses plotted in Figs. 1a–1c. Therefore, the "trunk" of a cascade tree typically follows the path traveled by a Fe^{2+} ion.

However, the number of pairs produced in a cascade is determined by branching and subcascading. In these processes, much lower energies are transferred (hundreds of electronvolts); i.e., elastic scattering plays the dominant role (see Fig. 1) and the projectiles are mostly atoms.

This trend is illustrated by Fig. 3, where the charge of a particle participating in a cascade with a PKA energy of 700 keV is plotted versus its energy. Almost every particle whose energy is higher than 5 keV is an ion with $Z = 2$. The number of neutral atoms and ions with $Z = 1$ and 3 is negligible. As the particle energy

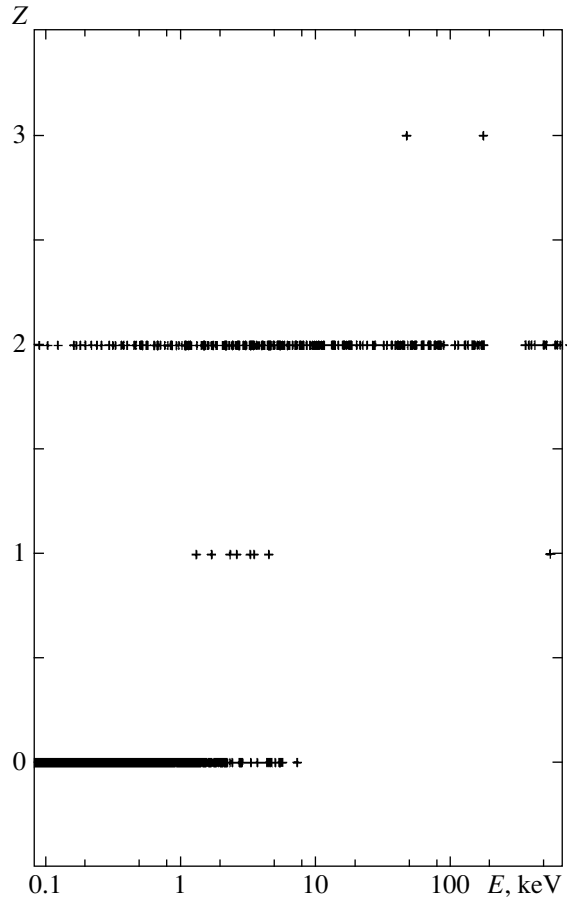


Fig. 3. Charge vs. energy for particles in a cascade with a PKA energy of 700 keV.

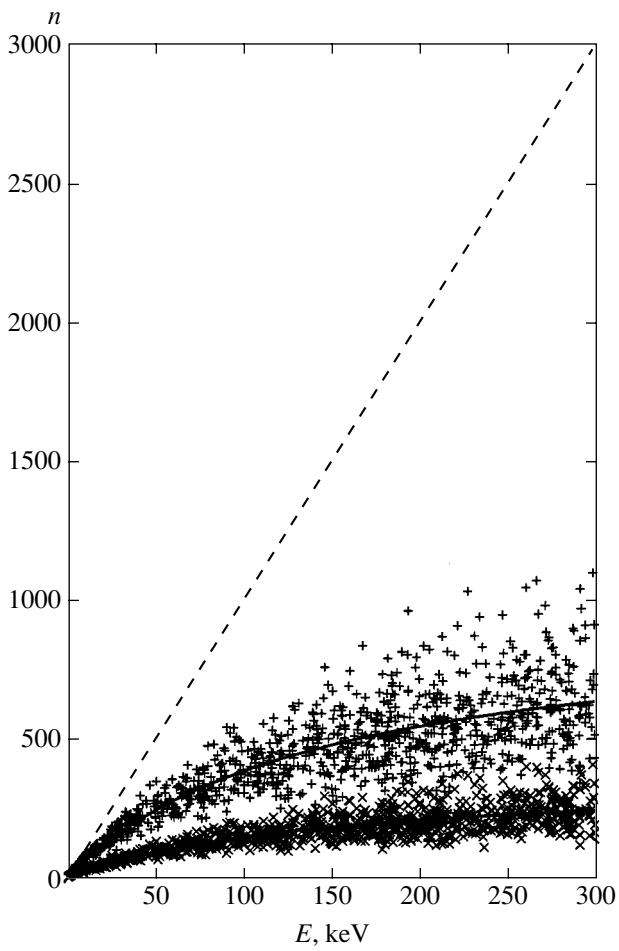


Fig. 4. Number of Frenkel pairs vs. PKA energy before recombination (+) and after recombination (×). Solid curves represent the mean number of pairs before recombination. The dashed curve is predicted by the NRT model (before recombination) [35, 36].

decreases in a developing cascade, neutral atoms tend to play the dominant role. Nevertheless, the contribution of ions with $Z = 2$ should not be neglected even if they have low energies. Figure 3 demonstrates the important role played by charge exchange in cascade evolution (particularly at high energies) and subcascade development.

4.2. Number of Frenkel Pairs in a Cascade

The dependence of the mean number n of Frenkel pairs on the initial PKA energy E is of primary importance. To find it, we simulated 10^4 cascades with $0.5 \le E \le 900$ keV. The number of Frenkel pairs produced at the consecutive steps of each cascade was recorded. Figure 4 shows the resulting number-of-pairs distribution plotted versus energy. The mean number of pairs is represented by solid curves. The scatter of data points around the curves widens toward high energies.[2]

[2] Note that each data point represents the number of pairs in a particular cascade and their mean number is obtained by averaging over cascades with equal initial PKA energies.

Our results agree with predictions of the Norgett–Robinson–Torres (NRT) model [35, 36] at very low energies ($E \le 5$ keV):

$$n \approx 0.8E/2E_{\mathrm{d}}. \tag{6}$$

The results based on the NRT model are shown by the dashed curve in Fig. 4. Our results predict a much lower number of Frenkel pairs at high energies. Note that the results of recent molecular-dynamics simulations of cascades of atomic displacements performed for different phenomenological potentials [37–39] were so disparate that no definite conclusion could be reached.

The energy dependence of the variance D of number of pairs before spontaneous recombination is well approximated by a power law with the exponent equal to 0.5036. Recall that the width of the Poisson distribution scales with the square root of mean number of Frenkel pairs. Since the number of pairs increases with PKA energy slower than a linear function, the distri-

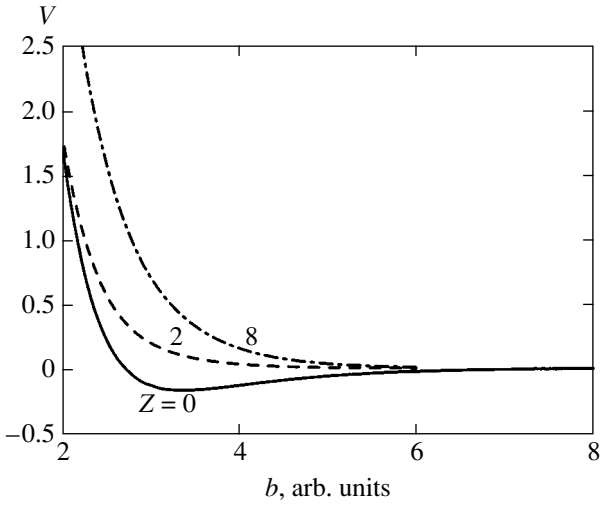


Fig. 5. Potentials of interaction of iron atoms ($Z = 0$) and ions ($Z = 2$ and 8) with a target iron atom vs. distance.

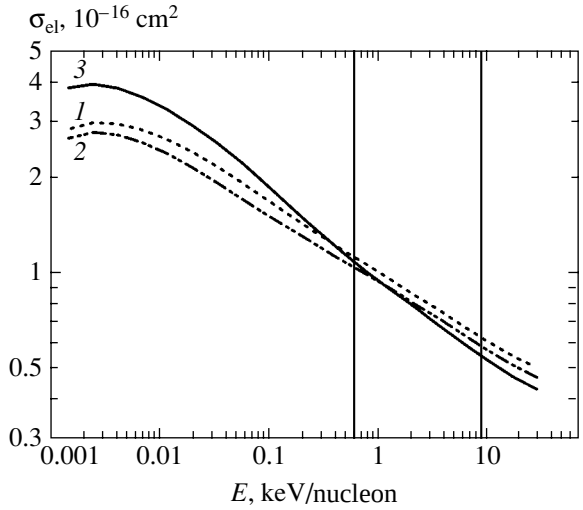


Fig. 6. Energy dependence of the cross section of elastic scattering of (*1*) iron atoms and (*2*) Fe^{2+} and (*3*) Fe^{8+} ions by a target iron atom.

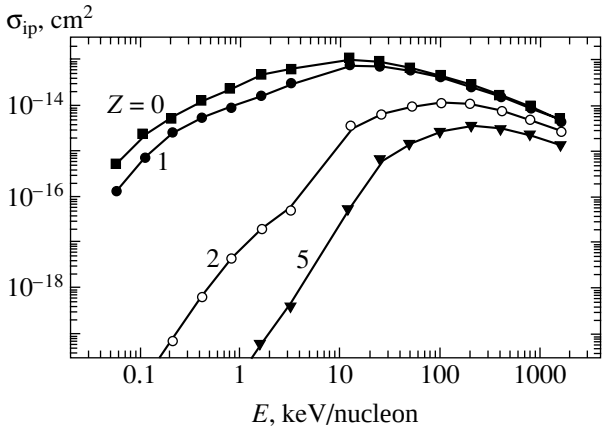


Fig. 7. Energy dependence of the cross section of ionization of iron atoms ($Z = 0$) and ions ($Z = 1, 2, 5$) by a target iron atom.

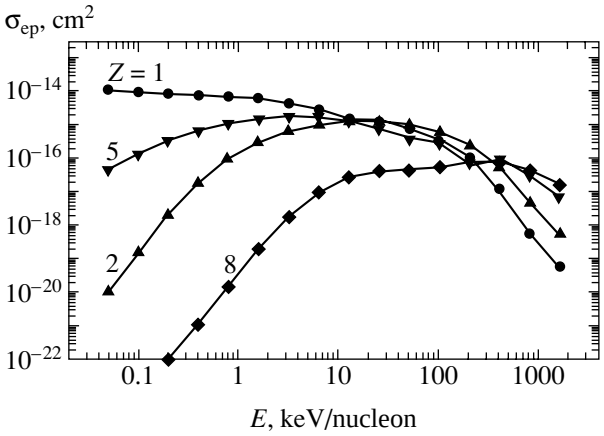


Fig. 8. Energy dependence of the cross section of electron capture of iron ions ($Z = 1, 2, 5, 8$) by a target iron atom.

bution shown here is wider than a corresponding Poisson one.

Once a pair is created, it can disappear via spontaneous recombination if the vacancy and interstitial are separated by a sufficiently small distance. Figure 2b shows the cascade obtained after all pairs with relative distances smaller than 6 Å have recombined in the cascade illustrated by Fig. 2a. The number of pairs that survive under the same condition for relative distance is shown in the lower part of Fig. 4 for all simulated cascades. It is clear that this number is several times smaller than the number of initially produced pairs.

Nevertheless, surviving subcascades make up large clusters of Frenkel pairs in cascade-tree branches. Even though the number of surviving pairs is smaller, the energy deposited in this region is the same. Quantitative analysis of subcascades and their characteristics should be based on an energy criterion, because the energy deposition (heat release) in subcascade regions determines the type and intensity of thermal spikes and phase particles.

Cross section for ionization of a target iron atom by an incident iron atom (or ion) at several values of projectile energy E_p in the laboratory system

E_p, keV/nucleon	σ_{it}, cm^2		
	Fe	Fe^{2+}	Fe^{8+}
0.1	3.1×10^{-15}	3.0×10^{-15}	3.2×10^{-15}
1.0	3.3×10^{-14}	3.0×10^{-14}	2.0×10^{-14}
10	3.2×10^{-13}	2.2×10^{-13}	2.1×10^{-13}
100	1.6×10^{-13}	1.2×10^{-13}	1.2×10^{-13}
1000	2.6×10^{-14}	1.9×10^{-14}	2.2×10^{-14}

5. CONCLUSIONS

We have described characteristics of the ballistic stage of evolution of cascades of atomic displacements initiated by primary knocked-on atoms with energies up to 900 keV. The energy limit imposed here is not substantial. It is dictated by the energy interval of practical importance rather than by the adopted computational scheme. The material was treated as amorphous in Monte Carlo computations (its crystalline structure was ignored), because we analyzed the large-scale structure of cascades. One essential distinction of this approach as compared to previous studies is that all necessary cross sections are calculated *ab initio* without invoking phenomenological adjustable potentials. This is done by taking into account the contributions of all electron shells of the projectile and target atoms and the dependence of the cross sections on the material density. We have used ionization and charge-exchange cross sections in analyzing the evolution of the charge state of the incident atom associated with a decrease in its energy. This process determines the structure of cascades (including subcascades of certain size). These computations were performed by running original codes developed at the Lebedev Physical Institute.

The calculated cross sections were used to formulate a model that admits algorithmic implementation in programs for computing cascade characteristics. The cascade characteristics were determined by simulating 10^4 cascades with PKA energy E ranging from 0.5 to 900 keV. We have calculated the number of Frenkel pairs as a function of the PKA energy at the end of the ballistic stage and after spontaneous recombination. The results obtained agree with those predicted by the standard NRT model [35, 36] for the low-energy range ($E \le 5$ keV), but differ substantially at higher energies.

It is also important that our approach can be used to describe the structure of a cascade, including separate treatment of the path traveled by the primary atom and the regions of subcascade formation. The definition of subcascade may depend on the ensuing process to be analyzed (one possible definition and the corresponding evolution at the quasi-thermodynamic stage are described in [4]). We emphasize here that approximate treatment of the material as amorphous, which is sufficient for characterizing the developing structure of cascades, provides a basis for classifying cascades. This makes it possible to compute the primary radiation damage caused by radiation with various energy spectra, e.g., in fusion and fission reactors.

Another important conclusion that follows from our analysis of cascade structure is the feasibility of computation of the hardening of VVER reactor pressure vessel steel as a function of the neutron dose.

APPENDIX 1

Elastic Scattering Cross Sections

Let us consider the elastic scattering cross sections, i.e., the cross sections characterizing processes that do not change the colliding partners (in particular, their charge). In such processes, only kinetic energy is transferred from the incident atom (or ion) to an atom of the material. The recoil atom participates in the development of a cascade. Therefore, these processes are important for estimating the energy lost by incident atoms and the energy spectra of recoil atoms. These cross sections strongly depend on the energy and the charge states of the colliding partners.

The cross sections were calculated according to classical mechanics. For an interaction potential $V(r)$, the impact parameter b is related to the scattering angle θ in the center-of-mass system as follows [2]:

$$\theta = \pi - 2b\int_{r_0}^{\infty}\frac{dr}{r^2 g(r)}, \tag{7}$$

with

$$g(r) = \sqrt{1 - \frac{2V(r)}{E_0} - \frac{b^2}{r^2}}, \tag{8}$$

where E_0 is the particle energy in the laboratory system. The lower limit of integration is the distance r_0 between the colliding particles at closest approach, which is found by solving the equation

$$g(r_0) = 0. \tag{9}$$

The differential elastic scattering cross section is given by the formula [2, 26]

$$d\sigma = 2\pi b db, \tag{10}$$

or, equivalently,

$$\frac{d\sigma}{d\Omega} = \frac{b}{\sin\theta}\left|\frac{db}{d\theta}\right|. \tag{11}$$

If the interaction potential is known, then the scattering cross section is obtained by integrating (10) or (11) with respect to the scattering angle from a certain minimum value to π. Thus, knowledge of interaction potentials is crucial for calculating elastic scattering cross sections.

We consider only binary elastic collisions. We use density functional theory to calculate the energy corresponding to a given electron density. The electron density is calculated by using a simple superposition approximation. The potential of interaction between

atoms and ions can be expressed in terms of the density functional $\mathscr{E}(\rho)$ as follows [27]:

$$V(R) = \mathscr{E}[\rho_1(R+r)+\rho_2(r)] - \mathscr{E}[\rho_1(R+r)] - \mathscr{E}[\rho_2(r)], \quad (12)$$

where ρ_i denotes the ith-particle electron density and R is the distance between particles. The coordinate r is related to the electron density and is used as an integration variable. Examples of application of this formula can be found in [28, 29], where it was shown that this approximation provides a sufficiently accurate description of properties of solids. In the cases considered here, the potentials are as shown in Fig. 5.

The potentials thus obtained are used in Eqs. (7)–(9) to calculate the impact parameter as a function of scattering angle and determine the behavior of differential elastic scattering cross sections as functions of scattering angle and initial energy by means of Eqs. (10) and (11).

The angular distributions of particles in elastic scattering processes are uniquely related to their energy spectra. The scattering angle is kinematically related to the energy lost by the incident particle as follows (e.g., see [2]):

$$\Delta E = E_0 \sin^2(\theta/2). \quad (13)$$

In the problem considered here, the minimum energy loss is either the energy $E_d \approx 40$ eV required to create a Frenkel pair in iron or an energy of about 100 eV; in the latter case, we neglect the processes in which the displaced atom cannot create another Frenkel pair and its energy dissipates in the material.

This model was used to calculate elastic scattering cross sections in the energy range of interest for the present study. The results are shown in Fig. 6 for the cut-off angle corresponding to a minimum energy loss of 40 eV (threshold displacement energy for iron). (The program employed in the computations was written in C++ at the Department of Theoretical Physics of the Lebedev Physical Institute.)

Note that one can easily determine the scope of the classical approach advocated here. The approach is not valid when the impact parameter is large and quantum-mechanical expressions should be used (e.g., see [40]). In what follows, we show that collisions with large impact parameters can be neglected in our analysis, because an atom of a solid cannot be knocked out and set in motion by such a collision. Accordingly, low energy losses (and, by Eq. (13), small scattering angles) cannot be associated with cascade initiation, even though they contribute to the energy lost by the incident particle. Since the integrals are cut off at small angles in our calculations, the quantum-mechanical contribution is not important.

A quantitative proof of the above assertion makes use of some estimates given in [40]. The classical scattering angle is written as

$$\theta_{cl} \approx \frac{p_t}{p} \approx \left|\frac{dV(b)}{db}\right| \frac{b}{pv}, \quad (14)$$

where p is the total momentum of the incident particle, p_t is its transverse component, and v is the particle's velocity. In quantum mechanics, the diffraction angle is estimated as

$$\theta_d \approx \frac{1}{pb}. \quad (15)$$

The minimum value of the impact parameter for which the classical approach is inapplicable is determined by equating these angles:

$$\left|\frac{dV(b)}{db}\right| b^2 \approx v. \quad (16)$$

This equation was solved numerically for each potential given by (12) to find the energy dependence of the minimum impact parameter. The angle θ_d was estimated by Eq. (15). We also used (13) to calculate the cut-off angles θ_{min} required to determine cross sections. It was shown that the validity criterion for the classical approach,

$$\theta_{min} > \theta_d, \quad (17)$$

holds within several orders of magnitude for the energy range of interest here.

APPENDIX 2

Cross Sections for Charge-Changing Ion–Atom Processes

Now, consider processes in which charge is transferred between the interacting particles: ionization of the projectile atom or ion by the target atom, which creates an ion with a higher charge; electron capture by the projectile ion, which reduces its charge; and ionization of the target atom. Allowing for these processes, one can evaluate, in particular, the fraction of the PKA energy transferred to electrons during the evolution of a cascade.

The ionization and charge-exchange cross sections for iron atoms or ions incident on iron atoms at rest in a target were calculated for energy ranging from 0.1 keV/nucleon to 1 MeV/nucleon, i.e., within four orders of magnitude. The cross sections of the following processes have been calculated:

$$Fe^{q+} + Fe \longrightarrow Fe^{(q+1)+} + Fe + e^-, \quad Z = 0\text{–}8; \quad (18)$$

$$Fe^{q+} + Fe \longrightarrow Fe^{(q-1)+} + Fe^+, \quad Z = 1\text{–}26; \quad (19)$$

$$Fe^{q+} + Fe \longrightarrow Fe^{q+} + Fe^+ + e^-, \quad Z = 0\text{–}26. \quad (20)$$

Analogous calculations were performed for atoms and ions of several different materials, but their results are not used in this paper.

The ionization cross sections were computed in the modified first Born approximation by using the LOSS code developed in the Optics Department of the Lebedev Physical Institute and described in [30]. The electronic structure of both incident and target particles was taken into account. The wave functions required to describe transitions of electrons from atomic bound states to continuum states were calculated in the partial wave representation by solving the Schrödinger equation numerically with an effective field of the atomic core.

The cross sections for ionization of a target neutral Fe atom by Fe ions increase from about 10^{-15} cm^2 at $E \approx 0.08$ keV/nucleon to about 10^{-13} cm^2 at $E \approx 10$ keV/nucleon (see table) and gradually decrease to about 10^{-14} cm^2 at $E \approx 400$ keV/nucleon. These cross sections are nonmonotonic functions of the charge of the incident ion.

The cross sections for ionization of incident iron ions by a target neutral Fe atom strongly depend on the energy and electronic structure of the colliding partners. They reach maximum values ($\sim 10^{-13}$ cm^2) for Fe and Fe^{1+} at $E \approx 15$ keV/nucleon. For ions with different degrees of ionization, the ionization cross sections are lower by an order of magnitude, and their maximum values are reached at $E \approx 100$–200 keV/nucleon.

The cross sections of charge exchange by electron capture were computed by using the CAPTURE code developed at the Optics Department of the Lebedev Physical Institute and described in [31]. The computations were performed by using the Brinkman–Kramers approximation and the impact parameter representation. Capture from each electron shell of the target atom was taken into account. We used normalized charge-exchange probabilities valid even for a large number of channels (up to 500). Hydrogen-like wave functions were also used, with the effective charge calculated by taking into account the screening of the target nucleus.

The charge-exchange cross sections calculated without allowing for the effect of target density (see below) also exhibit nonmonotonic behavior. They are almost constant at low energies and decrease according to power laws varying between E^{-2} and E^{-3} at higher energies. The largest cross section corresponds to Fe^{26+} + Fe collisions, reaching a maximum of 6×10^{-13} cm^2 at $E \approx 0.08$ keV/nucleon. Note that charge exchange in high-energy collisions is due solely to the capture of inner electrons from the target atom.

The effect of target density ($N \approx 0.8 \times 10^{23}$ cm^{-3}) is essential for our computations, since it can reduce charge exchange cross sections by several orders of magnitude (see [32] for details). This effect is also responsible for nonmonotonic behavior of charge-exchange cross sections as functions of the charge and energy of the incident ion, due to the contribution of electron capture from inner shells of the target atom.

In brief, the effect of target density on charge exchange can be explained as follows. The total charge-exchange cross section is the sum of the cross sections of electron capture into the ground state and all excited states of the scattered projectile. When the target density is low, these states vary from the ground state to those with infinitely large quantum numbers. As the density increases, excited atoms tend to be ionized by target atoms and the total charge-exchange cross section decreases accordingly. Note that ionization of highly excited ions is characterized by very high rates proportional to $Z_t^2 N_t v^{-1} n^2$, where Z_t is the nuclear charge of the target atom, N_t is the target density, v is the ion velocity, and n the principal quantum number of the excited state. When the density is very high, as in our case, the atom produced by charge exchange is in the ground state or in one of the lowest excited states. The corresponding cross sections are much smaller, and their dependence on the energy and charge of the incident ion is weaker. In other words, charge-exchange cross sections may exhibit nonmonotonic behavior when the target density is high.

These effects manifest themselves in numerical results: the computed cross sections rapidly decrease and exhibit nonmonotonic behavior. This substantially complicates computations, particularly in view of the fact that the electron-capture cross section must be found for every shell of the target atom when the energy is high.

Nevertheless, we computed the cross sections for incident ions of any degree of ionization, including neutral atoms, taking into account the effect of solid target density. In total, about 100 cross sections have been computed. Some of them are presented here.

First, let us consider the variation of the charge of the projectile atom (or ion) along the cascade path, which can be associated only with ionization of the atom (or ion) and electron capture. The cross sections of these processes (σ_{ip} and σ_{ep}, respectively) were computed, as functions of energy, for all possible initial charge numbers (from 0 to 26). The results are shown in Figs. 7 and 8 only for selected charge states to simplify presentation.

As mentioned above, we computed the cross section of ionization of a target atom σ_{it} by a projectile atom or ion for various initial charges of the incident particle. The results depend on energy, but are virtually independent of charge. This is demonstrated by the table, where results obtained for the iron atom and two ions are presented. The cross sections obtained for different ions are similar to those shown in the table.

The data presented in the first column of the table agree with the curve corresponding to $Z = 0$ in Fig. 7 since atom–atom collisions are symmetric. The curves

corresponding to ions lie close to it because the cross sections are virtually independent of ion charge.

REFERENCES

1. N. Bohr, *The Penetration of Atomic Particles through Matter*, 3rd ed. (Munksgaard, Copenhagen, 1960; Gostekhizdat, Moscow, 1948).
2. W. Eckstein, *Computer Simulation of Ion–Solid Interactions* (Springer, Berlin, 1991; Mir, Moscow, 1995).
3. J. Lindhard, V. Nielsen, and M. Scharff, K. Dan. Vidensk. Selsk. Mat. Fys. Medd. **36** (10), 1 (1968).
4. I. M. Dremin, O. V. Ivanov, V. A. Nechitaĭlo, and A. V. Subbotin, Preprint TD 13/09.03 (Physical Inst., Russian Academy of Sciences, Moscow, 2003).
5. J. A. Brinkman, J. Appl. Phys. **25**, 961 (1954).
6. M. I. Kaganov, I. M. Lifshits, and L. V. Tanatarov, Zh. Éksp. Teor. Fiz. **31**, 232 (1956) [Sov. Phys. JETP **4**, 173 (1956)]; At. Énerg. **6**, 391 (1959).
7. J. R. Beeler, Phys. Rev. **150**, 470 (1966).
8. G. J. Dienes and J. H. Vineyard, *Radiation Effects in Solids* (Interscience, New York, 1957; Inostrannaya Literatura, Moscow, 1960).
9. F. V. Nolfi, *Phase Transformations during Irradiation* (Appl. Sci., London and New York, 1983).
10. A. D. Brailsford and R. Bullough, Philos. Trans. R. Soc. London, Ser. A **302**, 87 (1981).
11. A. H. King and D. A. Smith, Radiat. Eff. **54**, 169 (1981).
12. L. K. Mansur, A. D. Brailsford, and W. G. Wolfer, J. Nucl. Mater. **105**, 36 (1982).
13. A. V. Subbotin, At. Énerg. **43**, 100 (1977).
14. V. M. Agranovich and V. V. Kirsanov, Usp. Fiz. Nauk **118**, 3 (1976) [Sov. Phys. Usp. **19**, 1 (1976)].
15. K. Nordlund and R. S. Averback, J. Nucl. Mater. **276**, 194 (2000).
16. R. S. Averback, J. Nucl. Mater. **216**, 49 (1994).
17. A. Souidi, M. Hou, C. S. Becquart, and C. Domain, J. Nucl. Mater. **295**, 179 (2001).
18. D. J. Bacon, F. Gao, and Yu. N. Osetsky, J. Nucl. Mater. **276**, 1 (2000).
19. A. Caro, M. Alurralde, R. Saliba, and M. Caro, J. Nucl. Mater. **251**, 72 (1997).
20. D. J. Bacon, A. F. Calder, and F. Gao, J. Nucl. Mater. **251**, 1 (1997).
21. A. B. Migdal, Zh. Éksp. Teor. Fiz. **9**, 1163 (1939); *Qualitative Methods in Quantum Theory* (Nauka, Moscow, 1975; Benjamin, Reading, Mass., 1977).
22. E. L. Feĭnberg, Dokl. Akad. Nauk **23**, 778 (1939); Candidate's Dissertation (Moscow State Univ., Moscow, 1939).
23. A. M. Dykhne and G. L. Yudin, *Sudden Disturbances and Quantum Evolution* (UFN, Moscow, 1996).
24. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory*, 2nd ed. (Fizmatgiz, Moscow, 1963; Pergamon, New York, 1977).
25. N. M. Sobolevskiĭ, É. A. Koptelov, S. G. Lebedev, and A. V. Subbotin, Report No. 070-082-038 (FGUP NIKIÉT im. N. A. Dollezhalya, Moscow, 2001).
26. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 1: *Mechanics* (Fizmatgiz, Moscow, 1958; Pergamon Press, Oxford, 1976).
27. W. Kohn and L. J. Sham, Phys. Rev. A **140**, 1133 (1965).
28. O. V. Ivanov and E. G. Maksimov, Phys. Rev. Lett. **69**, 108 (1992).
29. O. V. Ivanov and E. G. Maksimov, Zh. Éksp. Teor. Fiz. **108**, 1841 (1995) [JETP **81**, 1008 (1995)].
30. V. P. Shevelko, I. Yu. Tolstikhina, and T. Stohler, Nucl. Instrum. Methods Phys. Res. B **184**, 295 (2001).
31. V. P. Shevel'ko, Zh. Tekh. Fiz. **71** (10), 30 (2001) [Tech. Phys. **46**, 1225 (2001)].
32. O. Rozmej, I. Yu. Tolstikhina, and V. P. Shevel'ko, Zh. Tekh. Fiz. **73** (9), 31 (2003) [Tech. Phys. **48**, 1110 (2003)].
33. I. M. Sobol', *The Monte-Carlo Numerical Methods* (Nauka, Moscow, 1973).
34. *Standard E693, Annual Book of ASTM Standards* (Am. Society of Testing and Materials, Philadelphia, PA, 1994), part 45.
35. M. J. Norgett, M. T. Robinson, and J. M. Torres, Nucl. Eng. Des. **33**, 50 (1975).
36. *Standard E521, Annual Book of ASTM Standards* (Philadelphia, PA, 1989).
37. C. S. Becquart, C. Domain, A. Legris, and J. C. Van Duysen, J. Nucl. Mater. **280**, 73 (2000).
38. R. S. Averback and T. Diaz de la Rubia, Solid State Phys. **51**, 281 (1998).
39. K. Nordlund and R. S. Averback, Phys. Rev. B **56**, 2421 (1997).
40. A. B. Migdal and V. P. Kraĭnov, *Approximation Methods in Quantum Mechanics* (Nauka, Moscow, 1966; Benjamin, New York, 1969).

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 98, No. 2, 2004, pp. 334–337.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 2, 2004, pp. 377–381.

Delocalized Dislocations in Quantum Dots

I. A. Ovid'ko and A. G. Sheinerman

Institute for Problems of Mechanical Engineering, Russian Academy of Sciences, St. Petersburg, 199178 Russia
e-mail: ovidko@def.ipme.ru

Received June 24, 2003

Abstract—A theoretical model describing nucleation of the misfit dislocations with delocalized cores in island films is proposed. The parameters of nanoislands (quantum dots) with such delocalized misfit dislocations in the Ge/Si system are estimated. Within the framework of the proposed model, nucleation of the delocalized misfit dislocations at the quantum dots in a broad range of parameters is energetically more favorable than nucleation of the perfect misfit dislocations usually considered in the models of dislocated quantum dots. © *2004 MAIK "Nauka/Interperiodica".*

I. INTRODUCTION

In recent years, spatially ordered ensembles of semiconductor nanoislands (quantum dots) have been extensively studied in both basic and technological aspects (see, e.g., [1–15]). The functional properties of such nanoislands, opening wide prospects for applications in microelectronics and optoelectronics, significantly depend on the defect structure of these objects. In particular, nucleation of the misfit dislocations (MDs) in nanoislands leads to degradation of their unique properties [2]. This circumstance poses an important task of determining the critical parameters of MD nucleation in nanoislands.

The MDs in nanoislands can be perfect, partial, or split (see Figs. 1a–1c, respectively). These types of MDs are analogous to the perfect [16–25], partial, and split [26–30] MDs in continuous films. Each configuration of the partial and split MDs in a strained nanoisland includes one or two partial dislocations (with the Burgers vectors not coinciding with the basis vectors of crystal lattices of the nanoisland or the substrate) and a stacking fault (Figs. 1b and 1c). In the general case, continuous crystalline films may also feature nucleation of the MDs of another type, representing MDs with the Burgers vector "smeared" over an extended dislocation core [31]. Such MDs, referred to below as delocalized misfit dislocations, have been considered only in continuous films [31]. Until recently, theoretical models of the MD nucleation in nanoislands considered only perfect (undissociated) MDs (Fig. 1a) (see, e.g., [13–15]). Recently, we proposed new theoretical models describing the nucleation of partial split MDs in nanoislands [32, 33].

The main aim of this study was to provide a theoretical analysis of the conditions of nucleation of the delocalized MDs (Fig. 1d) in nanoislands.

2. THE MODEL OF A DELOCALIZED MISFIT DISLOCATION IN A NANOISLAND

Consider a heteroepitaxial system representing an island on a semiinfinite substrate (Fig. 1). Let the nanoisland have the shape of a regular pyramid with a square base of the side length l and the faces sloped at an angle α to the substrate surface. The nanoisland and the substrate are assumed to be isotropic elastic materials possessing the same values of the shear modulus G

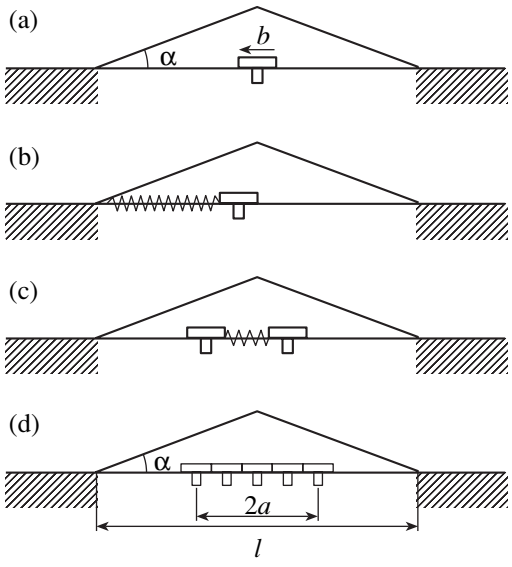


Fig. 1. Schematic diagrams showing the possible configurations of MDs in a pyramidal nanoisland on a substrate: (a) perfect dislocation with a linear core; (b) partial dislocation with a linear core and a stacking fault; (c) split dislocation composed of two partial dislocations (with linear cores) and a stacking fault between them; (d) delocalized dislocation with the Burgers vector uniformly distributed along an extended core representing a band of finite width.

and the Poisson ratio ν. In our model, the island–substrate boundary is characterized by a one-dimensional misfit $f = 2(a_i - a_s)/(a_i + a_s)$, where a_i and a_s are the crystal lattice parameters of the island and substrate, respectively. This misfit f is modeled by a system of edge dislocations (called coherent dislocations) with the Burgers vectors $db\mathbf{e}_x$ continuously distributed along the island–substrate boundary with a linear density of f/db.

The difference between the crystal lattice parameters of nanoisland and substrate gives rise to misfit stresses in the nanoisland. Relaxation of the misfit stresses in nanoislands and thin films usually takes place at the expense of nucleation and growth of the MDs and their combinations [13–33]. The most frequently encountered types are perfect MDs [13–25] and sometimes partial and split MDs [26–30, 32, 33]. In addition to these types, thin films also admit nucleation of delocalized MDs [31]. Each delocalized MD is characterized by the Burgers vector distributed along an extended dislocation core, in contrast to perfect and partial MDs with linear cores.

In this paper, we will consider delocalized MDs in nanoislands, where the nucleation conditions are different from those in the usual thin films (studied in [31]). The difference in the MD nucleation conditions is related to the fact that the geometry of a nanoisland differs from that of a continuous thin film. In particular, the island geometry makes possible the nucleation of delocalized MDs at the island–substrate boundary, near the line of intersection of this interface with the side face of the nanoisland (Fig. 1d). In this case, the free side face produces significant screening of the MD stress field, thus decreasing the elastic energy of MDs and facilitating their nucleation in the island.

Let us consider in more detail the formation of a perfect MD with a delocalized core at the island–substrate boundary (Fig. 1d). In the general case, this defect represents an inhomogeneous, continuous distribution of dislocations along a certain region of the boundary. For the sake of simplicity, we will model such an MD with a delocalized core by a homogeneous distribution of dislocations (called anticoherent dislocations).

The nucleation of an MD with a delocalized core is energetically favorable if the island energy difference ΔW between the state with an MD and the coherent state is negative. The exact analytical calculation of the energy of an island with MD on a substrate requires the knowledge of expressions for the edge dislocation stress field near a free surface formed by the free surfaces of the island and the substrate (Fig. 1d). Since no such expressions for the MD formation in the island are available, the energy of this process will be calculated for an island on a substrate modeled by an infinite elastic cylinder of radius R (Fig. 2). In this model, the island–substrate boundary represents a band intersecting the cylinder parallel to its axis and making an angle

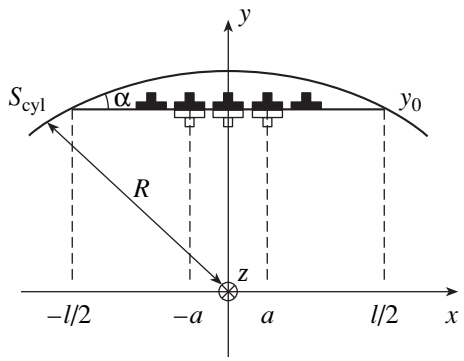


Fig. 2. A model MD with a delocalized core in a two-phase cylinder with an interface (island–substrate boundary). A misfit between the two crystal lattices and the delocalized MD are modeled by a continuous distribution of coherent (black symbols) and anticoherent (open symbols) dislocations.

α with the cylinder surface. The angle α corresponds to that between the base and side face of a real island. In the coordinate system adopted in Fig. 2, the island–substrate boundary corresponds to the segment ($|x| < l/2 = R\sin\alpha$, $y = y_0 = R\cos\alpha$). The MDs are modeled by anticoherent dislocations uniformly distributed along the region $|x| < a$ of the boundary at a linear density of $1/(2a)$.

3. THE ENERGY CHARACTERISTICS OF A DELOCALIZED MISFIT DISLOCATION IN A NANOISLAND

Let us determine a change in the model system energy related to the formation of an MD with a delocalized core at the island–substrate boundary. The energy change ΔW (per unit length of the boundary) can be represented as

$$\Delta W = W^d + W^{d-f} + W^c, \quad (1)$$

where W^d is the proper elastic energy of the MD, W^{d-f} is the energy of the elastic interaction between the dislocation and the misfit stress field, and W^c is the MD core energy.

The proper MD energy W^d is calculated as the energy of the elastic interaction between the anticoherent dislocation with the Burgers vectors $-b/(2a)dx\mathbf{e}_x$, where $b = a_i$ is the interatomic distance in the island:

$$W^d = \frac{1}{2}\left(\frac{b}{2a}\right)^2 \int_{-a}^{a}\int_{-a}^{a} g(x, x')dxdx'. \quad (2)$$

Here, $g(x, x')$ is the energy of interaction between two parallel dislocations (situated at points x and x') with the unit Burgers vectors $\mathbf{b} = -\mathbf{e}_x$. This quantity is calcu-

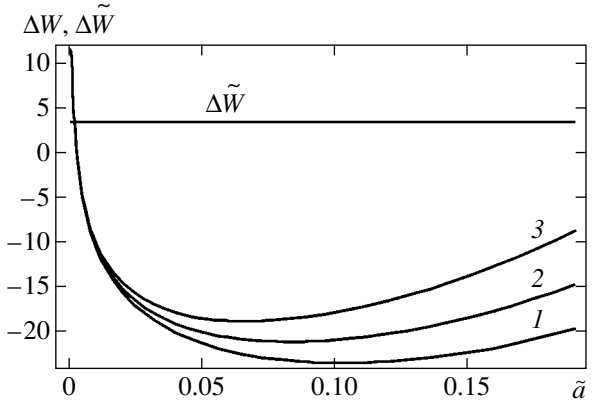


Fig. 3. Plots of the energy difference ΔW due to the formation of a delocalized MD in a nanoisland versus dimensionless halfwidth $\tilde{a}$ of this defect for $\Delta\gamma = 0$ (*1*), 0.03 (*2*), and 0.066 J/m^2 (*3*). The horizontal line shows the $\Delta\tilde{W}$ value corresponding to the formation of a perfect MD with a localized core at the center of the island base (ΔW and $\Delta\tilde{W}$ values are expressed in units of $Db^2/8$).

lated using an expression [23] for the stress function for a dislocation in a cylinder as $g(x, x') = \tilde{g}(2\sin\alpha x/l, 2\sin\alpha x'/l)$, where

$$g(\tilde{x}, \tilde{x}') = D\{(M-1)[1-2\tilde{y}_0^2(M+1)] - \ln M\}, \quad (3)$$

$$M = \frac{(\tilde{x}-\tilde{x}')^2}{\tilde{y}_0^4 + (\tilde{x}^2+\tilde{x}'^2-2)\tilde{y}_0^2 + (\tilde{x}\tilde{x}'-1)^2}, \quad (4)$$

$\tilde{y}_0 = 2\sin\alpha y_0/l$, and $D = G/[2\pi(1-\nu)]$.

The energy of the elastic interaction between the MD and the misfit stress field is calculated as the energy of interaction between coherent and anticoherent dislocations:

$$W^{d-f} = -\frac{b}{2a} f \int_{-a}^{a}\int_{-l/2}^{l/2} g(x, x')\,dx\,dx'. \quad (5)$$

The MD core energy is related to the distortion of chemical bonds at the interface as a result of the MD formation. Following [31], we represent this energy as

$$W^c = 2a\Delta\gamma, \quad (6)$$

where $\Delta\gamma$ is a change in the specific surface energy of the interface upon the formation of an MD with a delocalized core.

Substituting expressions (2), (5), and (6) into relation (1), we obtain

$$\Delta W = \frac{b^2}{8}\left\{\frac{1}{\tilde{a}^2}\int_{-\tilde{a}}^{\tilde{a}}\left[\int_{-\tilde{a}}^{\tilde{a}}\tilde{g}(\tilde{x}, \tilde{x}')d\tilde{x}' - \frac{4a}{b} f \int_{-\sin\alpha}^{\sin\alpha} g(\tilde{x}, \tilde{x}')d\tilde{x}'\right]d\tilde{x} + \frac{16a\Delta\gamma}{b^2}\right\}, \quad (7)$$

where $\tilde{a} = 2a\sin\alpha/l$.

The ΔW values were determined by analytically calculating the inner integrals and numerically calculating the outer integral in Eq. (7). Figure 3 shows the plots of $\Delta W(\tilde{a})$ in the interval $0 < a < l/2$ for various values of the parameter $\Delta\gamma$, $l = 100b$, and $\alpha = 11°$. Calculations were performed for Ge/Si nanoislands with the following parameters: $f = 0.042$; $b = 0.566$ nm; $G = 40$ GPa; $\nu = 0.26$. For comparison, the horizontal line in Fig. 3 shows the $\Delta\tilde{W}$ value corresponding to the formation of a perfect MD with a localized core.

As can be seen from Fig. 3, there exists a certain equilibrium defect length $\tilde{a} = \tilde{a}_0$ corresponding to a minimum of ΔW. The $\tilde{a}_0$ value decreases upon increase in $\Delta\gamma$. For any value of $\Delta\gamma$ not exceeding a specific surface energy of 0.066 J/m^2 for the stacking fault in the Ge/Si system, we have $\Delta W(\tilde{a}_0) < 0$ and $\Delta W(\tilde{a}_0) < \Delta\tilde{W}$, so that the formation of an MD with a delocalized core in the island becomes energetically favorable. Thus, the accommodation of misfit stresses via nucleation of delocalized MDs is an alternative to the formation of a localized dislocation that can take place even in small islands, where the formation of usual dislocations is energetically unfavorable.

4. CONCLUSIONS

We have performed a theoretical analysis of the conditions favoring the nucleation of delocalized MDs—misfit defects of the new type—in nanoislands (quantum dots). The theory shows that, in a broad range of structural and geometric parameters, the formation of such localized MDs (Fig. 1d) in nanoislands in the Ga/Si system can be energetically more favorable than the nucleation of usual perfect MDs (Fig. 1a). This result makes interesting the experiments on identification of the types of MDs in nanoislands, since various MDs must be, strictly speaking, sensitive to different parameters of the structure, chemical composition, and technology of nanoislands. In particular, the conditions for nucleation of a delocalized MD (Fig. 1d) significantly depend on the energy of distorted chemical bonds in the extended core of this dislocation, in contrast to the case of a perfect MD (Fig. 1a).

Identification of MDs of various types and determination of the critical parameters controlling their formation opens possibilities for developing a technology for obtaining ensembles of nanoislands possessing improved functional characteristics. In addition, the identification of various types of MDs in nanoislands will contribute to the development of basic notions about the physics of defects in nanostructured solids.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project no. 01-02-16853), the St. Petersburg Scientific Center of the Russian Academy of Sciences, and by the Federal Program "Integration" (project no. B0026).

REFERENCES

1. V. A. Shchukin and D. Bimberg, Rev. Mod. Phys. **71**, 1125 (1999).
2. N. N. Ledentsov, V. M. Ustinov, V. A. Shchukin, *et al.*, Fiz. Tekh. Poluprovodn. (St. Petersburg) **32**, 385 (1998) [Semiconductors **32**, 343 (1998)].
3. D. Kouris, A. Peralta, and K. Sieradzki, Surf. Sci. **445**, 420 (2000).
4. P. Sutter and M. G. Lagally, Phys. Rev. Lett. **84**, 4637 (2000).
5. A. Bouret, Surf. Sci. **432**, 37 (1999).
6. B. Voigtländer and N. Theuerkauf, Surf. Sci. **461**, L575 (2000).
7. I. A. Ovid'ko and A. G. Sheinerman, Appl. Phys. A **74**, 273 (2002).
8. D. E. Jesson, M. Kästner, and B. Voigtländer, Phys. Rev. Lett. **84**, 330 (2000).
9. T. I. Kamins, E. C. Karr, R. S. Williams, *et al.*, J. Appl. Phys. **81**, 211 (1997).
10. S. A. Chaparro, J. Drucker, Y. Zhang, *et al.*, Phys. Rev. Lett. **83**, 1199 (1999).
11. S. A. Chaparro, Y. Zhang, J. Drucker, *et al.*, J. Appl. Phys. **87**, 2245 (2000).
12. J. Tersoff, C. Teichert, and M. G. Lagally, Phys. Rev. Lett. **76**, 1675 (1996).
13. E. Pehlke, N. Moll, A. Kley, *et al.*, Appl. Phys. A **65**, 525 (1997).
14. H. T. Johnson and L. B. Freund, J. Appl. Phys. **81**, 6081 (1997).
15. R. V. Kukta and L. B. Freund, J. Mech. Phys. Solids **45**, 1835 (1997).
16. E. A. Fitzgerald, Mater. Sci. Rep. **7**, 87 (1991).
17. J. H. van der Merve, Crit. Rev. Solid State Mater. Sci. **17**, 187 (1991).
18. S. C. Jain, A. H. Harker, and R. A. Cowley, Philos. Mag. A **75**, 1461 (1997).
19. M. Yu. Gutkin and I. A. Ovid'ko, *Defects and Mechanisms of Plasticity in Nanostructural and Noncrystalline Materials* (Yanus, St. Petersburg, 2001).
20. T. J. Gosling and J. R. Willis, Philos. Mag. A **69**, 65 (1994).
21. F. Bailly, M. Barbé, and G. Cohen-Solal, J. Cryst. Growth **153**, 115 (1995).
22. I. A. Ovid'ko, J. Phys.: Condens. Matter **11**, 6521 (1999); **13**, L97 (2001).
23. M. Yu. Gutkin, I. A. Ovid'ko, and A. G. Sheinerman, J. Phys.: Condens. Matter **12**, 5391 (2000).
24. A. G. Sheinerman and M. Yu. Gutkin, Phys. Status Solidi A **184**, 485 (2001).
25. S. V. Bobylev, I. A. Ovid'ko, and A. G. Sheinerman, Phys. Rev. B **64**, 224507 (2001).
26. B. C. de Cooman and C. B. Carter, Acta Metall. **37**, 2765 (1989).
27. B. C. de Cooman, C. B. Carter, K. T. Chan, *et al.*, Acta Metall. **37**, 2779 (1989).
28. J. Zou and D. J. H. Cockayne, Appl. Phys. Lett. **69**, 1083 (1996).
29. M. Loubradou, R. Bonnet, A. Vila, *et al.*, Mater. Sci. Forum **207**–**209**, 285 (1996).
30. M. Tamura, Appl. Phys. A **63**, 359 (1996).
31. A. E. Romanov, T. Wagner, and M. Rühle, Scr. Mater. **38**, 869 (1998).
32. I. A. Ovid'ko, Phys. Rev. Lett. **88**, 046103 (2002).
33. I. A. Ovid'ko and A. G. Sheinerman, Phys. Rev. B **66**, 245309 (2002).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 98, No. 2, 2004, pp. 338–340.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 2, 2004, pp. 382–384.

SOLIDS
Structure

Effect of Long-Range Interactions on the Multicritical Behavior of Homogeneous Systems

S. V. Belim
Omsk State University, pr. Mira 55, Omsk, 644077 Russia
e-mail: belim@univer.omsk.su
Received August 29, 2003

Abstract—A field-theoretic approach is applied to describe behavior of homogeneous three-dimensional systems with long-range interactions defined by two order parameters at bicritical and tetracritical points. Renormalization-group equations are analyzed in the two-loop approximation by using the Padé–Borel summation technique. The fixed points corresponding to various types of multicritical behavior are determined. It is shown that effects due to long-range interactions can be responsible for a change from bicritical to tetracritical behavior. *© 2004 MAIK "Nauka/Interperiodica".*

It was shown in [1] that effects due to long-range interaction described by the power law $1/r^{-D-a}$ with $a < 2$ are responsible for a change in critical behavior. It was also revealed that three-dimensional systems with $a < 1.5$ are described by mean-field critical exponents.

This paper deals with effects of long-range interaction on systems described by two fluctuating order parameters. The phase diagrams of such systems can contain bicritical and tetracritical points. In the former case, a multicritical point is the point of intersection of two second-order phase transition curves and one first-order phase transition curve. In the latter case, it is the point of intersection of four second-order phase transition curves. In the neighborhood of a multicritical point, the system exhibits critical behavior characterized by competing types of ordering. Whereas one critical parameter is replaced by the other at a bicritical point, the ordering types can coexist in a mixed phase at a tetracritical point. Such systems can be described in terms of two order parameters that are transformed under different irreducible representations [2].

The model Hamiltonian is written as

$$H_0 = \frac{1}{2}\int d^D q(\tau_1 + q^a)\Phi_q\Phi_{-q} + \frac{1}{2}\int d^D q(\tau_2 + q^a)\Psi_q\Psi_{-q} + u_{01}\int d^D\{q_i\}(\Phi_{q1}\Phi_{q2})(\Phi_{q3}\Phi_{-q1-q2-q3}) + u_{02}\int d^D\{q_i\}(\Psi_{q1}\Psi_{q2})(\Psi_{q3}\Psi_{-q1-q2-q3}) + 2u_{03}\int d^D\{q_i\}(\Phi_{q1}\Phi_{q2})(\Psi_{q3}\Psi_{-q1-q2-q3}), \quad (1)$$

where Φ and Ψ are fluctuating order parameters, u_{01} and u_{02} are positive constants,

$$\tau_1 \propto |T - T_{c1}|/T_{c1}, \quad \tau_2 \propto |T - T_{c2}|/T_{c2},$$

T_{c1} and T_{c2} are the corresponding phase-transition temperatures, and a is the long-range parameter.

This Hamiltonian admits a wide diversity of multicritical points. The conditions for tetracritical and bicritical behavior are $u_3^2 < u_1u_2$ and $u_3^2 \geq u_1u_2$, respectively.

In the field-theoretic approach [3], the asymptotic critical behavior and structure of phase diagrams in the fluctuation region are governed by the Callan–Symanczyk equation for the vertex parts of irreducible Green functions. The β and γ functions contained in the Callan–Symanczyk equation for renormalized interaction vertexes u_1, u_2, and u_3 are calculated by a standard method based on the Feynman diagram technique and a renormalization procedure [4]. As a result, the following expressions for β functions are obtained in the two-loop approximation:

$$\beta_{u1} = -u_1 + 36J_0u_1^2 + 4J_0u_3^2 - 1728\left(2J_1 - J_0^2 - \frac{2}{9}G\right)u_1^3 - 192\left(2J_1 - J_0^2 - \frac{2}{9}G\right)u_1u_3^2 - 64(2J_1 - J_0^2)u_3^3,$$

$$\beta_{u2} = -u_2 + 36J_0u_2^2 + 4J_0u_3^2 - 1728\left(2J_1 - J_0^2 - \frac{2}{9}G\right)u_2^3 - 192\left(2J_1 - J_0^2 - \frac{2}{9}G\right)u_2u_3^2 - 64(2J_1 - J_0^2)u_3^3,$$

1063-7761/04/9802-0338$26.00 © 2004 MAIK "Nauka/Interperiodica"

$$\beta_{u3} = -u_3 + 16J_0u_3^2 + 12J_0u_1u_3 + 12J_0u_2u_3 - 320\left(2J_1 - J_0^2 - \frac{2}{5}G\right)u_3^3 - 288\left(2J_1 - J_0^2 - \frac{2}{3}G\right)u_1^2u_3$$

$$-288\left(2J_1 - J_0^2 - \frac{2}{3}G\right)u_2^2u_3 - 576(2J_1 - J_0^2)u_1u_3^2 - 576(2J_1 - J_0^2)u_2u_3^2, \quad (2)$$

$$J_1 = \int \frac{d^Dqd^Dp}{(1+|\mathbf{q}|^a)^2(1+|\mathbf{p}|^a)(1+|q^2+p^2+2\mathbf{p}\cdot\mathbf{q}|^{a/2})},$$

$$J_0 = \int \frac{d^Dq}{(1+|\mathbf{q}|^a)^2},$$

$$G = -\frac{\partial}{\partial|\mathbf{k}|^a}\int \frac{d^Dqd^Dp}{(1+|q^2+k^2+2\mathbf{k}\cdot\mathbf{q}|^a)(1+|\mathbf{p}|^a)(1+|q^2+p^2+2\mathbf{p}\cdot\mathbf{q}|^{a/2})}.$$

In terms of the new effective interaction vertices

$$v_1 = \frac{u_1}{J_0}, \quad v_2 = \frac{u_2}{J_0}, \quad v_3 = \frac{u_3}{J_0}, \quad (3)$$

the β functions are expressed as

$$\beta_1 = -v_1 + 36v_1^2 + 4v_3^2 - 1728\left(2\tilde{J}_1 - 1 - \frac{2}{9}\tilde{G}\right)u_1^3$$

$$-192\left(2\tilde{J}_1 - 1 - \frac{2}{9}\tilde{G}\right)v_1v_3^2 - 64(2\tilde{J}_1 - 1)v_3^3,$$

$$\beta_2 = -v_2 + 36v_2^2 + 4v_3^2 - 1728\left(2\tilde{J}_1 - 1 - \frac{2}{9}\tilde{G}\right)v_2^3$$

$$-192\left(2\tilde{J}_1 - 1 - \frac{2}{9}\tilde{G}\right)v_2v_3^2 - 64(2\tilde{J}_1 - 1)v_3^3,$$

$$\beta_{u3} = -v_3 + 16v_3^2 + 12v_1v_3 + 12v_2v_3$$

$$-320\left(2\tilde{J}_1 - 1 - \frac{2}{5}\tilde{G}\right)v_3^3 \quad (4)$$

$$-288\left(2\tilde{J}_1 - 1 - \frac{2}{3}\tilde{G}\right)v_1^2v_3 - 288\left(2\tilde{J}_1 - 1 - \frac{2}{3}\tilde{G}\right)v_2^2v_3$$

$$-576(2\tilde{J}_1 - 1)v_1v_3^2 - 576(2\tilde{J}_1 - 1)v_2v_3^2,$$

$$\tilde{J}_1 = \frac{J_1}{J_0^2}, \quad \tilde{G} = \frac{G}{J_0^2}.$$

This redefinition is meaningful for $a \le D/2$. In this case, J_0, J_1, and G are divergent functions. Introducing the cutoff parameter Λ, we obtain finite expressions for the ratios

$$\frac{J_1}{J_0^2} = \int_0^\Lambda\int_0^\Lambda \frac{d^Dqd^Dp}{(1+|\mathbf{q}|^a)^2(1+|\mathbf{p}|^a)(1+|q^2+p^2+2\mathbf{p}\cdot\mathbf{q}|^a)}\left[\int_0^\Lambda \frac{d^Dq}{(1+|\mathbf{q}|^\alpha)^2}\right]^{-2},$$

$$\frac{G}{J_0^2} = -\frac{\partial}{\partial|\mathbf{k}|^a}\int_0^\Lambda\int_0^\Lambda \frac{d^Dqd^Dp}{(1+|q^2+k^2+2\mathbf{k}\cdot\mathbf{q}|^a)(1+|\mathbf{p}|^a)(1+|q^2+p^2+2\mathbf{p}\cdot\mathbf{q}|^a)}\left[\int_0^\Lambda \frac{d^Dq}{(1+|\mathbf{q}|^\alpha)^2}\right]^{-2}, \quad (5)$$

as $\Lambda \longrightarrow \infty$.

The integrals are performed numerically. For $a \le D/2$, a sequence of J_1/J_0^2 and G/J_0^2 corresponding to various values of Λ is calculated and extrapolated to infinity.

It is well known that perturbation theory series are asymptotic and expressions (4) cannot be applied directly since the interaction vertexes for order-parameter fluctuations in the fluctuation region are too large. For this reason, the required physical information was extracted from these expressions by applying the Padé–

Fixed points and eigenvalues of the stability matrix

a	v_1^*	v_2^*	v_3^*	b_1	b_2	b_3
1.9	0.035842	0.035842	0.039202	0.069	0.505	0.702
1.8	0.033682	0.033682	0.034575	0.090	0.571	0.753
1.7	0.031287	0.031287	0.031334	0.113	0.629	0.809
1.6	0.027427	0.027427	0.026699	0.157	0.738	0.919
1.5	0.026514	0.026514	0.025973	0.171	0.762	0.949

Borel method extended to the four-parameter case. The appropriate direct and inverse Borel transforms have the form

$$f(v_1, v_2, v_3) = \sum_{i_1, i_2, i_3} c_{i_1 i_2 i_3} v_1^{i_1} v_2^{i_2} v_3^{i_3} = \int_0^\infty e^{-t} F(v_1 t, v_2 t, v_3 t, z_1 t, z_2 t, w_1 t, w_2 t) dt, \quad (6)$$

$$F(v_1, v_2, v_3) = \sum_{i_1, i_2, i_3} \frac{c_{i_1, i_2, i_3}}{(i_1 + i_2 + i_3)!} v_1^{i_1} v_2^{i_2} v_3^{i_3}. \quad (7)$$

To obtain an analytic continuation of the Borel transform of a function, a series in an auxiliary variable θ is introduced:

$$\tilde{F}(v_1, v_2, v_3, \theta) = \sum_{k=0}^{\infty} \theta^k \sum_{i_1, i_2, i_3} \frac{c_{i_1 i_2 i_3}}{k!} v_1^{i_1} v_2^{i_2} v_3^{i_3} \delta_{i_1 + i_2 + i_3, k}. \quad (8)$$

The Padé approximant $[L/M]$ is applied to this series at $\theta = 1$. The $[2/1]$ approximant is used to calculate the β functions in the two-loop approximation. The critical behavior is completely determined by the stable fixed point (v_1^*, v_2^*, v_3^*) satisfying the system of equations

$$\beta_i(v_1^*, v_2^*, v_3^*) = 0 \quad (i = 1, 2, 3). \quad (9)$$

The requirement of stability of a fixed point reduces to the condition that the eigenvalues b_i of the matrix

$$B_{i,j} = \frac{\partial \beta_i(v_1^*, v_2^*, v_3^*)}{\partial v_j} \quad (i = 1, 2, 3) \quad (10)$$

lie in the right half-plane.

The resulting system of summed β functions has a wide diversity of fixed points lying in the physical region of vertexes with $v_i \geq 0$. A complete analysis of the fixed points corresponding to the behavior of only one order parameter was presented in [1]. Here, the simultaneous critical behavior of both order parameters is considered. The stable fixed points and eigenvalues of the stability matrix are listed in the table.

An analysis of the critical points and their stability leads to certain conclusions. When $a > 1.6$, bicritical behavior is observed ($v_3^2 \geq v_1 v_2$), as in the case of a short-range interaction [2]. When $1.5 < a \leq 1.6$, tetracritical behavior is observed ($v_3^2 < v_1 v_2$).

Thus, effects due to long-range interaction are responsible for a change from bicritical to tetracritical behavior when $1.5 < a \leq 1.6$.

REFERENCES

1. S. V. Belim, Pis'ma Zh. Éksp. Teor. Fiz. **77**, 118 (2003) [JETP Lett. **77**, 112 (2003)].
2. V. V. Prudnikov, P. V. Prudnikov, and A. A. Fedorenko, Fiz. Tverd. Tela (St. Petersburg) **42**, 158 (2000) [Phys. Solid State **42**, 165 (2000)].
3. D. Amit, *Field Theory, the Renormalization Group and Critical Phenomena* (McGraw-Hill, New York, 1976).
4. J. Zinn-Justin, *Quantum Field Theory and Critical Phenomena* (Clarendon Press, Oxford, 1989).

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 98, No. 2, 2004, pp. 341–347.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 2, 2004, pp. 385–394.

SOLIDS
Electronic Properties

Coherent Phonons in $NdBa_2Cu_3O_{7-x}$ Single Crystals: Optical-Response Anisotropy and Hysteretic Behavior

O. V. Misochko[a], M. V. Lebedev[a], N. Georgiev[b], and T. Dekorsy[b]

[a]*Institute of Solid State Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia*

[b]*Institute for Ion Beam Physics and Materials Research, Forschungszentrum Rossendorf, P.O. Box 510119, D-01314 Dresden, Germany*

e-mail: misochko@issp.ac.ru

Received September 8, 2003

Abstract—Femtosecond pump–probe measurements of reflection from crystallographic planes are performed to investigate lattice relaxation dynamics in the $NdBa_2Cu_3O_{7-x}$ high-temperature superconductor. Ultrafast phonon response is examined over a wide temperature range for various orientations of the pump and probe polarization vectors with respect to particular crystallographic axes. The initial phases of coherent phonons are measured, and hysteretic behavior is revealed in the transition between two temperature regions above T_c for the *ac* plane. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The use of femtosecond laser pulses has opened up new vistas for real-time examination of lattice dynamics [1]. There is currently active interest in pump–probe measurements of lattice relaxation in high-temperature superconductors [2–8]. Time-domain studies of processes responsible for relaxation to equilibrium are frequently conducted to elucidate details of lattice dynamics that cannot be inferred from frequency-domain data.

Recently, pump–probe reflection spectroscopy of coherent phonons and carrier transport in $YBa_2Cu_3O_{7-x}$ has revealed crossovers characterized by dramatic change in the relaxation dynamics of the carriers and lattice in the pseudogap state [7, 8]. The crossovers exhibit hysteretic behavior with respect to temperature variation, thus indicating the inhomogeneity of the pseudogap state. This feature had been predicted theoretically [9–11], but lacked experimental support [12]. The experiments were conducted on films, which made it impossible to measure the anisotropy of optical response.

In this paper, we report the results of a femtosecond pump–probe experiment on bulk $NdBa_2Cu_3O_{7-x}$ single crystals, which are isomorphous to $YBa_2Cu_3O_{7-x}$. The study was conducted to obtain detailed information on lattice relaxation dynamics over a temperature range including both superconducting and pseudogap states. Dynamical optical responses from two crystallographic planes are compared for various orientations of the pump and probe polarization vectors with respect to one another and to particular crystallographic axes.

2. EXPERIMENTAL

The $NdBa_2Cu_3O_{7-x}$ single crystals under study were parallelepipeds whose faces lay in the crystallographic planes to within 3°. The crystals were oriented by x-ray diffraction. The critical temperature T_c was 95 K, indicating a doping level close to that of the $YBa_2Cu_3O_{7-x}$ films examined in [7, 8]. The crystals were placed in an optical helium cryostat, which made it possible to vary the temperature from 10 to 300 K.

The optical response was examined in the time and frequency domains. The time-domain measurements employed a 780-nm Ti:sapphire laser with a pulse width less than 50 fs and a repetition rate of 78 MHz. The response was detected with a fast scanning system [8, 13]. We measured $\Delta R(\tau) = R(\tau) - R_0$, where R is the reflectivity in the excited state, τ is the probe time delay, and R_0 is the reflectivity before excitation. To identify and evaluate phonon modes, the time-domain response was mapped into the frequency domain by numerical Fourier transform (FT). The response anisotropy was examined with respect to the *ab* and *ac* crystallographic planes. Two types of angular dependence were investigated. We aligned the pump vector with a crystallographic axis and varied the angle ϑ between the pump and probe polarization vectors. Alternatively, we set ϑ to 90° and varied the angle α between the pump polarization vector and a crystallographic axis.

We also measured spontaneous Raman scattering in the $x(zz)x$ and $z(x'x')z$ configurations (the prime denotes rotation through 45°). The Raman spectra were obtained with a 647-nm Kr^+ laser and a triple spectrometer equipped with a multichannel detector. This part

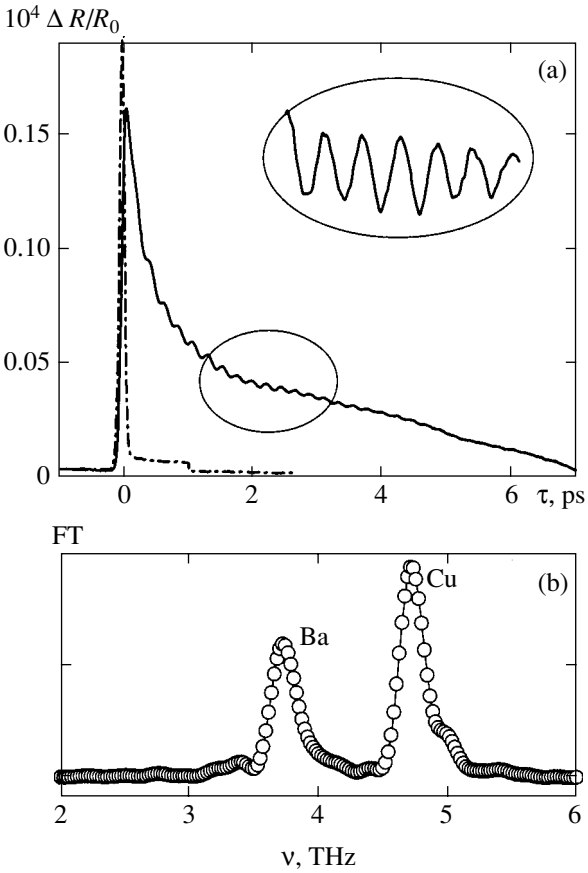


Fig. 1. (a) Time- and (b) frequency-domain representations of reflection from the *ac* plane of the $NdBa_2Cu_3O_{7-x}$ single crystal at room temperature. The solid curve is the optical response. The dot-and-dash curve is the autocorrelation function of the incident pulse. The inset is an enlarged view of coherent-phonon oscillations.

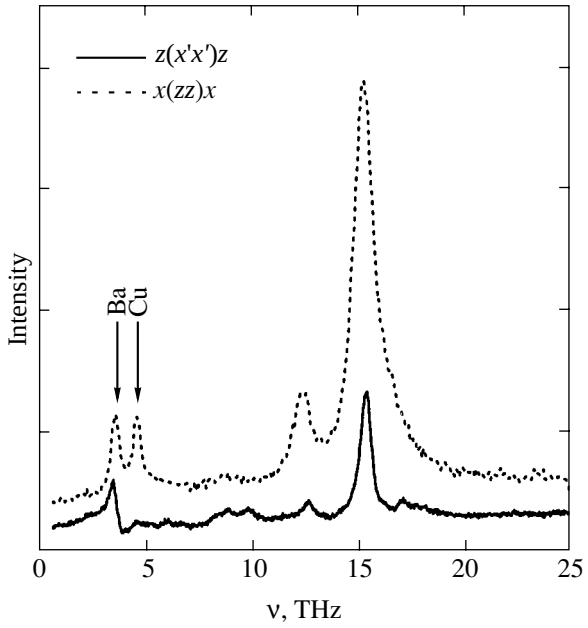


Fig. 2. Room-temperature Raman spectra of the $NdBa_2Cu_3O_{7-x}$ single crystal for two scattering configurations.

of the experiment was focused on fully symmetric phonons.

A detailed description of the apparatus employed can be found in [1, 14].

3. RESULTS

Figure 1 displays typical results obtained by measuring the optical response of a $NdBa_2Cu_3O_{7-x}$ single crystal. Within a relaxation time of several picoseconds, the monotonically decreasing nonexponential reflectivity due to electronic relaxation [8] is modulated by small-amplitude oscillations caused by coherent phonons [1, 15] (see also the inset to Fig. 1a).

The coherent-phonon spectra obtained for both crystallographic planes exhibit two modes (see Fig. 1b), which should be attributed to Ba and Cu ions by comparison with the Raman spectra shown in Fig. 2. The Ba and Cu modes of the Raman spectra are located at $\nu_1 = 3.6$ THz and $\nu_2 = 4.7$ THz, respectively [16]. Note that the Ba mode in the $z(x'x')z$ spectrum is much stronger, whereas the modes have comparable magnitudes in the spectra of $YBa_2Cu_3O_{7-x}$ that correspond to the basis polarizations [14] and in the $x(zz)x$ spectrum of $NdBa_2Cu_3O_{7-x}$.

Being fully symmetric, both phonon modes obey the transformation law for second-rank tensors of the form

$$A_g = \begin{pmatrix} a & & \\ & b & \\ & & c \end{pmatrix}.$$

In this paper, we analyze only the oscillatory component of the optical response as a function of ϑ, α, and temperature.

First, we consider the case when $\vartheta = 90°$ and both pump and probe polarization vectors are in the *ab* plane. The Cu mode predominates at room temperature. As T is decreased, the Ba mode grows, while the Cu mode remains almost invariant. At $T < T_c$, the Ba mode is almost as strong as the Cu one (see Fig. 3). The FT spectra shown in Fig. 3a indicate that the oscillation magnitudes are very sensitive to variation of ϑ. Furthermore, the patterns obtained for $\vartheta = 0$ and $\vartheta = 180°$ are dissimilar, whereas the corresponding crystallographic directions are equivalent.

However, the amplitude ratio $\beta = A_{Ba}/A_{Cu}$ is almost invariant over the entire range of ϑ (see Fig. 3b). This finding is consistent with the high degree of isotropy in the *ab* plane due to twinning and the fact that $|a| \approx |b|$. Figure 3b suggests that β varies as a periodic function

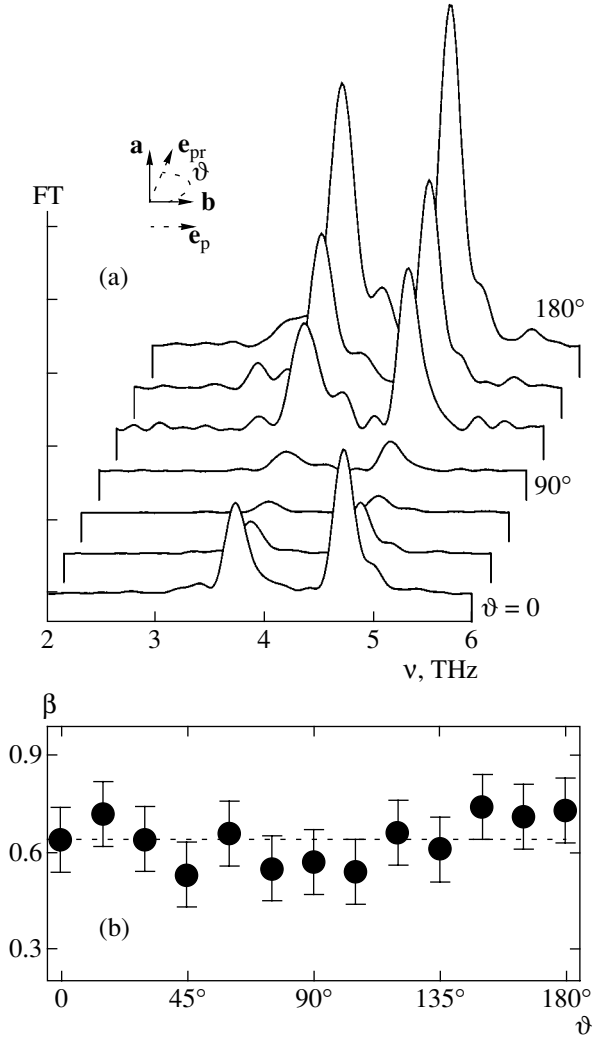


Fig. 3. Reflection from the *ab* plane at 10 K: (a) Fourier transform of the oscillatory component; (b) amplitude ratio of the Ba and Cu modes vs. relative orientation of polarizations.

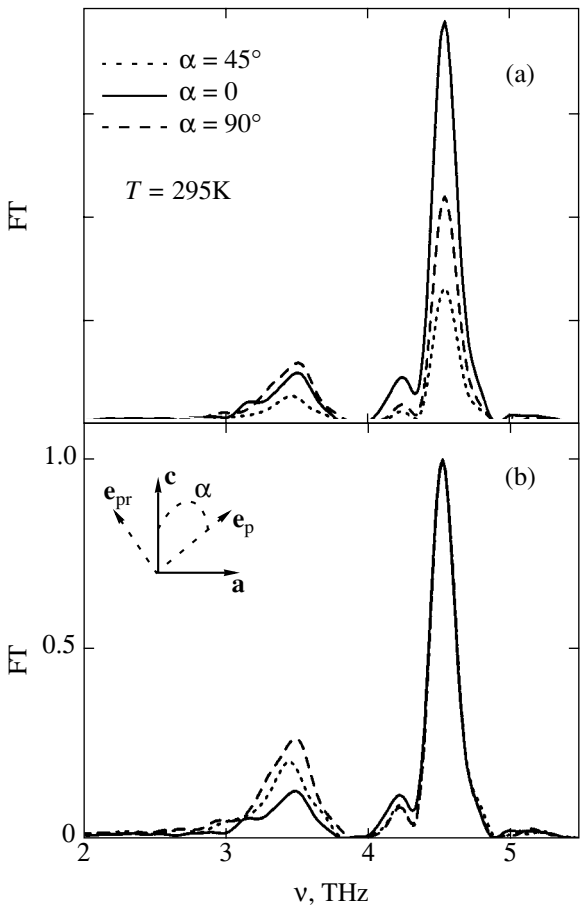


Fig. 4. Reflection from the *ac* plane at room temperature: (a) Fourier transform of the oscillatory component vs. angle between the pump polarization vector and the *c* axis; (b) FT data normalized to unity.

with a period close to 90°, but the data points available are too few to provide conclusive evidence.

Let us now consider the *ac*-plane measurements. Figures 4–7 present FT spectra obtained at room and liquid-helium temperature by varying α or ϑ, with α measured relative to the *c* axis.

Figure 4 shows the results of room-temperature measurements in which α was set to 0, 45°, or 90° while ϑ was held constant at 90°. (Note that the pump polarization vector is aligned with the *a* axis when α = 90° and makes 45° with each crystallographic axis when α = 45°.) Figure 4b shows FT spectra normalized to unity in order to visualize the behavior of β. In these measurements, the largest oscillation amplitude is observed when the pump beam is polarized along the *a* axis (α = 90°). According to Fig. 4, β also varies with α. This is explained by anisotropy with respect to the *ac* plane, which is produced both by the lattice and by the charge carriers.

Figure 5 illustrates the results obtained at 10 K for the same combinations of α and ϑ by electrooptical sampling. In this method, anisotropy is evaluated by measuring the squared difference of the matrix elements, which are uniquely determined by the corresponding angles with respect to the probe polarization vector, ±π/4. Again, both oscillation magnitude and β depend on α. Note, however, that their largest values correspond to α = 45°.

The Fourier spectra and the graph of β versus ϑ shown in Fig. 6 were obtained at 10 K for several values of ϑ. In contrast to the *ab*-plane case, β exhibits a strong angle dependence here. This provides additional evidence for the anisotropy of phonon response. Note also that the largest oscillation amplitude is observed when ϑ = 90°, whereas the response is much weaker when ϑ = 0°.

Figures 7a and 7b show the FT spectra obtained at 10 K by the standard method and by electrooptical sampling, respectively, for the same three values of α. Note

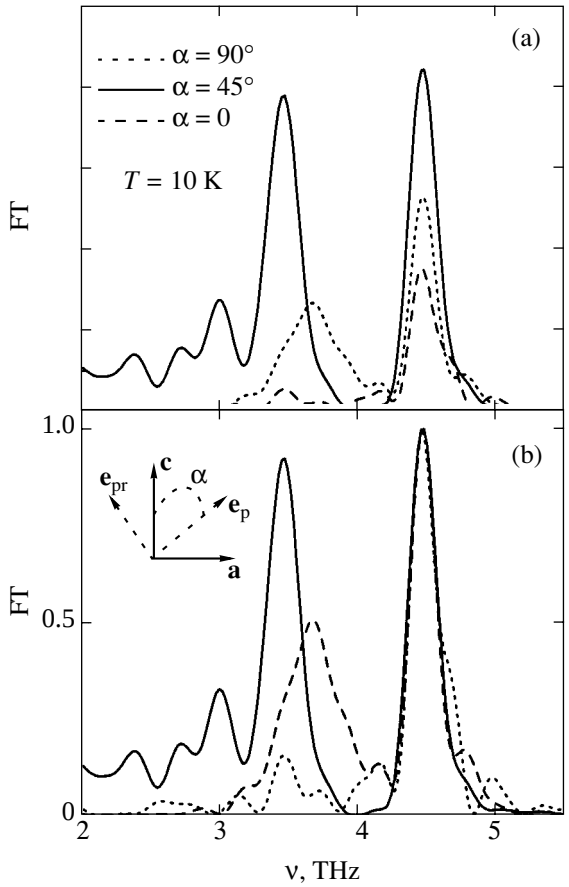


Fig. 5. Reflection from the *ac* plane at 10 K measured by electrooptical sampling: (a) Fourier transform of the oscillatory component vs. angle between the pump polarization vector and the *c* axis; (b) FT data normalized to unity.

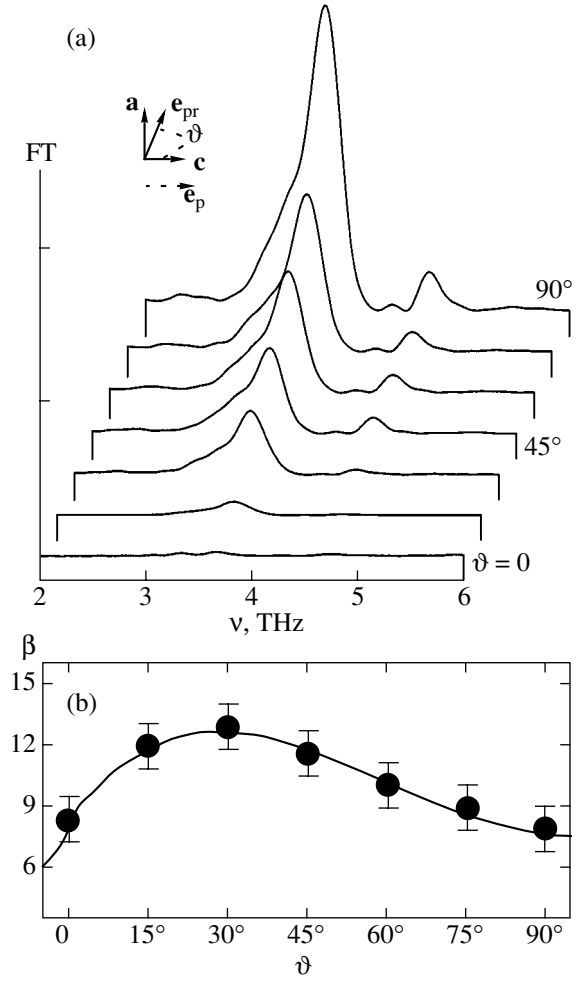


Fig. 6. Reflection from the *ac* plane at 10 K: (a) Fourier transform of the oscillatory component; (b) amplitude ratio of the Ba and Cu modes vs. relative orientation of polarizations.

that the largest oscillation magnitude measured by electrooptical sampling corresponds to $\alpha = 45°$, as in the room-temperature measurements.

The results presented above (particularly in Figs. 4 and 7) suggest that the amplitude ratio β depends not only on the measurement geometry, but also on temperature. Figure 8 shows the results of a more detailed study of this dependence in which the amplitude ratio was measured by varying temperature from about 250 to 10 K and back. Here, $\beta(T)$ exhibits hysteretic behavior at $T > T_c$ analogous to that observed for $YBa_2Cu_3O_{7-x}$ films [7, 8]. These results mean that the pseudogap phase affects the lattice polarizability perpendicular to the *ac* plane; i.e., they point to a three-dimensional nature of the pseudogap. They can be used to assess the accuracy of theoretical models describing the pseudogap state.

We evaluated the initial coherent-phonon phases for the *ac* plane at room temperature. The zero time point was found by measuring the autocorrelation function of the laser pulse with the use of a nonlinear crystal. The oscillatory component of the response was approximated by the function

$$\left(\frac{\Delta R}{R_0}\right)_{\text{osc}} = A_1 \exp(-t/\tau_1)\sin(\nu_1 t + \varphi_1) + A_2 \exp(-t/\tau_2)\sin(\nu_2 t + \varphi_2).$$

The initial values of the amplitudes, frequencies, and time constants were derived from FT spectra. The results shown in Fig. 9 imply that the initial phases are 0 and $\pi/2$ for the Cu and Ba modes, respectively. The phase shift was confirmed by analyzing the FT phases.

4. DISCUSSION

We start by comparing the pump–probe reflection measurements with the Raman spectra. As with $YBa_2Cu_3O_{7-x}$, the oscillatory component of reflectivity mapped into the frequency domain differs from the Raman spectrum. The difference in phonon line shape

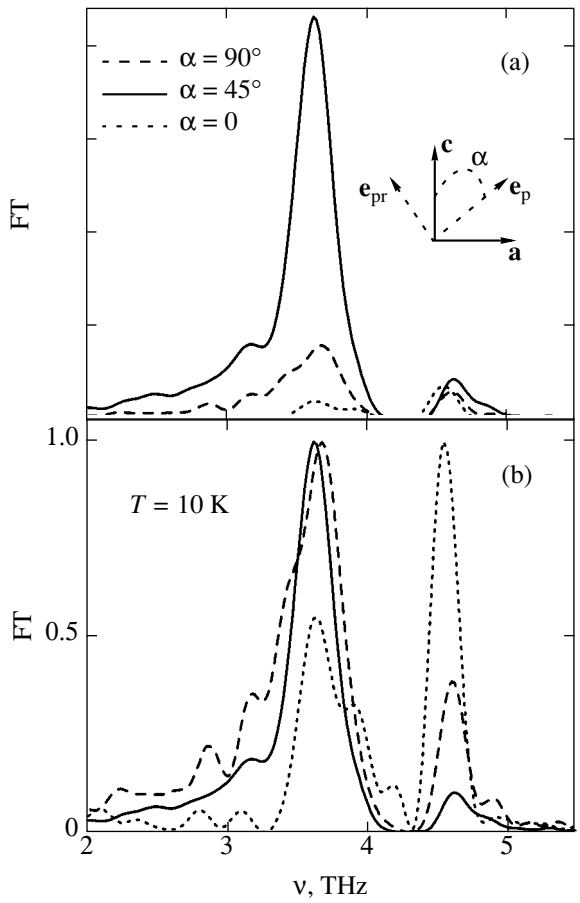


Fig. 7. Comparison between FT spectra of reflection from the *ac* plane at 10 K obtained for different angles between the pump polarization vector and the *c* axis by (a) standard method of detection and (b) electrooptical sampling. The data for $\alpha = 90°$ and $\alpha = 0$ in panel (a) are multiplied by 10^3 and 500, respectively. The data in panel (b) are normalized to unity.

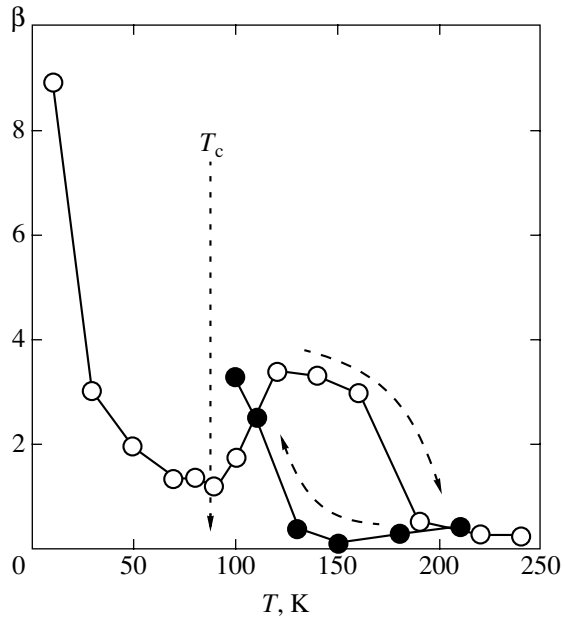


Fig. 8. Amplitude ratio of the Ba and Cu modes vs. temperature for the *ac* plane. Arrows indicate directions of temperature variation.

is predictable because the methods measure distinct states of the phonon system. In pump–probe measurements, the phase is well defined since phonons are brought into a coherent state by the pump, and the spectral range of frequency-domain response characterizes the rate of phase memory loss. In spontaneous Raman spectroscopy, the linewidth characterizes the phonon energy dissipation [1] via interaction with other phonon modes (acting as a thermal reservoir) and the attenuation of an optical phonon is mainly due to its anharmonic decay into two acoustic modes.

At room temperature, $\beta < 1$ and $\beta \geq 1$, respectively, in the time response and Raman spectrum of $YBa_2Cu_3O_{7-x}$ measured for a crystallographic plane. The same is true of $NdBa_2Cu_3O_{7-x}$. Below T_c, both methods yield $\beta \geq 1$ for $YBa_2Cu_3O_{7-x}$, but the β–T characteristics measured by these methods are different. For $NdBa_2Cu_3O_{7-x}$ in the superconducting state, $\beta \gg 1$ in the time response at liquid-helium temperature even though A_{Ba} grows with decreasing temperature, whereas the Raman spectrum gives $\beta \gg 1$. Therefore, it appears unlikely that the disagreement between the time- and the frequency-domain measurements can be fully explained in terms of the absolute values of Raman tensor components. It is also difficult to understand the dissimilarity between the FTs shown in Fig. 3a for $\vartheta = 0°$ and $\vartheta = 180°$. However, since the Raman tensor of a phonon can be expressed in terms of partial derivatives of the dielectric tensor with respect to the phonon normal coordinates, the Raman tensor components are, in general, complex numbers [14, 17]. Whereas standard Raman-spectroscopy measurements are insensitive to the phase factors of the tensor components, this is not the case when the incident and scattered polarization vectors are not parallel to the principal axes of the Raman tensor [17]. In such configurations, two tensor components contribute to the Raman spectrum and its intensity depends on their phases. Owing to the coherent nature of lattice response, phase factors are important for pump–probe measurements. Therefore, the discrepancy between pump–probe and Raman data must be explained in terms of the phase factors [5]. Mathematically, it is analogous to the distinction between a difference of squares and the square of difference.

The importance of the phase factors is illustrated by Fig. 3a. The dissimilarity between the curves obtained for $\vartheta = 0$ and $\vartheta = 180°$ ($[\bar{1}00]$ and $[100]$ directions) can be understood by noting that

$$|b_e\varphi_b + a_e\varphi_a|^2 \neq |b_e\varphi_b|^2 + |a_e\varphi_a|^2$$

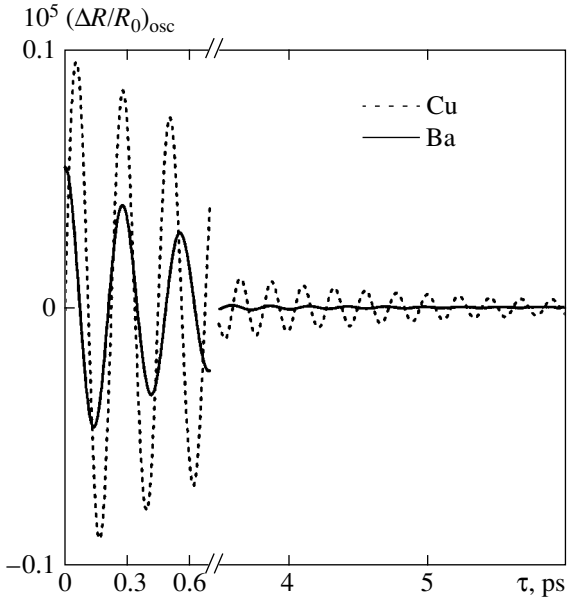


Fig. 9. Determination of the initial coherent-phonon phases for the Cu mode (dotted curve) and the Ba mode (solid curve).

since the interference term, being proportional to $\cos(\varphi_b + \varphi_a)$, has different values at $\varphi_a = 0$ and $\varphi_a = \pi$. This line of reasoning can also explain different types of angular dependence obtained for the *ac* plane, such as the discrepancy between the FT spectra measured by the standard method of detection and by electrooptical sampling for each α (Fig. 7).

The first observations of hysteretic behavior in the $ReBa_2Cu_3O_{7-x}$ system, where Re represents a rare-earth ion, were reported shortly after the discovery of high-temperature superconductivity [12, 18–22]. In most cases, the phenomenon was revealed by ultrasonic measurement of certain elastic constants. For optical modes, hysteresis was originally detected in $YBa_2Cu_3O_{7-x}$ in pump–probe laser measurements [7, 8]. Analogous hysteretic curves were obtained for $NdBa_2Cu_3O_{7-x}$ in the present experiment. Since a comparison of β obtained for $T > T_c$ and $T < T_c$ suggests that the corresponding coherent-phonon spectra are alike, the lattice in the normal state must have some properties of the superconducting state.

Hysteresis in $NdBa_2Cu_3O_{7-x}$ differs from that in $YBa_2Cu_3O_{7-x}$ by a wider range of β and a slight distinction between crossover points. In particular, the transition from the pseudogap state to the normal state with increasing temperature occurs at a lower temperature. However, detailed comparison is hindered by the wide difference in temperature resolution between this experiment and previous ones [7, 8].

The existence of crossovers in the normal state was interpreted as due to the aforementioned inhomogeneity of the pseudogap state [7, 8]. This phenomenon was attributed to local pairing and electron-pair delocalization [9], treated in terms of weak and strong pseudogap regimes of a nearly antiferromagnetic Fermi liquid (NAFL) [10], and explained by the formation of stripes [11]. However, it was noted that none of these models predicted the hysteresis. Moreover, the three-dimensional nature of the pseudogap established in the present experiment implies that essentially two-dimensional models of the pseudogap state (such as NAFL theory) are poorly suited to the study of hysteresis. However, the possibility that the hysteresis is caused by charge ordering was indicated in [23], where the hysteresis was attributed to the lattice anharmonicity associated with the double-well potential generated by the oxygen bridge.

Finally, some comments are in order concerning the initial phase of coherent phonons. In femtosecond pump–probe experiments, the phase can be stabilized by reducing the pump pulse width to below the inverse of the phonon mode frequency. Phase measurements are frequently used to single out the mechanism responsible for generation of coherent phonons, because the two competing phenomenological models proposed to describe coherent lattice dynamics tend to predict different values of the phase. In particular, the initial phase is $\pm\pi/2$ (the phonon mode behaves as a cosine) in the mechanism known as displacive excitation of coherent phonons (DECP), whereas the model of transient stimulated Raman scattering predicts zero phase (sinusoidal pattern) [1]. We note here that this distinction oversimplifies the relation between the phonon phase and the mechanism of phonon generation. In their pioneering work, Kütt *et al.* [3] successfully used calculations based on the DECP model to explain certain properties of coherent phonons in $YBa_2Cu_3O_{7-x}$, including temperature dependence, even though the phase was found to be zero. However, neither the initial phase value nor detection of off-diagonal Raman-active phonons can be explained by DECP calculations [1].

The phase shift of $\pi/2$ between the Cu and Ba phonon modes measured in this study should not be taken to imply that these modes are generated by essentially different mechanisms, because the modes have a variety of common features. In fact, the properties of coherent phonon as a quantum entity are largely determined by the measurement method [1, 25]. Also note that the absolute phase has no physical meaning because the quantity measured in any experiment is a phase difference. We plan to conduct an experimental comparison of coherent-phonon phases measured by the standard method of detection and by electrooptical sampling in the same configuration.

5. CONCLUSIONS

For a bulk $NdBa_2Cu_3O_{7-x}$ single crystal, the anisotropy of the coherent-phonon component of the optical response was examined by femtosecond pump–probe

measurements of reflection from crystallographic planes. The time-domain data were mapped into the frequency domain by numerical Fourier transform. The dependence of the amplitude ratio of the Ba and Cu modes on the angle between the pump and probe polarization vectors was found to be weak for the *ab* plane and strong for the *ac* plane. The anisotropy is explained by the difference between the absolute values of Raman tensor components and the influence of their phase factors. Hysteretic behavior of the optical response is revealed in a temperature range above T_c. For the same plane, it is shown that the initial phases of the fully symmetric Cu and Ba phonon modes are shifted by $\pi/2$ at room temperature.

ACKNOWLEDGMENTS

This work was supported by NATO under grant no. PST.GLG.978935 for joint studies and by the Russian Foundation for Basic Research, project nos. 01-02-16480 and 02-02-17074.

REFERENCES

1. O. V. Misochko, Zh. Éksp. Teor. Fiz. **119**, 285 (2001) [JETP **92**, 246 (2001)].
2. W. Albrecht, Th. Kruze, and H. Kurz, Phys. Rev. Lett. **69**, 1451 (1992).
3. A. Kütt, W. Albrecht, and H. Kurz, IEEE J. Quantum Electron. **28**, 2434 (1992).
4. C. J. Stevens, D. Smith, C. Chen, *et al.*, Phys. Rev. Lett. **78**, 2212 (1997).
5. O. V. Misochko, K. Kisoda, K. Sakai, and S. Nakashima, Phys. Rev. B **61**, 4305 (2000).
6. R. A. Kaindl, M. Woerner, T. Elsaesser, *et al.*, Science **287**, 470 (2000).
7. O. V. Misochko, N. Georgiev, T. Dekorsy, and M. Helm, Pis'ma Zh. Éksp. Teor. Fiz. **75**, 768 (2002) [JETP Lett. **75**, 642 (2002)].
8. O. V. Misochko, N. Georgiev, T. Dekorsy, and M. Helm, Phys. Rev. Lett. **89**, 067002 (2002).
9. V. J. Emery, S. A. Kivelson, and O. Zachar, Phys. Rev. B **56**, 6120 (1997).
10. J. Schmalian, D. Pines, and B. Stojkovic, Phys. Rev. B **60**, 667 (1999).
11. P. Devillard and J. Ranninger, Phys. Rev. Lett. **84**, 5200 (2000).
12. T. Timusk and B. Statt, Rep. Prog. Phys. **62**, 61 (1999).
13. T. Dekorsy, *Kohärente und Inkohärente Ladungsträger- und Gitterdynamik in Halbleitern und Halbleiterheterostrukturen* (Shaker, RWTH Aachen, 1997), p. 69.
14. E. Ya. Sherman, O. V. Misochko, and P. Lemmens, in *Spectroscopy of High-T_c Superconductors: A Theoretical View*, Ed. by N. M. Plakida (Taylor and Francis, London, 2003), p. 97.
15. O. V. Misochko, Fiz. Tverd. Tela (St. Petersburg) **42**, 1169 (2000) [Phys. Solid State **42**, 1204 (2000)].
16. O. V. Misochko, K. Kuroda, and N. Koshizuka, Phys. Rev. B **56**, 9116 (1997).
17. T. Strach, J. Brunen, B. Lederle, *et al.*, Phys. Rev. B **57**, 1292 (1998).
18. H. Yusheng, Z. Baiwen, L. Sihan, *et al.*, J. Phys. F: Met. Phys. **17**, L243 (1987).
19. S. Hoen, L. C. Bourne, C. M. Kim, and A. Zetl, Phys. Rev. B **38**, 11949 (1988).
20. V. Muller, C. Hucho, K. Groot, *et al.*, Solid State Commun. **72**, 997 (1989).
21. A. K. Yahya and R. Abd-Shukor, Supercond. Sci. Technol. **11**, 173 (1998).
22. A. K. Yahya and R. Abd-Shukor, Physica B (Amsterdam) **252**, 237 (1998).
23. A. P. Saĭko and V. E. Gusakov, Zh. Éksp. Teor. Fiz. **116**, 168 (1999) [JETP **89**, 92 (1999)].
24. I. I. Mazin, A. I. Liechtenstein, O. Jepsen, *et al.*, Phys. Rev. B **49**, 9210 (1994).
25. O. V. Misochko, M. Hase, and M. Kitajima, Pis'ma Zh. Éksp. Teor. Fiz. **78**, 85 (2003) [JETP Lett. **78**, 75 (2003)].

Translated by A. Sharshakov

Journal of Experimental and Theoretical Physics, Vol. 98, No. 2, 2004, pp. 348–351.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 2, 2004, pp. 395–398.

NONLINEAR PHYSICS

Effective Equation of Nonlinear Pulse Evolution in a Randomly Anisotropic Medium

I. V. Kolokolov[a,b] and K. S. Turitsyn[b]

[a]*Budker Institute of Nuclear Physics, Siberian Division, Russian Academy of Sciences, pr. Akademika Lavrent'eva 11, Novosibirsk, 630090 Russia*

[b]*Landau Institute for Theoretical Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia*

e-mail: kolokolov@itp.ac.ru, tur@itp.ac.ru

Received May 29, 2003

Abstract—Propagation of a light pulse through a weakly inhomogeneous optical fiber is analyzed. The nonlinear envelope equation describing the evolution of polarized pulses is determined by statistical properties of inhomogeneities in the optical fiber. The isotropic Manakov system of equations is shown to be applicable in the presence of high-frequency small-scale defects in the fiber. In the presence of only large-scale inhomogeneities, the signal dynamics are described by an anisotropic system of equations. © *2004 MAIK "Nauka/Interperiodica".*

Currently, fiber optic communication systems are considered the most promising in information transfer over long distances. Such a system is a sequence of optical fibers and amplifiers. The amplifiers are required to compensate for losses inside a fiber. In the linear regime (when pulse power is low), the channel capacity is primarily limited by the noise generated by amplifiers. Since the amplitude of spontaneous emission noise is independent of signal power, considerable effort is applied to develop soliton systems, where a sequence of digits is encoded into high-power soliton pulses. These systems are characterized by essentially nonlinear signal dynamics. In the case of an ideal fiber, the dynamics are described by the nonlinear Schrödinger equation [1]. In this study, we analyze the more realistic case of a fiber with random fluctuating index profile and polarization-dependent evolution of electric-field energy density. We show that the form of an averaged large-scale equation describing this system strongly depends on the statistics of fluctuations and their scale distribution.

The light pulses used in information transfer have narrow spectral widths $\delta\omega$ compared to the carrier frequency ω_0. They can be described in terms of the envelope defined by a two-component complex vector $\boldsymbol{\psi} = (\psi_1, \psi_2)$:

$$\mathbf{E} = \boldsymbol{\psi}(z, t)\exp(i\omega_0 t) + \boldsymbol{\psi}^*(z, t)\exp(-i\omega_0 t). \quad (1)$$

Here, $\mathbf{E}$ is the electric field of a pulse; z is the longitudinal coordinate in a fiber; and t is the retarded time related to the physical time t_{phys} as $t = t_{\text{phys}} - z/c$, where c is the group velocity of the packet. The evolution equation for the vector $\boldsymbol{\psi}$ is obtained by averaging Maxwell's equations for the electromagnetic field in the fiber medium over the fast-oscillation period $2\pi/\omega_0$. Taking into account the Kerr nonlinearity in chromatic dispersion and choosing appropriate units of $\boldsymbol{\psi}$, z, and t, one can reduce this equation to the following form [2]:

$$-i\partial_z\boldsymbol{\psi} = \partial_t^2\boldsymbol{\psi} + \frac{4}{3}(|\psi_1|^2 + |\psi_2|^2)\boldsymbol{\psi} + \frac{2}{3}(\psi_1^2 + \psi_2^2)\boldsymbol{\psi}^* + \hat{V}(z)\boldsymbol{\psi} + \ldots. \quad (2)$$

The matrix $\hat{V}(z)$ describing birefringence effects is a random function of z because the fiber shape is irregular. This irregularity can be caused by static stresses, technological defects, etc. In what follows, we assume that $V_{\alpha\beta} \gg 1$ unless stated otherwise. Physically, this means that the effects due to nonlinearity and chromatic dispersion are much weaker than birefringence for optical pulses of typical width Δ and amplitude A. This condition is satisfied in real communication lines [1]. In the units of measure used in Eq. (2), $\Delta \sim 1$ and $A \sim 1$.

In Eq. (2), we omit the terms containing time derivatives due to the same inhomogeneities, such as $\hat{m}(z)\partial_t\boldsymbol{\psi}$ and $\xi(z)\partial_t^2\boldsymbol{\psi}$, where $\hat{m}(z)$ and $\xi(z)$ are random matrix and scalar functions, respectively. These corrections for random dispersion are small (about $\delta\omega/\omega_0$) as compared to the terms retained in Eq. (2), and their contribution is significant only at large z. The effective deterministic equation describing unperturbed evolution (if this equation exists, see below) is determined by the statistical properties of the matrix $\hat{V}(z)$ at $z \leq 1$. The form of this averaged equation may depend on the parameters of the problem. In this paper, we refine the

applicability conditions both for specific effective equations and for deterministic description in general.

The term $V(z)\boldsymbol{\psi}$ in evolution equation (2) is responsible for strong dependence of the vector $\boldsymbol{\psi}$ on z. This dependence is eliminated by the transformation

$$\boldsymbol{\psi}(z, t) = \mathcal{T}\exp\left[i\int_0^z \hat{V}(\tau)d\tau\right]\boldsymbol{\Psi}(z, t). \quad (3)$$

The equation of motion for the field $\boldsymbol{\Psi}(z, t)$ contains rapidly oscillating functions of z. However, their amplitudes do not exceed unity, which means that the oscillation scale (about $1/V$) is much smaller than the scale of significant variation of signal amplitude (about 1). Therefore, an averaged description of the system's dynamics is possible.

The matrix $\hat{V}(z)$ is treated as traceless (this can be achieved by a phase transformation of the field $\boldsymbol{\psi}$). Furthermore, we consider fibers that do not exhibit natural optical activity. Therefore, $\hat{V}(z)$ can be represented as

$$\hat{V}(z) = b(\hat{\sigma}_3\cos\theta + \hat{\sigma}_1\sin\theta)$$

(see [3]), where $b(z)$ is the difference of the wave vectors for different polarizations and the angle $\theta(z)$ characterizes the orientation of these polarizations with respect to fixed coordinate axes. It is easy to check that the ordered exponential in Eq. (3) can be represented as

$$\mathcal{T}\exp\left[i\int_0^z \hat{V}(\tau)d\tau\right] = \exp\left[-\frac{i}{2}\hat{\sigma}_2\theta\right]\hat{W}(z), \quad (4)$$

with

$$\hat{W}(z) = \mathcal{T}\exp\left[i\int_0^z\left(b\hat{\sigma}_3 + \frac{\dot{\theta}}{2}\hat{\sigma}_2\right)d\tau\right], \quad (5)$$

where $\dot{\theta} \equiv d\theta/dz$. The matrix $\hat{W}(z)$ is the evolution operator for spin 1/2 in the varying magnetic field $\mathbf{h}(\tau) = (0, \dot{\theta}, b)$. Therefore, the explicit form of $\hat{W}(z)$ strongly depends on the ratio of the amplitude h to the characteristic scale l of its variation ($\dot{\theta}/\theta \sim \dot{h}/h \sim 1/l$). If the fluctuating amplitude $h = \sqrt{\dot{\theta}^2 + b^2}$ is much larger than $1/l$ (which is analogous to the characteristic frequency of the field $\mathbf{h}(\tau)$), then the following estimate for the operator $\hat{W}(z)$ holds up to values of z that are exponentially large in $hl \gg 1$ [4, 5]:

$$\hat{W}(z) = \exp\left[i\int_0^z h(\tau)d\tau + i\Gamma\hat{\sigma}_3\right](1 + \gamma\hat{\sigma}^+ - \gamma^*\hat{\sigma}^-), \quad (6)$$

$$\gamma \sim O\left(\frac{1}{hl}\right), \quad \Gamma \sim 1.$$

Here, Γ is the first correction in the adiabatic expansion for the spinor phase (which is sometimes called the Berry phase [6]) and $\hat{\sigma}^{\pm} = (\hat{\sigma}_1 \pm i\hat{\sigma}_2)/2$. Indeed, the varying profile $\mathbf{h}(\tau)$ can be represented in this case as a superposition of inhomogeneities of characteristic size l. First, consider one such fluctuation localized near the point $z = 0$. For $z \le l$, the off-diagonal elements in the matrix $\hat{W}(z)$ are determined by the "instantaneous" values $\dot{\theta}(z)$, $h(z)$, $\ddot{\theta}(z)$, $\dot{h}(z)$, … and are on the order of $\dot{\theta}(z)/h(z) \sim (hl)^{-1}$. It is easy to see that this parameter is an adiabaticity parameter: the first correction to the adiabatic approximation of $W(z)$ is proportional to $\dot{\theta}(z)/h(z)$. For $z \gg l$, all derivatives $\dot{\theta}(z)$, $\ddot{\theta}(z)$, … vanish and the off-diagonal elements γ are on the order of $\exp(iCh\tau_s)$, where the singular point (or zero) τ_s of the analytic continuation of $h(z)$ into the upper half-plane is nearest to the real axis (for details, see [4, 5]). If this function has no scales other than l, then $\mathrm{Im}\,\tau_s \sim l$ and $\gamma(z \gg l) \sim \exp(-\mathrm{const}\cdot hl)$. When inhomogeneities are repeatedly encountered by a pulse propagating along the fiber, such exponentially small corrections add up. Therefore, the applicability of estimate (6) is limited with respect to z. The inequality $hl \gg 1$ means that the scale of variation of h is much larger than the length $1/h$. Since $h \gg 1$, we can average over $1/h$-scale oscillations after substituting Eqs. (4) and (6) into Eq. (2). The resulting system of equations,

$$\begin{aligned} -i\partial_z\Psi_1 &= (1 + \xi_1)\partial_t^2\Psi_1 + 2\left(|\Psi_1|^2 + \frac{2}{3}|\Psi_2|^2\right)\Psi_1, \\ -i\partial_z\Psi_2 &= (1 + \xi_2)\partial_t^2\Psi_2 + 2\left(|\Psi_2|^2 + \frac{2}{3}|\Psi_1|^2\right)\Psi_2, \end{aligned} \quad (7)$$

was used in [7] to analyze the effects of small noise terms $\xi_{1,2}$ having a relative order of magnitude h^{-1}.

The above analysis is applicable when the Fourier components of the field $\mathbf{h}(z)$ with $k \sim h \gg 1/l$ are suppressed. For a random field $\theta(z)$, these conditions are satisfied when the correlation function $Q(z) = \langle\dot{\theta}(z)\dot{\theta}(0)\rangle$ is decreasing at $z \gg l$ and analytic at $z \longrightarrow 0$. If there exist regions of rapidly varying $\theta(z)$ (sharp bends, defects of structure, etc.), then the form of the matrix $\hat{W}(z)$ is determined by their statistical properties. For example, expression (6) is applicable at moderate distances z when the amplitudes of inhomogeneities are not too large, but with $\gamma \sim \sqrt{nz}$. Here, n is the number of such microscopic defects per unit length estimated as the asymptotic value of the Fourier transform $Q(z)$ at $k \sim h^{-1}$. A similar "Brownian" increase in γ is characteristic of intervals where the amplitude $h(z)$ is about $1/l$ (i.e., regions of nearly circular fiber cross section). The corresponding n is estimated as the frac-

tion of these intervals in the total distance z. We define z_c as the distance for which the off-diagonal elements of $\hat{W}(z)$ are about unity. For a fiber with weak defects, $z_c \sim 1/n$. For a fiber with sharp bends and jumps in θ, the length z_c is estimated as the characteristic distance between such defects. After averaging over scales exceeding z_c, only the identity representation of the group $SU(2)$ in the tensor product $\hat{W}(z) \otimes \hat{W}(z) \otimes \ldots$ is retained. Otherwise, the group $SU(2)$ would contain a subgroup invariant under multiplication by matrices $\hat{W}(z_1, z_2)$ with arbitrary z_1 and z_2. There is no such group unless the amplitude of fluctuations of the direction of $\mathbf{h}(z)$ is zero. This obviously follows from the fact that matrices $\exp(\mathbf{h}_1 \cdot \hat{\boldsymbol{\sigma}})$ and $\exp(\mathbf{h}_2 \cdot \hat{\boldsymbol{\sigma}})$ with noncollinear vectors $\mathbf{h}_1$ and $\mathbf{h}_2$ do not commute. Reduction to an identity representation is equivalent to averaging over an invariant measure on the group $SU(2)$ (e.g., see [8]).

Averaging over $SU(2)$ can be carried out in the equation for varying $\boldsymbol{\Psi}(z, t)$ if $z_c \ll 1$. In this case, the form of the effective equation is determined by the nonzero averages

$$\langle |W_{11}|^2 |W_{12}|^2 \rangle = 1/6, \quad \langle |W_{11}|^4 + |W_{12}|^4 \rangle = 2/3. \quad (8)$$

We conclude that the evolution of a light pulse in a fiber with a relatively high density of microscopic defects is described by the Manakov equations [9]

$$-i\partial_z \boldsymbol{\Psi} = (1 + \boldsymbol{\xi})\partial_t^2 \boldsymbol{\Psi} + \frac{16}{9}(|\Psi_1|^2 + |\Psi_2|^2)\boldsymbol{\Psi}, \quad (9)$$

where $\boldsymbol{\xi}$ denotes small chaotic perturbations (see above). Equations (9) were derived by various methods by Menyuk and Wai (see [10, 11] and references cited therein). However, these authors erroneously concluded that system (9) is universally applicable as a model of pulse evolution if the correlation length of fluctuations of fiber inhomogeneities is much less than both dispersion and nonlinearity length scales ($l \sim 1/\dot{\theta} \ll 1$ in the present units). It was shown above that one must take into account the relative values of $b \sim h$ and $1/l$, as well as the short-wavelength asymptotic behavior of the correlation function of these fluctuations, which is determined by rare events. The importance of the value of hl in the linear problem of evolution of polarization was emphasized in [12].

In principle, an averaged description based on Eq. (7) of Eq. (9) is applicable if $z_c \gg 1 \gg h^{-1}$ or $z_c \ll 1$, respectively. If $z \sim z_c$, the signal shape is determined by the detailed behavior of the functions $\hat{V}(z)$. Indeed, the averaging over $SU(2)$ can be performed only if the trajectory of $\hat{W}(z)$ has traversed the neighborhood of any point of the group manifold a sufficient number of times. The ratio $1/z_c$ is a good measure of this "covering density" on the nonlinearity scale (i.e., over lengths of about 1). When $z \sim z_c$, fluctuations of the moments of the ordered exponential $\hat{W}(z)$ are also on the order of unity and there is no self-averaging. In the limit of $z_c \gg z_c \sim 1$, fluctuating stresses inside the fiber combined with shape fluctuations destroy the pulse [7, 13]; i.e., the maximum amplitude falls well below its initial value. The values of z and z_c can be compared by measuring the ellipticity of a signal that is linearly polarized along one of the principal axes at $z = 0$ in the linear regime.

Finally, we present the basic conclusions of this work. Since we discuss signal propagation in a random medium, only statistics of various observables are generally meaningful. However, the system can be described by deterministic equations when the z_c (characteristic length of change in wave polarization of order unity) has either of two limit values. If $z \ll z_c$, where z is the fiber length, then polarization adiabatically follows the variations of the principal axes of the fiber and Eqs. (7) are applicable. In the opposite limit of $z \geq 1 \gg z_c$, effective self-averaging associated with uniform distribution of polarization over the Poincaré sphere is obtained, and pulse evolution is described by the Manakov equations (9). If $z_c \sim 1$, the system cannot be described by any deterministic model. We note that the fiber can be deformed intentionally to reduce z_c to $z_c \ll 1$ in the soliton regime of information transfer. The reason is that Manakov system (9) is integrable. This property is very important with regard to interaction between solitons via disorder-induced radiation: this interaction in integrable case (9) is much weaker than that in nonintegrable cases [7, 13, 14], and the signal structure is noticeably distorted at much longer distances.

ACKNOWLEDGMENTS

We are grateful to I.R. Gabitov, V.V. Lebedev, and M.V. Chertkov for numerous discussions and questions that stimulated this work. The work of I.V.K. was supported in part by the Russian Foundation for Basic Research (project no. 03-02-16147a) and the Russian Foundation for Support of Science. The work of K.S.T. was supported by the Dinastiya Foundation.

REFERENCES

1. G. P. Agrawal, *Applications of Nonlinear Fiber Optics* (Academic, New York, 2001).
2. A. L. Berkhoer and V. E. Zakharov, Zh. Éksp. Teor. Fiz. **58**, 903 (1970) [Sov. Phys. JETP **31**, 486 (1970)].
3. P. K. A. Wai and C. R. Menyuk, Opt. Lett. **19**, 1517 (1994).
4. A. M. Dykhne, Zh. Éksp. Teor. Fiz. **41**, 1326 (1961) [Sov. Phys. JETP **14**, 941 (1962)].

5. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory*, 3rd ed. (Nauka, Moscow, 1974; Pergamon, New York, 1977).
6. M. V. Berry, Proc. R. Soc. London, Ser. A **392**, 45 (1984).
7. M. Chertkov, I. Gabitov, I. Kolokolov, and V. Lebedev, Pis'ma Zh. Éksp. Teor. Fiz. **74**, 608 (2001) [JETP Lett. **74**, 535 (2001)].
8. M. I. Petrashen' and E. D. Trifonov, *Applications of Group Theory in Quantum Mechanics*, 2nd ed. (UrSS, Moscow, 2000; Butterworths, London, 1969).
9. S. V. Manakov, Zh. Éksp. Teor. Fiz. **65**, 505 (1974) [Sov. Phys. JETP **38**, 248 (1974)].
10. C. R. Menyuk and P. K. A. Wai, J. Light Technol. **14**, 148 (1996).
11. C. R. Menyuk and P. K. A. Wai, J. Opt. Soc. Am. B **11**, 1288 (1994).
12. C. D. Poole, J. H. Winters, and J. A. Nagel, Opt. Lett. **16**, 372 (1991).
13. M. Chertkov, Y. Chung, A. Dyachenko, *et al.*, Phys. Rev. E **67**, 036615 (2003).
14. Y. Chung, V. Lebedev, and S. S. Vergeles, Jr., Opt. Lett. (in press).

Translated by R. Tyapaev

Journal of Experimental and Theoretical Physics, Vol. 98, No. 2, 2004, pp. 352–358.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 2, 2004, pp. 399–405.

NONLINEAR PHYSICS

Structure of the Stochastic Layer in a Driven Pendulum

V. V. Vecheslavov

Budker Institute of Nuclear Physics, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia

e-mail: vecheslavov@inp.nsk.su

Received June 23, 2003

Abstract—An analysis of the stochastic layer in a driven pendulum is extended to the case when the separatrix map contains both single- and double-frequency harmonics. Resonance invariants of the first three orders are found for the double-frequency harmonic. Combined with the previously known single-frequency invariants, they can be used to obtain further information about the layer, in particular, to examine the neighborhoods of zeros of Melnikov integrals. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The problem of interaction between nonlinear resonances and the onset of dynamical chaos in typical (i.e., nonintegrable) Hamiltonian systems remains unsolved and is the subject of ongoing research. The problem can frequently be reduced to analysis of the behavior of a pendulum driven by a quasiperiodic perturbation (under initial conditions close to the main resonance):

$$H(y, p, t) = \frac{p^2}{2} + \omega_0^2 \cos y + V(y, t), \quad (1)$$

$$V(y, t) = \varepsilon_1 \cos(m_1 y - \Omega_1 t) + \varepsilon_2 \cos(m_2 y - \Omega_2 t), \quad (2)$$

where $\varepsilon_1, \varepsilon_2 \ll 1$.

Dynamics in the neighborhood of the main-resonance separatrices have been analyzed in many studies (e.g., see [1–3]). Recall that each separatrix of this kind is a combination of two trajectories connecting two saddle points and corresponding to forward and backward time evolution. Motion along separatrices is characterized by infinitely large periods and substantial interaction between resonances. This region has been commonly considered as the "origin" of chaos.

This is true when the main resonance corresponds to an analytic potential with exponentially decreasing Fourier amplitudes. In this case, each separatrix is split by any perturbing resonance into two distinct trajectories that do not return to their respective starting saddle points, but intersect at the so-called homoclinic points. The free ends of these trajectories make up an infinite number of loops of indefinitely increasing length filling a narrow region near the unperturbed separatrix. These loops are responsible for the formation of a stochastic layer, in which three parts should be distinguished: an upper one of width w_u (where the phase y rotates so that $p > 0$), a middle one of width w_m (where the phase oscillates), and a lower one of width w_l (where the phase y rotates so that $p < 0$).

However, entirely different behavior may be observed when the potential is a smooth function with Fourier amplitudes decreasing as a power of the harmonic number (see [4] and references therein). Striking examples of integer- and fractional-resonance separatrices preserved despite the presence of strong local chaos in perturbed piecewise linear systems were discussed in [5–7].

Analysis of the stochastic layer in a driven pendulum was developed in several steps. The first detailed study of layer formation was presented by Chirikov [1] for a symmetric high-frequency perturbation:

$$V(y, t) = \varepsilon \cos\left(\frac{m}{2} y - \Omega t\right) + \varepsilon \cos\left(\frac{m}{2} y + \Omega t\right), \quad (3)$$

where $\Omega \gg \omega_0$ and m is an integer. He used the standard map and an original criterion for resonance overlap to show that the three parts of the layer have equal widths in the high-frequency limit:

$$w_u = |w_m| = w_l = \lambda W, \quad \lambda \longrightarrow \infty, \quad (4)$$

where $\lambda = \Omega/\omega_0$ is the so-called adiabaticity parameter and $w = H(x, p, t)/\omega_0^2 - 1$ is the relative energy deviation from the unperturbed separatrix.

The quantity W in (4) is the harmonic amplitude of the separatrix map with frequency Ω for system (1), (3):

$$\bar{w} = w + W \sin x, \quad \bar{x} = x + \lambda \ln(32/|\bar{w}|), \quad x = \Omega T_0 \bmod (2\pi). \quad (5)$$

This map (originally introduced in [8]) approximately describes the near-separatrix dynamics at moments T_0 when the system passes through a stable equilibrium

(see also [1]). Iterating the map, one can find the energy width (4) of the layer.

The theoretical value $W = W_t$ is expressed as

$$W_t(\lambda) = \varepsilon\lambda A_m(\lambda) \quad (6)$$

in terms of the Melnikov integrals $A_m(\lambda)$ introduced in [1] (see also [9, 11]). It should be emphasized that W_t is an exponentially decreasing function of λ in the high-frequency limit $\lambda \longrightarrow \infty$.

Chronologically, the next step was to analyze the case of a high-frequency asymmetric perturbation (2). The first numerical experiments showed that the corresponding dynamics are entirely different from those observed in the symmetric case [9, 10]. In addition to the frequencies Ω_1 and Ω_2 of perturbation (2), the spectrum of the separatrix map contains harmonics with amplitudes proportional to $\varepsilon_1\varepsilon_2$ and frequencies equal to the sum and difference of the primary (perturbation) frequencies. Even more surprising is the fact that these secondary harmonics completely determine the stochastic-layer width under certain conditions. In the example considered in [10], the contribution of the sum-frequency secondary harmonic to the amplitude of the separatrix map is greater than the contributions of the primary harmonics by a factor of several hundreds. This is explained by the exponential dependence of W_t on frequency, which leads to an unexpectedly strong effect of weak low-frequency harmonics.

In [11], the resonance invariants of the first three orders were found for separatrix map (5) and the results obtained by using these invariants were compared with numerical findings. It was also shown that each of these invariants adequately represents the corresponding resonance in terms of location and geometry and provides a reliable estimate for the jump in the width of the stochastic layer merging with the resonance. However, note that no analytical tool, including the invariants, can be used to describe chaotic dynamics and, in particular, to characterize the stochastic-layer width.

Finally, the low-frequency asymptotics associated with symmetric perturbation (3) were analyzed in a recent study [12], where it was shown that the energy width of the corresponding stochastic layer is independent of frequency. This follows directly from the asymptotic properties of Melnikov integrals and from the fact that their derivation in [1] did not make use of any simplifying assumptions. Being valid for any perturbation frequency, they tend to constant values (depending only on m) as $\lambda \longrightarrow 0$, whereas the separatrix-map amplitude scales linearly with λ. This implies that the contribution of the secondary harmonics generated under an asymmetric perturbation vanishes in this limit, in contrast to the case of high-frequency perturbation.

It was also shown in [12] that the problem is most difficult to analyze in the intermediate frequency range $0.1 \lesssim \lambda \lesssim 5$, where the adiabaticity parameter cannot be treated as small or large. Moreover, it was noted that resonance invariants could play an important role in this range. An analysis of this possibility is presented below, after a discussion of the spectrum of the separatrix map.

2. RESONANCE INVARIANTS OF THE SEPARATRIX MAP

In the intermediate frequency range, the separatrix map for a pendulum driven by symmetric perturbation (3) actually has two harmonics, with frequencies Ω and 2Ω (see [12]). It should be rewritten as

$$\begin{aligned} \bar{w} &= w + W_1\sin x + W_2\sin 2x, \\ \bar{x} &= x + \lambda\ln(32/|\bar{w}|), \quad x = \Omega T_0 \bmod(2\pi), \end{aligned} \quad (7)$$

where W_1 and W_2 denote the single- and double-frequency amplitudes, respectively. As either low- or high-frequency asymptotic limit is approached, the double-frequency harmonic weakens and its contribution can be neglected. However, this cannot be done in the intermediate range, particularly in the neighborhoods of Melnikov integrals, since they are localized in this range and their number increases in proportion to the harmonic number m [1]. Moreover, when the value of the adiabaticity parameter λ corresponds to a zero of an integral, the single-frequency harmonic amplitude goes through zero (its sign changes) and the separatrix map in this neighborhood has again only one frequency, as in (5), but the frequency is 2Ω (which follows from (7) with $W_1 = 0$).

Using the technique proposed in [9], one can show that this harmonic is a secondary one, i.e., proportional to ε^2 (which is supported by numerical experiment), and corresponds to the frequency difference. We omit this analysis here because the resulting expressions are too cumbersome.

The resonance invariants for separatrix map (5) with single frequency were found in [11]. Here, the double-frequency invariants are found, which makes it possible to examine the neighborhoods of Melnikov integrals.

Recall that both the separatrix map and Chirikov's standard map belong to the class of so-called explicit rotation maps of the form

$$\bar{J} = J - \xi\frac{dQ(\theta)}{d\theta}, \quad \bar{\theta} = \theta + 2\pi\nu(\bar{J}), \quad (8)$$

where J is action, θ is angle, $\nu(J)$ is the frequency of unperturbed motion, and $Q(\theta + 2\pi) = Q(\theta)$ is a trigonometric polynomial in the general case (e.g., see (7)).

Difference equations (8) are known to be equivalent to the Hamiltonian equations of motion perturbed by kicking:

$$\dot{\theta} = \nu(J), \qquad \dot{J} = -\xi\delta_*(t)\frac{dQ(\theta)}{d\theta}, \tag{9}$$

where

$$\delta_*(t) = \left[1 + 2\sum_{n\geq 1}\cos nt\right]/2\pi$$

is a periodic delta function [1, 2].

For the separatrix map that has only the double-frequency harmonic,

$$Q(\theta) = \frac{\cos 2\theta}{2}, \quad \nu(J) = \frac{\lambda}{2\pi}\ln\frac{32}{|J|}.$$

In [13], the resonance invariant of order q was defined as the function of the form

$$S_q(J(t), \theta(t), t) = r_0(J) + \sum_{n=1}^{q}\frac{\xi^n}{n!}G_n(J, \theta, t), \tag{10}$$

which satisfies the following condition along the true trajectories of system (9):

$$\frac{D}{Dt}S_q(J(t), \theta(t), t) = \mathbb{O}(\xi^{q+1}), \tag{11}$$

where D/Dt is the generalized derivative equal to the sum of the classical derivative and a component associated with discontinuities [14]. It was also shown in [13] that each term in (10) is preserved within the intervals between kicks and changes only at the kicking moments. Thus, the classical derivative in (10) is zero everywhere, and condition (11) should be interpreted as a constraint on a jump of the first kind in a step function.

Following the technique detailed in [13], find the resonance invariants of the first three orders for the double-frequency harmonic:

$$G_1(J, \theta, t) = \frac{r_0'(J)}{2\sin(2\pi\nu(J))}\cos 2\Theta, \tag{12}$$

$$G_2(J, \theta, t) = \frac{a_1'(J)}{4\sin(4\pi\nu(J))}\cos 4\Theta, \tag{13}$$

$$G_3(J, \theta, t) = \frac{r_0'''(J) + 3a_2'(J)}{16\sin(6\pi\nu(J))}\cos 6\Theta - 3\frac{r_0'''(J) + a_2'(J)}{16\sin(2\pi\nu(J))}\cos 2\Theta. \tag{14}$$

Here, primes denote derivatives with respect to J, and use is made of the relations

$$\begin{aligned}\Theta &= \theta + (\pi - t_*)\nu(J),\\ a_1(J) &= r_0'\cot(2\pi\nu(J)),\\ a_2'(J) &= a_1'\cot(4\pi\nu(J)),\end{aligned} \tag{15}$$

where $0 < t_* \leq 2\pi$ is the local time between consecutive kicks.

Expressions (12)–(14) contain resonance denominators, which should be eliminated by the method described in [15]. To do this, the function r_0' contained in them must be defined so that the number and multiplicity of its zeros are equal to the number and multiplicity of poles of the resonance denominators, respectively. These functions and the invariants of the first three orders for the double-frequency harmonic are expressed in terms of x and w in the Appendix.

3. COMPARISON OF NUMERICAL RESULTS AND INVARIANTS

As an example, consider the symmetrically perturbed system (1), (3) with

$$\omega_0 = 1.0, \quad \varepsilon = 0.01, \quad m = 4, \quad \Omega = 1.8.$$

The corresponding separatrix map is (7) with the nearly equal amplitudes $W_1 = 3.81\times 10^{-3}$ and $W_2 = 2.93\times 10^{-3}$. The adiabaticity parameter is $\lambda = 1.8$.

The main stochastic layer is illustrated by the left-hand panel of Fig. 1, where the ordinate axis is the relative width $v = w/|W_1 + W_2|$. At first glance, the pattern in the upper part of the layer is reminiscent of a half-integer resonance. However, it represents two integer resonances whose interior regions (stability islands) are not related to one another.

The right-hand panel shows the contour map of the first-order invariant of the separatrix map (see Eq. (A.5)), which correctly describes the combined effect of both harmonics contained in (7).

The next two figures correspond to the second and third double-frequency invariants. Accordingly, they are plotted for the zero of a Melnikov integral, where the single-frequency harmonic vanishes ($A_4(\lambda) = 0$ at $\lambda = \lambda_1 = 1.41715$). The ordinate axis in these figures is $v = w/W_2$.

Figure 2 corresponds to system (1), (3) with $\varepsilon = 0.0287$ and $\Omega = 1.41715$, for which $W_1 = 0$ and $W_2 = 6.7\times 10^{-3}$ in separatrix map (7). The figure shows two chaotic trajectories. The upper one encompasses the half-integer resonance of the separatrix map, whose

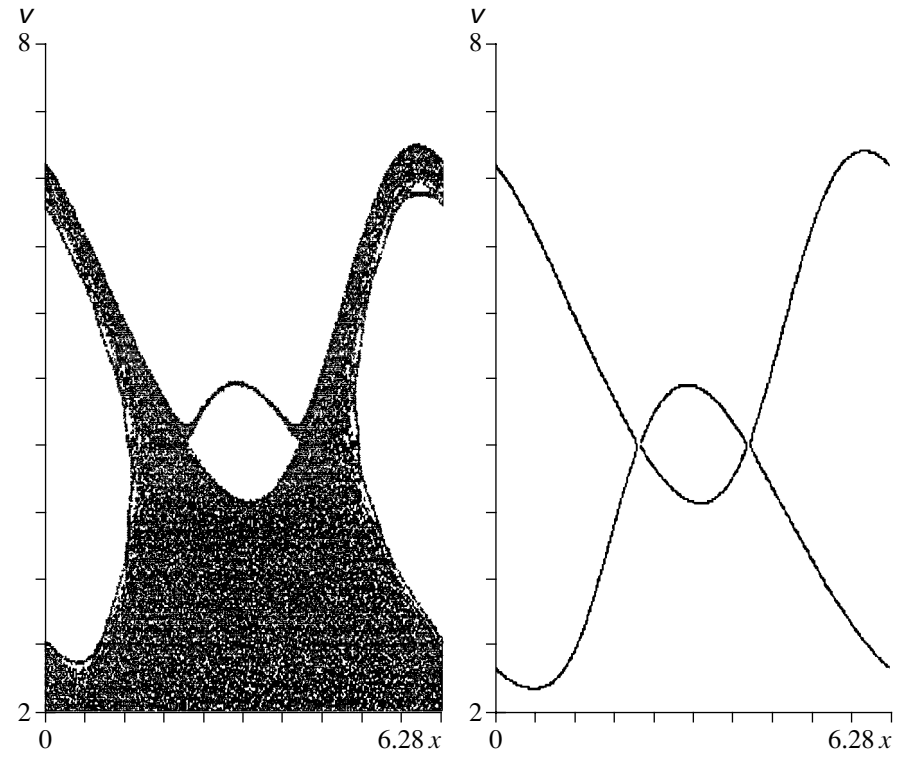


Fig. 1. Stochastic layer of the main resonance (left) and contour map of first-order invariant (A.5) (right) for harmonic amplitudes $W_1 = 3.81 \times 10^{-3}$ and $W_2 = 2.93 \times 10^{-3}$ in (7).

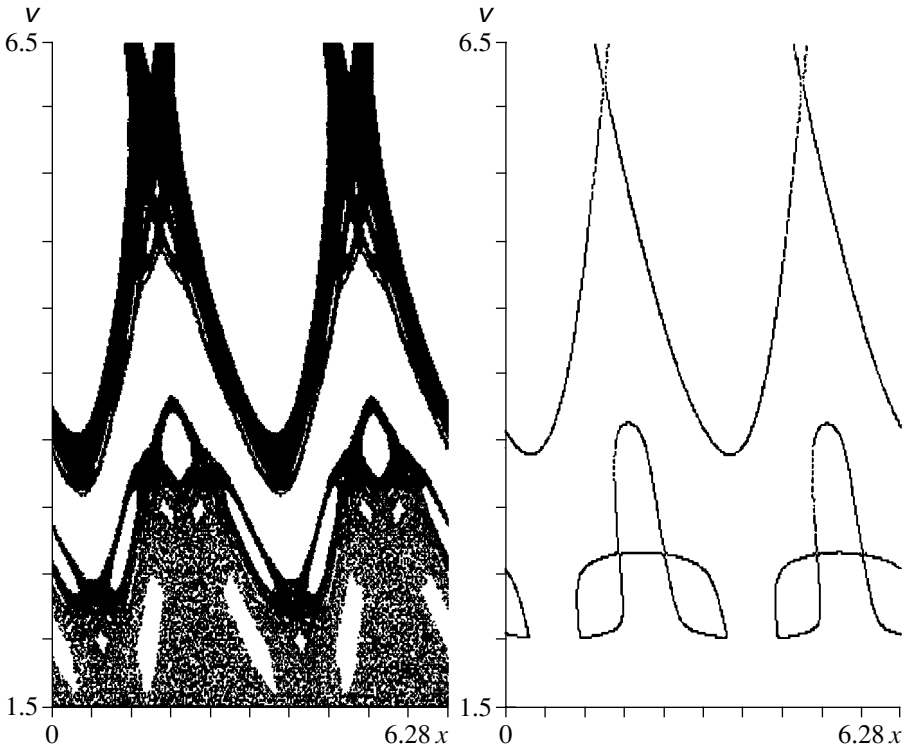


Fig. 2. Stochastic layer of the main resonance (lower trajectory of the two in left panel) and the first- and third-order invariants (A.5) and (A.7) (upper and lower contours in right panel) for harmonic amplitudes $W_1 = 0$ and $W_2 = 6.7 \times 10^{-3}$ in (7).

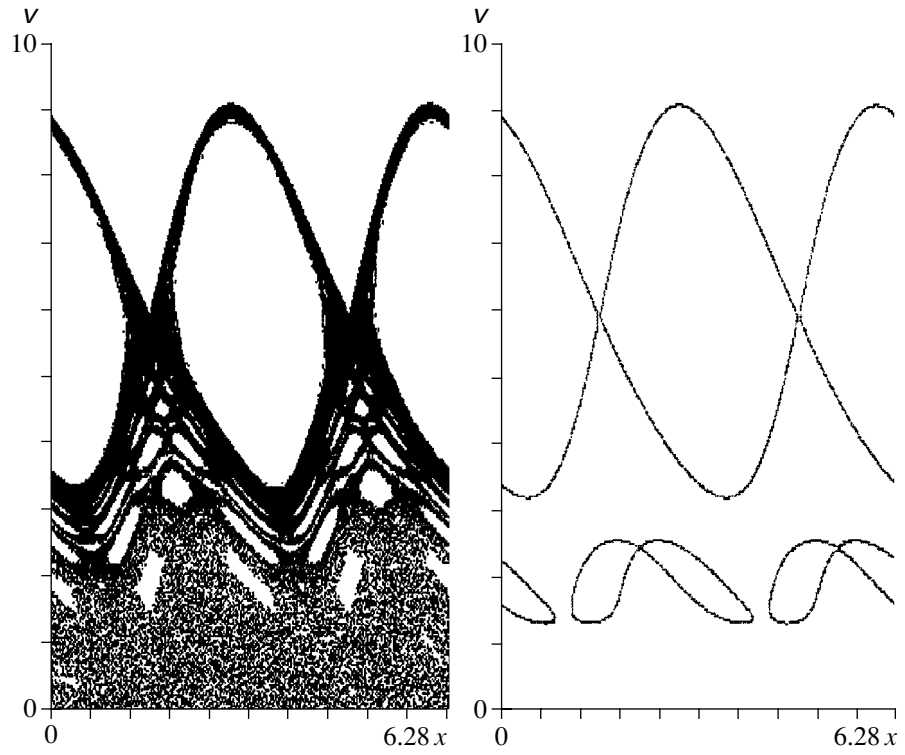


Fig. 3. Stochastic layer of the main resonance (left) and the first- and second-order invariants (A.5) and (A.6) (upper and lower contours in right panel) for harmonic amplitudes $W_1 = 0$ and $W_2 = 7.0 \times 10^{-3}$ in (7).

islands are mapped to one another. The lower trajectory is the main stochastic layer. The figure illustrates the pattern corresponding to a perturbation under which these objects are about to merge.

In the main layer, one can clearly see groups of four and six islands described by the second and third double-frequency invariants (Eqs. (A.6) and (A.7), respectively). An analysis of these groups shows that the former represents a fourth-order resonance and the latter corresponds to two separate third-order resonances with alternating islands. In the main layer, these resonances overlap. The curves of the corresponding invariants intersect, comprising an unsightly pattern. For this reason, we plotted only the curves of the third invariant here.

Finally, Fig. 3 illustrates the main resonance merging with the main stochastic layer at $\varepsilon = 0.03$. Here, the layer width increases stepwise by the phase volume of the merging resonance. Again, the curves representing the second and third invariants intersect, and only the second one is depicted as an example.

To demonstrate the accuracy of the second and third double-frequency invariants, we examined the neighborhood of the only zero of the Melnikov integral with $m = 4$. It was mentioned above that the number of zeros of such integrals increases with m. According to our preliminary measurements, the separatrix map contributes to each zero via the only double-frequency harmonic. We expect that the importance of the invariants found in this study will substantially increase with m.

4. CONCLUSIONS

Driven pendulum, as well as its discrete counterpart (the standard map), is a very popular model of nonlinear dynamics, investigated and employed in numerous studies.

The resonance invariants of the separatrix map considered in [11] and in this study can be used as a basis for more detailed studies of the stochastic layer of the pendulum in the intermediate-frequency range, which is most difficult to analyze. Practical work has shown that this task can be accomplished by using the invariants of the first three orders.

It is also advisable to develop the general theory of invariants further and to derive second- and third-order expressions allowing for the contributions of both harmonics contained in (7), as done for the first invariant (see (A.5)).

ACKNOWLEDGMENTS

This work was supported, in part, by the Russian Foundation for Basic Research, project no. 01-02-16836 and by the Russian Academy of Sciences under the multidisciplinary scientific program “Mathematical Methods in Nonlinear Mechanics.”

APPENDIX

Resonance Invariants of the Separatrix Map with Double Frequency

Practical calculation of invariants of the separatrix map should be performed after changing from the action–phase variables (J, θ) back to (x, w). Accordingly, primes in all equations written out below denote derivatives with respect to w. As pointed out in [11], straightforward use of expressions (12)–(14) is impeded by the numerical instability of calculation of the contour maps of the invariants. This impediment can be dealt with by changing from w to the new variable

$$z(w) = \frac{\lambda}{2}\ln\frac{32}{|w|}. \tag{A.1}$$

To obtain a correct representation of resonances in the (x, w) plane, one should set t_* equal to 2π in (15).

The particular functions r_0' proposed in [13] to eliminate the small denominators of the standard map were found to apply to the single-frequency separatrix map and were employed in [11]. However, new functions were required in the double-frequency case. In our notation,

$$r_0' = \sin 2z \tag{A.2}$$

for the first-order invariant,

$$r_0' = \sin^2 2z\sin 4z = \frac{\sin 4z}{2} - \frac{\sin 8z}{4} \tag{A.3}$$

for the second-order one, and

$$r_0' = \sin^2 2z\sin^2 4z\sin 6z = \frac{1}{16}(\sin 2z + 3\sin 6z - \sin 10z - 2\sin 14z + \sin 18z) \tag{A.4}$$

for the third-order one. The first three double-frequency invariants are written out below. Since the first invariant provides a good description of the two-frequency separatrix map (7) as well (see Fig. 1), we express it in terms of both W_1 and W_2:

$$S_1 = e^{-2z/\lambda}(\sin 2z + \lambda\cos 2z) + \frac{1+\lambda^2}{32} \times\left[W_1\cos z\cos(x-z) + \frac{W_2}{2}\cos 2(x-z)\right], \tag{A.5}$$

$$S_2 = 16e^{-2z/\lambda} \times\left[\frac{\sin 4z + 2\lambda\cos 4z}{1+4\lambda^2} - \frac{\sin 8z + 4\lambda\cos 8z}{2(1+16\lambda^2)}\right] + \frac{W_2}{2}\sin 2z\sin 4z\cos 2(x-z) - \frac{\lambda e^{2z/\lambda}W_2^2}{128}\cos 4z\cos 4(x-z), \tag{A.6}$$

$$S_3 = e^{-2z/\lambda}c_0(z) + \frac{W_2}{2}\sin 2z\sin^2 4z\sin 6z \times\cos 2(x-z) - \frac{3\lambda e^{2z/\lambda}}{2}\left(\frac{W_2}{8}\right)^2 \times\sin 4z(\sin 2z + 3\sin 10z) \times\cos 4(x-z) + \frac{\lambda}{6}\left(\frac{e^{2z/\lambda}}{256}\right)^2 \times W_2^3[c_1(z)\cos 6(x-z) - 3c_2(z)\cos 2(x-z)]. \tag{A.7}$$

In the last expression,

$$c_0(z) = g(2z) + 3g(6z) - g(10z) - 2g(14z) + g(18z),$$

$$g(nz) = \frac{2\sin(nz) + n\lambda\cos(nz)}{1 + (n\lambda/2)^2},$$

$$c_1(z) = \lambda(27 + 35\cos 4z + 49\cos 8z + 81\cos 12z) + 7\sin 4z + 7\sin 8z + 9\sin 12z,$$

$$c_2(z) = \lambda(23 + 51\cos 4z + 69\cos 8z + 49\cos 12z) + 5\sin 4z + 11\sin 8z + 7\sin 12z.$$

REFERENCES

1. B. V. Chirikov, Phys. Rep. **52**, 263 (1979).
2. A. J. Lichtenberg and M. A. Lieberman, *Regular and Chaotic Dynamics*, 2nd ed. (Springer, New York, 1992).
3. R. Z. Sagdeev and G. M. Zaslavsky, *Nonlinear Physics: from the Pendulum to Turbulence and Chaos* (Nauka, Moscow, 1988; Harwood, Chur, 1988).
4. B. V. Chirikov, Chaos, Solitons and Fractals **1**, 79 (1991).
5. S. Bullett, Comm. Math. Phys. **107**, 241 (1986).
6. V. V. Vecheslavov, Preprint No. 2000-27, IYaF SO RAN (Inst. of Nuclear Physics, Siberian Division, Russian Academy of Sciences, Novosibirsk, 2000); nlin.CD/0005048.

7. V. V. Vecheslavov and B. V. Chirikov, Zh. Éksp. Teor. Fiz. **120**, 740 (2001) [JETP **93**, 649 (2001)].

8. G. M. Zaslavskiĭ and N. N. Filonenko, Zh. Éksp. Teor. Fiz. **54**, 1590 (1965) [Sov. Phys. JETP **27**, 851 (1965)].

9. V. V. Vecheslavov, Zh. Éksp. Teor. Fiz. **109**, 2208 (1996) [JETP **82**, 1190 (1996)].

10. V. V. Vecheslavov, Pis'ma Zh. Éksp. Teor. Fiz. **63**, 989 (1996) [JETP Lett. **63**, 1047 (1996)].

11. V. V. Vecheslavov, Zh. Tekh. Fiz. **72** (2), 20 (2002) [Tech. Phys. **47**, 160 (2002)].

12. V. V. Vecheslavov, Preprint No. 2003-11, IYaF SO RAN (Inst. of Nuclear Physics, Siberian Division, Russian Academy of Sciences, Novosibirsk, 2003).

13. V. V. Vecheslavov, Zh. Tekh. Fiz. **58** (1), 20 (1988) [Sov. Phys. Tech. Phys. **33**, 11 (1988)].

14. I. M. Gel'fand and G. E. Shilov, *Generalized Functions and Operations with Them* (Gostekhizdat, Moscow, 1958).

15. D. A. Dunnet, E. W. Laing, and J. B. Taylor, J. Math. Phys. **9**, 1819 (1968).

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 98, No. 2, 2004, pp. 359–365.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 2, 2004, pp. 406–413.

NONLINEAR PHYSICS

Behavior of Reaction-Diffusion Waves with Fast Activator Diffusion near Propagation Threshold

A. Yu. Dovzhenko and É. N. Rumanov*

Institute of Structural Macrokinetics and Materials Science, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia

**e-mail: ed@ism.ac.ru*

Received July 11, 2003

Abstract—Numerical simulation is performed to analyze behavior of reaction-diffusion waves in a medium whose parameters are near both the propagation threshold and diffusive (oscillatory) instability boundary. The wave decays in the subthreshold parameter region and propagates at a constant velocity in the parameter region well above the threshold. Just above the threshold, the wave velocity exhibits alternate intervals of chaotic and constant-amplitude oscillations. The transition from steady to chaotic propagation is a sequence of period-doubling bifurcations that occupies a narrow interval of the bifurcation parameter. In the subthreshold region, the wave decay time is a random function of the bifurcation parameter increasing on average toward the threshold.

1. INTRODUCTION

Reaction-diffusion waves are frequently described by a mathematical model containing two diffusion equations with sources [1–3]. The source intensity increases nonlinearly with the concentration of one component (activator) and decreases as a function of the other component (inhibitor). An active medium with parameters near the propagation threshold for reaction-diffusion waves was considered in [4]. A quasi-stationary equation was obtained for the wave velocity. Its solutions corresponding to the suprathreshold and subthreshold values of parameters describe the onset of a steady propagation regime and a slow decay characterized by a certain characteristic time. Since near-critical waves are highly sensitive to disturbances (variations of parameters), small random inhomogeneities (always present in any real medium) induce strong chaotic oscillations of wave velocity and amplitude. The dispersion and correlation time of the oscillations increase as the threshold is approached.

The high near-threshold sensitivity promotes the development of instability. In a quiescent medium, stability is controlled by the inhibitor-to-activator diffusivity ratio. Diffusive instability can develop at both large and small values of this parameter. When the inhibitor diffusivity is high, the uniform distribution of concentrations is unstable. Local buildup of the activator increases the reaction rate, whereas the compensating effect of an accumulating inhibitor is impossible. Formation of stripe patterns, spots, etc., is observed. As the activator diffusivity increases, steadily propagating fronts and pulses transform into pulsating ones and the wave velocity tends to oscillate about an average value. Penetration of the activator into the substance ahead of the front is followed by a "burst" after which the wave decelerates until a new activator-rich layer forms.

In waves of exothermic reaction, the roles of activator and inhibitor are played by heat and reaction products, respectively. The instability parameter is the Lewis number L defined as the ratio of mass and thermal diffusivities. The numerical simulation reported in [4] was performed for $L = 1$, in which case a wave propagating with a constant velocity is stable in the entire parameter region above the threshold. Here, we consider diffusive (oscillatory) instability under conditions of high near-threshold sensitivity and set $L = 0$. The corresponding waves can be described by the equations

$$\frac{\partial \eta}{\partial t} = k(1-\eta)e^{-E/T} = \varphi(\eta, T),$$

$$\frac{\partial T}{\partial t} = \chi\frac{\partial^2 T}{\partial x^2} + \frac{Q}{c}\varphi - \frac{T - T_0}{\tau},$$

where η is the concentration of the reaction products, t is time, k is interpreted as a collision frequency, E is the activation energy, φ is assumed to scale with concentration of the initial substance, T is temperature, χ is thermal diffusivity, x is the coordinate along the direction of wave propagation, Q is the heat of reaction, c is specific heat, T_0 is the ambient temperature, and τ is the thermal relaxation time. The reaction-diffusion wave is a moving layer that separates the initial substance ($\eta = 0$) from the reaction products ($\eta = 1$). Waves of this type (traveling fronts), as well as traveling pulses, are characterized by propagation thresholds (see review [3] and references therein). Above the threshold, each combi-

nation of parameters corresponds to two wave solutions and only the fast wave is stable. These solutions merge and cease to exist at the threshold. Reaction-diffusion waves of different nature exhibit qualitatively similar behavior near the threshold. However, the phenomena associated with the exothermic reaction are more clear-cut and easier to analyze.

Owing to the exponential dependence on temperature, the nonlinearity of the source is essential only in a small neighborhood of the temperature maximum, i.e., in the reaction zone. The ratio of wave width to reaction-zone thickness is determined by the Zeldovich number

$$Z = E(T_b - T_0)/T_b^2, \quad (1)$$

$$T_b = T_0 + Q/c. \quad (2)$$

Since $Z \gg 1$ under typical conditions, the source term φ can be approximated by a delta function. Then, the wave velocity $u(t)$ is described by a nonlinear integral equation that is not easier to solve as compared to the starting equations. However, the velocity varies slowly near the threshold, which makes it possible to derive a simple first-order equation by using an expansion in powers of the acceleration du/dt [4]. An analysis of the applicability of the quasi-stationary equation has shown that it is not valid when L is small. This result is obviously due to diffusive instability: the wave moves unsteadily, and its acceleration is not low. The present analysis of a wave whose properties are determined both by proximity to the threshold and by oscillatory loss of stability relies on numerical simulation. Using the reaction time $\tau_R = (1/k)\exp(E/T_b)$ as the reference time and defining the reference length for x as $\sqrt{\chi\tau_R}$, we rewrite the equations for concentration and temperature as

$$\frac{\partial\eta}{\partial t} = \Phi(\eta, \Theta), \quad \Phi = (1-\eta)\exp\left(\frac{\Theta}{1+A\Theta}\right), \quad (3)$$

$$\frac{\partial\Theta}{\partial t} - \frac{\partial^2\Theta}{\partial x^2} = Z\Phi - \frac{\Theta + Z}{S}, \quad (4)$$

where Θ is the temperature measured from T_b in units of AT_b, with $A = T_b/E$. The Semenov number $S = \tau/\tau_R$ is the key varied parameter. We set $Z = 6$ because, according to [5, 6], the effects due to instability become noticeable at this value: the wave propagates at a constant velocity when $1/S = 0$, whereas the threshold value S_{th} may be adjoined by a narrow region of oscillatory behavior. The length of the interval on which our numerical experiments were performed, $x_0 = 1000$, exceeds the wave width by a factor of ten and is 300 times greater than the leading-edge zone that determines the wave velocity and amplitude. Therefore, we can use the simple boundary condition

$$\partial\Theta/\partial x = 0 \quad (5)$$

at $x = 0$ and $x = x_0$. The initial conditions (see details below) were such that the wave propagated leftwards. Every time when the point where $\eta(x) = 0.5$ entered the left half of the interval (i.e., once $x < x_0/2$), the wave profile was updated, i.e., shifted by ten grid points. Simultaneously, a segment of this length was added and removed at the left and right endpoints of the interval, respectively. The time interval between such consecutive shifts was found to be 15 to 20 time steps, at least. The time step was 0.01, and the mesh size was 0.05. These values are small as compared to the minimal time of reaction burst (in a periodic or chaotic regime) and the reaction-zone thickness, respectively.

When the value of S is substantially higher than S_{th}, the velocity of steady wave propagation is independent of S: $u_0 \approx 0.398$. It is close to the result $1/\sqrt{Z}$ obtained in the narrow-reaction-zone approximation, whereas $S_{th} \approx 326.2$ is much higher than the threshold value $2eZ^2$ obtained in the same approximation. This is a manifestation of oscillatory instability, which must be particularly strong near the threshold, where the wave is highly sensitive. The analysis that follows shows that diffusive instability not only raises the propagation threshold, but also substantially complicates wave propagation in the near-threshold parameter region.

2. ONSET OF PERIODIC REGIME

The wave propagation is steady when $S > S_1 \approx 381$, and pulsations correspond to a relatively small parameter interval of width about $0.1S_{th}$. When S is reduced to S_1, a hysteretic hard transition to pulsating propagation is observed: as S is increased, oscillations disappear at $S_0 \approx 384 > S_1$. Figure 1 shows the maximal and minimal values of velocity in the pulsating regime versus S. The curves plotted in the graph were obtained as follows. We took arbitrary values $S_i > S_0$ and $S_f < S_1$. We solved system (3) numerically for

$$\frac{1}{S} = \frac{1}{S_i} + \left(\frac{1}{S_f} - \frac{1}{S_i}\right)\frac{t}{t_0},$$

where t_0 is the run time for a particular variant. A computation was performed over the time interval t_0 as the value of S was increased from S_f to S_i according to this formula. We see that the location of the transition point depends on t_0 (retardation effect, see [7]). Increase in t_0 leads to saturation, and distinct equilibrium points are approached with increasing and decreasing S. Therefore, the hard transition to the periodic regime is not equivalent to a retarded soft transition due to a slow

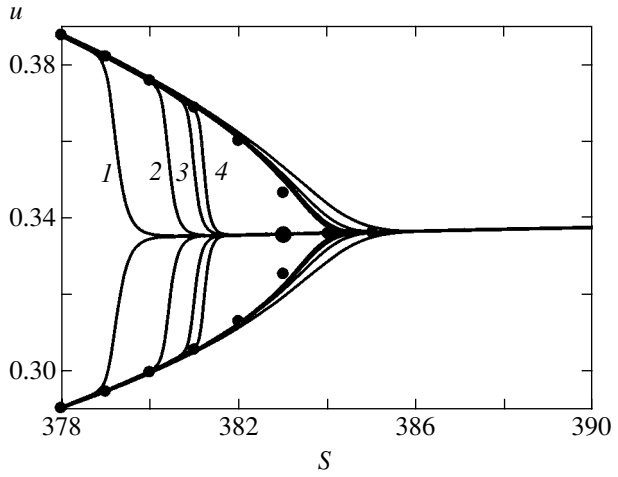


Fig. 1. Peak-to-peak amplitude of velocity oscillation vs. Semenov number. Curves *1–4* were obtained for slowly varying S with $t_0 = (1–4) \times 10^5$, respectively; symbols, for constant S.

decrease in S. These results were checked by performing computations at constant S. Once a chosen value of S was reached, its variation was stopped while the computation was continued until a steady state was obtained (the corresponding run time is 10^4, which is much longer than any reasonable relaxation time). In particular, we obtained a wave propagating at a constant velocity for $S = 383$ after the decrease in S was stopped. This regime is represented by a large circle in Fig. 1, and the other two circles determine the peak-to-peak amplitude of the oscillation observed when S was increased to the same value.

However, the oscillation amplitude is small near S_0. The transition from pulsating to steady propagation may be interpreted as a hard one that is close to a soft transition. This means that the boundary of the basin of attraction is close to the equilibrium point corresponding to steady propagation. In the region of hysteretic behavior, the steady regime is formally stable with respect to infinitesimal disturbances, but those of finite (even small) amplitude must destabilize this regime. This situation is reminiscent of the Poiseuille flow at high Reynolds numbers (e.g., see [8]). The flow can be sustained by removing perturbation sources at the pipe entrance, but a weak disturbance is sufficient to initiate a transition to a turbulent flow regime. The role of disturbance can be played by any small inhomogeneity encountered by a wave propagating in a real medium.

The parametric effect on a steadily propagating wave in the hysteresis region was examined by using a function $y(t)$ such that

$$S = \frac{S_a}{1 + y(t)}. \tag{6}$$

Figure 2 illustrates the oscillatory loss of stability for $S_a = 383$ and different y. First, this value was reached by slowly decreasing S. It was held constant over a time interval of 10^3 until the onset of a steady propagation regime, and then the perturbation $y(t)$ was "switched on." The graph shown in Fig. 2a was obtained by varying y randomly within (–0.1, 0.1) at each step over an interval of 10^3. This "white noise" induced chaotic oscillation of velocity. After it was switched off, the amplitude slowly increased to its value in the steady regime. Figure 2b illustrates the oscillation induced by a pulse having the form of one period of the function $y = k\sin\omega t$, and the period of the limit cycle corresponding to this value of S is $2\pi/\omega = 56.11$. When $k = 0.025$, the relaxation time is almost equal to the period. At lower k, the amplitude increases at a slower rate. At larger k, a large-amplitude oscillation develops in one period and then the amplitude decreases, approaching its steady-state value as illustrated by the graph. Note that the amplitude of oscillation at the steady-state frequency illustrated by Fig. 2a begins to increase after the

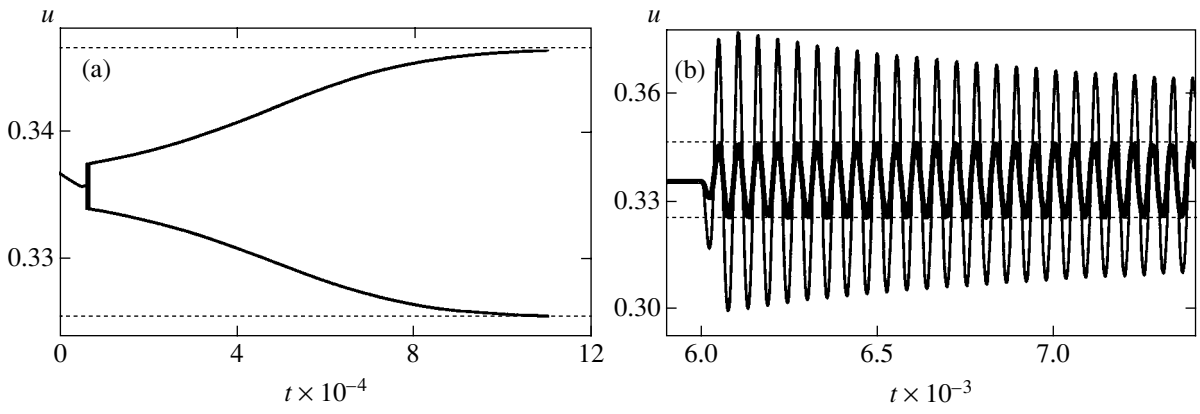


Fig. 2. Oscillations induced by (a) noise and (b) "resonant" pulse.

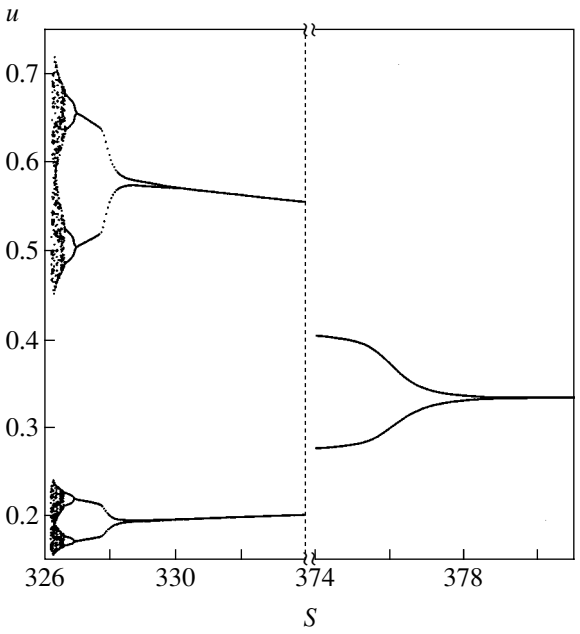


Fig. 3. Extrema of $u(t)$ vs. S.

white noise is switched off. The effect of white noise on steady wave propagation leads only to velocity fluctuations of increasing intensity whose spectrum shifts toward low frequencies as transition point S_1 is approached. Analogous behavior is exhibited by fluctuations near the propagation threshold when $L = 1$ [4].

3. PERIOD-DOUBLING SEQUENCE AND INTERMITTENCY

The regime whose incipience is described in the preceding section occupies almost the entire interval (S_{th}, S_1). The oscillation frequency decreases with S while the peak-to-peak amplitude increases. At S below $S_2 \approx 328$, a period-doubling sequence is observed. Figure 3 depicts the bifurcation diagram. The intervals between consecutive points of period doubling tend to decrease in a geometric progression with common ratio $1/\delta$ close to the tabular value of the Feigenbaum constant [9]. Using this value, we find $S_\infty \approx 326.611$. Figure 4 shows the distribution of maxima of u in the chaotic interval (S_{th}, S_∞). The segment enclosed in the frame is similar to the overall pattern up to a scaling transformation. One can see "gaps" in which the distribution concentrates near a limited number of points. The regions of chaotic behavior expand and overlap with decreasing S. Regions of condensation appear, and the graphs of $u(t)$ exhibit intermittency (see [8] and Fig. 5). Note that the numerical analysis performed for $1/S = 0$ in [10] revealed period-doubling bifurcations associated with increase in Z. No limit point was found for this period-doubling sequence, and the region of chaotic behavior has never been investigated. Intermittency was observed only when the model was modified by introducing a phase-transition front ahead of the reaction zone.

As the threshold S_{th} is approached, the chaotic amplitude distribution tends to widen (see Figs. 3 and 4) and the frequency spectrum becomes more complicated. Figure 6 shows examples of spectra observed in the period-doubling and chaotic intervals. One may characterize the general tendency as gradual transition from discrete to continuous spectra. However, instances of recurrent discrete spectra stand out against this background trend. These are obvious manifestations of frequency locking, which leads to concentration of velocity maxima near several points and the occurrence of "gaps" in Fig. 4. The regime observed just above the threshold can be characterized as well-developed

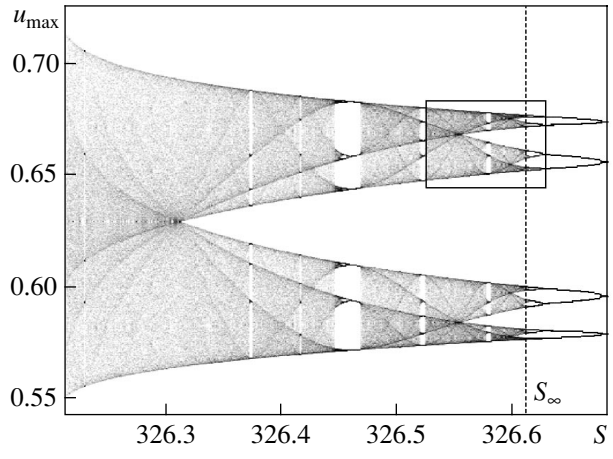


Fig. 4. Maxima of $u(t)$ vs. S (part of bifurcation diagram).

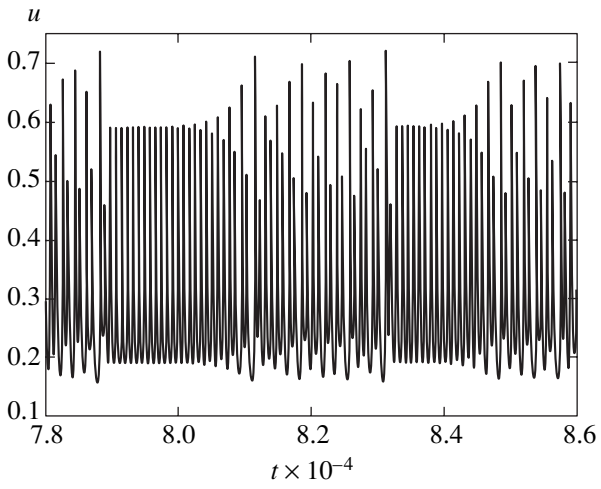


Fig. 5. Example of intermittency at $S = 326.313$, which corresponds to marginal overlapping of amplitude distributions, for a total run time of 10^5.

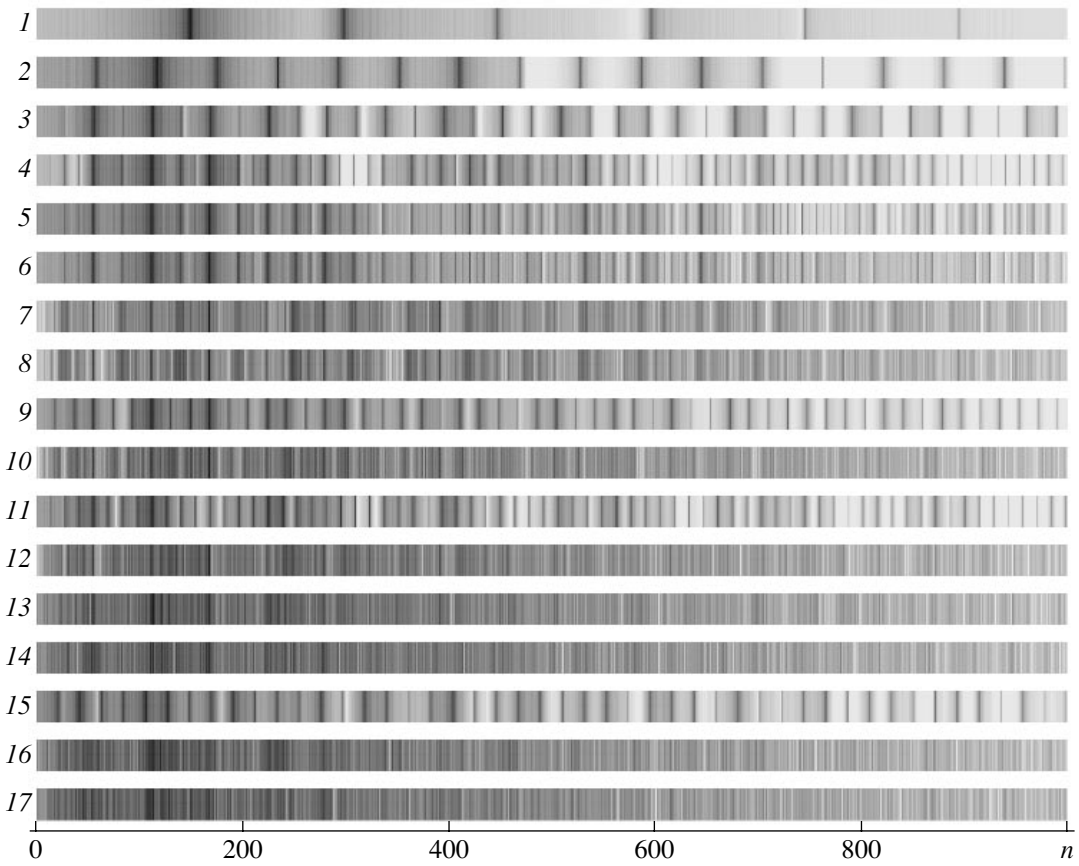


Fig. 6. Frequency spectra of velocity oscillations vs. n ($\omega_n = 2\pi n/t_0$ for a total run time of $t_0 = 10^5$) for $S = 350$ (*1*), 328 (*2*), 326.914 (*3*), 326.640 (*4*), 326.620 (*5*), 326.611 (*6*), 326.551 (*7*), 326.500 (*8*), 326.466 (*9*), 326.400 (*10*), 326.375 (*11*), 326.350 (*12*), 326.313 (*13*), 326.250 (*14*), 326.230 (*15*), 326.219 (*16*), and 326.213 (*17*).

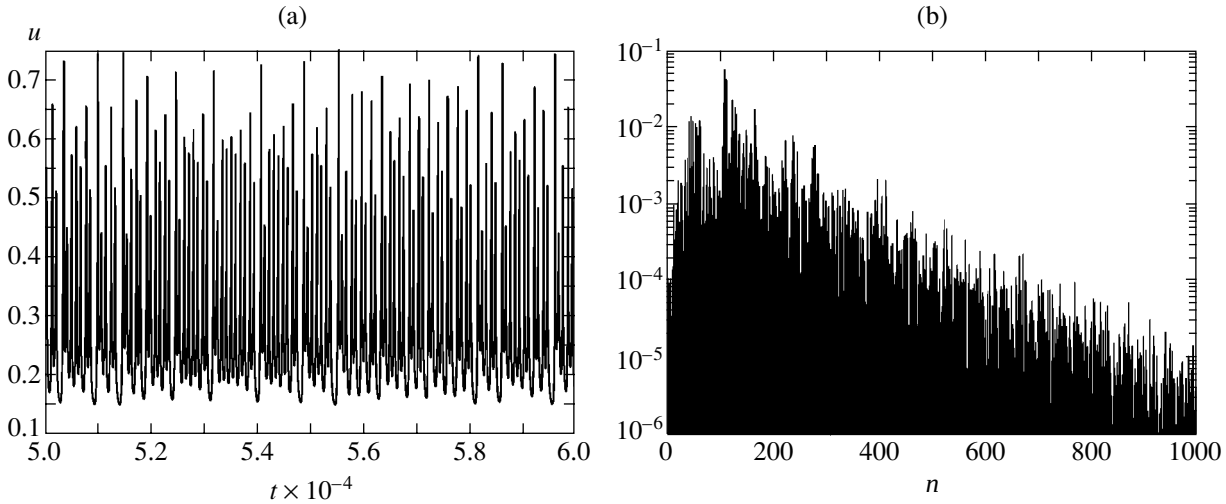


Fig. 7. Long-term wave propagation above the threshold at $S = 326.213$: (a) velocity vs. time; (b) frequency spectrum ($\omega_n = 2\pi n/t_0$, as in Fig. 6).

chaos. Figure 7 shows both function $u(t)$ and its spectrum. The velocity variance is similar to the mean velocity $\langle u \rangle$ in order of magnitude. The highest spectral density corresponds to $n \sim 100$, i.e., $\omega \sim 10^{-3}$, which is two orders of magnitude lower than the typical frequency, $\sim u^2$, and an order of magnitude less than the inverse average time interval between neighboring peaks in Fig. 7a. The shift of the spectrum toward lower

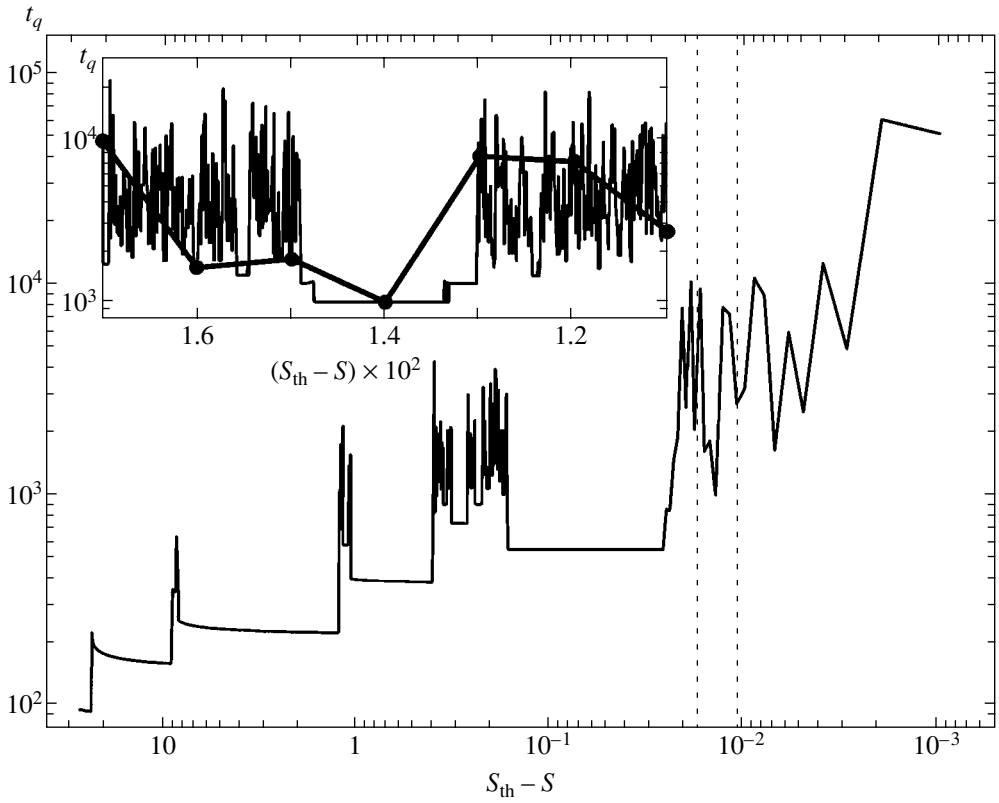


Fig. 8. Wave decay time vs. increment above the threshold. The inset shows an enlarged view of the segment between dashed vertical lines.

frequencies reflects the increase in the correlation time of pulsations toward the threshold [4].

4. DECAY TIME

When $S < S_{th}$, no steady propagation is observed; i.e., the waves decay. The characteristic decay time t_q was estimated in [4] as $t_q \propto (S_{th} - S)^{-\beta}$ with β nearly equal to 1/2, in agreement with solutions to the quasi-stationary equation. Numerical simulations of decay for $L = 0$ reveal an "average" increase in t_q close to that obtained in [4]. However, this monotonic trend is modulated by irregular variations of the decay time (see Fig. 8). These variations can be explained by noting that the long-term dynamics are chaotic just above the threshold, as shown above. In this regime, short intervals of fast propagation characterized by high activator concentrations in the reaction zone are followed by long intervals of "depression" characterized by low activator concentrations, and vice versa. These intervals occur at random. It is obvious that a "depressed" wave is more sensitive to activator losses [5]: a sufficiently long interval of "depression" is followed by an irreversible decline rather than activation. Under the present initial conditions, the onset of a prolonged and deep "depression" turns out to be a random function of S. It is clear that a change in initial conditions will result in different locations and magnitudes of the maxima and minima in Fig. 8. However, the general pattern is independent of initial conditions.

The results depicted in Fig. 8 were obtained by using the concentration distributions in a wave propagating with a constant velocity u as initial conditions:

$$x < x_1, \quad \eta = 0, \quad \Theta + Z = Z\exp[k_+(x - x_1)], \quad (7)$$

$$x > x_1, \quad \eta = 1, \quad \Theta + Z = Z\exp[k_-(x - x_1)], \quad (8)$$

where

$$k_\pm = \frac{u}{2} \pm \sqrt{\frac{u^2}{4} + \frac{1}{S}}.$$

These distributions are written for a reaction zone of zero thickness initially located at $x = x_1 = 500$. The value $S = 400$ corresponds to $u \approx 0.313$. According to (3), the instantaneous wave velocity is

$$u(t) = \int \Phi dx, \quad (9)$$

where the integral on the right-hand side is independent of its limits if they are separated from the reaction zone by distances much greater than the reaction-zone thickness. The decay time t_q was measured from the instant

of the last burst preceding an irreversible decrease in velocity.

In Fig. 8, the jumps correspond to the values of S at which the number of peaks in the curve of decaying $u(t)$ changes. When the number of bursts is constant, the function $t_q(S)$ is smooth. As the threshold is approached, the lengths of such smooth intervals decrease. Reducing the interval between the values of S used to compute the decay, we distinguish finer details of the graph (see inset, where the symbols used to plot the main graph are shown for comparison.) The rearrangement of decay pattern is the most obvious consequence of diffusive instability in the subthreshold parameter region.

5. CONCLUSIONS

The oscillatory instability of reaction-diffusion waves leads to periodic regimes and transition to chaotic behavior analogous to those observed in other well-studied systems (e.g., in flows with increasing Reynolds number). The wave instability in question is due to the temperature sensitivity of the reaction rate (see [3]), which increases with the parameter Z. However, this parameter was held constant in our numerical experiments, while the Semenov number (which measures proximity to the wave propagation threshold) was varied. Near the threshold, reaction-diffusion waves are highly sensitive [4] and the instability develops even at moderate values of Z.

Ignoring details of the pattern discussed here, such as consecutive transitions between chaotic regimes, we see that the prevalent trends observed as $S \longrightarrow S_{\text{th}}$ are the increase in velocity variance and shift of the oscillation spectrum toward lower frequencies above the threshold and the power-law growth of decay time below the threshold. All of these trends were revealed in [4] by simulating a wave that is not characterized by diffusive instability. They are explained by divergence of sensitivity at the low-frequency threshold, which is due to the weakness of the "restoring force" associated with deviations from a steady state that easily develop and take a long time to decay. In the state space, the basin of attraction of the attractor in question is small. This situation is analogous to dynamics in a shallow potential well, in which case a stable equilibrium is close to a flat one. Another example of a highly sensitive system is a substance near the point of a continuous phase transition, in which case critical behavior is caused by thermal noise. Active systems are typically perturbed by more intense sources.

ACKNOWLEDGMENTS

We thank A.G. Merzhanov for discussion of results. This work was supported by the Russian Foundation for Basic Research, project no. 01-02-17487.

REFERENCES

1. V. A. Vasil'ev, Yu. M. Romanovskiĭ, and V. G. Yakhno, Usp. Fiz. Nauk **128**, 625 (1979) [Sov. Phys. Usp. **22**, 615 (1979)].
2. B. S. Kerner and V. V. Osipov, Usp. Fiz. Nauk **157**, 201 (1989) [Sov. Phys.–Usp. **32**, 101 (1989)].
3. A. G. Merzhanov and E. N. Rumanov, Rev. Mod. Phys. **71**, 1173 (1999).
4. A. Yu. Dovzhenko, S. V. Maklakov, I. É. Rumanov, and É. N. Rumanov, Zh. Éksp. Teor. Fiz. **122**, 1125 (2002) [JETP **95**, 973 (2002)].
5. A. P. Aldushin, T. M. Martem'yanova, A. G. Merzhanov, *et al.*, Fiz. Goreniya Vzryva **9**, 613 (1973).
6. A. G. Merzhanov and B. I. Khaikin, Prog. Energy Combust. Sci. **14**, 1 (1988).
7. V. I. Arnol'd, *Catastrophe Theory*, 3rd ed. (Nauka, Moscow, 1990; Springer, Berlin, 1986).
8. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 6: *Fluid Mechanics*, 3rd ed. (Nauka, Moscow, 1986; Pergamon, New York, 1987).
9. M. J. Feigenbaum, J. Stat. Phys. **19**, 25 (1978).
10. A. Bayliss and B. J. Matkowsky, SIAM J. Appl. Math. **50**, 437 (1990).

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 98, No. 2, 2004, pp. 366–373.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 2, 2004, pp. 414–422.

MISCELLANEOUS

Heat Capacity of Isolated Clusters¶

R. S. Berry[a] and B. M. Smirnov[b]

[a]*Department of Chemistry, University of Chicago, Chicago, IL 60637 USA*
[b]*Institute for High Temperatures, Russian Academy of Sciences, Moscow, 127412 Russia*
e-mail: berry@uchicago.edu, smirnov@orc.ru

Received June 10, 2003

Abstract—The character of interaction between thermal (vibrational) and configurational cluster excitations is considered under adiabatic conditions when a cluster is a member of a microcanonical ensemble. The hierarchy of equilibration times determines the character of atomic equilibrium in the cluster. The behavior of atoms in the cluster can be characterized by two effective (mean) temperatures, corresponding to the solid and liquid aggregate states, because the typical time for equilibration of atomic motion is less than the transition time between aggregate states. If the cluster is considered for a time much longer than the typical dwell time in either phase, then it is convenient to characterize the system by only one temperature, which is determined from the statistical–thermodynamic long-time average. These three temperatures are not far apart, nor are the cluster heat capacities evaluated on the basis of these definitions of temperature. The heat capacity of a microcanonical ensemble may be negative for two coexisting phases if the mean temperature is defined in terms of the mean kinetic energy, rather than as the derivative of energy with respect to microcanonical entropy. However, if the configurational excitation energy is smaller than the total excitation energy separating the phases, then the two-state model predicts a positive heat capacity under either definition of temperature. Moreover, if the cluster is sufficiently large, then the maximum values of the microcanonical and canonical heat capacities are equal.

1. INTRODUCTION

The contemporary description of cluster evolution is based on saddle-crossing dynamics involving a large number of local minima for the potential energy of this system at zero temperature depending on the configuration of atoms [1–6]. Each local minimum corresponds to a locally stable configuration. A certain number of these lie at the point of lowest energy in a basin, while others, at higher energies in the same basin, correspond to configurational excitations from that lowest point. Neighboring local minima are separated by saddle points of the potential energy surface [7]. At low temperatures, most of the time the cluster is then found near the local energy minima and the durations of intermediate states during transitions across saddles between neighboring minima are brief. Taking the cluster aggregate states as a group of atomic configurations near local minima of the cluster potential energy [8], we obtain a precise picture of the cluster phase transitions, characterized by bands of dynamical coexistence of phases [9–12]. In the case of solid–liquid equilibrium, this means that the system is found in the solid aggregate state over certain periods; the remainder of the time, it is found in the liquid state, if the cluster can be located in two aggregate states. (It is also possible for a cluster to exhibit more than two phases in such a dynamic equilibrium, e.g., a "surface-melted" state together with a solid and a liquid state [13, 14].)

Using these concepts, one can generalize thermodynamics of bulk systems and relate these to clusters as systems of small, finite numbers of bound atoms or molecules. Within the framework of the saddle-crossing dynamics, one can define the cluster aggregate states as sets of atomic configurations near local minima of the cluster potential energy with nearby energies [8], and a cluster aggregate state can therefore include a finite number of elementary configurational excitations. In the classical thermodynamics of bulk systems, the aggregate state includes many configurational excitations, which leads to a uniform spatial distribution of atoms for a liquid. Next, the hierarchy of cluster times leads to the corresponding phenomenon of phase coexistence in clusters [9–12].

Thus, the properties and dynamics of cluster evolution allow one to apply thermodynamics to clusters. One must use a Gibbsian ensemble to describe the thermodynamics of clusters; some of the familiar concepts and characteristics of bulk systems disappear and some that are equivalent in common situations become inequivalent. For example, the distinction between phase and component is lost because phase equilibrium occurs on the same short time scale as the equilibrium among reacting components. Hence, the Gibbs phase rule loses its meaning for small systems [3]. Using this perspective, we here consider the heat capacity of a cluster that does not exchange energy with an environment; i.e., the cluster is in a microcanonical ensemble of atoms [15].

¶This article was submitted by the authors in English.

For clusters, the typical time for establishing vibrational thermal equilibrium is brief compared with that for establishing configurational equilibrium in the solid and liquid states [16]. Under microcanonical conditions, this time scale separation in clusters makes it possible to identify the temperatures of the solid and liquid states separately; the cluster is submitted to a two-temperature description. Likewise, under canonical conditions, the solid and liquid states can be assigned different mean energies and potential energies. If the time in which the cluster is observed is long compared to the typical time for dynamic equilibration between the aggregate states, it becomes appropriate to use a single cluster temperature, i.e., to model this cluster by one averaged aggregate state.

In an ensemble at a constant energy, the effective temperature of a cluster, solid or liquid, can be defined in either of two ways. One is the mean kinetic energy per degree of freedom; the other is the derivative of the internal energy with respect to the microcanonical entropy at constant volume. While these are equivalent for a canonical ensemble of macroscopic systems (with the conventional canonical entropy), they are not necessarily equivalent for microcanonical ensembles, particularly of small systems. Defined in terms of kinetic energy, the effective temperature of the solid is necessarily higher than that of a liquid at the same energy. Hence, increasing the energy in the zone of coexistence of the solid and the liquid (i.e., in the transition region) can lead to an effective temperature decrease. In the following, we consider this problem in terms of two aggregate states, with the additional simplifying assumptions that the separate caloric curves for the solid and liquid states are parallel straight lines, i.e., the transition energy is independent of the temperature, and that the difference of the solid and liquid temperatures is relatively small. These simplifications make it possible to understand the cluster properties near the melting point in a simple way. (We use the term "melting point" to mean the temperature at which the free energies of the solid and liquid clusters are equal. There is, of course, no sharp melting point for small clusters.)

Under these conditions, when equilibrium is established at each new cluster energy, each small increase of that energy near the melting point goes partly to excitation of thermal (vibrational) motion and partly to configuration excitation. Consequently, the heat capacity of an isolated cluster changes near the melting point. If the temperature is defined as the entropy derivative of the internal energy, then the heat capacity almost certainly remains positive and typically increases as more degrees of freedom absorb energy.[1] However, if the mean potential energy of the liquid form is significantly higher than that of the solid, and the temperature is defined in terms of mean kinetic energy per degree of freedom, then the system may exhibit a negative heat capacity and a region of a negative slope, an "S-bend," in its caloric curve. This behavior of cluster heating has been observed for clusters on the basis of theoretical [17–22] and experimental studies [23–28]. Below, we consider this problem in detail.

2. HIERARCHY OF CLUSTER TIMES

We first analyze the character of equilibrium in a cluster. We use the two-state approximation for cluster aggregate states [8], which extends the thermodynamic concept of the aggregate states from bulk to clusters, and we assume the existence—local stability and thermal equilibration—of two aggregate states, solid and liquid. Although clusters may exhibit several aggregate states in equilibrium, for example, associated with melting of different cluster shells [13, 14], the model that we use here involves the assumption that in a given range of parameters, the cluster can be found only in two aggregate states. The character of cluster equilibrium is determined by typical times for processes within the cluster. The typical time for establishing thermal equilibrium τ_{eq} between bound atoms is on the order of

$$\tau_{eq} \sim \frac{1}{\omega_D}, \tag{1}$$

where ω_D is the Debye frequency, roughly inversely proportional to the period of cluster oscillations ($\sim 10^{-14}$ s at room temperature). The typical dwell time of a cluster in the vicinity of the free energy minimum τ_{ag} associated with each aggregate state is long compared with τ_{eq},

$$\tau_{eq} \ll \tau_{aq}, \tag{2}$$

because transitions between aggregate states require that the cluster overcome a significant free energy barrier. We consider a cluster of bound atoms as a member of a microcanonical ensemble and neglect the interaction between the cluster and environment, i.e.,

$$\tau_{aq} \ll \tau_{\text{th}}, \tag{3}$$

where τ_{th} is the typical time for the exchange of energy between the cluster and its environment; for shorter times, the cluster can be considered as an isolated particle. We introduce a typical time τ of cluster observation such that

$$\tau_{aq} \ll \tau \ll \tau_{\text{th}}. \tag{4}$$

This hierarchy of cluster times leads to a particular pattern of cluster behavior. Indeed, during τ_{eq}, thermal equilibrium is established for the vibrational motion of the cluster atoms and the thermal motion of atoms can

[1] Only if the available phase space were to decrease with the energy for some pathological system could its caloric curve show a negative slope with this definition of temperature. Such a situation is logically possible, but physically almost unimaginable.

then be characterized by temperature [16]. Because of criterion (2), this temperature is different for the two aggregate states. We therefore introduce separate temperatures of atoms for the solid T_{sol} and liquid T_{liq} aggregate states. In particular, in the Dulong–Petit limit, the cluster energy is given by

$$E = (3n-6)T_{sol} = \Delta E + (3n-6)T_{liq}, \quad (5)$$

where n is the number of cluster atoms and ΔE is the fusion energy. This implies that

$$\Delta T = T_{sol} - T_{liq} = \frac{\Delta E}{3n-6}. \quad (6)$$

Along with these temperatures, one can introduce a general cluster temperature T for a large time on the order of τ, which can be expressed in terms of an average energy of an individual cluster atom if the average is taken for a time on the order of τ long enough for the cluster to change its aggregate state many times.

3. TWO-AGGREGATE APPROACH

Considering the approximation of two aggregate states [8], we express the total partition function of a cluster as

$$Z = Z_{sol} + Z_{liq}, \quad (7)$$

where Z_{sol} and Z_{liq}, and the partition functions for the solid and liquid cluster states, respectively, are related by

$$p(T) = \frac{Z_{liq}}{Z_{sol}}. \quad (8)$$

The respective probabilities w_{sol} and w_{liq} that the cluster is found in the solid and liquid states are

$$w_{sol} = \frac{1}{1+p}, \quad w_{liq} = \frac{p}{1+p}. \quad (9)$$

From the thermodynamic relation, we have

$$p = \exp\left[-\frac{\Delta E}{T} + S_{liq}(T_{liq}) - S_{sol}(T_{sol})\right] = \exp\left[-\frac{\Delta E}{T} + \Delta S\right], \quad (10)$$

where $S_{sol}(T)$ and $S_{liq}(T)$ are the entropies of the solid and liquid states at the given temperature, T is the effective temperature that characterizes the rates of transitions between the solid and liquid states, and ΔE and ΔS are the changes of the thermodynamic variables at the phase transition.

Although clusters exhibit bands of coexistence rather than the sharp melting points of bulk systems, we can, as mentioned above, define the melting point of a cluster by analogy with that of the bulk as the temperature of equal free energies of the two phases. In this way, the precise definition is that of the "equality" temperature T_{eq} such that

$$p(T_{eq}) = 1, \quad (11)$$

and hence,

$$w_{sol}(T_{eq}) = w_{liq}(T_{eq}) = 1/2.$$

As the general cluster temperature, the effective cluster temperature T tends to T_{sol} as $p \longrightarrow 0$ ($w_{sol} = 1$) and tends to T_{liq} in the limit as $p \longrightarrow \infty$, or when $w_{liq} = 1$. (It is sometimes convenient to use the quantity $(w_{liq} - w_{sol})/(w_{liq} + w_{sol})$ simply because it varies only between –1 and +1 [13, 14].)

4. ENTROPY OF AN ISOLATED CLUSTER IN THE TWO-STATE APPROACH

When a cluster does not interact with its environment, a thermodynamic equilibrium is established. In addition to the temperatures of the solid T_{sol} and liquid T_{liq} aggregate cluster states, this allows us to introduce the general cluster temperature from the thermodynamic relation

$$dE = TdS, \quad (12)$$

where E and S are the cluster energy and entropy. This definition can be used for the two aggregate states separately or for the long-time average over both aggregate states. Here, we use the latter option and evaluate the entropy S of a cluster in a long-term equilibrium (with or without the environment) between two aggregate states. Basing this analysis on a general entropy formula [29], we have

$$S = -\langle \ln w \rangle = -\sum_i w_i \ln w_i, \quad (13)$$

where i is a cluster state and w_i is the probability that the cluster is found in this state ($\sum_i w_i = 1$). Along with w_{sol} and w_{liq} in Eq. (9), we introduce the probability X_j for the cluster to be in the jth state if the cluster is first found in the solid aggregate state, and the probability Y_k for the cluster to be in kth state if it is initially in the liquid aggregate state. That is, we introduce a kind of conditional probability. According to the definition, we have

$$w_{sol} + w_{liq} = 1, \quad \sum_j X_j = \sum_k Y_k = 1. \quad (14)$$

From this, we obtain the cluster entropy

$$S = -w_{\rm sol}\sum_j X_j \ln(w_{\rm sol}X_j) - w_{\rm liq}\sum_k Y_k \ln(w_{\rm liq}Y_k) = w_{\rm sol}S_{\rm sol} + w_{\rm liq}S_{\rm liq} + S_{\rm conf}, \quad (15)$$

where

$$S_{\rm sol} = \sum_j X_j \ln X_j, \quad S_{\rm liq} = \sum_k Y_k \ln Y_k \quad (16)$$

are the entropies of the corresponding aggregate states. We thus express the entropy of a cluster with two aggregate states through entropies of each aggregate state and the entropy of the cluster configuration state $S_{\rm conf}$, equal to

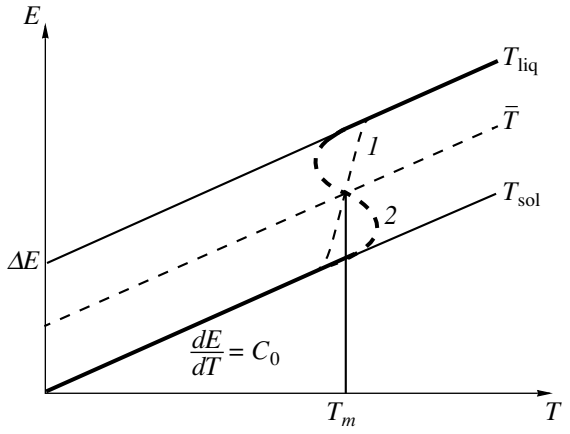
$$S_{\rm conf} = -\sum_i x_i \ln x_i = -w_{\rm sol}\ln w_{\rm sol} - w_{\rm liq}\ln w_{\rm liq} = \ln(1+p) - \frac{p}{1+p}\ln p, \quad (17a)$$

$$\frac{dS_{\rm conf}}{dp} = -\frac{\ln p}{(1+p)^2}, \quad (17b)$$

where x_i is the probability for the cluster to be in a given aggregate state, and we use Eqs. (9). We note that this expression is valid under the assumption that the cluster is observed in a long-term equilibrium; i.e., it can be located many times in each aggregate state during the observation time. Thus, expression (15) for the cluster entropy is a sum of terms corresponding to the solid and liquid aggregate states, and also of the term that accounts for configurational excitation.

5. TEMPERATURE OF A CLUSTER AS A MICROCANONICAL ENSEMBLE OF ATOMS

It follows from the above discussion that an isolated cluster with two aggregate states can be considered in the two-temperature approach if criterion (2) is satisfied; i.e., the typical time τ_{eq} for thermodynamic equilibration of the atomic thermal motion in each aggregate state is short compared with the dwell time τ_{ag} of the cluster in each aggregate state. When we observe a cluster over a time long compared with τ_{ag}, we can model the cluster with two aggregate states by a cluster with one aggregate state, and thus introduce a single average cluster temperature T. For this purpose, we can use the connection between the kinetic energy of cluster atoms and their temperature. Connecting the mean kinetic energy of cluster atoms averaged over a time long enough to reflect the kinetic energies of atoms in both aggregate states, we then define the (long-term) cluster temperature as



Caloric curves of an isolated cluster with two aggregate states in the one-temperature approach: (*1*) the case of a positive heat capacity; (*2*) the case of a negative heat capacity near the melting point.

$$T = w_{\rm sol}T_{\rm sol} + w_{\rm liq}T_{\rm liq}. \quad (18)$$

This is the definition of temperature one would use in the context of traditional statistical physics, i.e., on the basis of a very long time average. This remains a useful and valid approach, but the availability of measurements fast enough for one to observe the individual aggregate states justifies the extension of the conceptual framework to describe each aggregate state by itself, in order to supplement our long-time-average description.

Turning to the two-temperature approach for a cluster, we assume the cluster heat capacity to be independent of the temperature in the range of phase coexistence; in other words, the caloric curves for the solid and liquid states are straight lines, as shown in the figure. We use the parameters of this curve

$$\bar{T} = \frac{T_{\rm sol} + T_{\rm liq}}{2}, \quad \Delta T = T_{\rm sol} - T_{\rm liq}, \quad (19a)$$

and for simplicity assume that

$$\Delta T \ll \bar{T}. \quad (19b)$$

According to Eq. (18), the statistical temperature T is expressed via these parameters as

$$T = \bar{T} + \frac{\Delta T}{2}\frac{1-p}{1+p}. \quad (20)$$

Evidently, within the framework of the general temperature, it follows from (10) that

$$p = \exp\left[-\frac{\Delta E}{T} + \Delta S\right]. \quad (21)$$

For simplicity, we assume that the entropy jump ΔS at the phase transition and the transition energy ΔE are independent of the temperature. In addition, we have for the cluster energy E under our assumptions (see figure)

$$E = C_0\bar{T}, \quad \Delta E = C_0\Delta T, \tag{22}$$

where C_0 is the cluster heat capacity far from the melting point.

We now use Eq. (12) as the thermodynamic definition of the cluster temperature T,

$$\frac{1}{T} = \left[\frac{dS}{dE}\right]_V. \tag{23}$$

That is, we use the entropy–energy definition, rather than the mean kinetic energy definition for temperature. Because this formula is also valid for each aggregate state, we have the relations

$$\frac{1}{T_{\text{sol}}} = \frac{dS_{\text{sol}}}{dE}, \quad \frac{1}{T_{\text{liq}}} = \frac{dS_{\text{liq}}}{dE}$$

for the cluster temperature of a given aggregate state. On the basis of these formulas and formula (17) for the entropy of a cluster with two aggregate states, we then express the cluster temperature as

$$\begin{aligned}\frac{1}{T} &= \frac{w_{\text{sol}}(T_{\text{sol}})}{T_{\text{sol}}} + \frac{w_{\text{liq}}(T_{\text{liq}})}{T_{\text{liq}}} + \frac{dS_{\text{conf}}}{dE} \\ &= \frac{1}{(1+p)T_{\text{sol}}} + \frac{p}{(1+p)T_{\text{liq}}} - \frac{\ln p}{(1+p)^2}\frac{dp}{dE},\end{aligned} \tag{24}$$

where

$$\frac{dS_{\text{conf}}}{dE} = -\frac{\ln p}{(1+p)^2}\frac{dp}{dE} \tag{25}$$

in accordance with Eqs. (17). It can be seen that the statistical and thermodynamic temperature definitions (20) and (23) are different. But the difference between them is small in the present case because we have imposed criterion (19b). In particular, at the equality point ($p = 1$), we have $T_{eq} = \bar{T}$ in accordance with (20), and Eq. (24) then gives

$$T_{eq} = \frac{2T_{\text{sol}}T_{\text{liq}}}{T_{\text{sol}} + T_{\text{liq}}} = \bar{T}\left(1 - \frac{\Delta T^2}{4\bar{T}^2}\right). \tag{26}$$

Thus, although the definitions of the cluster temperature are different, under assumption (19b) this difference is only of the second order in terms of the expansion in the small parameter $\Delta T/\bar{T}$.

6. HEAT CAPACITY OF A CLUSTER AS A MICROCANONICAL ENSEMBLE OF ATOMS

We now construct caloric curves for a large cluster, supposing that the caloric curves for the solid and liquid aggregate states are parallel straight lines and the distance between these lines satisfies criterion (19b). The cluster state corresponds to the solid caloric curve at low temperatures below the equality point, and to the liquid caloric curve at high temperatures above that temperature, T_{eq}. The intermediate part of the caloric curve near the melting point can have two forms, as shown in the figure; in case 1, the cluster heat capacity is positive at any temperature, and in case 2, it is negative near T_{eq}. In principle, both cases are possible. Based on their experimental study of sodium clusters of hundreds of atoms, Haberland *et al.* [30, 31] infer that the case of a negative cluster heat capacity near T_{eq} is more representative. Initially, the accuracy of the experimental data [23–26] left some possibility to question that inference, but more recent, independent measurements have made the instance of certain microcanonical negative heat capacities much more plausible [27, 28]. All these experiments, in effect, base the evaluation of temperature on the kinetic energy of the atoms of the clusters, consistent with this definition being the one that allows negative heat capacities. Below, we analyze this problem using the above approach. Within the framework of the statistical and thermodynamic considerations, we introduce the general temperature for an isolated cluster that can be found in both the solid and liquid states.

We evaluate the heat capacity C of a cluster of bound atoms as a member of a microcanonical ensemble near the melting or equality point. When we introduce one cluster temperature T, its heat capacities are

$$C = \frac{dE}{dT}, \quad C_0 = \frac{dE}{d\bar{T}} = \frac{dE}{dT_{\text{sol}}} = \frac{dE}{dT_{\text{liq}}},$$

where E is the internal cluster energy and $\bar{T}$ is given by Eq. (19a). We assume in the discussion here that T_{sol} and T_{liq} are almost equal; hence, we are not dealing with the general case. Relation (21) is valid under our assumption and gives

$$\frac{dp}{dE} = \frac{p\Delta E}{T^2 C}. \tag{27}$$

For simplicity, we assume here that the cluster parameters ΔE and ΔS are independent of the cluster temperature T. Under these conditions, for the statistical definition of the temperature, we have, taking the differential of (20) and using Eq. (22),

$$\frac{1}{C} = \frac{1}{C_0} - \frac{\Delta T}{(1+p)}\frac{p\Delta E}{T^2 C}.$$

This implies that

$$C = C_0(1 + Z), \quad Z = \frac{p}{(1 + p)^2}\frac{\Delta T \Delta E}{T^2}. \qquad (28)$$

We now consider the case of the thermodynamic definition of the cluster temperature when a cluster with two aggregate states is modeled by a cluster with one average aggregate state. Taking the differential of (24) and expanding it in a small parameter $\Delta T/\bar{T}$, we then have

$$-\frac{dT}{T^2} = -\frac{d\bar{T}}{\bar{T}^2} + \frac{\Delta T d\bar{T}}{\bar{T}^3}\frac{1-p}{1+p} + \frac{\Delta T dp}{\bar{T}^2(1+p)^2} + d\left(\frac{\ln p}{(1+p)^2}\frac{dp}{dE}\right).$$

Ignoring the second term in the right-hand side of this equation in comparison with the first one because of (19b), we then obtain

$$-\frac{1}{CT^2} = -\frac{1}{C_0\bar{T}^2} + \frac{\Delta T}{\bar{T}^3(1+p)^2}\frac{dp}{dE} + \frac{d}{dE}\left(\frac{\ln p}{(1+p)^2}\frac{dp}{dE}\right). \qquad (29)$$

Because the maximum heat capacity of the cluster corresponds to the equality or melting point (if its value is positive), we consider Eq. (29) at T_{eq}, where $p = 1$, $T = \bar{T} = T_{eq}$, and therefore,

$$\frac{1}{C} = \frac{1}{C_0} - \frac{Z}{C} - \frac{ZC_0}{C^2}. \qquad (30)$$

We note that Eq. (28) follows from this if we ignore the last term, i.e., if we neglect the configurational part S_{conf} in expression (15) for the cluster entropy. The physical solution of Eq. (30) is given by

$$C = C_0\left[\frac{1+Z}{2} + \sqrt{\left(\frac{1+Z}{2}\right)^2 + Z}\right]. \qquad (31)$$

Formulas (28) and (31) for the cluster heat capacity, based on the two different definitions of the cluster temperature, lead to identical results in the limiting cases $Z = 0$ and $Z = \infty$. The maximum ratio of the heat capacities according to formulas (31) and (28) corresponds to $Z = 1$ and is $(1 + \sqrt{2})/2 \approx 1.2$. Thus, if an isolated cluster with two aggregate states is modeled by a cluster with one temperature, the values of its heat capacity depend on the definition of the cluster temperature. However, for the statistical and thermodynamic definitions of the cluster temperatures, the values of the cluster heat capacity coincide within the limits of 20% in the cases considered here.

We now analyze the character of the consumption of energy that is transferred to an isolated cluster very slowly, such that equilibrium is established for each input of energy. We divide the total energy of the ensemble of bound atoms into the kinetic energy of atoms, the potential energy of the interaction between atoms, and the energy of configurational excitation. For simplicity, we take the ratio between the kinetic and potential energy to be governed by the virial relation; this ratio is therefore independent of excitation. Moreover, these excitations involve sufficiently low energies such that they can be supposed to be harmonic oscillations. Therefore, the excitation energy can be thought to be consumed in two channels, thermal motion of atoms and configurational excitation. It is clear that the cluster total heat capacity must be greater than if configuration excitation were absent, because only a part of the input energy is consumed by the thermal motion of atoms. A significant part must go toward increasing excitation of "new" degrees of freedom that are unexcited at lower temperatures but fully excited at higher temperatures. Hence, in the range of the phase change, the heat capacity exceeds that in the temperature ranges far from the phase change.

The configurational and vibrational contributions may nonetheless be strongly linked in the following way. If the configurational excitations bring the cluster to a region of high potential energy and, thus, to a region of low kinetic energy, then, during those intervals in which the configurational excitation is high, the kinetic temperature is necessarily low. This situation does not conform to our above assumption that T_{sol} and T_{liq} are almost equal. In accordance with this assumption, configurational excitation requires relatively little energy. Consequently, we can suppose that clusters of sizes far from the "magic number" or closed-shell sizes are likely to conform to the assumption, but that the "magic number" or closed-shell clusters are least likely to satisfy it.

If the configurational excitation energy requires a significant part of the cluster internal energy, the heat capacity may become negative because an increase in the total cluster energy leads to a decrease in its thermal (or vibrational) energy. Just this situation is proposed by Haberland [30, 31] for sodium clusters consisting of a hundred or more atoms. Evidently, the conditions favorable for clusters to have a negative heat capacity apply to such systems. Based as it is on Eqs. (28) and (31), the analysis here, with its strong assumptions, leads to a positive cluster heat capacity at any cluster temperature. We can therefore interpret the assumptions leading to this conclusion as sufficient conditions for a positive heat capacity. This should not be interpreted as implying that negative heat capacities cannot occur in microcanonical systems; they certainly can

The parameters of the isolated Lennard–Jones cluster consisting of 13 atoms at the melting point. The data are obtained on the basis of [10]

Parameter	Value
E_m	13.6
ΔE	2.46
$\eta(E_m)$	0.39
T_{sol}	0.32
T_{liq}	0.26
T_m	0.29
$Z(T_m)$	0.46
$\frac{C_{max}}{C_0}$, (28)	1.46
$\frac{C_{max}}{C_0}$, (31)	1.73

occur if the kinetic definition of temperature is used. In fact, we can now say that the next challenge in this field is finding sufficient conditions for a negative heat capacity, in terms of the relative energies and phase space volumes of the solid and liquid phases, and then finding what classes of systems best satisfy those conditions—or those for strictly positive heat capacities.

Equations (28) and (31) characterize the increase in the cluster heat capacity near the melting point. We now consider this increase for a large cluster with $Z \gg 1$. Because for a large cluster $C_0 \sim n$, where n is the number of atoms comprising the cluster, criterion (19b) becomes

$$\left(\frac{T_m}{\Delta T}\right)^2 \ll n, \tag{32}$$

such that the increase in the heat capacity for such large clusters is strong. In this limit, Eqs. (28) and (31) give the maximum heat capacity that corresponds to the equality or melting point

$$C_{max} = C_0 Z = \frac{\Delta E^2}{4T_{eq}^2}. \tag{33}$$

This expression also involves the assumption that the transition thermodynamic parameters ΔE and ΔS are independent of the temperature. The maximum heat capacity has the same value in the case of isothermal heating [8].

We now apply Eq. (30) to the Lennard–Jones cluster of 13 atoms, taking its parameters from the computer modeling [10] of this cluster. Assuming the Dulong–Petit law to be valid at melting, we have the heat capacity of this cluster $C_0 = 3n - 6 = 33$. The temperatures of the solid T_{sol} and liquid T_{liq} states are

$$T_{sol} = \frac{2\eta E}{C_0}, \quad T_{liq} = \frac{2\eta(E - \Delta E)}{C_0}, \tag{34}$$

where η is the part of the cluster excitation energy E that is transformed into kinetic energy of the atoms. Thus, we express the temperature of cluster atoms through the total kinetic energy. Parameters in the table refer to the equality or melting point ($p(E_m) = 1$, $T = T_{eq}$), and we use reduced energy units with the energy unit given by the binding energy per bond.

The data in the table, based on cluster computer simulation [10], show the validity of criterion (19b); the small parameter $\Delta T/T_{eq}$ used above is equal to 0.2 for the Lennard–Jones cluster of 13 atoms. This small parameter determines the accuracy of using one cluster temperature. We note that the increase of the heat capacity near the melting point is not strong for this cluster, and the heat capacity is positive at any temperature. This does not agree with the supposition of a negative heat capacity of clusters near T_{eq} [30, 31] of the sort inferred from experiments with sodium clusters [23–26]. A more exacting test with Lennard–Jones clusters would require examining the behavior of a larger system, e.g., a closed-shell icosahedral structure of 55 or 137 atoms.

7. CONCLUSIONS

It follows from the above analysis that the behavior of a real isolated cluster with two aggregate states can be described using either one aggregate state with a general temperature or two aggregate states with two temperatures. The latter requires that the vibrational modes of the cluster equilibrate rapidly compared with the rate of passage between aggregate states, such that temperatures of those states can be well defined. In reality, such thermodynamic equilibrium is established for many kinds of clusters and other small systems during the time the system resides in each aggregate state. The single-temperature description requires that, during an observation time, a cluster changes its aggregate state many times. Although the statistical and thermodynamic definitions of the temperature lead to different temperatures and heat capacities, the differences between these values are not large under the assumptions used in this work. In addition, under these assumptions, the maximum heat capacities of a large cluster at the melting point coincide for the adiabatic and isothermal regimes of energy input, and in the case of the adiabatic regime (or for an isolated cluster), the heat capacities of a large cluster coincide for the statistical and thermodynamic definitions of temperature. A system satisfying the conditions invoked here does not exhibit a negative heat capacity of the type reported for an isolated cluster [30, 31]. The analysis provides con-

ditions sufficient for the heat capacity of an isolated cluster in a microcanonical ensemble with two aggregate states to be positive at any size and temperature. The question is now open to find comparable general conditions that produce the negative heat capacities that have been seen in experiment and theory.

ACKNOWLEDGMENTS

B.M.S. thanks the Russian Foundation for Basic Research for partial support (project no. 03-02-16059). R.S.B. acknowledges the support of a grant from the National Science Foundation.

REFERENCES

1. M. R. Hoare and P. Pal, Adv. Phys. **20**, 161 (1971).
2. F. H. Stillinger and T. A. Weber, Phys. Rev. A **25**, 978 (1982).
3. R. S. Berry, in *Theory of Atomic and Molecular Clusters*, Ed. by J. Jellinek (Springer, Berlin, 1999), p. 1.
4. K. D. Ball and R. S. Berry, J. Chem. Phys. **111**, 2060 (1999).
5. T. Komatsuzaki and R. S. Berry, J. Chem. Phys. **110**, 9160 (1999).
6. D. J. Wales, J. P. K. Doye, M. A. Miller, *et al.*, Adv. Chem. Phys. **115**, 1 (2000).
7. K. D. Ball, R. S. Berry, R. E. Kunz, *et al.*, Science **271**, 963 (1996).
8. R. S. Berry and B. M. Smirnov, J. Chem. Phys. **114**, 6816 (2001).
9. R. S. Berry, J. Jellinek, and G. Natanson, Phys. Rev. A **30**, 919 (1984).
10. J. Jellinek, T. L. Beck, and R. S. Berry, J. Chem. Phys. **84**, 2783 (1986).
11. R. S. Berry, T. L. Beck, H. L. Davis, and J. Jellinek, Adv. Chem. Phys. **90**, 75 (1988).
12. D. J. Wales and R. S. Berry, J. Chem. Phys. **92**, 4283 (1990).
13. R. E. Kunz and R. S. Berry, Phys. Rev. Lett. **71**, 3987 (1993).
14. R. E. Kunz and R. S. Berry, Phys. Rev. E **49**, 1895 (1994).
15. D. ter Haar, *Elements of Thermostatics* (Addison-Wesley, New York, 1966).
16. B. Vekhter, K. D. Ball, J. Rose, and R. S. Berry, J. Chem. Phys. **106**, 4644 (1997).
17. M. Bixon and J. Jortner, J. Chem. Phys. **91**, 1631 (1989).
18. P. Labastie and R. L. Whetten, Phys. Rev. Lett. **65**, 1567 (1990).
19. D. J. Wales, Mol. Phys. **78**, 151 (1993).
20. D. J. Wales and R. S. Berry, Phys. Rev. Lett. **73**, 2875 (1994).
21. I. H. Umirzakov, Phys. Rev. E **60**, 7550 (1999).
22. O. Mülken, H. Stamerjohanns, and P. Borrmann, Phys. Rev. E **64**, 047105-1 (2001).
23. M. Schmidt, R. Kusche, W. Krommüller, *et al.*, Phys. Rev. Lett. **79**, 99 (1997).
24. M. Schmidt, R. Kusche, B. von Issendorf, *et al.*, Nature **393**, 238 (1998).
25. M. Schmidt, R. Kusche, T. Hippler, *et al.*, Phys. Rev. Lett. **86**, 1191 (2001).
26. M. Schmidt, T. Hippler, J. Donges, *et al.*, Phys. Rev. Lett. **87**, 203402 (2001).
27. E. Gobet, B. Farizon, M. Farizon, *et al.*, Phys. Rev. Lett. **89**, 183403 (2002).
28. J. A. Reyes-Nava, I. L. Garzón, and K. Michaelian, Phys. Rev. B **67**, 165401 (2003).
29. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 5: *Statistical Physics*, 3rd ed. (Nauka, Moscow, 1976; Pergamon Press, Oxford, 1980), Part 1.
30. H. Haberland, in *Metal Clusters*, Ed. by W. Ekardt (Wiley, New York, 1999).
31. H. Haberland, in *Proceedings of Les Houches Summer 2000 School on Atomic Clusters and Nanoparticles* (2000).

Journal of Experimental and Theoretical Physics, Vol. 98, No. 2, 2004, pp. 374–389.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 2, 2004, pp. 423–440.

MISCELLANEOUS

On the Transmission Rate of Classical Information through Quantum Communication Channels

S. N. Molotkov[a,b]

[a]*Institute of Solid State Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia*

[b]*Department of Computational Mathematics and Cybernetics, Moscow State University, Vorob'evy gory, Moscow, 119991 Russia*

e-mail: molotkov@issp.ac.ru

Received March 1, 2003

Abstract—The coding of quantum communication channels in real time is considered as applied to the situation when information is coded into continuous quantum degrees of freedom (into the shape of the amplitude of quantum states with an arbitrary number of photons). It is shown that the nonlocalizability of states in quantum field theory requires that the identity of particles should be taken into account. This, together with the finiteness of the limit speed of propagation, leads to the fact that the formulas for the transmission rate of nonrelativistic communication channels have an asymptotic character; i.e., these formulas are formally valid only when the separation between messages is infinite (when the identity of particles can be neglected) and, hence, when the transmission rate in [bit/message s] is infinitely small. A real-time information capacity of a sequential relativistic quantum communication channel is obtained that takes into account the identity of particles for pure signal states with an arbitrary number of photons. An explicit analytic expression is obtained for the transmission rate of a quantum channel of finite bandwidth for one-photon input states. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The transmission capacity is an important characteristic of a communication channel that determines the speed of information transmission through the channel. This parameter defines an asymptotic bound for long messages up to which information can be transmitted without error (more precisely, with arbitrarily small error probability). Classical information theory and the coding theorems that determine this asymptotic bound for communication channels are formulated in terms of statistical ensembles. A source of discrete messages is described by a finite or infinite set of signals (an alphabet) $x = \{x_i\}$ each of which is sent by the source into a communication channel with a priori probabilities $\{p\{x_i\}\}$. A discrete memoryless communication channel is defined by an output alphabet $y = \{y_i\}$, which may partially or completely coincide with the input alphabet, and by the transition probabilities $\{p(y_j|x_i)\}$ that actually describe the transmission characteristics of the communication channel.

The mutual information of a source is described by the Shannon entropy [1], which is given by

$$H(x) = -\sum_i p(x_i)\log p(x_i) \qquad (1)$$

for an ensemble x; here, the logarithm is to the base 2. The quantity $H(x)$ is equal to the amount of information in bits per symbol. More precisely, for a sufficiently long sequence of length n ($n \longrightarrow \infty$), information in bits per symbol of a sequence is equal to $H(x)$ with a probability of one.

The model of statistical ensembles corresponds to a real-life situation. For example, the frequency of occurrence of letters x_i (symbols of alphabet) in a sufficiently long fragment of a standard English text tends to a priori probabilities of $p(x_i)$. The amount of information in bits per symbol in a text is equal to 1.3 bit per byte (eight bits), $H(x) = 1.3/8$. Accordingly, the redundancy of the text is $1–1.3/8 \approx 84\%$. This means that a sufficiently long fragment of the text can be compressed to 84% without loss of useful information.

The amount of information per symbol that can be transmitted through a communication channel is given by mutual information,

$$I(x; y) = H(x) - H(x|y),$$

$$H(x|y) = -\sum_{i,j} p(y_j)\frac{p(y_j|x_i)p(x_i)}{p(y_j)}\log\frac{p(y_j|x_i)p(x_i)}{p(y_j)}, \qquad (2)$$

where $H(x|y)$ is the conditional entropy of the input with respect to the output, which describes the loss of information in the communication channel, for example, due to noise. Accordingly, information that can be

transmitted by a sufficiently long sequence tends to $I_n(x; y) = nI(x; y)$.

The maximum (more precisely, the smallest upper bound) of mutual information over all possible a priori input probabilities $p(x_i)$,

$$C = \max_{p(x_i)} I(x; y) = \lim_{n \to \infty} \max_{p(x_i)} \frac{I_n(x; y)}{n}, \qquad (3)$$

represents the transmission capacity of the communication channel and determines the largest possible amount of information per symbol that can be transmitted through a communication channel without error (with an arbitrarily small error probability) [1].

Notice that the transmission capacity C determines the speed of information transmission in terms of the number of bits (≤ 1) per message (a symbol of the alphabet) [bit/message] rather than in terms of the real-time transmission rate [bit/message s], which is of great practical interest.

The arguments given above are purely mathematical and do not apply to physical reality; the information carriers themselves, which are assigned certain symbols of the alphabet, are also not specified, although information carriers in any real-life situation are represented by specific physical objects (either classical or quantum). In nonrelativistic classical physics, when one interprets the transmission capacity of discrete channels as the rate of information transmission, one implies the following. Since no fundamental constraints are imposed on the preparation time of a (localized or spatially extended) classical object, the transmission rate is limited only by the frequency of messages of information carriers sent into a communication channel, which, in essence, may be arbitrarily large. If the frequency of messages sent into a communication channel is $1/T$, then the information transmission rate is

$$\frac{C}{T} \left[\frac{\text{bit}}{\text{message s}}\right]. \qquad (4)$$

All of the aforesaid applies to classical discrete communication channels. In general, a classical signal is described by a real function of coordinates and time. The possibility of transmitting information by time-continuous signals of finite bandwidth is given by the well-known Whittaker–Shannon–Kotelnikov sampling theorem [2]. This theorem states that, for a signal $f(t)$ with finite bandwidth ($\text{supp}\, f(\omega) \in \Delta\omega$), the number of independent degrees of freedom (symbols of a continuous alphabet) on a finite time interval T is equal to $2\Delta\omega T$. If there is no noise in the channel and no constraints are imposed on the amplitudes of the function at the sampling points, then, formally, the transmission capacity tends to infinity because each of the $2\Delta\omega T$ degrees of freedom may take arbitrary values.

The transmission capacity of a classical channel with additive Gaussian noise, finite bandwidth, and Gaussian distribution of a continuous input signal of limited power is given by the well-known Wiener–Shannon formula [1].

The description of classical communication channels where signal depends only on time is an idealization. If we assume that information is transmitted through space by electromagnetic waves (it is this situation that we consider here), which are described by the Maxwell equations, then we cannot consider signals that are functions of coordinates only. Therefore, the approaches described above are idealistic and do not fully apply to the situation. Moreover, at low signal levels, one inevitably has to take into account quantum phenomena. For example, in quantum cryptographic systems (where a quantum communication channel is given, as a rule, by an optical-fiber communication line), when a signal level is close to the one-photon level, the situation requires a relativistic quantum description because photons, being massless particles, are essentially relativistic objects. Determination of the fundamental restrictions imposed by the quantum character of a signal and by the special theory of relativity on the transmission rate of a signal has already attracted practical interest. Since all real prototypes of quantum cryptosystems employ an optical fiber, which has a finite transmission bandwidth due to the properties of the material, the derivation of formulas for the transmission capacity, which describe the rate of information transmission, is also of practical interest.

2. NONRELATIVISTIC QUANTUM COMMUNICATION CHANNELS WITH DISTINGUISHABLE STATES

Up to now, we have been dealing with classical communication channels. As the signal level decreases to the level of individual photons, one inevitably encounters fundamental constraints imposed on the transmission capacity and the transmission rate of classical information by means of quantum states. This question is of practical interest in relation to the problems of quantum cryptography.

Profound and good results have recently been obtained concerning the coding in nonrelativistic quantum communication channels [3–6]. Significant progress has been made in the understanding of the transmission of quantum states proper, as well as the transmission of classical information by means of distributed quantum communication channels (superdense coding; see, for example, [7]). In spite of the progress made, a clear answer has not been obtained to the question about the real-time transmission rate of classical information by means of quantum states. This is primarily associated with the fact that the above-mentioned coding theorems are formulated as communication protocols in the space of states of a quantum system (in a Hilbert space), without explicitly introducing spacetime into the problem and, hence, without explicitly taking into account the spacetime structure of the quan-

tum states. Without taking into account these circumstances, one cannot obtain the constraints imposed both by the quantum character of the states and by the presence of the limit speed of propagation.

By analogy with the classical case, the coding theorems that determine the channel capacities when transmitting classical information by means of quantum states are formulated in terms of quantum statistical ensembles [3–7]. A classical alphabet $\{x_i\}$ is assigned a quantum alphabet that is described by density matrices $\{\rho_i\}$ prepared by a source with a priori probabilities $\{p_i\}$. The state space of an individual quantum carrier is a Hilbert space $\mathcal{H}$ of states in which the operator of the density matrix acts. A message of length n is described by the tensor product of separate symbols of the quantum alphabet (density matrices) in separate messages:

$$p_{\mathbf{i}}^{(n)} = p_{i_1} \otimes \rho_{i_2} \otimes \ldots \otimes \rho_{i_n}, \quad \mathbf{i} = (i_1, i_2, \ldots, i_n). \quad (5)$$

The state space of messages of length n is the tensor product

$$\mathcal{H}^{\otimes n} = \mathcal{H} \otimes \mathcal{H} \otimes \ldots \otimes \mathcal{H}. \quad (6)$$

The quantum communication channel itself is described by the transformation (map) of the input operators of density matrices into the output operators of density matrices. According to the Kraus representation theorem [8], the transformations of the density matrices ρ into the density matrices ρ', which are admissible by quantum-mechanical laws, are given by the map

$$\rho' = \mathcal{T}[\rho], \quad (7)$$

which is a linear, trace-preserving, and completely positive map. Any such transformation is given by the representation

$$\mathcal{T}[\ldots] = \sum_k V_k[\ldots]V_k^+, \quad (8)$$

where the operators V_k form a partition of unity in $\mathcal{H}$:

$$\sum_k V_k^+ V_k = I. \quad (9)$$

Such a mapping of operators into operators is called either a tool (according to the terminology of [8]) or a superoperator. By specifying a superoperator, one completely describes a quantum communication channel, as in the case when the description of a channel in the classical case is specified by transition probabilities.

Accordingly, a particular message of length n (5) at the output of a channel is described as follows:

$$\mathcal{T}[\rho^{(n)}] = \mathcal{T}[\rho_{i_1}] \otimes \mathcal{T}[\rho_{i_2}] \otimes \ldots \otimes \mathcal{T}[\rho_{i_n}]. \quad (10)$$

The extraction of classical information from quantum states at the receiving end is performed by measurements performed over the output density matrix $\mathcal{T}[\rho^{(n)}]$. Any measurement is described by a certain partition of unity in $\mathcal{H}^{\otimes n}$:

$$I^{(n)} = I^{\otimes n} = \sum_{\mathbf{k}} \mathcal{M}_{\mathbf{k}}^{(n)}, \quad (11)$$

where the subscript $\mathbf{k}$ enumerates the results of measurements. Note that we do not assume that the measurements are performed immediately over a message of length n as a whole (the so-called collective measurements) rather than over a state in each message. It is these measurements on which the mutual information between input and output attains its maximum. Note that there is nothing of the kind in classical discrete memoryless communication channels (it suffices to perform a measurement over each individual message; see [5–7]).

The amount of mutual information between output and input for an ensemble of length n is given by

$$I_n(p, \mathcal{M}) = \sum_{\mathbf{j}} \sum_{\mathbf{i}} P_{\mathbf{i}} P(\mathbf{j}|\mathbf{i}) \log \frac{P(\mathbf{j}|\mathbf{i})}{\sum_{\mathbf{k}} P_{\mathbf{k}} P(\mathbf{j}|\mathbf{k})}, \quad (12)$$

where

$$P(\mathbf{j}|\mathbf{i}) = \mathrm{Tr}\{\rho_{\mathbf{i}}^{(n)'} \mathcal{M}_{\mathbf{j}}^{(n)}\} \quad (13)$$

is the probability that a particular message $\mathbf{i}$ of length n at the transmitting end is interpreted as a message $\mathbf{j}$ at the receiving end. The amount of mutual information between the input and output of a communication channel depends on which measurements are used for extracting classical information from a quantum ensemble.

The transmission capacity (the number of bits) per message is defined by

$$C = \lim_{n \to \infty} \frac{C_n}{n} = \lim_{n \to \infty} \max_{\{p_i, \mathcal{M}^{(n)}\}} \left(\frac{1}{n} I_n(p, \mathcal{M}) \right), \quad (14)$$

which is the maximum of mutual information over all possible a priori input probabilities and measurements at the receiving end (the states ρ_i of the carriers themselves are assumed to be fixed). The quantity C represents the amount of classical information in bits (≤ 1) per message (per symbol of the quantum alphabet ρ_i) that can be transmitted through a communication chan-

nel in the limit of long sequences with arbitrarily small error probability.

It should be noted that the expression of messages (5) and (10) in terms of a tensor product of density matrices automatically presumes that quantum states in each message are sent into a communication channel independently of each other. However, the maximum of mutual information is attained for collective measurements at the receiving end when one measures the state of individual messages as a single composite quantum object rather than the states in each separate message. At this point, we have a fundamental difference between quantum and classical channels.

The fact that the source sends quantum states that are described by the density matrices ρ_i and are chosen by the source according to a priori probabilities p_i implies that the state of the source is described by the density matrix

$$\rho = \sum_i p_i \rho_i. \quad (15)$$

Accordingly, the state at the output of the channel results from the action of the superoperator on the density matrix ρ (15) of the source. Then, we obtain

$$\rho' = \mathcal{T}[\rho] = \sum_i p_i \rho_i' = \sum_i p_i \mathcal{T}[\rho_i]. \quad (16)$$

The amount of classical information (in bits) that can be extracted from a quantum ensemble at the output of the channel as a result of measurements is essentially bounded by a quantity given by the fundamental inequality proved by Holevo [3]:

$$\chi \le H\left(\sum_i p_i \rho_i'\right) - \sum_i p_i H(\rho_i'), \quad (17)$$

where the equality is attained if and only if the density matrices in the ensemble commute, $[\rho_i', \rho_i'] = 0$ $(i \ne j)$. Here, $H(\rho)$ is the von Neumann entropy, which is a quantum analog of the Shannon entropy:

$$H(\rho) = -\mathrm{Tr}\{\rho \log \rho\}. \quad (18)$$

If the density matrices ρ_i' correspond to pure states, $\rho_i' = |\varphi_i\rangle\langle\varphi_i|$, then the condition that the density matrices should commute implies that the states are orthogonal. Orthogonal quantum states are certainly distinguishable and, in this sense, are analogous to classical states. In this case, the von Neumann entropy reduces to the Shannon entropy:

$$\chi \le H(p) = -\sum_i p_i \log p_i. \quad (19)$$

A remarkable result is that the Holevo fundamental limit is attainable and represents the transmission capacity of a quantum communication channel [3, 5–7],

$$C = \max_{\{p_i\}} \left\{ H\left(\sum_i p_i \mathcal{T}[\rho_i]\right) - \sum_i p_i H(\mathcal{T}[\rho_i]) \right\}. \quad (20)$$

Up to now, we have not specified the type of the object that was assigned a quantum state ρ_i. Moreover, the problem involves neither time nor space. Therefore, we can say nothing about the information transmission rate in real time in terms of [bit/message s]. In nonrelativistic quantum mechanics, such an approach is not self-contradictory in principle. Although the problem does not involve time, it is assumed (although is not explicitly stipulated) that, because there are no fundamental constraints on the preparation time of the quantum states (even spatially extended ones) and, hence, on the measurement time of these states, the transmission rate is bounded only by the sending frequency of states into the communication channel. The latter frequency may be arbitrarily large.

3. RELATIVISTIC QUANTUM COMMUNICATION CHANNELS WITH IDENTICAL PARTICLES

A quite different situation occurs in the relativistic case. Since photons are the only acceptable information carriers for transmitting information to large distances and, because they have zero mass, are essentially relativistic particles, one cannot determine the transmission rate of classical information by means of quantum states without explicitly introducing Minkowski spacetime.

Any real transmission of information occurs in time and space. Moreover, the very existence of spacetime suggests that there exist only certain types of elementary quantum systems (particles). This means that the basis vectors of irreducible unitary representations of the Poincaré group that act in a Hilbert space of states are associated with various types of particles (electrons, positrons, photons, neutrinos, etc.). The basis vectors corresponding to different irreducible representations (different types of particles) have different transformation properties under the transformations of Minkowski spacetime [9–11].

Further, we will consider a one-dimensional situation. This approximation is physically justified because optical-fiber systems, which play the role of quantum communication channels, are quasi-one-dimensional systems. A quantized photon field has a transverse character; for further consideration, the fact that the photon field has zero mass is significant. We will ignore the polarization degrees of freedom since we will focus on the coding into the shape of the quantized state of the photon field. Earlier [12], we considered a quantum binary communication channel when the coding was

performed into polarization degrees of freedom of a one-photon state. The coding into continuous degrees of freedom (not necessarily of a one-photon state) that is considered below is more complicated.

The symbols of the classical (finite or infinite) alphabet $\{x_i\}$ are assigned pure states of a quantized field. A symbol of the classical alphabet is assigned the amplitude φ_i (a shape, a smoothing function) of an n_i-photon state $|\varphi_i\rangle$. The state of a quantized field is represented as (the number of photons n_i in the state is determined by the number of birth operators $\varphi^+(\hat{x}_k)$)

$$|\varphi_i\rangle = \int\ldots\int d\hat{x}_1\ldots d\hat{x}_{n_i} \times \varphi_i(\hat{x}_1, \ldots, \hat{x}_{n_i})\varphi^+(\hat{x}_1)\ldots\varphi^+(\hat{x}_{n_i})|0\rangle, \tag{21}$$

where the field operators (more precisely, operator-valued generalized functions) for a massless field are given by

$$\varphi^+(\hat{x}) = \int d\hat{k}e^{i\hat{k}\hat{x}}a^+(\hat{k})\theta(k_0)\delta(\hat{k}^2), \quad \hat{x} = (x, t), \quad \hat{k} = (k, k_0). \tag{22}$$

The creation and annihilation operators satisfy the Bose commutation relations

$$[a^-(\hat{k}), a^+(\hat{k}')] = k_0\delta(k - k'). \tag{23}$$

The operator-valued generalized functions $\varphi^+(\hat{x})$ cannot be regarded as mere operators acting in $\mathcal{H}$, even if they are unbounded operators. If we assume that $\varphi^+(\hat{x})$ are mere operators, then, according to Jaffe [13], the requirement that a scalar product in $\mathcal{H}$ should be Lorentz invariant implies that the matrix element $\langle 0|\varphi^-(\hat{x}')\varphi^+(\hat{x})|0\rangle$, interpreted as the creation amplitude of a particle at $\hat{x} = (x, t)$, its propagation in spacetime, and annihilation at $\hat{x}' = (x', t')$, $t' > t$, is equal to a constant independent of $\hat{x}$ and $\hat{x}'$, which contradicts the relativistic causality.

The generalized basis vectors (more precisely, continuous linear functionals in $\mathcal{H}$) are given by

$$a^+(\hat{k})|0\rangle = |k\rangle, \quad \langle k|k'\rangle = k_0\delta(k - k'), \quad \varphi^+(\hat{x})|0\rangle = |\hat{x}\rangle, \tag{24}$$

where $|\hat{x}\rangle$ and $|k\rangle$ belong to Ω^* (to the conjugate space of the space of test functions $\varphi \in \Omega$). Physical states (normalized vectors in $\mathcal{H}$) are obtained from Ω—a space of infinitely differentiable functions that decrease at infinity faster than an inverse power of any polynomial—by smoothing the basis vectors with the test functions (amplitudes). For an n_i-photon state corresponding to symbol i of the alphabet, we have

$$|\varphi_i\rangle = \int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}\frac{dk_1\ldots dk_{n_i}}{k_{01}\ldots k_{0n_i}} \times \varphi(k_1, k_{01} = |k_1|, \ldots, k_{0n_i}, k_{0n_i} = |k_{n_i}|)|k_1, \ldots, k_{n_i}\rangle. \tag{25}$$

The amplitude $\varphi(k_1, \ldots)$ is defined by its values on the mass surface $k_0 = |k|$. The generalized basis vectors are $|k_1, \ldots\rangle \in \Omega^*$, $\varphi \in \Omega$, $|\varphi\rangle \in \mathcal{H}$. The construction $\Omega \subset \mathcal{H} \subset \Omega^*$ is called an equipped Hilbert space (the Gelfand triple) [11, 14].

Further, we will consider the states that propagate in one direction $k > 0$; it is these states that carry information between remote users. In this case, the amplitude of a state is given by

$$|\varphi_i\rangle = \int_0^{\infty}\ldots\int_0^{\infty}\frac{dk_1\ldots dk_{n_i}}{k_1\ldots k_{n_i}} \times \tilde{\varphi}_i(k_1, \ldots, k_{n_i})|k_1, \ldots, k_{n_i}\rangle, \tag{26}$$

where

$$\tilde{\varphi}_i(k_1, \ldots, k_{n_i}) = \varphi(k_1, k_{01} = k_1, \ldots, k_{n_1}, k_{0n_i} = k_{n_i}).$$

Below, we will need a coordinate–time representation of states. In this representation, a state is expressed as

$$|\varphi_i\rangle = \int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} d(x_1 - t_t)\ldots d(x_{n_i} - t_{n_i}) \times \varphi_i(x_1 - t_1, \ldots, x_{n_i} - t_{n_i})|x_i - t_1, \ldots, x_{n_i} - t_{n_i}\rangle = |\varphi_i\rangle = \int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} d\tau_1\ldots d\tau_{n_i} \times \varphi_i(\tau_1, \ldots, \tau_{n_i})|\tau_1, \ldots, \tau_{n_i}\rangle. \tag{27}$$

For the states that propagate in one direction, the amplitude depends only on the difference $\tau_l = x_l - t_l$ ($l_1, \ldots, n_i$). Physical states in $\mathcal{H}$ are defined by the values of the amplitude on the mass surface. The dependence of the amplitude of the states propagating in one direction on the difference $\tau = x - t$ reflects the fact that, if a measurement result was obtained at a moment t in the neighborhood of a point $(x, x + dx)$, then a similar result

can be obtained at a moment t' in the neighborhood of a point $(x', x' - x + t + dx)$. Next, we have

$$\varphi_i(\tau_1, \ldots, \tau_{n_i}) = \frac{1}{(2\pi)^{n_i}} \int_0^\infty \ldots \int_0^\infty \frac{dk_1 \ldots dk_{n_1}}{\sqrt{k_1 \ldots k_{n_i}}} \times \exp(i(k_1\tau_1 + \ldots + k_{n_i}\tau_{n_i}))\varphi_i(k_1, \ldots, k_{n_i}), \quad (28)$$

$$|\tau_1, \ldots, \tau_{n_i}\rangle = \int_0^\infty \ldots \int_0^\infty \frac{dk_1 \ldots dk_{n_1}}{\sqrt{k_1 \ldots k_{n_i}}} \times \exp(-i(k_1\tau_1 + \ldots + k_{n_i}\tau_{n_i}))|k_1, \ldots, k_{n_i}\rangle. \quad (29)$$

Since photons are bosons, the amplitude of a state should be symmetric with respect to the permutation of particles. This result follows automatically if the generalized basis vectors $|k_1, \ldots, k_{n_i}\rangle$ themselves (more precisely, continuous linear functionals) are symmetric. By definition, we have [10–12]

$$|k_1, \ldots, k_{n_i}\rangle = a^+(\hat{k}_1)\ldots a^+(\hat{k}_{n_i})|0\rangle = \sqrt{\frac{k_1 \ldots k_{n_i}}{n_i!}} \sum_{\{j\}} \delta(k_i - q_{j_1})\ldots\delta(k_{n_i} - q_{j_{n_i}}), \quad (30)$$

where the symbol $\{j\}$ under the sign of summation implies all possible permutations of indices. Accordingly, a scalar product of generalized basis vectors is given by

$$\langle k'_1 \ldots k'_m | k_1 \ldots k_n\rangle = \delta_{n,m} k_1 \ldots k_n \sum_{\{j\}} \delta(k_1 - k'_{j_i})\ldots\delta(k_n - k'_{j_n}). \quad (31)$$

These relations represent a definition of generalized basis vectors (continuous linear functionals—generalized functions). This orthogonality relation means that the basis vectors with different numbers of particles are orthogonal.

Now, let us consider a transmission of classical information by means of quantum states. Different symbols of the classical alphabet $\{x_i\}$ are assigned the states of a (not necessarily one-photon) quantized field with different amplitudes (shapes of multiphoton packets) $|\varphi_i\rangle$. Let us calculate the transmission rate in terms of [bit/message s] for a channel with finite bandwidth Δk (here, we use the system of units in which $\hbar = c = 1$).

By analogy with [1, 3–7], we will apply the method of random coding (random choice of code words), which is reduced to the following. Randomly choose N symbols $|\varphi_i\rangle$ of the alphabet with probabilities p_i. Generate M such random vectors of states, each of length N. Here, we have an essential difference from quantum channels in the nonrelativistic case, where it is assumed that the states are distinguishable and a message of length N is described by a vector of states as the tensor product

$$|\varphi_{i_1}\rangle \otimes |\varphi_{i_2}\rangle \otimes \ldots \otimes |\varphi_{i_N}\rangle \in \mathcal{H}^{\otimes N}. \quad (32)$$

This expression implies that quantum states in each separate message are sent into a communication channel independently of the preceding and subsequent states.

In the relativistic case, the situation is essentially different for two reasons. Because the amplitudes of the states defined on the mass surface are essentially nonlocalizable, one has to take into account the identity of particles. Due to the identity of particles, the states in individual messages cannot be assumed independent. In quantum field theory, nonlocalizability means that the amplitude of a state (a smoothing function) is different from zero in the entire space and cannot be exactly equal to zero outside any compact domain of the space. Moreover, the amplitude cannot decay exponentially at infinity. The square integrability (normalization condition) of $\varphi(k)$ imposes constraints on the admissible decrease rate of $\varphi(\tau)$ at infinity ($\tau \longrightarrow \infty$). An answer is given by the Wiener–Paley theorem [15]. For a square-integrable function $\varphi(k)$ that is equal to zero on the half-line $k < 0$ but is not identically zero, the following integral must converge:

$$\int_{-\infty}^{\infty} \frac{\ln|\varphi(\tau)|}{1 + \tau^2} d\tau < \infty.$$

Hence, the amplitude $\varphi(\tau)$ on the light cone cannot decrease exponentially but can decrease by a law arbitrary close to an exponential function:

$$|\varphi(\tau)| \propto \exp\{-\alpha|\tau|/\ln(\ln(\ln\ldots\ln|\tau|))\},$$

where α is any positive number. This implies that the states are nonlocalizable (different from zero outside any compact domain). A fundamental cause of such nonlocalizability is the fact that physical states are determined by the values of the amplitude on the mass surface. Moreover, as was shown by Hegerfeldt [16], the possibility of strict localizability of the amplitude would contradict the relativistic causality.

It is also known that the nonlocalizability is closely related to the causality in the quantum relativistic field. As was pointed out in [16], if one could strictly localize the amplitude of a state in a finite domain at the initial moment, then, at subsequent moments, the amplitude of the state would be different from zero in the domains separated from the original domain by a spacelike interval; this would lead to a situation when the speed of information transmission would be faster than the speed of light.

Thus, because the states of a quantized field are nonlocalizable, the amplitudes of states in different messages inevitably overlap. In this case, one can speak only of a general vector of state of the whole message

of length N and take into account the identity of particles. It is this fact that imposes a fundamental constraint on the speed of information transmission through quantum communication channels. In the nonrelativistic description (which, strictly speaking, is inapplicable to photons), these constraints (especially the identity of particles) have not been taken into account up to the present. Note that taking into account the identity of particles is also essential for the processes of teleportation [17] and when realizing quantum algorithms in real physical systems [18, 19].

The states in individual messages can be made distinguishable by increasing the time separation between successive messages by a large value (in principle, up to infinity). However, the speed of information transmission in this case will tend to zero. Therefore, to clear up the question about the transmission rate of a quantum communication channel (the real-time speed of information transmission), one cannot, in general, apply formulas obtained without taking into account the identity of particles because the latter formulas are valid only for infinite separation between messages, which corresponds to an infinitely slow transmission of information. In this sense, the formulas for the transmission rate for distinguishable particles have an asymptotic character. The decrease in the time interval between separate messages suggests that the identity of particles should be taken into account.

Our immediate task is to clear up the question about the transmission speed of classical information by means of quantum states, depending on the shape of these states, the separation between individual messages, the channel bandwidth, and the observation time window at the receiving end.

The propagation of states toward the receiving end is determined by the action of a unitary operator that describes the translation. In order to avoid cumbersome calculations, we do not write out a general expression for the translation operator; however, we will assume that measurements at the receiving end are performed over states that are translated in spacetime.

The state vector corresponding to a code word of length N from a random set of M code words is given by

$$\begin{aligned}
&|\varphi_{i_1};\ \varphi_{i_2};\ \ldots;\ \varphi_{i_N}\rangle \\
&= \int_0^\infty \ldots \int_0^\infty d\mathbf{k}_{i_1} \ldots d\mathbf{k}_{i_N} \exp(i\mathbf{k}_{i_1}\cdot\boldsymbol{\tau}_{0i_1}) \\
&\times \varphi_{i_1}(\mathbf{k}_{i_1}) \ldots \exp(i\mathbf{k}_{i_N}\cdot\boldsymbol{\tau}_{0i_N})\varphi_{i_N}(\mathbf{k}_{i_N})|\mathbf{k}_{i_1}\ldots\mathbf{k}_{i_N}\rangle \qquad (33)\\
&= \int_{-\infty}^\infty \ldots \int_{-\infty}^\infty d\boldsymbol{\tau}_{i_1}\ldots d\boldsymbol{\tau}_{i_N}\varphi_{i_1}(\boldsymbol{\tau}_{i_1}-\boldsymbol{\tau}_{0i_1})\ldots \\
&\times \varphi_{i_N}(\boldsymbol{\tau}_{i_N}-\boldsymbol{\tau}_{0i_N})|\boldsymbol{\tau}_{i_1}\ldots\boldsymbol{\tau}_{i_N}\rangle \in \mathrm{Sym}\,\mathscr{H}^{\otimes N},
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{k}_{i_m} &= (k_1, k_2, \ldots, k_{n_{i_m}}),\\
\boldsymbol{\tau}_{i_m} &= (\tau_1, \tau_2, \ldots, \tau_{n_{i_m}}),\\
\boldsymbol{\tau}_{0i_m} &= (\tau_{0i_1}, \tau_{0i_2}, \ldots, \tau_{0i_m}),\\
\varphi_{i_1}(\mathbf{k}_{i_1}) &= \varphi_{i_1}(k_{1,i_1}, k_{2,i_1}, \ldots, k_{2,i_{n_i}}).
\end{aligned} \qquad (34)$$

The amplitudes of the states $\varphi_{i_m}(\mathbf{k}_{i_m})$ in two successive messages represent amplitudes (packets in separate messages) that are displaced by $\tau_{0i_{m-1}}$ and $\tau_{0i_{m+1}}$.

The fact that independent states in separate messages do not exist follows from the identity of particles. More formally, the states in all messages are generated from the common vacuum operator $|0\rangle$ that describes an empty communication channel. The state vector of the entire message represents a symmetrized vector whose space of states is a symmetrized vector product $\mathrm{Sym}\,\mathscr{H}^{\otimes N}$ rather than the tensor product $\mathscr{H}^{\otimes N}$.

Note once again that the state of an entire message can be represented as a tensor product of separate independent messages only if the time interval between separate messages is sufficiently large (formally, when this interval is infinite, $\tau_{0i} \longrightarrow \infty$).

The density matrix corresponding to the code word from a random set has the form

$$\begin{aligned}
\rho_{\mathbf{i}}^{(N)} &= |\varphi_{i_1};\ \varphi_{i_2};\ \ldots;\ \varphi_{i_N}\rangle\langle\varphi_{i_1};\ \varphi_{i_2};\ \ldots;\ \varphi_{i_N}|,\\
\mathbf{i} &= (i_1, i_2, \ldots, i_N),
\end{aligned} \qquad (35)$$

which, after averaging $\mathbf{E}$ over all possible random sets of code words, yields a complete density matrix of all possible messages of length N:

$$\begin{aligned}
\rho^{(N)} &= \mathbf{E}(\rho_{\mathbf{i}}^{(N)})\\
&= \sum_{\mathbf{i}} P_{\mathbf{i}}|\varphi_{i_1};\ \varphi_{i_2};\ \ldots;\ \varphi_{i_N}\rangle\langle\varphi_{i_1};\ \varphi_{i_2};\ \ldots;\ \varphi_{i_N}|,
\end{aligned} \qquad (36)$$

$$P_{\mathbf{i}} = \sum_{\mathbf{i}} p_{i_1}\ldots p_{i_N},$$

where $P_{\mathbf{i}}$ is the probability of occurrence of a particular message that corresponds to the state vector $|\varphi_{i_1};\ \varphi_{i_2};\ \ldots;\ \varphi_{i_N}\rangle$.

If the measurement time at the receiving end was unlimited, then the amount of classical information that can be extracted from an ensemble of messages of length N (36) would be given by the following expression (which is obtained from the Holevo inequality [3]):

$$H(\rho^{(N)}) - \sum_{\mathbf{i}} P_{\mathbf{i}} H(\rho_{\mathbf{i}}^{(N)}). \qquad (37)$$

However, actually, all measurements are bound and are performed in finite (even though arbitrarily large) domains of spacetime. Since the states of the massless field of photons depend only on the difference $\tau = x - t$, one can perform measurements either at a fixed moment of time in a certain spatial domain or in the neighborhood of a certain point of space but during a certain finite period of time, waiting until the entire state reaches this point. More precisely, in the first case, measurements are performed by a spatially distributed device, so that an interaction between this device and a state is switched on at all points of the domain at the moment t. In the second case, when a measurement is performed in a bounded domain of space, the state of a field is unitarily transformed into a state of a localized atomic system (detector); moreover, a unitary transformation lasts a certain finite period of time that is necessary for the state to eventually reach a local device.

Further, for short, we will say that measurements are performed within a time window, which we denote by T; here, by T, we mean a spacetime domain (in the above-mentioned sense of $T = \Delta(x - t)$).

Both situations are formally described identically and, according to the general theory of quantum-mechanical measurements, correspond to the introduction of measuring operators that form a partition of unity. This means that each spacetime domain is assigned a measuring operator, so that the sum of operators assigned to the entire space is equal to unity.

As applied to our situation, there are two domains (two time windows),

$$T + \bar{T} = (-\infty, \infty).$$

Information is extracted by an observer from the time window T of measurement.

The measurements in a finite time window correspond to the introduction of a certain superoperator that gives a density matrix that an observer at the receiving end can see in the time window T. The introduction of such a superoperator allows one to apply formula (37) to calculate the transmission capacity. We are interested in the limit when the length of a particular message tends to infinity (the limit of long sequences as $N \longrightarrow \infty$). In this case, the time window T_N necessary for the measurement also increases with N. We are interested in the limit of the ratio of the amount of classical information extracted from a message of length N to the size of the time window T_N. It is this limit that represents the transmission rate of the communication channel in terms of the number of classical bits per one message per unit time, [bit/message s].

Next, we will need a tool that transforms a state defined in an infinite time window into a state that an observer can see in a finite time window. One can easily construct an operator-valued measure that describes measurements in a finite time window. Any measurement is described by a partition of unity. The full identity operator is a direct sum of symmetrized identity operators in subspaces with different numbers of photons $m = 0, 1, \ldots, \infty$:

$$I = \bigoplus_{m=0}^{\infty} I^{(m)}, \tag{38}$$

where

$$I^{(m)} = \int_0^\infty \ldots \int_0^\infty \frac{dk_1 \ldots dk_m}{k_1 \ldots k_n} \times |k_1, \ldots, k_n\rangle\langle k_1, \ldots, k_m|. \tag{39}$$

If the measurements are performed in a finite time window T, then the partition of unity that describes this measurement is

$$I^{(m)} = \mathcal{M}^{(m)}(T) + \mathcal{M}^{(m)}(\bar{T}), \quad T + \bar{T} = (-\infty, \infty), \tag{40}$$

where

$$\mathcal{M}^{(m)}(T) = \int_{-T}^{T} \ldots \int_{-T}^{T} \frac{d\tau_1 \ldots d\tau_m}{(2\pi)^m} \left(\int_0^\infty \ldots \int_0^\infty \frac{dk_1 \ldots dk_m}{\sqrt{k_1 \ldots k_m}} \times \exp(i(k_1\tau_1 + \ldots + k_m\tau_m))|k_1, \ldots, k_m\rangle \right) \times \left(\int_0^\infty \ldots \int_0^\infty \frac{dk'_1 \ldots dk'_m}{\sqrt{k'_1 \ldots k'_m}} \exp(-i(k'_1\tau_1 + \ldots + k'_m\tau_m))\langle k'_1, \ldots, k'_m| \right). \tag{41}$$

The operator-valued measure (40), (41) is sufficient to obtain the probabilities of measurement results in a finite time window. However, to calculate the transmission rate, one needs the density matrix of the ensemble of messages that an observer can see in the finite time window $(-T, T)$ in which the measurements are performed. An operator-valued measure alone is insufficient for these purposes; one needs the very tool (a superoperator) that leads to this measure. According to the Kraus theorem [8], any superoperator can be represented as

$$\mathcal{T}[\ldots] = \bigoplus_{m=0}^{\infty} \mathcal{T}^{(m)} = \bigoplus_{m=0}^{\infty} \sum_j V(T)_j^m [\ldots] V(T)_j^{(m)+} + \bigoplus_{m=0}^{\infty} \sum_j V(\bar{T})_j^m [\ldots] V(\bar{T})_j^{(m)+}, \tag{42}$$

where the operators $V(T)_j^{(m)}$ and $V(\bar{T})_j^{(m)}$ must satisfy

the relation

$$\sum_j V(T)_j^{(m)+} V(T)_j^{(m)} + \sum_j V(\bar{T})_j^{(m)+} V(\bar{T})_j^{(m)} = I^{(m)}. \quad (43)$$

The operator-valued measure generated by a superoperator is expressed in terms of the operators $V(T)_j^{(m)}$ in (42) and (43) as follows:

$$\mathcal{M}^{(m)}(T) = V(T)_j^{(m)+} V(T)_j^{(m)}. \quad (44)$$

As the operators $V_j^{(m)}[\ldots]$, one can choose the following operators:

$$\mathcal{T}^{(m)}[\ldots] = \sqrt{\mathcal{M}^{(m)}(T)}[\ldots]\sqrt{\mathcal{M}^{(m)+}(T)} + V(\bar{T})_j^{(m)}[\ldots]V(\bar{T})_j^{(m)+}. \quad (45)$$

The density matrix of the ensemble of messages that an observer can see in the finite time window $(-T, T)$ is obtained by the action of a superoperator and is given by

$$\mathcal{T}[\rho^{(N)}] = \bigoplus_{m=0}^{\infty} \sqrt{\mathcal{M}^{(m)}(T)}[\rho^{(N)}]\sqrt{\mathcal{M}^{(m)+}(T)} + \bigoplus_{m=0}^{\infty} V(\bar{T})_j^{(m)}[\rho^{(N)}]V(\bar{T})_j^{(m)+}. \quad (46)$$

Now, we have to determine the explicit form of the operators entering in (42) and (43); to this end, we should specify the properties of the quantum communication channel since the operators in (42) and (43) represent a formalized description of the properties of the quantum communication channel.

So far, we have not imposed any restrictions on the transmission bandwidth of the quantum communication channel. In any real-life situation, the transmission bandwidth is finite. In what follows, we will assume that the bandwidth is finite and equal to Δk (this is a free parameter that describes the channel and can be arbitrary, even arbitrarily large). In this case, the integrals in (39)–(41) should be taken with respect to each k_i over a finite interval Δk rather than over the interval $(0, \infty)$. If we have to determine the properties of a channel with infinite bandwidth, we can let the bandwidth in the final expressions tend to infinity.

Now, our task is reduced to the extraction of the square root from the operators $\mathcal{M}$. Such an operator exists and is given by

$$\sqrt{\mathcal{M}^{(m)}(T)} = \frac{1}{\pi}\int_0^\infty \frac{d\xi}{\sqrt{\xi}} \frac{\mathcal{M}^{(m)}(T)}{(\xi I^{(m)}(\Delta k) + \mathcal{M}^{(m)}(T))} = \sum_{\mathbf{j}=0}^{\infty} \sqrt{\lambda_{\mathbf{j}}^{(m)}} |u_{\mathbf{j}}^{(m)}\rangle\langle u_{\mathbf{j}}^{(m)}|, \quad (47)$$

$$\mathbf{j} = (j_1, j_2, \ldots, j_m), \quad j_k = 0, 1, \ldots, \infty$$

$$\lambda_{\mathbf{j}}^{(m)} = \lambda_{j_1}\ldots\lambda_{j_m}.$$

The vectors $|u_{\mathbf{j}}^{(m)}\rangle$ satisfy the integral equation

$$\lambda_{\mathbf{j}}^{(m)}|u_{\mathbf{j}}^{(m)}\rangle = \hat{\mathcal{K}}^{(m)}(T)|u_{\mathbf{j}}^{(m)}\rangle, \quad (48)$$

where the integral operator is expressed as

$$\hat{\mathcal{K}}^{(m)}(T) = \int_{\Delta k}\ldots\int_{\Delta k} dk_1\ldots dk_m dk_1'\ldots dk_m' \times \mathcal{K}(T; k_1 - k_1')\ldots\mathcal{K}(T; k_m - k_m') \times |k_1\ldots k_m\rangle\langle k_1'\ldots k_m'|, \quad (49)$$

where

$$\mathcal{K}(T; q - q') = \frac{1}{\pi}\frac{\sin(q-q')T}{q-q'}. \quad (50)$$

The eigenvectors of the integral operator are given by

$$|u_{\mathbf{j}}^{(m)}\rangle = \int_{\Delta k}\ldots\int_{\Delta k} \frac{dq_1 dq_m}{\sqrt{q_1\ldots q_m}} \times u_{j_1}(q_1)\ldots u_{j_m}(q_m)|q_1, \ldots, q_m\rangle, \quad (51)$$

the functions entering in (47), (48), and (51) satisfy the following integral equation:

$$\lambda_n u_n(q) = \frac{1}{\pi}\int_{\Delta k} dq' \frac{\sin(q-q')T}{q-q'} u_n(q'). \quad (52)$$

The maximal eigenvalue corresponds to the maximum of the functional, and the eigenfunction of this eigenvalue yields an optimal shape of the state. This equation was earlier studied in [20, 21]; the eigenvalues of the equation are positive and form a sequence that decreases as the number n increases ($1 > \lambda_0 > \lambda_1 > \ldots > 0$, $n = 0, 1, \ldots, \infty$). The eigenvalues depend on the parameter $\Delta k \cdot T$; the first several eigenvalues for different values of the parameter $\Delta k \cdot T$ have been obtained in [20] (for

large values of $\Delta k \cdot T$, they rapidly tend to 1; for example, $\lambda_0 = 0.99589$ for $\Delta k \cdot T = 4$). The asymptotic behavior of the eigenvalues for $\Delta k \cdot T \gg 1$ for fixed n is also known and given by

$$\lambda_n(\zeta) \approx 1 - \frac{4\sqrt{\pi}8^n}{n!}\zeta^{n+1/2}e^{-2\zeta}, \quad \zeta = \Delta k \cdot T; \qquad (53)$$

i.e., the eigenvalues are exponentially close to unity. The eigenvalues give the maximum possible degree of localization of states in spacetime within a given channel bandwidth Δk. Recall that $T = \Delta(x - t)$ is a distance on the light cone.

The density matrix that an observer can see in a finite time window consists of two terms (46). The first term represents an effective density matrix that an observer can see in the window $(-T, T)$. The second term represents a density matrix that corresponds to the results of measurements performed outside the observation time window. In other words, this term is constructed so that it corresponds to the measurement results obtained outside the observation time window $(-T, T)$. When a measurement is performed in a finite window, any measurement result takes place either in this window or outside it. For the observer, the measurement results outside the window look like the absence of recording in the window $(-T, T)$. Therefore, the observer should associate the absence of a result in the observation window (formally, the result is obtained but is not available to the observer) with an uncertain result. Therefore, the density matrix in (46) can be represented as

$$\mathcal{T}[\rho^{(N)}] = \bigoplus_{m=0}^{\infty}\sqrt{\mathcal{M}^{(m)}(T)}[\rho^{(N)}]\sqrt{\mathcal{M}^{(m)+}(T)} + \bigoplus_{m=0}^{\infty}\int_{\bar{T}} d\tau|?\rangle\langle\Psi^{(m)}(\tau)|[\rho^{(N)}]|\Psi^{(m)}(\tau)\rangle\langle?|, \qquad (54)$$

where

$$|\Psi^{(m)}(\tau)\rangle = \int_0^{\infty}\cdots\int_0^{\infty}\frac{dk_1\ldots dk_m}{\sqrt{k_1\ldots k_m}} \times \exp(i(k_1\tau_1 + \ldots + k_m\tau_m))|k_1, \ldots, k_m\rangle. \qquad (55)$$

Here, we introduced a formal state $|?\rangle$ that describes the results outside the window. This state is orthogonal to all the other states. It implies that the result of any measurement takes place either in $(-T, T)$ or outside it. The result of any measurement within the window $(-T, T)$ can be represented in terms of a partition of unity in this window. Such unity is given by $\mathcal{M}(T)$, while the vectors $|u_{\mathbf{j}}^{(m)}\rangle$ form a complete orthogonal basis (orthogonal partition of unity) within the observation time window. Therefore, any (orthogonal or nonorthogonal) measurements within a window are reduced to the measurements over the density matrix described by the first term in (54). The second term in (54) is responsible only for the probability of the absence of a result within the window (or for the probability of a result outside the window), which can be expressed in terms of a state orthogonal to any other state in the first term. The probability of such results is defined as

$$|?\rangle\left(\bigoplus_{m=0}^{\infty}\int_{\bar{T}} d\tau\langle\Psi^{(m)}(\tau)|[\rho^{(N)}]|\Psi^{(m)}(\tau)\rangle\right)\langle?| = |?\rangle\left(\bigoplus_{m=0}^{\infty}\mathrm{Tr}\{\mathcal{M}^{(m)}(\bar{T})\rho^{(N)}\}\right)\langle?|, \qquad (56)$$

as is prescribed by the action of the second part of the superoperator. Here, the summation over j is replaced by the integration with respect to $d\tau$.

The difficulties described that are associated with the construction of an orthogonal basis in a finite time window arise due to the nonorthogonality of the original basis in the τ representation (the generalized basis vectors $|\tau\rangle$ for different τ are not orthogonal even for the infinite bandwidth). If the generalized basis vectors were orthogonal, then the effective density matrix in a finite time domain would be obtained simply by projecting this matrix onto this domain. In our case, we have to apply a slightly more complicated calculation procedure.

If we have the density matrix that is available to the observer who performs measurements in a finite time window, then we can apply the Holevo formula. Using the method of doubly typical sequences (see [3, 6] for details), we can obtain the following inequality for the mean decoding error at the receiving end $\bar{P}_{err}(N, M)$ for arbitrarily small $\varepsilon > 0$ and $\delta > 0$:

$$\bar{P}_{err}(N, M) \le \varepsilon + Me^{-(\chi^{(N)} - \delta)}, \qquad (57)$$

where

$$\chi^{(N)} = H(\mathcal{T}[\rho^{(N)}]) - H\left(\sqrt{\mathcal{M}^{(m)}(T)}[\rho^{(N)}]\sqrt{\mathcal{M}^{(m)+}(T)}\right) - H\left(|?\rangle\langle?|\left(\bigoplus_{m=0}^{\infty}\mathrm{Tr}\{\mathcal{M}^{(m)}(\bar{T})\rho^{(N)}\}\right)\right). \qquad (58)$$

The decoding error at the receiving end tends to zero if the number of code words M (that are uniquely identified with probability one) for an ensemble of messages of length N does not exceed

$$M = e^{NR^{(N)}}, \quad R^{(N)} < \frac{\chi^{(N)}}{N} - \delta', \quad \delta' = \frac{\delta}{N} \longrightarrow 0. \qquad (59)$$

What is important here is the limit of $\chi^{(N)}/N$ as $N \longrightarrow \infty$. This limit is always less than $\chi^{(1)N}$ because the overlap of individual packets decreases the distinguishability of the code words due to the identity of particles. For example, if one uses orthogonal states of separate packets and the spatial carriers of these packets do not overlap (the packets of individual messages are separated by a large interval; formally, by an infinite interval), then the quantity

$$H^{(1)N} = NH(p) = -N\sum_i p_i \log p_i$$

tends to the classical limit. In the case of overlapping spatial amplitudes (such overlapping always takes place due to the essential nonlocalizability of the amplitudes of states), separate code words, even if each of them is constructed from orthogonal packets, cease to be orthogonal, i.e., certainly distinguishable. Even in the case of arbitrarily strongly localized carriers, we have

$$H^{(N)} \longrightarrow NH(p) = -N\sum_i p_i \log p_i$$

only for the infinite separation of states, $\tau_0 = \infty$.

The quantity

$$\chi^{(N)} \longrightarrow N\chi^{(1)} = -N\mathrm{Tr}\{\rho^{(1)}\log\rho^{(1)}\}, \quad \tau_0 = \infty,$$

tends to the limit of distinguishable particles, which is attained in the limit of an infinitely small transmission rate.

The ratio $\chi^{(N)}/N$ is a decreasing function of N for fixed separation τ_0 and shape of the amplitude. For $\tau_0 = 0$, we have $\chi^{(N)} = 0$ because all states from a typical subspace become identical. Suppose that at least one packet is shifted in time to infinity ($\tau_0 = \infty$). In this case, the density matrix is

$$\rho^{(N)} = \rho^{(N-1)} \otimes \rho^{(1)}$$

and

$$\frac{\chi^{(N)}}{N} \le \frac{\chi^{(N-1)}}{N} + \frac{\chi^{(1)}}{N} \le \frac{\chi^{(N-1)}}{N-1}. \tag{60}$$

The sequence $\chi^{(N)}/N$ decreases and is bounded from below; therefore, the limit exists and is equal to

$$\chi_\infty = \lim_{N\to\infty} \frac{\chi^{(N)}}{N}. \tag{61}$$

The quantity χ_∞ is the transmission capacity and defines the rate of information transmission in terms of the number of bits per message per photon; however, it is not the bit rate because it requires access to the entire branch of the light cone. The transmission capacity χ_∞ cannot be calculated analytically; however, it can be determined numerically for given shape of a packet and separation τ_0 between messages:

$$\begin{aligned}\chi^{(N)} = H\Bigg(&\bigoplus_{m,m'=0}^{\infty}\sum_{\mathbf{j},\mathbf{j}'=0}^{\infty}\sqrt{\lambda_{\mathbf{j}}^{(m)}\lambda_{\mathbf{j}'}^{(m')}}|u_{\mathbf{j}'}^{(m')}\rangle\langle u_{\mathbf{j}}^{(m)}| \\ &\times\Bigg(\sum_{\mathbf{i}} P_{\mathbf{i}}\rho_{\mathbf{i}}^{(N)}\Bigg)|u_{\mathbf{j}}^{(m)}\rangle\langle u_{\mathbf{j}'}^{(m')}| \\ &+ |?\rangle\Bigg(\bigoplus_{m=0}^{\infty}\mathrm{Tr}\{\mathcal{M}^{(m)}(\bar{T})\rho^{(N)}\}\Bigg)\langle ?|\Bigg) \\ -\sum_{\mathbf{i}} P_{\mathbf{i}} H\Bigg(&\bigoplus_{m,m'=0}^{\infty}\sum_{\mathbf{j},\mathbf{j}'=0}^{\infty}\sqrt{\lambda_{\mathbf{j}}^{(m)}\lambda_{\mathbf{j}'}^{(m')}}|u_{\mathbf{j}'}^{(m')}\rangle \\ &\times\langle u_{\mathbf{j}}^{(m)}|\rho_{\mathbf{i}}^{(N)}|u_{\mathbf{j}}^{(m)}\rangle\langle u_{\mathbf{j}'}^{(m')}|\Bigg).\end{aligned} \tag{62}$$

Suppose that the time taken for sending a message of length N is chosen equal to

$$T_N = N\cdot T = \sum_i^N \tau_{0i} \tag{63}$$

(here, T is time per one message). Recall that τ_{0i} are the separation intervals between individual messages; i.e., the amplitudes of states (packets) are given by

$$\varphi_i(k_{1_i}, \ldots, k_{n_i})\exp(-i(k_{1_i}\tau_{0i} + \ldots + k_{n_i}\tau_{0i})).$$

In the spacetime representation, this situation corresponds to a shift by τ_{0i} between two successive messages i and $i+1$.

The transmission rate of a channel with finite bandwidth Δk is equal to

$$C(\Delta k, T) = \lim_{N\to\infty}\max_{\rho_i}\left\{\frac{\chi^{(N)}}{NT}\right\}\left[\frac{\text{bit}}{\text{message s}}\right], \tag{64}$$

which represents the number of bits per one message per unit time. Formulas (62)–(64) give a closed-form expression for the transmission rate of a sequential quantum channel as a function of the bandwidth (Δk), packet shape (ρ_i), separation τ_{0i} between individual messages, and the observation time window at the receiving end. Recall that the time window $T = \Delta(x - ct)$ represents a distance on the light cone for the photon field.

For a given channel bandwidth Δk, the maximum degree of localization in spacetime is determined by the solutions of the integral equation (52) and its eigenvalues (53). For a given bandwidth Δk, the size of the time window T outside which the normalization of packets is exponentially small is determined by (53). Here, successive messages can be assumed independent and the packet overlap and identity of particles can be neglected. Formally, the overlapping disappears completely only for infinite separation of messages. In this limit, one can apply formulas for the transmission capacity, which describes the amount of information per message, for nonrelativistic quantum communication channels with distinguishable states. However, in terms of speed of information transmission, the transmission capacity will tend to zero.

If the separation between successive messages is not infinite, then one cannot neglect the identity of particles. In this case, one should apply formulas (62) and (63). The order of magnitude of the transmission rate is given by

$$C(\Delta k, T) \le \frac{H(\rho)}{T}, \quad \rho = \sum_i^n p_i \rho_i,$$

where the density matrix ρ_i is constructed from pure states (27), which are chosen with a priori probabilities p_i. The value of T for which the messages can be assumed independent (overlapping) is determined by formula (53), $T > 1/\Delta k$. For the states that are optimal in spacetime (very short for a given bandwidth), one can choose T to be several times greater than $1/\Delta k$ due to the fact that part of the normalization of the state in the time window T exponentially rapidly reaches unity. The decrease in the time window increases the overlapping and, hence, decreases the distinguishability of states in individual messages; therefore, the estimate given above for the transmission rate is an upper estimate. The exact value of the transmission rate is given by (62) and (64) and is no greater than this estimate.

4. TRANSMISSION RATE OF A PARALLEL QUANTUM CHANNEL WITH ORTHOGONAL ONE-PHOTON INPUT STATES

The formulas obtained in the preceding section describe a general situation for a communication channel, which can be conventionally called a parallel channel. In this situation, the states are physically sent into a communication channel successively in time.

Another limiting situation is also possible when all the states are sent in the entire bandwidth of the channel in the same time window. Such a method of transmission is used in the so-called multiplexed mode, when each state (respectively, each symbol of the alphabet) in the message corresponds to its own frequency bandwidth. If the frequency bandwidths for separate states do not overlap, then the state space of messages is a tensor product of Hilbert spaces of states for each frequency bandwidth, rather than a symmetrized tensor product as in the case of the sequential communication channel. In this case, one can obtain final analytic formulas for orthogonal input states with a certain spacetime shape. These states represent a quantum analog of sampling functions [2].

To clear up the general formulas and their relation to the classical case, when the alphabet is coded into a time shape of a signal, we give an example of calculating the transmission capacity of a communication channel with finite bandwidth when the input signal states are one-photon states. The symbols of the classical alphabet $\{x_i\}$ are assigned pure one-photon states with different shapes of packets (amplitudes).

Let us return for a while to the classical case and consider the coding of a source on a finite time interval $(-T, T)$ for the case of a continuous classical input signal defined in a finite frequency band Δk.

A classical signal is defined by the time function

$$u(t) = \sum_i \sqrt{p_i} u_i(t), \tag{65}$$

where the subscript i runs through $2\Delta kT$ values corresponding to independent degrees of freedom of the signal; hence, we obtain an orthogonal set of functions that describe independent degrees of freedom of the signal:

$$u_i(t) = \sqrt{\Delta k} \frac{\sin 2\pi\Delta k[t - i/(2\Delta k)]}{2\pi\Delta k[t - i/(2\Delta k)]}. \tag{66}$$

Often, these functions are also called sampling functions [2]. The quantity $\sqrt{p_i}$ describes the amplitude, and p_i^2 makes the sense of the signal power at a certain harmonic $u_i(t)$.

Amplitudes $\sqrt{p_i}$ are random variables with certain a priori input probabilities; the probability distributions for each sampling test function are independent. In this case, the problem is reduced to the coding of $2\Delta kT$ independent sources. Formally, each p_i describes an a priori distribution of input probabilities in one of independent (parallel) channels. Physically, this situation is realized when narrow pulses of intensity p_i are applied to the input of a channel with finite bandwidth Δk successively over equal sampling times $[t - i/(2\Delta k)]$.

In the quantum case, there is no direct analogy with the classical case considered above. Consider an example of coding a quantum source in a channel with finite transmission bandwidth Δk, which is most similar to the classical case. As input signal states of the source, we choose the states that are orthogonal in a finite time window T (which is assumed to be fixed at the receiving

end). These states satisfy the integral equation (52); they represent analogs of orthogonal sampling test functions that are defined in a finite time window and have a finite frequency bandwidth. These states depend only on the product $\Delta k \cdot T$:

$$|u_i(\Delta k \cdot T)\rangle = \int_{\Delta k} \frac{dk}{\sqrt{k}} u_i(T, k)|k\rangle,$$
$$\langle u_i(\Delta k \cdot T)|u_{i'}(\Delta k \cdot T)\rangle = \delta_{i,i'}. \quad (67)$$

The a priori probabilities of the states are $\{p_i\}$. We assume that the number of symbols in the alphabet is bounded by a certain number $N_{\max}$. In the classical case, such a restriction implies that the signal intensity for each sampling function is bounded by a certain value (by the intensity of each sampling harmonic).

A significant difference between the quantum and classical cases manifests itself during a physical realization. In the classical case, one can apply narrow pulses at sampling times to the input of a channel with a finite frequency bandwidth. These pulses are added up in the channel to form a complete time-continuous signal (65). At the receiving end, one measures the amplitude of signal (65) at equidistant moments of time $[t - i/(2\Delta k)]$ and extracts the values of p_i into which information is coded. In this case, due to the classical nature of the signal, such a process is continuous in time, and the signal itself is not destroyed under the measurements. Therefore, in formal mathematical consideration, one can assume that the quantities p_i are transmitted through independent parallel channels due to the orthogonality of the sampling harmonics.

In the ultraquantum case, when the input states are one-photon states, such a realization is certainly impossible. Arguments analogous to those in the classical case immediately encounter the following obstacles. Suppose that strongly localized states that can be assumed independent are applied to the input of a channel successively over equal sampling intervals. Suppose that this sequence can be described by the tensor product of separate messages rather than by a symmetrized tensor product. The states are so strongly localized that one can neglect the overlapping of adjacent states. Although a quantum state of the photon field cannot be strictly localized, it can be made arbitrarily strongly localized. In this case, the states at the output of the channel essentially overlap. The general state at the output cannot be described (due to the identity of particles) as a tensor product of separate states of the form (5) but should be described by a unified tensor of states, as is done in Section 3 for a sequential communication channel. There is an interesting question concerning the transition from the classical to the quantum case. Such a transition is performed by decreasing the intensity (the number of photons) of each sampling function. Below, we consider a case when the number of photons is maximally reduced, namely, reduced to the one-photon level.

To make the analysis maximally close to that in the classical case, consider the following method of coding. We apply the following arguments to preserve the structure of independent states in each message and, hence, to preserve the structure of the tensor product. Any derivation of the coding theorems and of the expression for the transmission capacity suggests the use of long typical sequences whose length in the final formulas tends to infinity. Since the chosen orthogonal states in a finite time window T depend only on the product of the channel bandwidth multiplied by the time window, to form a sequence of messages of length n (where n is arbitrary and tends to infinity), we divide the frequency bandwidth of the channel into n identical nonoverlapping frequency bands of width $\Delta k/n$ each of which corresponds to one of parallel channels. The quantum states in parallel channels with nonoverlapping frequency bandwidths can be represented by a tensor product.

The n-fold decrease in the frequency bandwidth leads to the effective increase in the observation time window by a factor of nT. Such a scaling of the time window is necessary to guarantee that the input states corresponding to the symbols of the alphabet in sequences of different lengths are invariant (independent of n). However, since the signal amplitude depends only on the product $(\Delta k/n) \cdot nT$, the signal input states themselves within the bandwidth $\Delta k/n$ will have the same localization in the time window nT (67).

The physical realization of the coding reduces to the following. Choose a number n and divide a channel into n independent parallel channels. The quantum states (68) are simultaneously (in parallel) sent into each frequency channel with a priori probabilities p_i. The total number of input states in all parallel channels is given by the tensor product of states in separate independent channels. At the receiving end, one performs a collective measurement, in the time window nT, over the tensor product of the output density matrices in all the channels. The transmission rate is calculated as the limit of C_n/nT.

In view of the aforesaid, we have

$$\rho_{i_k} = \left|u_{i_k}\left(\frac{\Delta k}{n}nT\right)\right\rangle\left\langle u_{i_k}\left(\frac{\Delta k}{n}nT\right)\right| \quad (68)$$

and the a priori probabilities $\{p_i\}$ in n independent parallel channels. The subscript i_k refers to the states in the kth parallel channel. There is no need to introduce this subscript into the argument of $u_{i_k}((\Delta k/n)nT)$. At the receiving end, the measurements over n states are performed in the time window nT. Now, the density matrix

of a particular message of length n can be represented as the tensor product

$$\rho^{(n)} = \rho_{i_1} \otimes \rho_{i_2} \otimes \ldots \otimes \rho_{i_n}, \tag{69}$$

and state ρ_{i_k} is sent into the kth channel with probability p_i. This means that, since the argument of a state is independent of n, the density matrix of each channel can be represented as

$$\begin{aligned}\rho &= \sum_i p_i \rho_i \\ &= \sum_i p_i \left| u_i\left(\frac{\Delta k}{n} nT\right)\right\rangle\left\langle u_i\left(\frac{\Delta k}{n} nT\right)\right| \\ &= \sum_i p_i |u_i(\Delta kT)\rangle\langle u_i(\Delta kT)|.\end{aligned} \tag{70}$$

The superoperator that sends the input density matrix (70) to the density matrix at the output in the finite time window nT acts as follows:

$$\begin{aligned}\mathcal{T}[\rho] &= \sqrt{\mathcal{M}(nT)}[\rho]\sqrt{\mathcal{M}(nT)^+} \\ &+ |?\rangle\langle ?|\mathrm{Tr}\left\{\left(I\left(\frac{\Delta k}{n}\right) - \mathcal{M}(nT)\right)\rho\right\},\end{aligned} \tag{71}$$

where

$$I\left(\frac{\Delta k}{n}\right) = \int\limits_{\frac{\Delta k}{n}} \frac{dk}{k}|k\rangle\langle k|. \tag{72}$$

The first term of the density matrix in (71) is responsible for the measurement results in the time window T that is available for observation. The second term is responsible for the measurement results outside T. For an observer, these results imply the absence of a state (the absence of recording in the time window T). Formally, the state $|?\rangle$ is orthogonal to all the other states. For a dummy state $|?\rangle$, it suffices to preserve only a diagonal term with respect to the new index ? to preserve the unit trace of the density matrix.

Strictly speaking, one should introduce an index that enumerates parallel channels into the identity operator in (72). However, since all these channels are identical and the results are independent of the position of the bandwidth, we omit the index. Next, we have

$$\begin{aligned}\mathcal{M}(nT) &= \sum_j^\infty \lambda_j\left(\frac{\Delta k}{n} nT\right) \\ &\times \left| u_j\left(\frac{\Delta k}{n} nT\right)\right\rangle\left\langle u_j\left(\frac{\Delta k}{n} nT\right)\right|,\end{aligned} \tag{73}$$

$$\begin{aligned}\sqrt{\mathcal{M}(nT)} &= \sum_j^\infty \sqrt{\lambda_j\left(\frac{\Delta k}{n} nT\right)} \\ &\times \left| u_j\left(\frac{\Delta k}{n} nT\right)\right\rangle\left\langle u_j\left(\frac{\Delta k}{n} nT\right)\right| .\end{aligned} \tag{74}$$

The transmission capacity is calculated as the limit as $n \longrightarrow \infty$. When the number of channels tends to infinity and, hence, $nT \longrightarrow \infty$ (for fixed initial T), we have the following expression for the transmission rate

$$C(\Delta k \cdot T) = \lim_{n\to\infty} \max_{p_i}\left\{\frac{nH\left(\sum_i p_i \mathcal{T}[\rho_i]\right) - n\sum_i p_i \mathcal{T}[\rho_i]}{nT}\right\}\left[\frac{\text{bit}}{\text{message s}}\right]. \tag{75}$$

Using (71)–(74) and the orthogonality of the functions u_i, we obtain

$$\begin{aligned}&\mathcal{T}[\rho_i]\left(\frac{\Delta k}{n} nT\right) \\ &= \lambda_i\left(\frac{\Delta k}{n} nT\right)\left| u_j\left(\frac{\Delta k}{n} nT\right)\right\rangle\left\langle u_j\left(\frac{\Delta k}{n} nT\right)\right| \\ &+ \left(1 - \lambda_i\left(\frac{\Delta k}{n} nT\right)\right)|?\rangle\langle ?|.\end{aligned} \tag{76}$$

Next,

$$\begin{aligned}&H\left(\sum_i^{N_{\max}} p_i \mathcal{T}[\rho_i]\right) \\ &= -\sum_i^{N_{\max}} \mu_i \log \mu_i - \mu_\perp \log \mu_\perp,\end{aligned} \tag{77}$$

where

$$\mu_i(\Delta k \cdot T) = p_i\lambda_i\left(\frac{\Delta k}{n}nT\right) = p_i\lambda_i(\Delta k \cdot T), \quad (78)$$

$$\mu_\perp(\Delta k \cdot T) = \sum_i^{N_{\max}} \mu_{i\perp}, \quad (79)$$

$$\begin{aligned}\mu_{i\perp} &= p_i\left(1 - \lambda_i\left(\frac{\Delta k}{n}nT\right)\right)\\ &= p_i(1 - \lambda_i(\Delta k \cdot T)),\end{aligned} \quad (80)$$

$$\begin{aligned}&\sum_i^{N_{\max}} p_i H(\rho_i)\\ &= -\sum_i^{N_{\max}} (p_i\lambda_i\log\lambda_i + p_i(1-\lambda_i)\log(1-\lambda_i)).\end{aligned} \quad (81)$$

Taking into account (75)–(81), we have the following final expression for the transmission rate in terms of the number of bits per message (per alphabet symbol) per unit time:

$$\begin{aligned}&C(\Delta k \cdot T)\\ &= \max_{\{p_i\}}\left\{\left[-\left(\sum_i^{N_{\max}} \mu_i\log\mu_i + \mu_\perp\log\mu_\perp\right)\right.\right.\\ &\left.\left. + \sum_i^{N_{\max}} \mu_i\log\lambda_i + \mu_{i\perp}\log(1-\lambda_i)\right]T^{-1}\right\}\left[\frac{\text{bit}}{\text{message s}}\right].\end{aligned} \quad (82)$$

Let us analyze the expression obtained. The first term in the numerator describes the entropy of the source with orthogonal (distinguishable) signals that correspond to a priori probabilities μ_i, $\mu_\perp$ ($i = 1, \ldots, N_{\max}$). The second term represents an analog of the conditional entropy between the input and output and describes a decrease in the entropy of the source due to the finiteness of the observation window. For a given channel bandwidth Δk, when the observation time window tends to infinity, $T \longrightarrow \infty$, we have $\mu_i \longrightarrow p_i$ and $\mu_\perp \longrightarrow 0$, and the numerator tends to the transmission rate of an ideal (noiseless) classical channel with a priori probabilities p_i:

$$C = -\sum_i^{N_{\max}} p_i\log p_i. \quad (83)$$

This quantity represents the rate of information transmission in terms of [bit/message]; however, in the sense of the speed of transmission [bit/message s], we have

$$C(\Delta k \cdot T) = \frac{C}{T} \longrightarrow 0, \quad T \longrightarrow \infty. \quad (84)$$

This result is associated with the fact that, for the input states $|u_i\rangle$ (67) that are orthogonal in a finite time window, the quantities $\lambda_i(\Delta k \cdot T)$ with moderately large i tend to unity exponentially rapidly with respect to the parameter $\Delta k \cdot T$ (53). Therefore, in a real-life situation, one can choose the time window T several times greater than the inverse of the frequency bandwidth of the channel, $T \geq 1/\Delta k$. In this case, the states "interfere" in this time window with exponential accuracy in this parameter. For a given channel bandwidth, one can always choose the time window T so that the tails of the states outside this window are exponentially small, less than any physically meaningful level (for example, e^{-80}); in this case, $T \approx 40/\Delta k$. Recall that the number of atoms in the visible part of the universe is estimated as 10^{77}. The aforesaid implies that, for a given channel bandwidth, in the ultraquantum limit of one-photon states, it suffices to separate successive messages by $T \approx 10/\Delta k$ in time so that the messages could be assumed independent. Each message (in the limit of long sequences) carries C bits, while the time interval between messages is T.

In this situation, it does not matter how to send the states into the channel, either sequentially or in parallel, because one can neglect the overlap of these states when sending them successively over intervals of T. Accordingly, one can also neglect the identity of particles and can apply formulas for the transmission capacities of communication channels with distinguishable (not identical) particles [3–7] that were obtained earlier. In this case, one should bear in mind that the transmission rate is given by C/T (T is chosen with regard to the aforesaid); i.e., the time interval between successive messages of states cannot be less than T. Sometimes, it is convenient to use the dimensionless transmission rate

$$\tilde{C}(\Delta k \cdot T) = \frac{C}{\Delta k \cdot T}. \quad (85)$$

This quantity represents the transmission rate per one message per unit time and within unit bandwidth.

If the transmission bandwidth of a channel is unbounded ($\Delta k \longrightarrow \infty$; although such a situation does not occur in reality), then the states can be sent into a channel with any frequency (since the time window can be chosen as $T \approx 1/\Delta k \longrightarrow 0$); in this case, to obtain the transmission capacities, one can apply formulas for the channels where the quantum states in each message are assumed to be independent (nonoverlapping) and, hence, are described by a tensor product rather than by a symmetrized tensor product.

In conclusion, note that the dimensionless transmission rate (85) is Lorentz invariant; it does not change when the observation at the receiving end is made in a frame of reference that moves with respect to the frame of reference of the source. In other words, the transmission rate is invariant with respect to the measurements in a moving system of coordinates. This result is intu-

itively clear since a transition to a moving frame of reference leads to the effective compression of the frequency spectrum of a state due to the Doppler effect,

$$\Delta k \longrightarrow \Delta k \sqrt{\frac{1-\beta}{1+\beta}}$$

(if the observer moves in the same direction as the state), and, consequently, to the effective spatial dilation of the state. The latter circumstance requires larger time,

$$T \longrightarrow T \sqrt{\frac{1+\beta}{1-\beta}},$$

to guarantee the same proportion of the state in the time window. However, since an answer depends only on the product $\Delta k \cdot T$, the result remains the same and does not depend on the direction of motion of the observer (on the sign of $\beta = v/c$). This conclusion can be explained as follows. Since the scalar product $\hat{k} \cdot \hat{x}$ is Lorentz invariant, $k = k_0$ for a photon propagating in one direction, and the spacetime variables enter only in the combinations $\tau = x - t$, the quantity $k\tau$ which determines the answer, is also Lorentz invariant (see, for example, [22]).

5. CONCLUSIONS

Thus, we have demonstrated that the nonlocalizability of states in quantum field theory leads to a situation when separate messages cannot generally be described as independent because they inevitably overlap, which requires that one should take into account the identity of particles. This fact, combined with the finiteness of the limit speed of propagation, suggests that the formulas for the transmission rate of nonrelativistic communication channels have an asymptotic character (i.e., they are formally valid only for the infinite separation between messages, when one can neglect the identity of particles). Formulas (62) and (82) give the transmission rate as a function of the parameters of the communication channel and the shape of the signal states.

ACKNOWLEDGMENTS

I am grateful to S. S. Nazin for useful discussions and critical remarks.

This work was supported by the Russian Foundation for Basic Research (project no. 02-02-16289) and the Ministry of Industry and Science of the Russian Federation (project nos. 40.020.1.1.1170 and 37.029.1.1.0031).

REFERENCES

1. C. E. Shannon, Bell Syst. Tech. J. **27**, 397 (1948); **27**, 623 (1948).
2. V. A. Kotel'nikov, in *Proceedings of First All-Union Meeting on Problems of Technical Reconstruction of the Communication Job and Weak-Current Development* (Vsesoyuz. Énerg. Komissiya, Moscow, 1933).
3. A. S. Kholevo, Probl. Peredachi Inf. **8**, 63 (1972); **15**, 3 (1979); Usp. Mat. Nauk **53**, 193 (1998).
4. R. Jozsa and B. Schumacher, J. Mod. Opt. **41**, 2343 (1994).
5. P. Hausladen, R. Jozsa, B. Schumacher, *et al.*, Phys. Rev. A **54**, 1869 (1996).
6. B. Schumacher and M. D. Westmoreland, Phys. Rev. A **56**, 131 (1997).
7. A. S. Kholevo, *Introduction to the Quantum Theory of Information* (MTsNMO, Moscow, 2002), Sovrem. Mat. Fiz., No. 5.
8. K. Kraus, *States, Effects, and Operations* (Springer, Berlin, 1983).
9. N. N. Bogolyubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields*, 3rd ed. (Nauka, Moscow, 1973; Wiley, New York, 1980).
10. N. N. Bogolyubov, A. A. Logunov, and I. T. Todorov, *Introduction to Axiomatic Quantum Field Theory* (Nauka, Moscow, 1969; Benjamin, New York, 1974).
11. N. N. Bogolyubov, A. A, Logunov, A. I. Oksak, and I. T. Todorov, *General Principles of the Quantum Field Theory* (Nauka, Moscow, 1987).
12. S. N. Molotkov, Pis'ma Zh. Éksp. Teor. Fiz. **76**, 683 (2002) [JETP Lett. **76**, 584 (2002)]; Pis'ma Zh. Éksp. Teor. Fiz. **77**, 51 (2003) [JETP Lett. **77**, 47 (2003)].
13. A. M. Jaffe, Phys. Rev. **158**, 1454 (1967).
14. I. M. Gel'fand and N. Ya. Vilenkin, *Generalized Functions*, Vol. 4: *Some Applications of the Harmonic Analysis. The Equipped Hilbert Spaces* (Fizmatgiz, Moscow, 1961; Academic, New York, 1964).
15. N. Wiener and R. Paley, *Fourier Transform in the Complex Domain* (American Mathematical Society, New York, 1934; Nauka, Moscow, 1964).
16. G. C. Hegerfeldt, Phys. Rev. D **10**, 3320 (1974); G. C. Hegerfeldt and S. N. M. Ruijsenaar, Phys. Rev. D **22**, 377 (1980).
17. R. Laiho, S. N. Molotkov, and S. S. Nazin, Phys. Lett. A **275**, 36 (2000); Phys. Lett. A **278**, 9 (2000).
18. S. B. Bravyi and A. Yu. Kitaev, Ann. Phys. (N.Y.) **298**, 210 (2002).
19. Yu. I. Ozhigov, Pis'ma Zh. Éksp. Teor. Fiz. **76**, 799 (2002) [JETP Lett. **76**, 675 (2002)]; quant-ph/0205038.
20. D. Slepian and H. O. Pollak, Bell Syst. Tech. J. **40**, 40 (1961).
21. W. H. Fuchs, J. Math. Anal. Appl. **9**, 317 (1964).
22. S. N. Molotkov, Pis'ma Zh. Éksp. Teor. Fiz. **74**, 477 (2001) [JETP Lett. **74**, 436 (2001)].

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 98, No. 2, 2004, pp. 390–393.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 2, 2004, pp. 441–444.

MISCELLANEOUS

Bound States of an Electron and a Macroscopic Cluster at a Liquid Helium Surface

E. V. Lebedeva[a], A. M. Dyugaev[b,c], and P. D. Grigor'ev[b,d]

[a]*Institute of Solid State Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia*
[b]*Landau Institute of Theoretical Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia*
[c]*Max Planck Institute for the Physics of Complex Systems, Dresden D-01187, Germany*
[d]*High Magnetic Field Laboratory, MPI–FKF and CNRS, Grenoble F-38042, France*
e-mail: pashag@itp.ac.ru

Received June 10, 2003

Abstract—In a strong electric field, there are bound states of an electron at the surface of liquid helium, interacting with a large cluster of atoms in the bulk of liquid. This phenomenon is related to long-range interaction between the electron and the dipole moment of the cluster. The electron, holding the cluster under the liquid surface, is localized at this surface. One electron is capable of binding a cluster of up to 10^6 atoms. The value of the binding energy may reach up to several kelvins. © *2004 MAIK "Nauka/Interperiodica".*

A two-dimensional electron gas at the surface of liquid helium is characterized by very high mobility. This is related to the ideal purity of this fluid: at cryogenic temperatures, all foreign particles are frozen out of helium, which is experimentally observed as a snowfall of small crystals of hydrogen and air [1, 2]. In recent years, there has been extensive investigation of the properties of foreign particles formed in liquid helium, for example, by laser sputtering. There are well-developed experimental techniques for introducing such impurities through the surface into the bulk of liquid helium [3, 4].

An interesting problem is related to the existence of charged complexes at the liquid helium surface, formed between electrons and macroscopic clusters. Recently, we have proposed and theoretically studied a model cluster formed below the liquid helium surface, comprising an electron tightly bound to the cluster by the polarization attraction forces [5]. Previously, the bound states of electrons at the surface of macroscopic clusters were studied by Khaikin [6] with respect to the influence of the cluster surface curvature on the stationary energy levels of an electron at this surface. We have also investigated [7] ions of large radius at the surface of liquid helium.

In this study, we will consider the bound states of an electron at the surface of helium, interacting with a macroscopic cluster situated below the surface. It will be demonstrated that in a sufficiently strong electric field, such bound states can form even for giant clusters with dimensions on the order of $R \sim 10^3$ Å. The essence of this phenomenon consists in that the liquid helium surface plays the role of the third body reducing dimensionality of the problem. The electron does not penetrate into helium because the energy required for this incorporation is very high, $V_0 \sim 1$ eV, which corresponds to an electric breakdown field strength of $E_0 \approx 10^8$ V/cm. On the other hand, the cluster is held within the liquid because of repulsion from the surface under the action of van der Waals forces. Switching on the electron–cluster interaction in a strong electric field gives rise to the self-localization effect, whereby the cluster is positioned at a certain depth z_{B_0} under the liquid helium surface and the electron performs a limited oscillatory motion along the surface.

The potential energy of the electron–cluster system under consideration can be expressed as

$$V = V_e(z_e) + V_B(z_B) + V_{eB}(\rho_e, z_e, z_B), \qquad (1)$$

where ρ_e, z_e are the electron coordinates and z_B is the cluster depth. The cluster is modeled by a sphere of radius R with the dielectric permittivity ϵ_B (see figure).

The electron potential energy $V_e(z_e)$ corresponds to an ansatz wave function $\varphi_0(z_e)$ such that [8]

$$V_e(z_e) = eEz_e - \frac{\epsilon_h - 1}{\epsilon_h + 1}\frac{e^2}{4z_e} + V_0\theta(-z_e),$$
$$\varphi_0^2(z_e) = \frac{z_e^2}{2z_0^3}\exp\left(-\frac{z_e}{z_0}\right), \qquad (2)$$

where E is the external electric field strength, $\epsilon_h = 1.054$ is the permittivity of He^4, and z_0 is a characteristic scale of electron motion along the z axis. For $E = 0$, we have

$z_0 \approx 100$ Å; in a strong field E, this parameter is estimated as $z_0 \approx (\hbar^2/meE)^{1/3}$.

Let us consider the potential energy of the cluster $V_B(z_B)$. With neglect of the gravity force, this energy can be represented as a sum of two parts:

$$V_B(z_B) = V_{BE}(z_B) + V_{\beta h}(z_B). \tag{3}$$

The first (electrostatic) part $V_{BE}(z_B)$ describes repulsion of a dipole with the moment **d** from the surface [9]:

$$V_{BE}(z_B) = \frac{d^2}{8z_B^3}\frac{\epsilon_h - 1}{\epsilon_h + 1}, \quad \mathbf{d} = \mathbf{E}R_*^3,$$
$$R_*^3 = R^3\frac{\epsilon_B - \epsilon_h}{\epsilon_B + 2\epsilon_h}. \tag{4}$$

Relation (4) is derived assuming that $z_B > R$. In a homogeneous electric field E, the cluster acquires the dipole moment **d** expressed in terms of the effective radius R_* by Eq. (4).

The second (van der Waals) part $V_{\beta h}(z_B)$ comprises a sum of the term independent of z_B, having the sense of the work of escape from the liquid $V_{\beta h}(z_\infty)$, and the term describing repulsion from the surface of liquid helium. The latter term is determined for $z_B \gg R$ as

$$V_{\beta h}(z_B) = V_{\beta h}(\infty) - n_B n_h \frac{4}{3}\pi R^3$$
$$\times \int v_{hB}(\mathbf{r}_h - \mathbf{r}_B)d\mathbf{r}_h, \quad z_h > 0, \tag{5}$$

where n_B and n_h are the atomic densities of the cluster and liquid helium, respectively, and v_{hB} is the potential of interaction between cluster and helium molecules. The integration in Eq. (5) is performed over the half-space $z_h > 0$. Since the z_B value is assumed to be macroscopically large, the potential v_{hB} in Eq. (5) decreases with increasing distance r between the interacting molecules as $1/r^7$ and depends only on their polarizabilities β_B and β_h [10]:

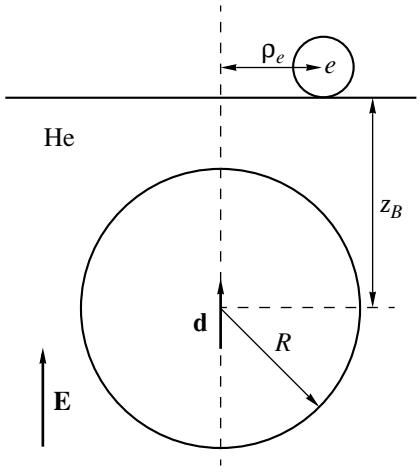
$$v_{hB} = -\frac{23}{4\pi}\frac{e^2}{\alpha}\frac{\beta_h\beta_B}{|\mathbf{r}_h - \mathbf{r}_B|^7}, \tag{6}$$

where α is the fine structure constant.

Using Eqs. (5) and (6) and taking into account definition of the dielectric constant $\epsilon = 1 + 4\pi\beta n$, we eventually obtain

$$V_{\beta h}(z_B) = \frac{23(\epsilon_h - 1)}{480\pi\alpha}\frac{e^2}{z_B^4}R_*^3(2\epsilon_h + \epsilon_B) + V_{\beta h}(\infty), \tag{7}$$

$$V_{\beta h}(z_B) \approx 0.11\frac{e^2R_*^3}{z_B^4}(2\epsilon_h + \epsilon_B) + V_{\beta h}(\infty), \tag{8}$$



Schematic diagram of an electron–cluster complex at the surface of liquid helium in an external electric field E.

where the relation between R_* and R is determined by (4).

The interaction between the electron and the cluster is also described as a sum of two components:

$$V_{eB}(\rho_e, z_e, z_B) = -\frac{eER_*^3(z_B + \bar{z}_e)}{[(z_B + \bar{z}_e)^2 + \rho_e^2]^{3/2}}$$
$$-\frac{e^2R_*^3}{2[(z_B + \bar{z}_e)^2 + \rho_e^2]}, \tag{9}$$

where ρ_e is the electron coordinate in the horizontal plane. The polarized cluster generates an electric field that can be described for $R \ll z_B$ by the point dipole potential $\Phi(\mathbf{r}) = (\mathbf{d}\cdot\mathbf{r})/\mathbf{r}^3$. This potential restricts the electron motion in the horizontal plane (the first term in (9)). The electron also polarizes the cluster representing a dielectric sphere. In the limit of $r \gg R$, this influence can be described considering the electron field $\mathbf{E} = -e\mathbf{r}/r^3$ as homogeneous. The second term in Eq. (9) describes the additional energy of the dipole in the electron field.

The problem of electron motion in the system under consideration was solved in the adiabatic approximation used previously for electrons at the liquid helium surface in an external electric field [8]. First, let us consider the motion in the z direction perpendicular to the liquid helium surface. This motion is determined by the potential energy component V_e (2). The electron is "pressed" to the surface by the external field and by the image forces, so that the lowest energy level corresponds to an average height of $\bar{z}_e \approx 50$ Å above the surface, which is much smaller as compared to z_B and ρ_e ($\bar{z}_e \ll z_B, \rho_e$). For this reason, in (9), we may neglect $\bar{z}_e$ in the sum $z_B + \bar{z}_e$. Assuming also that $\rho_e \ll z_B$, we may

expand the energy (9) into a series with respect to a small parameter $(\rho_e/z_B)^2$:

$$V_{eB}(\rho_e, z_e, z_B) = -\frac{eER_*^3}{z_B^2}\left[1-\frac{3}{2}\frac{\rho_e^2}{z_B^2}\right]-\frac{e^2R_*^3}{2z_B^4}\left[1-2\frac{\rho_e^2}{z_B^2}\right],$$

$$V_{eB}(\rho_e, z_e, z_B) = -\left[\frac{eER_*^3}{z_B^2}+\frac{e^2R_*^3}{2z_B^4}\right] \quad (10)$$

$$+\frac{\rho_e^2}{z_B^2}\left[\frac{3eER_*^3}{2z_B^2}+\frac{e^2R_*^3}{z_B^4}\right].$$

The second term in (10) is the potential of a harmonic oscillator with the zero-order frequency

$$\omega_0^2 = \frac{2}{mz_B^2}\left[\frac{3eER_*^3}{2z_B^2}+\frac{e^2R_*^3}{z_B^4}\right]. \quad (11)$$

A similar expansion was used for determining the energy levels of an electron at the liquid helium surface in the potential field of a positive ion situated below the surface [11] or a positively charged surface impurity at the substrate supporting the liquid helium film [12, 13].

Expansion (10) can only be performed provided that the zero-order oscillation energy is small as compared to the electron–cluster binding energy:

$$\hbar\omega \ll U_0, \quad U_0 = \frac{eER_*^3}{z_B^2}+\frac{e^2R_*^3}{2z_B^4}. \quad (12)$$

For a cluster with the effective size $R_* = 500$ Å, situated at a distance on the order of $z_B = 10^4$ Å below the liquid helium surface, condition (12) will be fulfilled in external electric fields of $E \gg 300$ V/cm.

Now let us estimate the critical distance z_{B_0} corresponding to the minimum energy of the system under consideration. With neglect of the zero-order oscillations and the gravity, this energy is expressed as

$$V(z_B) \approx \frac{(E_{\text{tot}}R_*^3)}{8z_B^3}\frac{\epsilon_h-1}{\epsilon_h+1}-\frac{eER_*^3}{z_B^2}$$
$$+\frac{e^2R_*^3}{z_B^4}[0.11(2\epsilon_h+\epsilon_B)-1/2], \quad (13)$$

where $E_{\text{tot}} = E + e/z_B^2$. For an ice cluster, $\epsilon_B = 80$ and the expression in square brackets is positive. Since $(\epsilon_h - 1)/(\epsilon_h + 1) \approx 1/300$, we may ignore the first term on the right-hand side of (13). From the condition of minimum for $V(z_B)$, we obtain the following relation between z_{B_0} and the external electric field strength (in e.s.u.):

$$z_{B_0} = \frac{10^{-4}}{\sqrt{E}}. \quad (14)$$

For a cluster with $R_* = 500$ Å, the condition $z_{B_0} \gg R$ is valid for $E \ll 30000$ V/cm, that is, in the entire range of field strengths used in experiments. For such a cluster, the energy of binding to an electron localized at the liquid helium surface in a field of $E = 3000$ V/cm is $U_0 \approx 6 \times 10^{-17}$ erg ≈ 0.5 K. It can be shown that, for a cluster of smaller effective size $R_* = 100$ Å, the interval of electric field strengths featuring stable charged macroscopic clusters in liquid helium is also sufficiently wide. In a field of $E = 30000$ V/cm, we have $z_{B_0} \sim 10^{-5}$ cm and a binding energy of $U_0 \approx 5$ K.

It would be interesting to determine the critical concentration of such complexes at the liquid helium surface. As is known, the maximum surface density of free electrons over a thick liquid helium film is restricted to 10^9 cm^{-2}. Above this level, the charged surface becomes unstable [14]. This instability is related to the ability of electrons to move freely over the surface. In a strong electric field "pressing" electrons to the surface, it is more energetically favorable for electrons to group over a scale on the order of the instability wavelength and to sink down in the form of many-electron droplets [8].

The system of charged complexes described above possesses a much lower mobility that depends on the cluster size and the "pressing" field strength. In addition, the external electric field can be reduced as compared to the case of free electrons because each electron is additionally pressed to the liquid helium surface by the force of polarization attraction to the bound cluster. However, if the field E is much smaller than $4\pi\sigma$, where σ is the surface density of electrons, these electrons will occur in a quasistationary state and move upward from the liquid helium surface. Under these conditions, the mechanism leading to the development of instability on a charged liquid helium surface may become inoperative (or suppressed) in the system of charged complexes. At the surface of such a liquid helium film, the critical electron density is much greater than that related to renormalization of the effective acceleration of gravity g due to the van der Waals attraction of atoms to the substrate [15]. The presence of a macroscopic cluster below the surface may lead to an additional gravity renormalization effect.

Using the proposed system of charged complexes, it is possible to study the Wigner crystallization of electrons at the liquid helium surface by tracing the ordering of clusters below the surface, which can be monitored (for sufficiently large clusters) by optical methods.

ACKNOWLEDGMENTS

The authors are grateful to V. B. Shikin for fruitful discussions.

This study was supported by the Russian Foundation for Basic Research (project no. 03-02-16121), the INTAS Foundation (grant no. 01-0791), and the European Community's Human Potential Program (grant no. HPRN-CT-2000-00157, Surface Electrons).

REFERENCES

1. E. B. Gordon, L. P. Mezhov-Deglin, and O. F. Pugachev, Pis'ma Zh. Éksp. Teor. Fiz. **19**, 103 (1974) [JETP Lett. **19**, 63 (1974)].
2. L. P. Mezhov-Deglin and A. M. Kokotin, J. Low Temp. Phys. **119**, 385 (2000).
3. L. P. Mezhov-Deglin and A. M. Kokotin, Pis'ma Zh. Éksp. Teor. Fiz. **70**, 744 (1999) [JETP Lett. **70**, 756 (1999)].
4. L. P. Mezhov-Deglin and A. M. Kokotin, Fiz. Nizk. Temp. **28**, 235 (2002) [Low Temp. Phys. **28**, 165 (2002)].
5. A. M. Dyugaev, P. D. Grigoriev, and P. Wyder, Phys. Status Solidi B **237**, 260 (2003).
6. M. I. Khaĭkin, Pis'ma Zh. Éksp. Teor. Fiz. **27**, 706 (1978) [JETP Lett. **27**, 668 (1978)].
7. P. D. Grigor'ev and A. M. Dyugaev, Zh. Éksp. Teor. Fiz. **115**, 593 (1999) [JETP **88**, 325 (1999)].
8. V. B. Shikin and Yu. P. Monarkha, *Two-Dimensional Charged Systems in Helium* (Nauka, Moscow, 1989).
9. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media*, 4th ed. (Nauka, Moscow, 2001; Pergamon Press, Oxford, 1984).
10. L. D. Landau, E. M. Lifshitz and L. P. Pitaevskiĭ, *Course of Theoretical Physics*, Vol. 5: *Statistical Physics* (Nauka, Moscow, 1978; Pergamon Press, New York, 1980), Part 2.
11. Yu. P. Monarkha and Yu. Z. Kovdrya, Fiz. Nizk. Temp. **8** (2), 215 (1982) [Sov. J. Low Temp. Phys. **8**, 107 (1982)].
12. G. A. Farias and F. M. Peeters, Phys. Rev. B **55**, 3763 (1997).
13. P. D. Grigor'ev, Pis'ma Zh. Éksp. Teor. Fiz. **66** (9), 599 (1997) [JETP Lett. **66**, 630 (1997)].
14. L. P. Gor'kov and D. M. Chernikova, Pis'ma Zh. Éksp. Teor. Fiz. **18** (2), 119 (1973) [JETP Lett. **18**, 68 (1973)].
15. V. B. Shikin and Yu. P. Monarkha, Fiz. Nizk. Temp. **1**, 957 (1975) [Sov. J. Low Temp. Phys. **1**, 459 (1975)].

Translated by P. Pozdeev