Journal of Experimental and Theoretical Physics, Vol. 93, No. 5, 2001, pp. 1004–1016.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 5, 2001, pp. 1157–1171.

ATOMS, SPECTRA, RADIATION

Optical Bistability¶

Yu. N. Ovchinnikov[a,*] and I. M. Sigal[b,]**

[a]*Laudau Institute for Theoretical Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia*
[b]*Department of Mathematics, University of Toronto*
**e-mail: ovc@itp.ac.ru*
***e-mail: sigal@math.toronto.edu*
Received May 23, 2001

Abstract—We consider the problem of the wave propagation through a nonlinear medium. We derive a dynamic system that governs the behavior of standing (or solitary) waves. The form of this system alone suffices to understand the qualitative dependence of solutions of the original equation on the intensity of the incident wave. We solve this dynamical system in the leading order in the nonlinearity strength. We find multiple solutions of the original problem for a given incoming wave and turning points of these solutions as a function of the wave intensity. We briefly investigate stability of different branches. Our results yield an analytic description of the optical bistability phenomenon. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The problem of light propagation through nonlinear media is of a great theoretical and practical interest. At large intensities, the dielectric constant ε is not a constant but varies as the intensity does. Even a tiny dependence of ε on the intensity can produce significant effects over large distances. Controlling and utilizing these effects is one of the main challenges of the theory of wave propagation. In this paper, we consider the propagation of light through a slab of medium whose dielectric constant depends on the intensity of light. A key effect of interest here is the bistability phenomenon—existence of several solutions (with different transmission coefficients) with alternating stability properties for a given intensity of the incoming beam. This phenomenon was predicted about 20 years ago in [1] and it has been a subject of intensive research since then. The research in this area further intensified about 10 years ago with the theoretical discovery of gap solitons in [2]. See [2–10] for some of the important original works and [11–13] for recent reviews and the background material.

In this paper, we address the problem of the propagation of electromagnetic waves through a nonlinear dialectic slab in a systematic way. To keep the exposition as simple as possible, we consider the simplest possible dependence of the slab dialectic constant on the intensity of light. Our goal is to clarify some conceptual points and to perform concrete computations. Specifically, we (i) establish a minimum action principle and consequently a Hamiltonian structure for the basic (phenomenological) equation; (ii) find a criterion of bistability in terms of linear resonances, which offers a possibility for multidimensional extensions; (iii) find the location of turning points; (iv) estimate the number of solutions for incoming waves of high intensities; and (v) discuss general features of the stability analysis.

To our knowledge, the results summarized above are new.

2. THE MODEL

In the local and nondissipative approximation, the equation describing the propagation of light through a medium with the dielectric constant ε and without charges or currents is given by (see, e.g., [14–18])

$$\partial_t^2(\varepsilon\mathbf{E}) = \Delta\mathbf{E}, \tag{2.1}$$

where $\mathbf{E}(\mathbf{x}, t)$ is the electric field at a point $\mathbf{x} \in \mathbb{R}^3$ at time t: the speed of light c is set to 1. This equation arises from the principle of minimum (or more precisely, stationary) action. We write the action for the electromagnetic field in a medium whose dielectric constant ε depends on the amplitude of the electric field $\mathbf{E}$ (i.e., $\varepsilon = \varepsilon(|\mathbf{E}|^2)$) as

$$S(\mathbf{A}) = \frac{1}{2}\int\int (f(|\mathbf{E}|^2, x) - |\mathbf{B}|^2),$$

where $\mathbf{A}$ is the transverse vector potential, $\mathrm{div}\mathbf{A} = 0$ (we work in the Coulomb gauge), $\mathbf{E} = -\partial\mathbf{A}/\partial t$, $\mathbf{B} = \mathrm{curl}\,\mathbf{A}$ (magnetic field), and

$$f(s, \mathbf{x}) = \int_0^s \varepsilon(u, \mathbf{x})du.$$

Moreover, we set the magnetic permeability μ to 1. Here, we modified only the part of the action related to the electric field $\mathbf{E}$, leaving the part connected to the magnetic field $\mathbf{B}$ unchanged. The reason for this is that

¶This article was submitted by the authors in English.

the electric susceptibility $\chi_e = \varepsilon - 1$ can take relatively large values, even much larger than 1, while the magnetic susceptibility $\chi_m = \mu - 1$ is always much smaller than 1 in nonmagnetic materials, namely of the order 10^{-5}–10^{-8}.

The critical points of the above functional are given by the Euler–Lagrange equation

$$-\frac{\partial}{\partial t}\left(\varepsilon(|\mathbf{E}|^2, \mathbf{x})\frac{\partial \mathbf{A}}{\partial t}\right) + \Delta \mathbf{A} = 0. \tag{2.2}$$

Differentiating this equation with respect to t and utilizing that $\partial \mathbf{A}/\partial t = -\mathbf{E}$, we arrive at (2.1). Conversely, Eq. (2.1) implies Eq. (2.2) if we require that

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T \mathbf{A}\,dt = 0;$$

i.e., the vector potential $\mathbf{A}$ has no zero harmonic. The latter is consistent with Eq. (2.2) because that equation contains only odd powers of $\mathbf{A}$.

A reformulation of Eq. (2.1) (or (2.2)) in terms of the minimum action principle immediately leads to energy conservation with the energy functional

$$\mathscr{E}(\mathbf{A}) = \int (\mathbf{A}\partial_{\mathbf{A}} L) - L,$$

where

$$L = \frac{1}{2}\int (f(|\mathbf{A}|^2, x) - |\mathrm{curl}\mathbf{A}|^2).$$

This functional can be explicitly computed as

$$\mathscr{E}(\mathbf{A}) = \frac{1}{2}\int \{f(|\mathbf{E}|^2, \mathbf{x}) + |\mathrm{curl}\mathbf{A}|^2\}.$$

Moreover, the variational formulation given above shows that Eq. (2.2) is Hamiltonian, with the standard Poisson brackets and with the Hamiltonian functional found via the Legendre transform as

$$H(\mathbf{A}, \boldsymbol{\pi}) = \frac{1}{2}\int (\phi(|\mathbf{E}|^2) + |\mathrm{curl}\mathbf{A}|^2),$$

where the momentum field $\boldsymbol{\pi}(\mathbf{x})$ is related to the electric field $\mathbf{E}(\mathbf{x})$ as $\boldsymbol{\pi} = -f'(|\mathbf{E}|^2)\mathbf{E}$ and we set $\phi(s) = f'(s)s - f(s)/2$.

In what follows, we consider the simplest model of the nonlinear wave propagation. We assume that

(a) the medium in question is uniform in the y and z directions, i.e., ε does not explicitly depend on y and z;

(b) apart from its dependence on x, ε depends on $\mathbf{E}$ only through the amplitude $|\mathbf{E}|^2$:

$$\varepsilon = \varepsilon(|\mathbf{E}|^2, x);$$

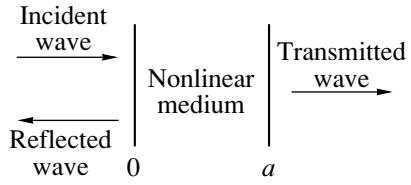


Fig. 1.

(c) the nonlinear part of the medium forms a slab of the thickness a perpendicular to the x axis (see Fig. 1):

$$\varepsilon(|\mathbf{E}|^2, x) = \begin{cases} 1 & \text{if } x < 0 \text{ or } x > a \\ n^2\bar{\varepsilon}(|\mathbf{E}|^2) & \text{if } 0 \le x \le a, \end{cases} \tag{2.3}$$

where n is the refractive index. The function $\bar{\varepsilon}(|\mathbf{E}|^2)$ is taken in the simplest possible form

$$\bar{\varepsilon}(|\mathbf{E}|^2) = 1 + \bar{\mu}|\mathbf{E}|^2. \tag{2.4}$$

In real materials, $\bar{\mu} \sim |\mathbf{E}_0|^{-2}$, where $\mathbf{E}_0$ is the internal (atomic) electric field. Because the electric breakdown already occurs when $|\mathbf{E}| \ll |\mathbf{E}_0|$ the second term in the right-hand side of (2.4), which is of the order $(|\mathbf{E}|^2/|\mathbf{E}_0|)^2$, is indeed very small,

$$\bar{\mu}|\mathbf{E}|^2 \ll 1. \tag{2.5}$$

We consider only waves of a fixed polarization, i.e., assume that $\mathbf{E}(\mathbf{x}, t)$ in Eq. (2.1) can be written as

$$\mathbf{E}(\mathbf{x}, t) = E(\mathbf{x}, t)\mathbf{e}, \tag{2.6}$$

where $\mathbf{e}$ is a fixed vector (the polarization vector) propagating in the x direction; i.e., $\mathbf{e}$ is perpendicular to the x axis. In this case, $\mathbf{E}(\mathbf{x}, t)$ can be assumed to depend on x only, and therefore, Eq. (2.1) reduces to the equation

$$\frac{\partial^2 E}{\partial x^2} = \frac{\partial^2}{\partial t^2}(\varepsilon E), \tag{2.7}$$

where $E = E(x, t)$ and $\varepsilon = \varepsilon(|E|^2, x)$.

This equation is subject to the boundary conditions

$$E(x, t) = \begin{cases} f(x-t) + f_1(x+t) & \text{if } x < 0 \\ f_2(x-t) & \text{if } x > a, \end{cases} \tag{2.8}$$

with the function f given and the functions f_1 and f_2 unknown. These boundary conditions say that the field on the left of the sample consists of the incoming wave $f(x-t)$ and some reflected wave $f_1(x+t)$, while the field on the right of the sample consists of only the outgoing wave $f_2(x-t)$. In addition, we specify the incoming wave as

$$f(x-t) = \mathrm{Re}(e^{-ik_0(x-t)}) \tag{2.9}$$

for some $k_0 > 0$.

3. SOLITARY WAVES

We study solitary waves for Eq. (2.7), i.e., waves of the form

$$E(x, t) = \mathrm{Re}(e^{-ik_0 t}\psi_0(x)), \tag{3.1}$$

where ψ_0 is a complex function. In the leading approximation in the nonlinearity parameter $\bar{\mu}$, it then follows that ψ_0 satisfies the stationary equation (see Appendix 1)

$$\Delta\psi_0 + k_0^2\varepsilon_0\psi_0 = 0, \tag{3.2}$$

with $\varepsilon_0 = \varepsilon((3/4)|\psi_0|^2, x)$. We are interested in the problem of the solitary-wave passage through the nonlinear slab, which amounts to taking the solutions ψ_0 such that

$$\psi_0 = \begin{cases} Ae^{ik_0 x} + RAe^{-ik_0 x} & \text{for } x < 0 \\ TAe^{ik_0 x} & \text{for } x > a, \end{cases} \tag{3.3}$$

with a given A and for some R and T. Here $Ae^{ik_0 x}$, $RAe^{-ik_0 x}$, and $TAe^{ik_0 x}$ are the incident, reflected, and transmitted waves, respectively (see Fig. 2); and R and T are the reflection and transmission coefficients.

The flux conservation (see below) implies that R and T satisfy

$$|R|^2 + |T|^2 = 1. \tag{3.4}$$

In the linear case, R and T are independent of A, and the amplitude A drops out of the equation. This is not so in the nonlinear case. The goal of this paper is to find the dependence of $|R|$ (or $|T|$) on A. The main point here is that although two initial conditions uniquely define a solution of a second-order ODE, two boundary conditions can be satisfied by several (a finite number of) solutions of a nonlinear second-order ODE. Figure 3 shows two solutions satisfying the boundary conditions $\psi(0) = 0$ and $\psi(a) = 0$. In constrast, in the linear case, two boundary conditions determine a solution of a second-order ODE uniquely (modulo eigen-functions).

Fig. 2. Reflection and transmission coefficients.

Fig. 3. Two solutions satisfying boundary conditions $\psi(0) = 0$ and $\psi(a) = 0$.

4. THE BOUNDARY VALUE PROBLEM

Instead of considering Eq. (3.2) and conditions (3.3) on the entire real axis, we study this problem on the interval $[0, a]$,

$$\psi_0'' + k_0^2\bar{\varepsilon}\left(\frac{3}{4}|\psi_0|^2\right)\psi_0 = 0, \quad 0 \le x \le a, \tag{4.1}$$

and use conditions (3.3) to set the boundary conditions at $x = 0$ and $x = a$ as

$$\psi_0(0) = A(1 + R), \quad \psi_0'(0) = ik_0A(1 - R) \tag{4.2}$$

and

$$\psi_0(a) = ATe^{ik_0 a}, \quad \psi_0'(a) = ik_0ATe^{ik_0 a}. \tag{4.3}$$

We thus arrive at a boundary value problem on $[0, a]$. It is convenient to rescale this problem as

$$\psi_0(x) = A\psi(kx),$$

where $k = k_0 n$ the wave vector in the medium with the refraction coefficient n. The new boundary value problem is given by

$$\psi'' + \varepsilon_1(|\psi|^2)\psi = 0, \quad 0 \le x \le b, \tag{4.4}$$

where $b = ka$ and $\varepsilon_1(|\psi|^2) = \bar{\varepsilon}((3/4)A^2|\psi|^2)$, with the boundary conditions

$$\psi(0) = 1 + R, \quad \psi'(0) = \frac{i}{n}(1 - R) \tag{4.5}$$

and

$$\psi(b) = Te^{ib/n}, \quad \psi'(b) = \frac{i}{n}Te^{ib/n}. \tag{4.6}$$

Because Eq. (4.4) is invariant under the gauge transformation $\psi(x) \longrightarrow e^{i\beta}\psi(x)$, we can assume $T \ge 0$.

Recalling expression (2.4) for $\bar{\varepsilon}$, we find

$$\varepsilon_1(|\psi|^2) = 1 + 2\mu|\psi|^2, \quad \mu = 3\bar{\mu}|A|^2/8, \tag{4.7}$$

and therefore, the incident beam amplitude A enters the new equation only through the parameter μ and varying A is the same as varying μ.

We note that Eq. (2.5) implies that in real material, $\mu \ll 1$.

Although Eqs. (4.5)–(4.6) appear to represent four (complex) constraints, these equations in fact constitute only two conditions because R and T are unknown. Eliminating the unknowns R and T from boundary conditions (4.5)–(4.6), we obtain the conditions

$$\psi(0) - in\psi'(0) = 2, \tag{4.8}$$

$$\psi(b) + in\psi'(b) = 0. \tag{4.9}$$

Equation (4.6) shows that a solution of Eq. (4.4) with boundary conditions (4.8), (4.9) determines the transmission coefficient $T = |\psi(b)|$; on the other hand, knowing T determines the solution of (4.4). Our goal in

what follows is to find $T = |\psi(b)|$, where ψ solves (4.4), (4.8), and (4.9) as a function of μ.

5. RESONANCES AND THE EFFECTIVE WAVE VECTOR

We now describe the physical mechanism underlying the nonlinear phenomenon under consideration. We begin with the linear component of this mechanism, and therefore set $\mu = 0$. In this case, Eq. (4.4) can be solved explicitly with the result

$$\psi^{\text{lin}} = \frac{1}{2}\left(1 + R^{\text{lin}} + \frac{1 - R^{\text{lin}}}{n}\right)e^{ix} + \frac{1}{2}\left(1 + R^{\text{lin}} - \frac{1 - R^{\text{lin}}}{n}\right)e^{-ix} \tag{5.1}$$

and

$$R^{\text{lin}} = \frac{(n^2 - 1)(e^{ib} - e^{-ib})}{-(n-1)^2 e^{ib} + (n+1)^2 e^{-ib}}. \tag{5.2}$$

The last equation shows that as a function of $b = ka$, R^{lin} has a series of minima and maxima,

$$b = ka = \pi m \longrightarrow |R^{\text{lin}}| = 0 \quad (= |R^{\text{lin}}|_{\min}),$$

$$b = ka = \pi\left(m + \frac{1}{2}\right) \longrightarrow |R^{\text{lin}}| = \frac{n^2 - 1}{n^2 + 1} \quad (= |R^{\text{lin}}|_{\max}).$$

For $n \gg 1$, this resonance behavior is rather sharp: if the width a of the slab contains an integer number of the half-wave lengths, $\lambda/2 = 2\pi/2k$, then the transmission is perfect and the slab is therefore transparent. If the width of the slab contains an odd number of quarter-wave lengths, then there is almost no transmission and the slab is opaque.

The resonance structure of the linear case plays a crucial role in the peculiar behavior of the nonlinear solution. This solution can be considered as a linear one with a varying effective wave vector,

$$k_{\text{eff}} \equiv k\varepsilon_1^{1/2} = k(1 + 2\mu|\psi|^2)^{1/2}. \tag{5.3}$$

As the intensity (i.e., μ) varies, so does the effective wave-length and the medium goes through a series of resonances in which it is either perfectly transparent, $|T| = 1$, or almost opaque, $|T| \approx 0$.

Thus, the presence of sharp minima and maxima of the reflection (or transmission) coefficient offers a simple criterion for the occurrence of the bistability phenomenon. One way to extend this criterion to the multidimensional case is to relate it to the resonance structure of the scattering process considered above. Indeed, we observe that R^{lin} (and therefore, ψ^{lin}) display a resonance structure in the sense that it has complex poles at

$$b(=ka) = \pi m - i\ln\frac{1 + 1/n}{1 - 1/n}, \quad m = 0, \pm 1, \ldots. \tag{5.4}$$

The real parts of these poles exactly give the position of maxima of the transmission coefficient. If we recall that $b = ka$, we can rewrite (5.4) as

$$k = \frac{\pi m}{a} - \frac{i}{a}\ln\frac{1 + 1/n}{1 - 1/n}, \quad m = 0, \pm 1, \ldots. \tag{5.5}$$

The real part of this expression takes the values $2\pi m/a$ that coincide with the eigenvalues of the operator $\sqrt{-d^2/dx^2}$ on the interval $[0, a]$ with the periodic boundary conditions.

To obtain resonance solutions, we must solve the original wave equations

$$\frac{\partial^2 E}{\partial x^2} = \frac{\partial^2}{\partial t^2}(\varepsilon E) \tag{5.6}$$

(where $\varepsilon = n^2$) on the interval $[0, a]$ with the boundary conditions representing an outgoing wave. For this, we set $E = e^{-i\sigma t}\xi$ with $\sigma > 0$ and

$$\xi = \begin{cases} A_1 e^{-i\sigma x} & \text{for } x < 0 \\ A_2 e^{i\sigma x} & \text{for } x > a, \end{cases} \tag{5.7}$$

where A_1 and A_2 are arbitrary constants. Eliminating these constants, we obtain

$$\frac{\xi'(0)}{\xi(0)} = -i\sigma \text{ and } \frac{\xi'(a)}{\xi(a)} = i\sigma. \tag{5.8}$$

On the other hand, Eq. (5.6) implies the equation for ξ,

$$-\varepsilon\sigma^2\xi = \frac{\partial^2}{\partial x^2}\xi, \tag{5.9}$$

on the interval $[0, a]$. Solving Eqs. (5.8), (5.9) and recalling that $\varepsilon = n^2(>1)$, we find

$$\sigma = \frac{\pi m}{an} - \frac{i}{an}\ln\frac{1 + 1/n}{1 - 1/n}, \quad m = 0, \pm 1, \ldots \tag{5.10}$$

and the corresponding expression for ξ which we omit here.

We have thus arrived at the following conclusion: resonances of the transmission coefficient, which are responsible for the bistable behavior of our nonlinear system, coincide with the resonances of the linear wave equation (5.6) (in (5.6), ε is n^2 times the characteristic function of the interval $[0, a]$). This is important because there are well developed techniques for finding resonances in multidimensional linear systems. Thus, we have a possibility of identifying the bistability phenomenon in the multidimensional case.

We indicate the connection between the above resonance solutions and the stability problem for Eqs. (2.7)–(2.9) with $\mu = 0$ (the linear problem). In this case, we seek solutions to Eqs. (2.7)–(2.9) in the form

$$E = \text{Re}(e^{-ik_0 t}\psi_0(x) + \eta(x, t)), \tag{5.11}$$

where $|\eta| \ll |\psi_0|$. From (2.8), (2.9) and (3.3), we can assume that η satisfies the boundary conditions

$$\eta = \begin{cases} e^{-i\sigma(x+t)} & \text{for } x < 0, \\ e^{i\sigma(x-t)} & \text{for } x > a, \end{cases} \tag{5.12}$$

with σ that is complex but close to k_0. Clearly, η is of the form $\eta = e^{-i\sigma t}\xi$, where ξ satisfies Eqs. (5.8) and (5.9), and consequently, σ is given in (5.10). The resonance eigenvalues therefore serve as the stability exponents for solution (3.1) in the linear case (because $\mathrm{Im}\sigma < 0$, the solution $e^{-ik_0 t}\psi_0(x)$ is stable).

6. CONSERVATION LAWS

In this section, we describe conservation laws obeyed by Eq. (4.4). We consider x as a time variable. We first define the "energy" density

$$e(\psi) = |\psi'|^2 + G(|\psi|^2), \tag{6.1}$$

where

$$G(u) = \int_0^u \varepsilon_1(v)dv = u + \mu u^2. \tag{6.2}$$

Using Eq. (4.4), we conclude that $\partial e(\psi)/\partial x = 0$, and, therefore,

$$e(\psi) \equiv |\psi'|^2 + |\psi|^2 + \mu|\psi|^4 = C > 0. \tag{6.3}$$

In the same way, it follows that the flux density $j = \mathrm{Im}(\bar{\psi}\,\partial\psi/\partial x)$ is also conserved, $j = C_1$. To combine these two conservation laws, it is convenient to pass to the polar representation

$$\psi = \sqrt{\rho}e^{i\alpha}. \tag{6.4}$$

The conservation of the flux the gives

$$\rho\alpha' = C_1. \tag{6.5}$$

In classical mechanics, this equation expresses the angular momentum conservation or the Kepler law: the rate of change of the area swept by the radius vector of a particle in a central potential is constant. Together with the energy conservation equation, this equation gives

$$\rho'^2 = 4(-C_1^2 + C\rho - g(\rho)), \tag{6.6}$$

where $g(\rho) = \rho G(\rho) = \rho^2 + \mu\rho^3$. Starting with this equation and boundary conditions (4.8), (4.9), we derive our main equations in the next section.

7. THE MAIN EQUATIONS

In this section, we derive the equations for ρ and $|T|^2 = \rho(b)$ on which we base our analysis. We observe that the right-hand side of Eq. (6.6) contains two integration constants (or conservation constants). We use boundary conditions (4.8) and (4.9) to express these constants in terms of $\rho(0)$ and $\rho(b)$.

The boundary condition at $x = b$ gives

$$\left.\frac{\frac{1}{2}\frac{\rho'}{\sqrt{\rho}} + i\alpha'\sqrt{\rho}}{\sqrt{\rho}}\right|_{x=b} = \frac{i}{n}, \tag{7.1}$$

which implies

$$\alpha'(b) = \frac{1}{n} \text{ and } \rho'(b) = 0. \tag{7.2}$$

Equation (6.5) and the first equation in (7.2) yield $C_1 = \rho(b)/n$, and therefore,

$$\alpha' = \frac{\rho(b)}{n\rho}. \tag{7.3}$$

Equations (6.6) and $C_1 = \rho(b)/n$ give

$$\rho'^2 = f(\rho), \tag{7.4}$$

where

$$f(\rho) = 4\left(-\frac{\rho(b)^2}{n^2} + C\rho - g(\rho)\right). \tag{7.5}$$

Equations (7.2) and (7.4), (7.5) imply that $\rho(b) = |T|^2$ is a root of

$$f_1(\rho(b)) = 0, \tag{7.6}$$

where

$$f_1(u) = Cu - \frac{1}{n^2}u^2 - g(u). \tag{7.7}$$

Equations (7.4)–(7.7) constitute all the basic equations of our analysis except one equation. We now find the constant C as a function of $\rho(b)$ by solving Eqs. (7.6), (7.7) for C,

$$C = \left(1 + \frac{1}{n^2}\right)\rho(b) + \mu\rho(b)^2. \tag{7.8}$$

This is a quadratic relation between $\rho(0)$ and $\rho(b)$. Substituting Eq. (7.8) in Eq. (7.5), we find

$$\begin{aligned} f(\rho) = 4\Bigg(&-\frac{\rho(b)^2}{n^2} \\ &+ \rho(b)\left(1 + \frac{1}{n^2}\right)\rho + \mu\rho(b)^2\rho - \rho^2 - \mu\rho^3\Bigg). \end{aligned} \tag{7.9}$$

Finally, we find the remaining basic equation using the boundary condition at $x = 0$. Using Eqs. (6.4), (7.3), and (7.4), we rewrite Eq. (4.9) as

$$\left(\rho(0) \mp \frac{in}{2}\sqrt{f(\rho(0))} + \rho(b)\right)\frac{e^{i\alpha(0)}}{\sqrt{\rho(0)}} = 2, \tag{7.10}$$

which implies, after taking the absolute value,

$$(\rho(0) + \rho(b))^2 + \frac{1}{4}n^2 f(\rho(0)) = 4\rho(0). \quad (7.11)$$

Inserting expression (7.9) for $f(\rho)$ in (7.11), we obtain, after simple transformations,

$$\varphi(\rho(b), \rho(0)) = 0, \quad (7.12)$$

where

$$\varphi(u, v) = \mu n^2 v^2 + (n^2 - 1)v - (n^2 + 3)u - n^2\mu u^2 + 4. \quad (7.13)$$

We now can formulate the problem as follows. We must solve the differential equation $\rho'^2 = f(\rho)$ with the boundary values $\rho(0)$ and $\rho(b)$ satisfying the equation $\varphi(\rho(b), \rho(0)) = 0$, where f and φ are given by the respective equations (7.9) and (7.13).

We split our task as follows:

(i) Using Eqs. (7.4) and (7.9), we first determine $\rho(0)$ as a function of $\rho(b)$. Here, we consider $\rho(b)$ as an initial condition for the dynamic system

$$\rho' = \pm\sqrt{f(\rho)_+}, \quad (7.14)$$

where $x_+ = \max(x, 0)$. We then solve (7.14) backwards, from $x = b$ to $x = 0$ (with the change of signs at turning points, $f(\rho) = 0$!) and find $\rho(0)$ as a function of $\rho(b)$, n, and μ. Here, the cases where $\rho'(0) > 0$ and $\rho'(0) < 0$ can be considered separately.

(ii) We then insert $\rho(0)$ found in step (i) in Eq. (7.12). The result is an algebraic equation for $\rho(b) = |T|^2$. In general, this algebraic equation has several solutions depending on μ and n.

The essence of this analysis can be inferred from the form of Eqs. (7.9) and (7.14), without solving them. Indeed, let ρ_1, ρ_2, and ρ_3 be the roots of the equation $f(\rho) = 0$. As we already know, one of these roots is $\rho(b)$, for example $\rho_1 = \rho(b)$. We let $\rho_3 < \rho_2$. It is not difficult to show (see below) that $\rho_2 < \rho_1$ (in fact, $\rho_2 < (1 + \mu\rho_1)^{-1}n^{-2}\rho_1$) and $\rho_3 < 0$. The roots of $f(\rho)$ are equilibria and turning points of the dynamic system in Eq. (7.14). At these roots, $\rho' = 0$ and $\sqrt{f(\rho)_+}$ changes its sign. The behavior of this dynamic system is shown in Fig. 4.

Open problem. Dynamic system (7.14) is parameterized by ρ_1. Its phase portrait should qualitatively change as ρ_1 goes through a turning point. How?

8. EQUATION (7.6)

We now show that we can infer much information from Eq. (7.6) without solving differential equation (7.4). We consider the two simplest cases:

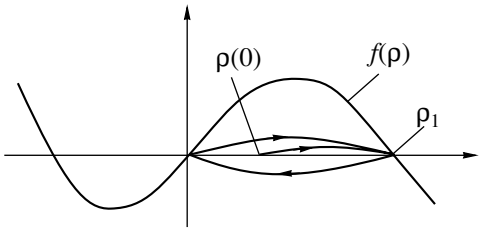


Fig. 4. A trajectory of $\rho' = \pm\sqrt{f(\rho)}$.

(a) Linear case: $G(u) = u$, and hence, $g(u) = u^2$. In this case,

$$f_1^{(\mathrm{lin})}(u) = Cu - \left(1 + \frac{1}{n^2}\right)u^2,$$

and therefore, the equation $f_1^{(\mathrm{lin})}(\rho(b)) = 0$ has, in addition to the trivial solution $\rho(b) = 0$, one nontrivial solution

$$\rho(b) = \frac{C}{1 + n^{-2}}.$$

(b) Cubic nonlinearity: $G(u) = u + \mu u^2$, and therefore, $g(u) = u^2 + \mu u^3$. In this case,

$$f_1(u) = Cu - \left(1 + \frac{1}{n^2}\right)u^2 - \mu u^3.$$

Thus, depending on the coefficients, the equation $f_1(\rho(b)) = 0$ has, in addition to the trivial solution $\rho(b) = 0$, either none or one or two nontrivial solutions. All the possibilities are listed in Fig. 5.

Conclusion. In the linear case, we always have one nontrivial solution (after the division by u, the function $f_1^{(\mathrm{lin})}(u)$ becomes linear). In the simplest, cubic nonlinear case, depending on C and μ, there can be none, one or two nontrivial solutions for $\rho(b) = |T|^2$.

9. SOLUTION OF THE NONLINEAR PROBLEM

We first solve differential equation (7.4), $\rho' = \pm\sqrt{f(\rho)_+}$. Recalling that $\rho(b)$ solves $f(\rho) = 0$, we express $f(\rho)$ as

$$f(\rho) = 4(\rho(b) - \rho)\left(\mu\rho^2 + (1 + \mu\rho(b))\rho - \frac{\rho(b)}{n^2}\right). \quad (9.1)$$

Integrating the equation $\rho' = \pm\sqrt{f(\rho)_+}$, we find

$$\pm\int_{\rho_0}^{\rho_1} \frac{d\rho}{\sqrt{f(\rho)}} = b \quad (9.2)$$

(we recall that $\rho_1 = \rho(b)$ and $\rho_0 = \rho(0)$). It is shown in Appendix 2 that Eq. (9.2) is equivalent, modulo $O(\mu)$

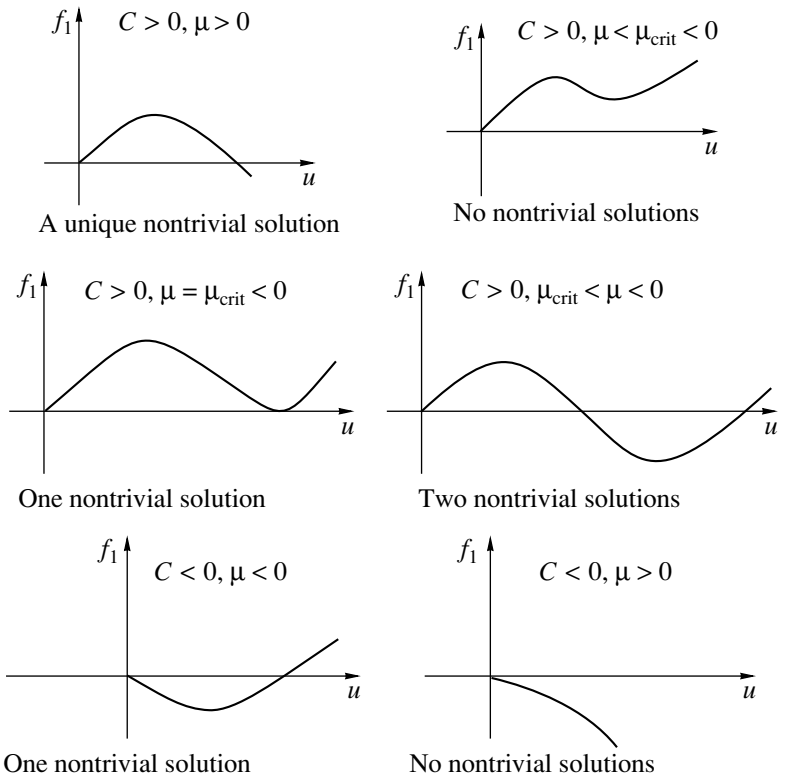


Fig. 5.

(but keeping terms of the order $O(\mu b)$), to the equation

$$\frac{4(1-\rho_1)}{\rho_1}\frac{n^2}{(n^2-1)^2} = \sin^2\frac{b}{1-\frac{3}{4}\mu\rho_1\frac{n^2+1}{n^2}}. \quad (9.3)$$

This equation defines $\rho_1 (=\rho(b) = |T|^2)$ as a multivalued function of μ for given values of the parameters n and b.

The turning points of this function are of primary interest. To find them, we differentiate Eq. (9.3) with respect to ρ_1,

$$-\frac{4}{\rho_1^2}\frac{n^2}{(n^2-1)^2} = \frac{3\mu b}{4}\frac{n^2+1}{n^2}\sin\frac{2b}{1-\frac{3}{4}\mu\rho_1\frac{n^2+1}{n^2}}, \quad (9.4)$$

and then solve the resulting two equations for ρ_1 and μ. Thus, the turning points of ρ_1 as a function of μ are given by solutions of Eqs. (9.3) and (9.4).

10. TURNING POINT IN THE LARGE-n CASE (SEMICLASSICAL LIMIT)

We investigate Eqs. (9.3) and (9.4) in the case where $n \gg 1$. In this case, the factor $\frac{n^2}{(n^2-1)^2}$ in both equations can be replaced by $1/n^2$. We consider two cases:

(a) $|\mu|b \ll 1$. Because $n \gg 1$, Eq. (9.4) shows that either $\rho_1 \sim 1/n^2$ or b is close to πm, where m is an integer. We consider the latter case and set

$$b = \pi m + \delta \text{ with } |\delta| \ll 1. \quad (10.1)$$

Equation (9.3) can then be reduced to the equation

$$4(1-\rho_1) = n^2\rho_1\left(\delta + \frac{3\mu b}{4}\rho_1\right)^2. \quad (10.2)$$

Differentiating this equation with respect to ρ_1, we obtain the equation for the turning points,

$$-4 = \frac{3\mu b n^2\rho_1}{2}\left(\delta + \frac{3}{4}b\mu\rho_1\right) + n^2\left(\delta + \frac{3\mu b}{n}\rho_1\right)^2. \quad (10.3)$$

We now pass from ρ_1 to the variable z defined by

$$n\left(\delta + \frac{3\mu b}{4}\rho_1\right) = -\frac{8}{3z}.$$

The resulting equation for z has the solutions

$$z_{1,2} = \frac{2}{3}(-\delta n \pm \sqrt{\delta^2 n^2 - 12}). \quad (10.4)$$

Substituting in Eqs. (10.2) and (10.3) $-8/3z$ for $n(\delta + 3\mu b\rho_1/4)$ and then solving Eq. (10.2) for ρ_1 and

Eq. (10.3), for μnb we find

$$\rho_1^{(k)} = \frac{9z_k^2}{16 + 9z_k^2} \text{ and } \mu nb = \frac{z_k}{(\rho_1^{(k)})^2} \tag{10.5}$$

for $k = 1, 2$. In the region between these turning points, all the three solutions of Eq. (10.2) can be represented as

$$\rho_1 = -\frac{4\delta(1+z)}{3\pi\mu m} \tag{10.6}$$

and

$$z = -\frac{1}{3} + \sqrt{\frac{4}{9} - \frac{16}{3\delta^2 n^2}}\cos\varphi, \tag{10.7}$$

where $\varphi = \varphi_j$, $j = 1, 2, 3$, is given by

$$\varphi_j = \frac{1}{3}\left(\frac{\pi}{2} + 2\pi j + \arcsin\left\{\frac{1 + \dfrac{36}{\delta^2 n^2} + \dfrac{81b\mu}{2n^2\delta^3}}{\left(1 - \dfrac{12}{\delta^2 n^2}\right)^{3/2}}\right\}\right). \tag{10.8}$$

Because $\rho_1 > 0$, solutions (10.6) exist only in the region $\delta\mu < 0$.

(b) $n|\mu|b \gg 1$. In this case, n can take an arbitrary value larger than one. Eliminating trigonometric functions from Eqs. (9.3) and (9.4), we arrive at the equation

$$\left(\frac{4n^3}{3\mu b(n^2+1)^2}\right)^2 = \rho_1^2(1-\rho_1)\left(\rho_1 - \frac{4n^2}{(n^2+1)^2}\right). \tag{10.9}$$

This equation can have either two positive solutions for ρ_1 or none. In the first case, one of the solutions is close to 1,

$$\rho_1^{(1)} = 1 - \frac{1}{2}\left(\frac{n^2-1}{n^2+1}\right)^2 + \sqrt{\frac{1}{4}\left(\frac{n^2-1}{n^2+1}\right)^4 - \left(\frac{4n^3}{3\mu b(n^2+1)^2}\right)^2}, \tag{10.10}$$

while the other one, $\rho_1^{(2)}$, is a solution of a cubic equation. For $\mu bn \gg 1$ and $n \gg 1$ the cubic equation is given by

$$\rho_1^3 - \frac{4\rho_1^2}{n^2} - \left(\frac{4}{3\mu bn}\right)^2 = 0.$$

A positive solution to the last equation is

$$\rho_1^{(2)} = \frac{8}{3n^2}\left(\frac{1}{2} + \cosh\beta\right), \tag{10.11}$$

where $\cosh 3\beta = 1 + \dfrac{3}{8}\left(\dfrac{n^2}{\mu b}\right)^2$. A graphical solution to Eq. (10.9) is shown in Fig. 6.

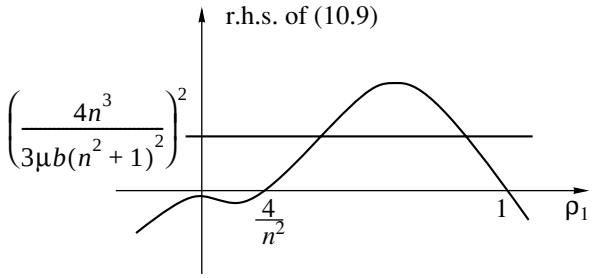


Fig. 6. Graphical solution to Eq. (10.9).

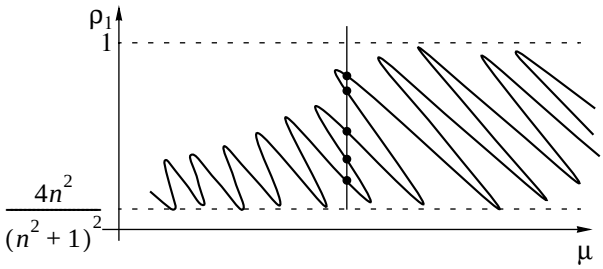


Fig. 7. Dependence of ρ_1 on μ; the number of solutions for a given μ.

Using values (10.10) and (10.11) for ρ_1, we obtain the values for μ from Eq. (9.3) as

$$\mu^{(1)} = \frac{4}{3}\frac{\pi m - b}{b}\frac{n^2}{n^2+1} \tag{10.12}$$

and

$$\mu^{(2)} = \frac{n^2+1}{3}\frac{\pi(m+1/2) - b}{b}, \tag{10.13}$$

provided $|\pi m - b| \ll b$.

Equations (10.10)–(10.13) give the top (Eqs. (10.10) and (10.12)) and bottom (Eqs. (10.11) and (10.13)) turning points. The distance between the neighboring turning points in the first and the second sets is

$$\delta\mu_1 = \frac{4}{3}\frac{\pi}{b}\frac{n^2}{n^2+1} \text{ and } \delta\mu_2 = \frac{n^2+1}{3}\frac{\pi}{b}. \tag{10.14}$$

The dependence of ρ_1 on μ is shown graphically in Fig. 7.

We finally compute the number of solutions for a given μ in the region $|\mu|b \gg 1$. It is given by

$$N(\mu) = 2(N_{\text{top}}(\mu) - N_{\text{bottom}}(\mu)) + 1, \tag{10.15}$$

where $N_{\text{top}}(\mu)$ and $N_{\text{bottom}}(\mu)$ is the number of the top and bottom turning points in the interval $[0, \mu]$ and the coefficient 2 accounts for the fact that there are two solutions corresponding to each turning point (see Fig. 7).

We have roughly $N_{\text{top}}(\mu) \approx \mu/\delta\mu_1$ and $N_{\text{bottom}}(\mu) \approx \mu/\delta\mu_2$, and, therefore,

$$N(\mu) \approx 2\mu\left(\frac{1}{\delta\mu_1} - \frac{1}{\delta\mu_2}\right) \approx \frac{3}{2\pi}\mu b. \tag{10.16}$$

11. STABILITY

In this section, we study the general stability properties of solutions to boundary value problem (4.4)–(4.6). A detailed analysis will be given elsewhere. Clearly, given $\rho_1 \equiv \rho(b) \equiv |\psi(b)|^2 = T^2$, the problem in Eqs. (4.4)–(4.6) has a unique solution. In other words, solutions of the latter problem can be parameterized by ρ_1. (This can be done explicitly by expressing μ in terms of ρ_1.) In what follows, we tacitly assume that the curve (= multivalued function) $\psi = \psi(\mu)$ is parameterized by ρ_1. With this parameterization in mind, we sometimes speak about stability of a point (ρ_1, μ) understanding it as the stability of the corresponding point (ψ, μ).

Our task is to find stability of the solutions of Eqs. (2.7)–(2.9) of form (3.1)–(3.3). To fit this problem into the standard framework, one would have to rewrite (2.7) as a system of the first order Hamiltonian equations and apply to it a rather subtle stability theory for solitary waves (see, e.g., [19] and references therein). We adopt a direct approach instead. We seek solutions of Eqs. (2.7)–(2.9) in the form

$$E = \begin{cases} \mathrm{Re}[e^{-i\omega t}(\psi_0 + e^{\lambda t}\xi)] \quad \text{if } \lambda \text{ is real} & (11.2) \\ \mathrm{Re}[e^{-i\omega t}(\psi_0 + e^{\lambda t}\xi_1 + e^{\bar\lambda t}\xi_2)] \\ \text{if } \lambda \text{ is complex,} & (11.3) \end{cases}$$

where $\omega = k_0$, ψ_0 satisfies Eqs. (3.2), (3.3), and ξ are small and such that

$$\begin{gathered} e^{-i(\omega + i\lambda)t}\xi \quad \text{is an outgoing wave} \\ \text{for } x < 0 \text{ and } x > a \end{gathered} \tag{11.4}$$

and similarly for λ complex. This implies

$$\xi = \begin{cases} A_1 e^{-i(\omega + i\lambda)x} & \text{for } x < 0 \\ A_2 e^{i(\omega + i\lambda)x} & \text{for } x > a \end{cases} \tag{11.5}$$

for some constants A_1 and A_2, which gives

$$\frac{\xi'(0)}{\xi(0)} = -i(\omega + i\lambda) \text{ and } \frac{\xi'(a)}{\xi(a)} = i(\omega + i\lambda) \tag{11.6}$$

for λ real. For λ complex, the boundary conditions are

$$\frac{\xi_1'(0)}{\xi_1(0)} = -i(\omega + i\lambda), \quad \frac{\xi_1'(a)}{\xi_1(a)} = i(\omega + i\lambda) \tag{11.7}$$

and

$$\frac{\xi_2'(0)}{\xi_2(0)} = -i(\omega + i\bar\lambda), \quad \frac{\xi_2'(a)}{\xi_2(a)} = i(\omega + i\bar\lambda). \tag{11.8}$$

For simplicity, we deal only with the case in Eq. (11.2), the case (11.3) is treated in a similar way. Substituting (11.2) in (2.7), we derive the linearized equation for ξ (see Appendix 1 for a similar derivation),

$$L_\lambda(\xi) = 0, \tag{11.9}$$

where, with $\sigma = \omega + i\lambda$ and $\varepsilon'(s, x) = \partial\varepsilon(s, x)/\partial s$,

$$\begin{aligned} L_\lambda(\xi) &= \partial_x^2\xi + \sigma^2\varepsilon(|\psi_0|^2, x)\xi \\ &+ \sigma^2\varepsilon'(|\psi_0|^2, x)\psi_0\mathrm{Re}(\bar\psi_0\xi). \end{aligned} \tag{11.10}$$

Equation (11.9) is a nonlinear eigenvalue problem. We observe that the operator family L_λ satisfies, $L_\lambda^* = L_{-\bar\lambda}$, with respect to the inner product

$$(\xi, \eta) = \mathrm{Re}\int\bar\xi\eta. \tag{11.11}$$

A crucial role in our analysis is played by the following result which is stated directly for the rescaled function $\psi(x) = A^{-1}\psi_0(x/k)$, $k = k_0 n$.

Theorem. $(\bar\psi, \bar\mu)$ is a turning point iff

$$\frac{\partial\psi}{\partial\rho_1} \text{ solves } L_0(\xi) = 0 \tag{11.12}$$

at that point.

Proof. We write Eq. (4.1) as $F(\psi, \mu) = 0$ and let $\psi = \psi(\rho_1)$ and $\mu = \mu(\rho_1)$. Differentiating the last equation with respect to ρ_1, we obtain

$$F_\psi(\psi, \mu)\frac{\partial\psi}{\partial\rho_1} + \frac{\partial F}{\partial\mu}(\psi, \mu)\frac{\partial\mu}{\partial\rho_1} = 0, \tag{11.13}$$

where $F_\psi(\psi, \mu)$ is the variational derivative of $F(\psi, \mu)$ with respect to ψ. We note that $F_\psi(\psi, \mu)$ is equal to L_0 up to a rescaling. Now, $(\bar\psi, \bar\mu)$ is a turning point iff $\partial\mu/\partial\rho_1 = 0$ at that point, and therefore, iff $F_\psi(\bar\psi, \bar\mu)\,\partial\psi/\partial\rho_1 = 0$.

This theorem implies that

$$\begin{gathered} \lambda = 0 \text{ is an eigenvalue of (rescaled)} \\ \text{Eq. (11.9)} \longleftrightarrow (\rho_1, \mu) \text{ is a turning point.} \end{gathered} \tag{11.14}$$

We claim that λ changes its sign as ρ_1 passes a turning point,

$$\frac{\partial\lambda}{\partial\rho_1} \begin{cases} < 0 & \text{at a top turning point} \\ > 0 & \text{at a bottom turning point,} \end{cases} \tag{11.15}$$

where top and bottom refers to the turning points of $\rho_1 = \rho_1(\mu)$.

Equations (11.14) and (11.15) suggest that the real eigenvalue λ, is negative on the top branches of $\rho_1 = \rho_1(\mu)$ (see Fig. 8), while $\lambda > 0$ on the bottom ones. Hence, the bottom branches are unstable. In order to

understand the stability properties of the top branches, one has to envoke the remaining, complex eigenvalues. We expect that they are stable near the top turning points and unstable elsewhere.

Equation (11.15) is proved by a perturbation theory, which requires the information about solutions to (11.9) only at the turning point. The details will be presented elsewhere.

Comparing Eqs. (11.6) and (11.9) with Eqs. (5.8) and (5.9), i.e., with the equations for the resonance solution in the linear case, we conclude that the former equations describe the resonance solution in the nonlinear case (the nonlinear resonance). It is remarkable that while the corresponding problem is always stable in the linear case (see Eq. (5.10)), the stable and unstable branches alternate in the nonlinear case.

12. EXPRESSION FOR ψ

In this section, we find an approximate form of the solution ψ to Eq. (4.4). This information is needed, in particular, for a more detailed study of the stability of various branches.

We first find the function $\rho(x)$ for $0 < x < b$. For this, we replace Eq. (9.2) with the equation

$$\pm\int_{\rho_0}^{\rho(x)} \frac{d\rho}{\sqrt{f(\rho)}} = x,$$

which is integrated in the same way as (9.2) to yield (see Appendix 2)

$$\rho(x) = \frac{1}{2}\rho_1\left\{1 + \frac{1}{n^2} + \left(1 - \frac{1}{n^2}\right)\cos 2\gamma(x)\right\}, \quad (12.1)$$

where

$$\gamma(x) = \frac{b - x}{1 - \frac{3\mu\rho_1}{4}\left(1 + \frac{1}{n^2}\right)}. \quad (12.2)$$

We next find the expression for $\alpha = \arg(\psi)$ that matches (12.1). Observing that

$$1 + \frac{1}{n^2} + \left(1 - \frac{1}{n^2}\right)\cos 2\gamma = \frac{(n+1)^2}{2n^2}\left|e^{i\gamma} + \frac{n-1}{n+1}e^{-i\gamma}\right|^2,$$

we seek α in the form

$$\alpha(x) = \bar{\alpha}(x) + \beta(x), \quad (12.3)$$

where $\bar{\alpha}(x) = -\arctan[n^{-1}\tan\gamma(x)]$ and where the function β is to be found using Eq. (7.3) for α. The latter gives, modulo $O(\mu)$,

$$\beta(x) = -\frac{3\mu\rho_1}{4}\left(1 + \frac{1}{n^2}\right)(\bar{\alpha}(x) - \bar{\alpha}(0)) + \beta(0). \quad (12.4)$$

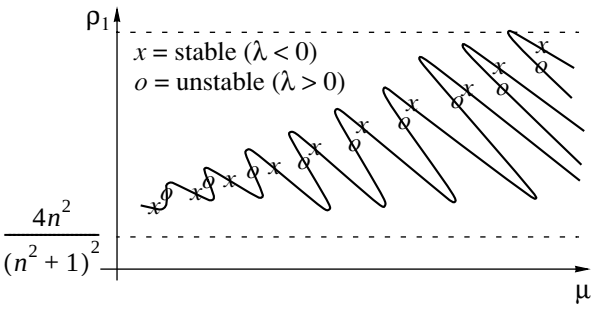


Fig. 8.

The initial condition $\beta(0)$ is found from

$$\beta(0) = \alpha(0) - \arctan\left(\frac{1}{n}\tan\gamma(0)\right) \quad (12.5)$$

and

$$\alpha(0) = \operatorname{arccot}\left(\frac{n\rho'}{2(\rho + \rho_1)}\right)\Bigg|_{x=0}. \quad (12.6)$$

The function $\beta(x)$ can also be represented as (again modulo $O(\mu)$)

$$\beta(x) = \beta(0) - \frac{3\mu\rho_1}{4n}\left(1 + \frac{1}{n^2}\right) \times \int_0^x \frac{dy}{1 + \frac{1}{n^2} + \left(1 - \frac{1}{n^2}\right)\cos 2\gamma(y)}. \quad (12.7)$$

Putting Eqs. (12.1) and (12.3) together, we write $\psi(x) = \sqrt{\rho(x)}\, e^{i\alpha(x)}$ as

$$\psi(x) = \frac{n+1}{2n}\sqrt{\rho_1}e^{i\delta(x)}\left[e^{-i\gamma(x)} + \frac{n-1}{n+1}e^{i\gamma(x)}\right], \quad (12.8)$$

where $\gamma(x)$ is given by Eq. (12.2) and

$$\delta(x) = \text{const} + \frac{\mu\rho_1}{2n}(b - x).$$

The explicit form of ψ reflects the picture of a nonlinear wave propagation: it is a superposition of two waves travelling in opposite directions with slightly different speed. The nonlinearity leads to a renormalization of the wave vector,

$$k_0 \longrightarrow k_0\left(1 - \frac{3\mu\rho_1}{4}\left(1 + \frac{1}{n^2}\right)\right)^{-1},$$

and to the appearance of a slowly varying phase $\beta(x)$.

Expression (12.8) for ψ will be used in the study of the stability problem which will appear elsewhere.

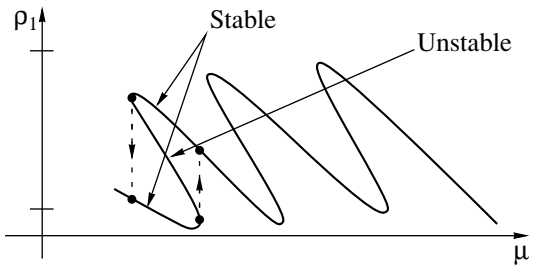


Fig. 9. Switching of solutions.

13. CONCLUSION

As we see from Fig. 8, a small change of the light intensity (i.e., of μ) near a turning point is capable of switching the system from one state (solution ψ) to another. Namely, moving around a turning point changes a stable solution into an unstable one; under the action of a random perturbation, the system then jumps to a stable solution as shown in Fig. 9. This either turns on or turns off the light passing through the slab.

At the next step, we would like to address the nonlinear stability problem, in other words, to study solutions of Eqs. (2.7)–(2.9) with the initial conditions close to the solitary wave $e^{-ik_0t}\psi_0(x)$ in the cases where ψ_0 is on a stable branch and, more interestingly, in the case where ψ_0 is on the unstable branch (see Eq. (5.11)). In the latter case, it is desirable to find a mathematical description to the processes described above (see Fig. 9).

The second author is grateful to Doug Mills for fruitful discussions and encouragement. Supported by NSERC under Grant NA7901.

APPENDIX 1

One-Mode Approximation

In this appendix, we derive Eq. (3.2), which is a one-mode approximation to Eq. (2.7). We seek a solution to Eq. (2.7) in the form

$$E(x,t) = \sum_{n=0}^{\infty} \mathrm{Re}(e^{-i(2n+1)k_0t}\psi_n(x)). \qquad \text{(A1.1)}$$

Substituting (A1.1) in (2.7), we obtain an infinite system of coupled equations for the coefficients $\psi_n(x)$, $n = 0, 1, \ldots$, the $\pm n$-th equation read off from the coefficient in front of $e^{\mp i(2n+1)k_0t}$. Because of boundary conditions (2.8), (2.9), it is easy to show that $\psi_n = O(\bar{\mu}^n)$. Hence, the contribution of ψ_n, $n \geq 1$, to the $n = 0$ equation is of the order $O(\bar{\mu}^2)$, and we drop this contribution in the leading-order approximation. Finally, to derive the $n = 0$ equation, we use the relation

$$\begin{aligned}
&\frac{\partial^2}{\partial t^2}\left(\left|\mathrm{Re}(e^{-ik_0t}\psi_0)\right|^2\mathrm{Re}(e^{-ik_0t}\psi_0)\right)\\
&= \frac{\partial^2}{\partial t^2}\left(\frac{1}{4}2|\psi_0|^2\left(\frac{1}{2}e^{-ik_0t}\psi_0\right)\right)\\
&+ \frac{\partial^2}{\partial t^2}\left(\frac{1}{4}e^{-2ik_0t}\psi_0^2\left(\frac{1}{2}e^{ik_0t}\bar{\psi}_0\right)\right)\\
&+ \text{terms proportional on } e^{ik_0t} \text{ and } e^{\pm 3ik_0t}\\
&= \frac{\partial^2}{\partial t^2}\left(\frac{3}{4}|\psi_0|^2\left(\frac{1}{2}e^{-ik_0t}\psi_0\right)\right)\\
&+ \text{terms proportional to } e^{ik_0t} \text{ and } e^{\pm 3ik_0t}.
\end{aligned} \qquad \text{(A1.2)}$$

In the above approximation, this expression immediatly implies Eq. (3.2).

APPENDIX 2

Computation of $\int \frac{d\rho}{\sqrt{f(\rho)}}$

The derivation of Eq. (9.3). We set $\rho_1 = \rho(b)$ and recall expression (9.1) for $f(\rho)$,

$$f(\rho) = 4(\rho_1 - \rho)\left[\mu\rho^2 + (1+\mu\rho_1)\rho - \frac{\rho_1}{n^2}\right]. \qquad \text{(A2.1)}$$

We observe that ρ_1 is a root of the equation $f(\rho) = 0$. We find the other two roots,

$$\begin{aligned}
\rho_{2,3} &= (2\mu)^{-1}\left[-(1+\mu\rho_1) \pm \sqrt{(1+\mu\rho_1)^2 + \frac{4\mu\rho_1}{n^2}}\right]\\
&= \frac{1+\mu\rho_1}{2\mu}\left[-1 \pm \sqrt{1 + \frac{4\mu\rho_1}{n^2(1+\mu\rho_1)^2}}\right].
\end{aligned} \qquad \text{(A2.2)}$$

The expansion in powers of μ shows that

$$\begin{aligned}
\rho_2 &= \frac{\rho_1}{n^2(1+\mu\rho_1)}\left(1 - \frac{\mu\rho_1}{n^2}\right) + O(\mu^2)\\
&= \frac{\rho_1}{n^2}\left(1 - \mu\rho_1\left(1 + \frac{1}{n^2}\right)\right) + O(\mu^2)
\end{aligned} \qquad \text{(A2.3)}$$

and

$$\rho_3 = -\frac{1}{\mu}\left[1 + \mu\rho_1\left(1 + \frac{1}{n^2}\right)\right] + O(\mu). \qquad \text{(A2.4)}$$

Clearly, for μ small and $n \geq 1$, a safe way to expand $f^{-1/2}(\rho)$ is by expanding

$$(\mu\rho - \mu\rho_3)^{-1/2} = \left[1 + \mu\rho_1\left(1 + \frac{1}{n^2}\right) + \mu\rho + O(\mu^2)\right]^{-1/2}$$
$$= 1 - \frac{1}{2}\mu\left(\rho_1 + \frac{\rho_1}{n^2} + \rho\right) + O(\mu^2).$$

Hence, for

$$f(\rho) = -4(\rho - \rho_1)(\rho - \rho_2)(\mu\rho - \mu\rho_3) \qquad \text{(A2.5)}$$

we obtain, modulo $O(\mu^2)$,

$$2f^{-1/2}(\rho) = \frac{1 - \frac{1}{2}\mu\left(\rho_1 + \frac{\rho_1}{n^2} + \rho\right)}{[-(\rho - \rho_1)(\rho - \rho_2)]^{1/2}}. \qquad \text{(A2.6)}$$

Our aim is to integrate this expression. For this, we present it as

$$2f^{-1/2}(\rho) = \frac{1 - \frac{1}{2}\mu\left(\frac{3}{2}\rho_1 + \frac{1}{2}\rho_2 + \frac{\rho_1}{n^2}\right)}{\sqrt{-(\rho - \rho_1)(\rho - \rho_2)}} + \frac{1}{2}\mu\frac{-\rho + \frac{1}{2}(\rho_1 + \rho_2)}{\sqrt{-(\rho - \rho_1)(\rho - \rho_2)}}. \qquad \text{(A2.7)}$$

We let $\rho_0 = \rho(0)$ and recall that $\rho_1 = \rho(b)$. We then have

$$\int_{\rho_0}^{\rho_1} d\rho \frac{-\rho + \frac{1}{2}(\rho_1 + \rho_2)}{\sqrt{-(\rho - \rho_1)(\rho - \rho_2)}} = \sqrt{-(\rho - \rho_1)(\rho - \rho_2)}\Big|_{\rho_0}^{\rho_1}$$
$$= -\sqrt{(\rho_1 - \rho_0)(\rho_0 - \rho_2)}. \qquad \text{(A2.8)}$$

Utilizing that

$$\left(\arcsin\frac{x-p}{q}\right)' = \frac{1}{\sqrt{1 - \frac{(x-p)^2}{q^2}}}\frac{1}{q}$$
$$= \frac{1}{\sqrt{q^2 - (x-p)^2}} = \frac{1}{\sqrt{(p+q-x)(x-p+q)}}$$

and setting $p = (\rho_1 + \rho_2)/2$ and $q = (\rho_1 - \rho_2)/2$, we obtain

$$\int_{\rho_0}^{\rho_1} \frac{d\rho}{\sqrt{(\rho_1 - \rho)(\rho - \rho_2)}} = \arcsin 2\left(\frac{\rho - \frac{\rho_1 + \rho_2}{2}}{\rho_1 - \rho_2}\right)\Bigg|_{\rho_0}^{\rho_1} \qquad \text{(A2.9)}$$
$$= \frac{\pi}{2} - \arcsin\left[\frac{2\rho_0 - (\rho_1 + \rho_2)}{\rho_1 - \rho_2}\right].$$

Combining Eqs. (A2.7)–(A2.9), we obtain

$$2\int_{\rho_0}^{\rho_1} \frac{d\rho}{\sqrt{f(\rho)}} = -\frac{1}{2}\mu\sqrt{(\rho_1 - \rho_0)(\rho_0 - \rho_2)} + \left[1 - \frac{1}{4}\mu\left(3\rho_1 + \rho_2 + 2\frac{\rho_1}{n^2}\right)\right] \times \left[\frac{\pi}{2} - \arcsin\left[\frac{2\rho_0 - (\rho_1 + \rho_2)}{\rho_1 - \rho_2}\right]\right]. \qquad \text{(A2.10)}$$

Together with Eqs. (9.2) and (A2.3), this gives, modulo $O(\mu^2)$,

$$\left[1 - \frac{3}{4}\mu\left(1 + \frac{1}{n^2}\right)\rho_1\right]\left[\frac{\pi}{2} - \arcsin\left[\frac{2\rho_0 - (\rho_1 + \rho_2)}{\rho_1 - \rho_2}\right]\right] - \frac{1}{2}\mu\sqrt{(\rho_1 - \rho_0)(\rho_0 - \rho_2)} = \pm 2b. \qquad \text{(A2.11)}$$

This can be rewritten as

$$\frac{2\rho_0 - (\rho_1 + \rho_2)}{\rho_1 - \rho_2} = \cos\frac{\pm 2b + \mu\sqrt{(\rho_1 - \rho_0)(\rho_0 - \rho_2)}/2}{1 - \frac{3}{4}\mu\left(1 + \frac{1}{n^2}\right)\rho_1}. \qquad \text{(A2.12)}$$

Equations (7.11)–(7.12) now imply

$$\rho_0 = \frac{(n^2+3)\rho_1 - 4}{n^2 - 1} + \frac{n^2\mu}{n^2 - 1}\left[\rho_1^2 - \frac{((n^2+3)\rho_1 - 4)^2}{(n^2-1)^2}\right] + O(\mu^2). \qquad \text{(A2.13)}$$

Inserting this expression for ρ_0 and Eq. (A2.3) for ρ_2 in Eq. (A2.12), we finally obtain an equation for ρ_1 only,

$$\frac{4(1-\rho_1)}{\rho_1}\frac{n^2}{(n^2-1)^2} \times \left[1 - \frac{2n^2\mu}{(n^2-1)^2}((n^2+1)\rho_1 - 2) - \frac{\mu\rho_1(n^2+1)}{n^2(n^2-1)}\right] = \sin^2\left[\frac{b + \frac{\mu}{2(n^2-1)}\sqrt{(1-\rho_1)n^{-2}((n^2+1)^2\rho_1 - 4n^2)}}{1 - \frac{3\mu}{4}\left(1 + \frac{1}{n^2}\right)\rho_1}\right], \qquad \text{(A2.14)}$$

modulo $O(\mu^2)$ terms. We simplify this equation by dropping terms of the order $O(\mu)$ (but keeping terms of the order $O(\mu b)$!) to obtain Eq. (9.3).

The derivation of Eqs. (12.1), (12.2). Proceeding as above, we obtain

$$\int_{\rho_0}^{\rho(x)} \frac{d\rho}{\sqrt{f(\rho)}} = \left[1 - \frac{3\mu}{4}\left(1 + \frac{1}{n^2}\right)\rho_1\right] \times \left[\arcsin\left[\frac{2\rho(x) - (\rho_1 + \rho_2)}{\rho_1 - \rho_2}\right] - \arcsin\left[\frac{2\rho_0 - (\rho_1 + \rho_2)}{\rho_1 - \rho_2}\right]\right] \quad \text{(A2.15)}$$

$$+ \frac{1}{2}\mu\left[\sqrt{(\rho_1 - \rho(x))(\rho(x) - \rho_2)} - \sqrt{(\rho_1 - \rho_0)(\rho_0 - \rho_2)}\right]$$

(Eq. (A2.15) with $x = b$ yields, as it should, Eq. (A2.10)). Using this expression, we find an approximate solution of the equation

$$\int_{\rho_0}^{\rho(x)} \frac{d\rho}{\sqrt{f(\rho)}}$$

for $\rho(x)$ by dropping terms of the order $O(\mu)$ but keeping terms of the order $O(\mu b)$. This yields

$$\rho(x) = \frac{1}{2}(\rho_1 + \rho_2) + \frac{1}{2}(\rho_1 - \rho_2)\cos 2\gamma(x), \quad \text{(A2.16)}$$

where

$$\gamma(x) = \frac{b - x}{1 - \frac{3\mu\rho_1}{4}\left(1 + \frac{1}{n^2}\right)}.$$

Inserting expression (A2.3) for ρ_2 in the right-hand side, we arrive at Eqs. (12.1) and (12.2).

APPENDIX 3

In this appendix, we outline another derivation of the expression for the solution ψ of Eq. (4.4). In this derivation, we consider (4.4) as a linear equation for ψ by assuming that $|\psi|^2 = \rho$ in this equation is given by (12.1) and (12.2). We seek two linearly independent solutions of the resulting equation in the Bloch function the form

$$[Ae^{i\gamma(x)} + Be^{-i\gamma(x)}]e^{i\nu(b-x)} + O(\mu) \quad \text{(A3.1)}$$

for some ν, A, and B. Inserting this in the equation in question and using the solvability condition for the constants A and B, we obtain, after a simple calculation,

$$\nu = \pm\lambda, \text{ where } \lambda = \frac{\mu\rho_1}{2n}. \quad \text{(A3.2)}$$

With these values for ν, we solve for A as

$$A = \frac{n \pm 1}{n \pm 1}B.$$

As a result, the general solution of the above linear equation is given by

$$\psi = C\left\{\left(1 + \frac{1}{n}\right)e^{i\gamma(x)} + \left(1 - \frac{1}{n}\right)e^{-i\gamma(x)}\right\}e^{-i\lambda(b-x)} + D\left\{\left(1 + \frac{1}{n}\right)e^{-i\gamma(x)} + \left(1 - \frac{1}{n}\right)e^{i\gamma(x)}\right\}e^{i\lambda(b-x)}. \quad \text{(A3.3)}$$

From boundary condition (4.9), we find

$$C = \frac{\mu\rho_1}{4}\left(\frac{1}{2} + \frac{3}{2n^2}\right)D. \quad \text{(A3.4)}$$

Hence, the last term in (A3.3) is $O(\mu)$ and can therefore be omitted. Finally, we use that $|\psi(b)| = \sqrt{\rho_1}$, to find that $|D| = \sqrt{\rho_1}/2$, and therefore,

$$\psi = \frac{n+1}{2n}\sqrt{\rho_1}\left[e^{-i\gamma(x)} + \frac{n-1}{n+1}e^{i\gamma(x)}\right]e^{i\beta + i\lambda(b-x)}, \quad \text{(A3.5)}$$

where λ is given by Eq. (A3.2) and β is some constant related to $\alpha(0)$ in a simple way.

The value of $|\psi|$ that follows from (A3.5) is the same as that given by Eq. (12.1).

REFERENCES

1. H. G. Winful, J. H. Marburger, and E. Germire, Appl. Phys. Lett. **35**, 379 (1979).
2. Wei Chen and D. L. Mills, Phys. Rev. Lett. **58**, 160 (1987).
3. Wei Chen and D. L. Mills, Phys. Rev. B **36**, 6269 (1987).
4. Wei Chen and D. L. Mills, Phys. Rev. B **35**, 524 (1987).
5. D. L. Mills and S. E. Trullinger, Phys. Rev. B **36**, 947 (1987).
6. J. E. Sipe and H. G. Winful, Opt. Lett. **13**, 132 (1988).
7. C. M. de Sterke and J. E. Sipe, Phys. Rev. A **38**, 5149 (1988).
8. D. N. Christodoulides and R. I. Joseph, Phys. Rev. Lett. **62**, 1746 (1989).
9. A. B. Aceves and S. Wabmitz, Phys. Lett. A **141**, 37 (1989).
10. N. O. Sankey, D. F. Prelewitz, and T. G. Brown, Appl. Phys. Lett. **60**, 1427 (1992).
11. C. M. de Sterke and J. E. Sipe, in *Progress in Optics XXXIII*, Ed. by E. Wolf (North-Holland, Amsterdam, 1994), p. 203.
12. S. Dutta Gupta, in *Progress in Optics XXXIII*, Ed. by E. Wolf (North-Holland, Amsterdam, 1998).
13. D. Henning and G. P. Tsironis, Phys. Rep. **307**, 333 (1999).
14. G. P. Agrawal, *Nonlinear Fiber Optics* (Academic, Boston, 1989; Mir, Moscow, 1996).
15. N. Bloembergen, *Nonlinear Optics: a Lecture Note and Reprint Volum* (W. A. Benjamin, New York, 1965; Mir, Moscow, 1966).
16. H. M. Gibbs, *Optical Bistability: Controlling Light with Light* (Academic, New York, 1985; Mir, Moscow, 1988).
17. D. L. Mills, *Nonlinear Optics: Basic Concepts* (Springer-Verlag, New York, 1991).
18. Y. R. Shen, *The Principles of Nonlinear Optics* (Wiley, New York, 1984; Nauka, Moscow, 1989).
19. W. A. Strauss, *Nonlinear Wave Equations* (American Mathematical Society, Providence, 1989), CBMS Regional Conference Series in Mathematics, No. 73, 1987.

Journal of Experimental and Theoretical Physics, Vol. 93, No. 5, 2001, pp. 1017–1027.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 5, 2001, pp. 1172–1184.

ATOMS, SPECTRA, RADIATION

Two-Photon Jaynes–Cummings Systems Interacting with a Classical Field

R. A. Ismailov and A. Ya. Kazakov

St. Petersburg State University of Aerospace Instrumentation, St. Petersburg, 190000 Russia
e-mail: a_kazak@mail.ru

Received May 30, 2001

Abstract—If a two-level atom is in the two-photon resonance with a quantized mode and simultaneously interacts with a quasi-resonance classical field, then a photon exchange is observed in this system between the quantized and classical modes. It is demonstrated that such a physical system can serve as a source of squeezed radiation in the quantized mode. The squeezing can be arbitrarily close to unity, while the radiation amplitude can be relatively large. A situation is discussed when N atoms are in the two-photon resonance with a quantized mode and simultaneously interact with a classical field. The phenomenon of exponential superradiation is described when the number of photons in the quantized mode exponentially depends on the number N of atoms.

1. INTRODUCTION

Recently, physical systems based on the simultaneous interaction of an atom with a quantized mode and classical fields have been intensively studied (see [1–12]). In these studies, classical fields, whose parameters can easily be specified in practice, provide an efficient tool for controlling the interaction between an atom and a quantized field.

There exists a natural hierarchy of scales in such physical systems; usually, classical fields are much more intense than quantized fields. This means that, if the appropriate interaction constants of an atom with fields are of the same order of magnitude, then the effective Rabi parameter of the classical field is much greater than the effective Rabi parameter of the quantized field. Thus, the dynamics of the system is divided into two parts: a "fast" one, connected with the interaction of an atom (or atoms if there are several of them) with the classical field, and the "slow" one, which corresponds to the interaction of the atom with the quantized field. This fact leads to the description of the dynamics of a system that is "averaged" with respect to fast oscillations. As is shown below, in this case, "closed" and "open" types of resonance may occur when the dynamics averaged over fast oscillations is nontrivial. In the present paper, we discuss open resonances, so that, under an appropriate resonance condition, there is no conservation law for excitations in the averaged system (in contrast to the standard Jaynes–Cummings model).

The dynamics of a two-photon Jaynes–Cummings model was earlier discussed in [13–18]. The study of the dynamics of a two-photon Jaynes–Cummings model interacting with a classical field was carried out in [9, 10] mainly by numerical experiments. Various aspects of the dynamics of one-photon Jaynes–Cummings models interacting with classical fields were studied in [1–5] for closed resonances. A one-photon Jaynes–Cummings system for an open resonance was considered in [6], the case of N two-level atoms was discussed in [7], and the case of N three-level atoms, in [8]. The papers [11, 12] were devoted to the study of the dynamics of a two-photon Jaynes–Cummings system (the case of a single atom was considered) interacting with a classical field.

In the present paper, we restrict the discussion to two-level atoms. Similar results can also be obtained for three-level atoms. From the viewpoint of applications, the use of three-level atoms can be even more attractive, since, in the three-level case, it seems to be easier to find a real atom with transitions corresponding to relevant resonance conditions. However, the technical details in the case of three-level atoms are somewhat more cumbersome, although the physics of the processes is essentially the same. Therefore, we restrict ourselves to the case of two-level atoms. Here, we will not discuss a possible influence of an intermediate level assuming that this level is eliminated adiabatically (as is usually assumed when analyzing a two-photon interaction). In addition, we will neglect the decay of the quantized mode; i.e., we assume that the resonator considered is ideal.

This paper is organized as follows. In Section 2, we consider the dynamics of a single two-level atom that simultaneously interacts with a quasi-resonance classical field and (in a two-photon approximation) with a quantized mode. For an open resonance, applying an appropriate averaging procedure, we obtain a Hamiltonian averaged over fast oscillations and analyze the dynamics of the corresponding system. Our methods make it possible to obtain a full analytic description of

such dynamics and calculate the basic quantum-statistical characteristics of the quantized field. Then, using these results (which were partly obtained in [11, 12]), we discuss the interaction of N atoms with classical and quantized fields. We demonstrate that situations are possible when the population of the quantized mode exponentially grows with time, where the exponent is a linear function of N, the number of atoms.

2. A SINGLE ATOM INTERACTING WITH A QUANTIZED MODE AND A CLASSICAL FIELD

Let us begin our analysis with the case when a single atom interacts with quantized and classical fields. Here, our goal is to describe the quantum-statistical properties of a quantized radiation and to construct a necessary formalism for studying the case of N atoms. We will discuss the solution of a problem for an appropriate Schrödinger equation; i.e., we will use the Schrödinger representation.

2.1. Basic Concepts

The starting point of our analysis is the Hamiltonian describing the interaction of a two-level atom with a quantized mode (two-photon interaction) and a classical monochromatic field (quasi-resonance interaction). Within a dipole approximation and the rotating wave approximation, this Hamiltonian is expressed as

$$\begin{aligned}\mathbf{H} = \omega\mathbf{a}^+\mathbf{a} + \kappa\mathbf{J}_0 + \zeta\{(\mathbf{a}^+)^2\mathbf{J}_- + \mathbf{a}^2\mathbf{J}_+\}\\ + \mu[\mathbf{J}_-\exp(i\Omega t) + \mathbf{J}_+\exp(-i\Omega t)],\end{aligned} \tag{1}$$

where $\mathbf{a}^+$ and $\mathbf{a}$ are the creation and annihilation operators for a quantized field of frequency ω, Ω stands for the frequency of the classical field, and the operators

$$\mathbf{J}_0 = \mathrm{diag}\{1, -1\}, \quad \mathbf{J}_- = \mathbf{J}_+^T = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

describe a two-level atom interacting with external fields; here, the known commutation relations hold:

$$[\mathbf{a}, \mathbf{a}^+] = 1, \quad [\mathbf{J}_0, \mathbf{J}_-] = -2\mathbf{J}_-, \quad [\mathbf{J}_0, \mathbf{J}_+] = 2\mathbf{J}_+.$$

The parameter ζ in (1) describes the coupling constant of the atom and the quantized mode, and μ is the (scaled) amplitude of the classical field. It follows from standard commutation relations that the following equations hold for any analytic function $q(s)$:

$$\begin{aligned}\mathbf{a}q(\mathbf{a}^+\mathbf{a}) &= q(\mathbf{a}^+\mathbf{a} + 1)\mathbf{a},\\ \mathbf{a}^+q(\mathbf{a}^+\mathbf{a}) &= q(\mathbf{a}^+\mathbf{a} - 1)\mathbf{a}^+,\\ \mathbf{J}_-q(\mathbf{J}_0) &= q(\mathbf{J}_0 + 2)\mathbf{J}_-,\\ \mathbf{J}_+q(\mathbf{J}_0) &= q(\mathbf{J}_0 - 2)\mathbf{J}_+.\end{aligned} \tag{2}$$

We will focus on the dynamics of the wave function defined by the Schrödinger equation

$$i\frac{\partial}{\partial t}\Psi = \mathbf{H}\Psi. \tag{3}$$

As we have already noted, the Hamiltonian (1) is written in the rotating wave approximation. This means that the optical frequency is much greater than all the other frequency parameters of the problem. Similar Hamiltonians (for a one-photon interaction, under the conditions of the type $\omega = \Omega$ or $\omega = 2\kappa = \Omega$) were used in [1–3].

Hamiltonian (1) contains several frequency parameters defined by the characteristics of the atom and the fields. These are the optical frequency $\omega \sim \kappa \sim \Omega/2$, the Rabi parameter $R \sim \mu$ of the classical field, and the Rabi parameter R_q of the quantized mode, $R_q \sim \zeta\sqrt{n(t)}$ ($n(t)$ is the population of the quantized mode, i.e., the average number of photons in the quantized mode). Henceforth, we assume that

$$R \gg R_q. \tag{4}$$

Note that the population $n(t)$ is not known *a priori*. Therefore, we either have to check the validity of (4) *a posteriori*, or restrict ourselves to the consideration of those time intervals for which this condition is satisfied.

The functions that substantially vary within the time intervals R^{-1} (R_q^{-1}) are called fast (slow) functions. Our immediate goal is to separate the fast oscillations, associated with the interaction of the atom with the classical field, from the slow evolution, associated with the interaction with the quantized mode. In other words, we derive a Hamiltonian averaged over fast oscillations. As the first step, we apply the known transformation that separates optical frequencies:

$$\Psi(t) = \exp[-i\omega t(\mathbf{a}^+\mathbf{a} + \mathbf{J}_0)]\Phi(t). \tag{5}$$

It reduces the Schrödinger equation to the form

$$\begin{aligned}i\frac{\partial}{\partial t}\Phi = \{(\kappa - \omega)\mathbf{J}_0 + \zeta[(\mathbf{a}^+)^2\mathbf{J}_- + \mathbf{a}^2\mathbf{J}_+]\\ + \mu[\mathbf{J}_-\exp[-i(2\omega - \Omega)t]\\ + \mathbf{J}_+\exp[i(2\omega - \Omega)t]]\}\Phi.\end{aligned} \tag{6}$$

Next, suppose that the 2×2 matrix function $\Xi(t)$ is a solution to the following auxiliary initial problem:

$$\begin{aligned}i\frac{\partial}{\partial t}\Xi = \{(\kappa - \omega)\mathbf{J}_0 + \mu[\mathbf{J}_-\exp[-i(2\omega - \Omega)t]\\ + \mathbf{J}_+\exp[i(2\omega - \Omega)t]]\}\Xi,\end{aligned} \tag{7}$$

$$\Xi(0) = I, \tag{8}$$

where I is a 2×2 identity matrix. We can easily show that

$$\Xi(t) = \exp[i(2\omega - \Omega)t\mathbf{J}_0/2]\mathbf{U}\exp[-iRt\mathbf{J}_0]\mathbf{U}^{-1},$$

$$\mathbf{U} = \begin{pmatrix} \mu & \Delta - R \\ R - \Delta & \mu \end{pmatrix},$$

$$R = \sqrt{\mu^2 + \Delta^2}, \quad \Delta = \kappa - \frac{\Omega}{2}.$$

We seek a solution to (6) in the form

$$\Phi(t) = \Xi(t)\varphi(t). \tag{9}$$

It follows from the results given below that formula (9) describes a wave function as a product of fast and slow cofactors.

Using relation (7), we obtain the following equation for the function $\varphi(t)$:

$$i\frac{\partial}{\partial t}\varphi(t) = \zeta\Xi^{-1}(t)[(\mathbf{a}^+)^2\mathbf{J}_- + \mathbf{a}^2\mathbf{J}_+]\Xi(t)\varphi(t). \tag{10}$$

Thus, we pass from the wave function $\Psi(t)$ to $\varphi(t)$ using the substitution

$$\Psi(t) = \exp[-i\omega t(\mathbf{a}^+\mathbf{a} + \mathbf{J}_0)]\Xi(t)\varphi(t),$$

here, as follows from (8),

$$\varphi(0) = \Psi(0).$$

Equation (10) is simpler than Eq. (3) since its right-hand side is proportional to a (relatively) small parameter R_q. This fact allows one to apply an appropriate asymptotic procedure for constructing a solution to (10) in the leading asymptotic order. Such a procedure was earlier discussed in [6–8] and is similar to the standard averaging procedure. In this case, we seek a solution to the equation

$$i\frac{\partial}{\partial t}\varphi(t) = \zeta\langle\langle\Xi^{-1}(t)\{(\mathbf{a}^+)^2\mathbf{J}_- + \mathbf{a}^2\mathbf{J}_+\}\Xi(t)\rangle\rangle\varphi(t). \tag{11}$$

Here, $\langle\langle\ldots\rangle\rangle$ denotes that we have rejected the fast harmonics with frequencies of order R on the right-hand side of (10).

Note that the operators $\Xi(t)$ and $\exp[-i\omega t(\mathbf{a}^+\mathbf{a} + \mathbf{J}_0)]$ are unitary operators. The Fock operators $\mathbf{a}$ and $\mathbf{a}^+$ commute with the operators $\exp(-i\omega t\mathbf{J}_0/2)$ and $\Xi(t)$. Hence, if we calculate $\langle G\rangle$, where G is a purely Fock operator that is polynomial in $\mathbf{a}^+\mathbf{a}$, then

$$\langle G\rangle = \langle\overline{\Psi(t)}, G\Psi(t)\rangle = \langle\overline{\varphi(t)}, G\varphi(t)\rangle. \tag{12}$$

The atoms interacting with quasi-resonant and classical fields can be considered as atoms dressed by the field with a quantized mode. In these terms, our problem can be considered as a two-photon interaction of atoms dressed by the field with a quantized mode. Relation (9) can be interpreted as a transition to the basis of states of atoms dressed by the field.

2.2. Averaged Hamiltonian

Simple calculations with the use of our results imply that the operator

$$\Xi^{-1}(t)[(\mathbf{a}^+)^2\mathbf{J}_- + \mathbf{a}^2\mathbf{J}_+]\Xi(t)$$

contains the following harmonics:

$$\pm(2\omega - \Omega), \quad \pm(2\omega - \Omega \pm 2R).$$

Thus, there are three situations when the averaging yields a nontrivial (i.e., nonzero in the leading asymptotic order) result.

1. $|R - (2\omega - \Omega)/2| = |\sigma| \ll R$. In this case, the averaged Hamiltonian is given by

$$\mathbf{H}_1 = \zeta\mu^2 D^{-1}\mathbf{U} \times \begin{pmatrix} 0 & \mathbf{a}^2\exp[2i\sigma t] \\ (\mathbf{a}^+)^2\exp[-2i\sigma t] & 0 \end{pmatrix}\mathbf{U}^{-1},$$

where

$$D = \det\mathbf{U} = \mu^2 + (R - \Delta)^2.$$

2. $|R + (2\omega - \Omega)/2| = |\eta| \ll R$. In this case, the averaged Hamiltonian on the right-hand side of (11) is given by

$$\mathbf{H}_2 = -\zeta(R - \Delta)^2 D^{-1}\mathbf{U} \times \begin{pmatrix} 0 & (\mathbf{a}^+)^2\exp[2i\eta t] \\ \mathbf{a}^2\exp[-2i\eta t] & 0 \end{pmatrix}\mathbf{U}^{-1}.$$

In these cases, we obtain a Hamiltonian that coincides (after trivial transformations) with the ordinary Hamiltonian of a two-photon Jaynes–Cummings model. Here, as is well-known, the conservation law for excitations holds in the system. Therefore, we refer to these resonance conditions as closed conditions.

3. $|2\omega - \Omega)|/2 = |\nu| \ll R$. The averaging of the right-hand side of (11) yields

$$\mathbf{H}_3 = \zeta\mu(R - \Delta)D^{-1}((\mathbf{a}^+)^2 e^{2i\nu t} + \mathbf{a}^2 e^{-2i\nu t})\mathbf{U}\mathbf{J}_0\mathbf{U}^{-1}.$$

It follows from the results below that the conservation law for excitations does not hold for this Hamiltonian (just as for the original Hamiltonian). This fact reflects a sharp difference between the dynamics of this system and the systems controlled by ordinary Jaynes–Cummings two-photon Hamiltonian. In view of this fact, we can refer to this resonance condition as open condition. Below, we will discuss in detail the characteristic features of this very case. Note that conditions 1–3 can be satisfied by choosing appropriate fre-

quencies of external fields and the amplitude of the classical field.

The averaged Hamiltonian thus constructed represents a product of the Fock and matrix operators. This fact allows us to make further simplifications. Note that the eigenvectors of the matrix operator $\mathbf{U}\mathbf{J}_0\mathbf{U}^{-1}$ are

$$\mathbf{e}_1 = \frac{1}{\sqrt{D}}\begin{pmatrix}\mu\\ R-\Delta\end{pmatrix}, \quad \mathbf{e}_2 = \frac{1}{\sqrt{D}}\begin{pmatrix}\Delta-R\\ \mu\end{pmatrix},$$

while $\varepsilon_k = (-1)^{k+1}$, $k = 1, 2$, are the corresponding eigenvalues. As follows from the last relation, in this case, the Hamiltonian of the system is split (up to an elementary rotation) into two one-dimensional Hamiltonians that differ only by sign. Thus, we can restrict the analysis to a system controlled by the Hamiltonian

$$\mathbf{H}_4 = \rho((\mathbf{a}^+)^2 e^{2i\nu t} + \mathbf{a}^2 e^{-2i\nu t}), \tag{13}$$

where $\rho = \zeta\mu(R-\Delta)/D$, which corresponds to the solution of the initial problem with a special vector structure of the initial condition. This fact allows us to calculate the basic quantum-statistical characteristics of the quantized radiation corresponding to the special initial condition. The formulas thus obtained allow us to describe the wave function in the case of an arbitrary initial condition as well. These results will allow us to study the situation when N identical two-level atoms simultaneously interact with quantized and classical fields.

2.3. Solution of the Initial Problem

Let us rewrite the Schrödinger equation corresponding to the Hamiltonian (13) in the Fock–Bargmann representation (see, for example, [19]) under the substitutions $\mathbf{a} \longrightarrow d/dz$ and $\mathbf{a}^+ \longrightarrow z$:

$$i\frac{\partial\varphi}{\partial t} = \rho(z^2 e^{2i\nu t}\varphi + e^{-2i\nu t}\varphi_{zz}). \tag{14}$$

After the change of variables

$$\tau = t, \quad y = z\exp(i\nu t)$$

we arrive at the equation

$$i\varphi_\tau = \rho[\varphi_{yy} + y^2\varphi] + \nu y\varphi_y. \tag{15}$$

One can solve these equations by separating the variables (see [11]). However, we will apply a more convenient formalism involving an appropriate integral transformation. The detailed description of this approach was given in [12]. Omitting cumbersome details, we can rewrite a solution to Eq. (15) as

$$\begin{aligned}\varphi(z,t) &= \exp\left[\frac{\beta z^2}{2}\exp(2i\nu t) + i(\nu+\beta\rho)t\right]\\ &\times\int ds\exp\left[iz\exp(i\nu t)s - \frac{\rho s^2}{2(\nu+2\beta\rho)}\right]\\ &\times\Pi(s\exp[i(\nu+2\beta\rho)t]),\end{aligned} \tag{16}$$

where

$$\beta = -(\nu + \sqrt{\nu^2 - 4\rho^2})(2\rho)^{-1},$$

and the function $\Pi(x)$ is defined by the initial data, i.e., by the equation

$$\begin{aligned}&\exp\left[\frac{\beta z^2}{2}\right]\int ds\,\exp\left[izs - \frac{\rho s^2}{2(\nu+2\beta\rho)}\right]\\ &\times\Pi(s) = \varphi(0,z) = Q(z).\end{aligned} \tag{17}$$

If the quantized mode is in a coherent state at the initial moment, $Q(z) = \exp(\alpha z)$ we obtain

$$\exp\left[\frac{\beta z^2}{2}\right]\int ds\,\exp\left[izs - \frac{\rho s^2}{2(\nu+2\beta\rho)}\right]\Pi(s) = \exp(\alpha z).$$

Simple calculations yield the following relation for the initial problem:

$$\Pi(s) = \frac{1}{\beta\sqrt{\pi}}\exp\left[\frac{\alpha^2}{2\beta} - \frac{i\alpha s}{\beta} - \frac{s^2}{2\beta} + \rho\frac{s^2}{2(\nu+2\beta\rho)}\right],$$

and, accordingly,

$$\begin{aligned}\varphi(z,t) = &\sqrt{\frac{2\sqrt{\nu^2-4\rho^2}}{\kappa(t)}}\exp\left\{\frac{i\nu t}{2}\right.\\ &\left.+\frac{2\alpha z\sqrt{\nu^2-4\rho^2}\exp(i\nu t) - \alpha^2\rho\chi(t) - \rho z^2\exp(2i\nu t)\chi(t)}{\kappa(t)}\right\},\end{aligned} \tag{18}$$

where

$$\begin{aligned}\chi(t) &= \exp(i\sqrt{\nu^2-4\rho^2}t) - \exp(-i\sqrt{\nu^2-4\rho^2}t),\\ \kappa(t) &= (\nu+\sqrt{\nu^2-4\rho^2})\exp(i\sqrt{\nu^2-4\rho^2}t)\\ &-(\nu-\sqrt{\nu^2-4\rho^2})\exp(-i\sqrt{\nu^2-4\rho^2}t).\end{aligned}$$

2.4. Quantum-Statistical Characteristics of Generated Quantized Radiation

As the initial state of the field, we only consider the vacuum state, setting $\alpha = 0$, because the case of an arbitrary coherent state leads to much more cumbersome results. More precisely, our initial state is described by the wave function

$$\Psi(0) = |0\rangle\mathbf{e}_1,$$

where $|0\rangle$ is the vacuum state of the quantized mode and $\mathbf{e}_1$ is the eigenvector of the matrix $\mathbf{U}$ defined above (i.e., we consider the initial condition with the special vector structure). We will concentrate on purely Fock physical variables calculated by using an appropriate wave function, so that, as was noted above, one can trivially take into account the vector structure of the wave function. We will calculate the number of photons in the quantized mode,

$$n(t) = \langle \mathbf{a}^+ \mathbf{a} \rangle$$

and the quantities

$$D_\pm(t) = \langle (\Delta X_\pm(t))^2 \rangle,$$

where

$$X_+(t) = \frac{\mathbf{a}\exp(i\omega t) + \mathbf{a}^+ \exp(-i\omega t)}{2},$$

and

$$X_-(t) = \frac{\mathbf{a}\exp(i\omega t) - \mathbf{a}^+ \exp(-i\omega t)}{2i}$$

are quadrature components of the quantized field [20]. More precisely,

$$\begin{aligned} D_+(t) &= \frac{1}{4}[2\langle \mathbf{a}^+\mathbf{a}\rangle + 1 + \langle \mathbf{a}^2 \exp(2i\omega t)\rangle \\ &\quad + \langle (\mathbf{a}^+)^2 \exp(-2i\omega t)\rangle \\ &\quad - \langle [\mathbf{a}\exp(i\omega t) + \mathbf{a}^+ \exp(-i\omega t)]\rangle^2], \end{aligned} \tag{19}$$

$$\begin{aligned} D_-(t) &= \frac{1}{4}[2\langle \mathbf{a}^+\mathbf{a}\rangle + 1 - \langle \mathbf{a}^2 \exp(2i\omega t)\rangle \\ &\quad - \langle (\mathbf{a}^+)^2 \exp(-2i\omega t)\rangle \\ &\quad - \langle i[\mathbf{a}\exp(i\omega t) - \mathbf{a}^+ \exp(-i\omega t)]\rangle^2]. \end{aligned} \tag{20}$$

The values of these physical quantities describe the squeezing of the quantized field. One can easily verify that the corresponding Heisenberg inequality is reduced to the inequality

$$D_-(t)D_+(t) \geq 1/16.$$

If one of these quantities is less than 1/4, then the corresponding quantized field is called squeezed. When calculating these quantities, we will use the results of the preceding sections and take into account commutation relation (2). Thus,

$$\begin{aligned} n(t) &= \int dz(d\bar{z}) \exp(-z\bar{z}) \overline{\varphi(z,t)} z \frac{d}{dz} \varphi(z,t) \\ &= \int dz d\bar{z} (z\bar{z} - 1) \exp(-z\bar{z}) \overline{\varphi(z,t)} \varphi(z,t), \end{aligned}$$

$$\begin{aligned} D_\pm(t) &= \frac{1}{4}\int dz d\bar{z} [2z\bar{z} - 1 \pm (z^2 + \bar{z}^2)] \\ &\quad \times \exp(-z\bar{z}) \overline{\varphi(z,t)} \varphi(z,t). \end{aligned}$$

Taking into account (18), we can reduce these integrals to Gaussian ones. We obtain the following results.

(A) If $|\nu| > 2\rho$, then

$$n(t) = \frac{4\rho^2}{\nu^2 - 4\rho^2} \sin^2(\sqrt{\nu^2 - 4\rho^2}\, t), \tag{21}$$

$$\begin{aligned} D_\pm(t) &= \frac{1}{4} \pm \frac{\rho}{4(\nu^2 - 4\rho^2)} \\ &\quad \times \{(\nu + \sqrt{\nu^2 - 4\rho^2})\cos[2(\nu + \sqrt{\nu^2 - 4\rho^2})t] \\ &\quad + (\nu - \sqrt{\nu^2 - 4\rho^2})\cos[2(\nu - \sqrt{\nu^2 - 4\rho^2})t] \\ &\quad - 2\nu\cos(2\nu t)\} + \frac{2\rho^2}{\nu^2 - 4\rho^2} \sin^2(\sqrt{\nu^2 - 4\rho^2}\, t). \end{aligned} \tag{22}$$

(B) If $|\nu| < 2\rho$, then

$$n(t) = \frac{4\rho^2}{4\rho^2 - \nu^2} \{\sinh(\sqrt{4\rho^2 - \nu^2}\, t)\}^2, \tag{23}$$

$$\begin{aligned} D_\pm(t) &= \frac{1}{4} + \frac{2\rho^2}{4\rho^2 - \nu^2} \{\sinh(\sqrt{4\rho^2 - \nu^2}\, t)\}^2 \\ &\quad \mp \frac{\rho \sinh(\sqrt{4\rho^2 - \nu^2}\, t)}{4\rho^2 - \nu^2} \\ &\quad \times \{\sqrt{4\rho^2 - \nu^2}\cosh(\sqrt{4\rho^2 - \nu^2}\, t)\sin 2\nu t \\ &\quad + \nu \sinh(\sqrt{4\rho^2 - \nu^2}\, t)\cos 2\nu t\}. \end{aligned} \tag{24}$$

It follows from relations (21) and (23) that, when $|\nu| > 2\rho$, the population of the quantized mode oscillates and its amplitude may take arbitrarily large values, whereas, when $|\nu| < 2\rho$, the population grows exponentially with time. However, here one should take into account constraint (4). In other words, these results are valid on the time intervals where constraint (4) is satisfied. The behavior of $D_+(t)$ for variations of the parameter $\zeta = |\rho/\nu|$ is illustrated by diagrams in Fig.1. These diagrams show that the quantized radiation is squeezed ($D_+(t) < 0.25$) within certain time intervals and the maximum squeezing increases up to 1 as ζ tends to 1/2 from below and above. The quantity $D_-(t)$ exhibits similar behavior. Note that, as follows from the relations obtained, squeezed radiation can be intense at the same time (i.e., $n(t)$ can assume arbitrarily large values for the squeezed radiation).

The parameter ρ differs from the parameter ζ describing the relation between the quantized mode and an atomic transition in a two-photon resonance only by a factor of the order of unity. Therefore, this parameter

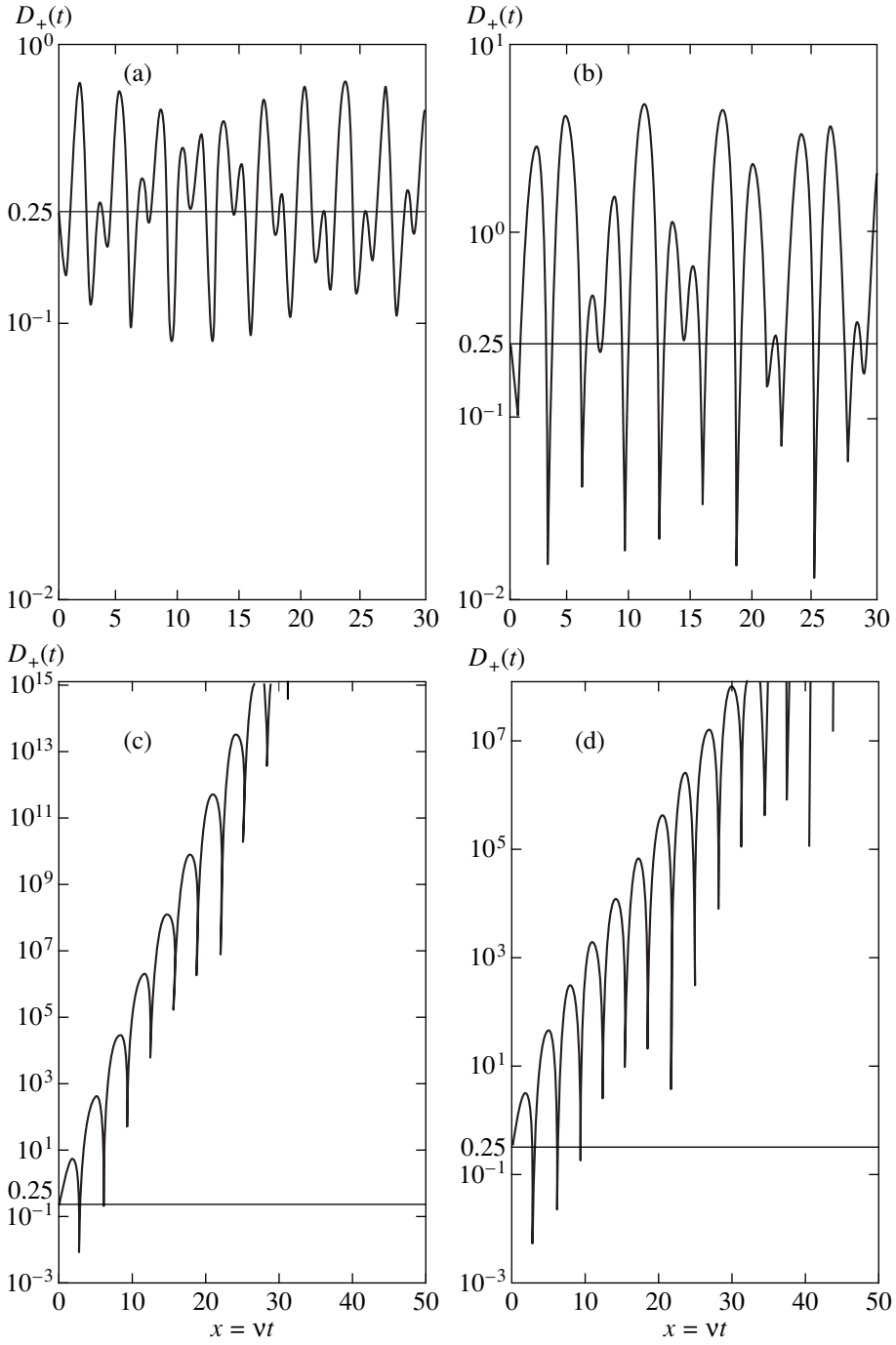


Fig. 1. The dynamics of the quadrature component $D_+(t)$ for various values of the parameter ζ: (a) 0.25, (b) 0.45, (c) 0.60, and (d) 0.52.

is rather small, and it is rather difficult to satisfy condition (B) in practice. Condition (B) can be interpreted as the condition of fine tuning of the frequencies of the classical and quantized modes, whereas the open resonance condition given above (condition 3) makes the sense of coarse tuning condition for these frequencies. The case of N atoms discussed below makes it possible to substantially weaken condition (B).

2.5. A Special Case

A special situation arises when $\nu = 2\varepsilon\rho$, $\varepsilon = \pm 1$. Recall that the parameter ν is related to the frequency of the classical field and, in general, can be chosen arbitrarily. This case can be considered using the passage to the limit as $\zeta \longrightarrow 1/2$. However, one can also apply a direct approach. Repeating the arguments above, we

obtain the following integral representation for the initial problem:

$$\varphi(z, t) = \exp\left[\frac{\varepsilon z^2 \exp(2i\nu t)}{2}\right]$$
$$\times \int \exp\left[izs\exp(i\nu t) + i\rho t s^2 - \frac{i\nu t}{2}\right]\Pi(s)ds,$$

where the function $\Pi(s)$ is determined from the initial data. If the quantized mode was in a coherent state at the initial moment, $\varphi(z, 0) = \exp(\alpha z)$, then

$$\varphi(z, t) = \frac{1}{\sqrt{1 - i\nu t}}$$
$$\times \exp\left\{\frac{2\alpha z\exp(i\nu t) - i\nu t z^2\exp(2i\nu t) - i\nu\alpha^2 t}{2(1 - i\nu t)} - \frac{i\nu t}{2}\right\}$$

(for definiteness, here we set $\varepsilon = 1$). Simple calculations yield the following results in the case $\alpha = 0$:

$$N(t) = \nu^2 t^2,$$
$$D_{\pm}(t) = \frac{1}{4} + \frac{\nu^2 t^2}{2} \tag{25}$$
$$\mp \frac{1}{2}[\nu^2 t^2\cos(2\nu t) + \nu t\sin(2\nu t)].$$

Note that the population of the quantized mode indefinitely increases with time, so that we have to use an appropriate reservation concerning the time intervals on which our results hold. The behavior of $D_+(t)$ in this case is illustrated in Fig. 2. In this case, the squeezing of the generated quantized radiation can also be indefinitely close to 1. One can readily check that expressions (22) and (24) tend to (25) as $\zeta \longrightarrow 1/2$.

Using the results obtained, we can easily describe the solution to the initial problem for $\Psi(0) = |0\rangle\mathbf{e}$, where $\mathbf{e}$ is an arbitrary vector describing the initial state of the atom. This fact will be used below when discussing the dynamics in the case of N atoms.

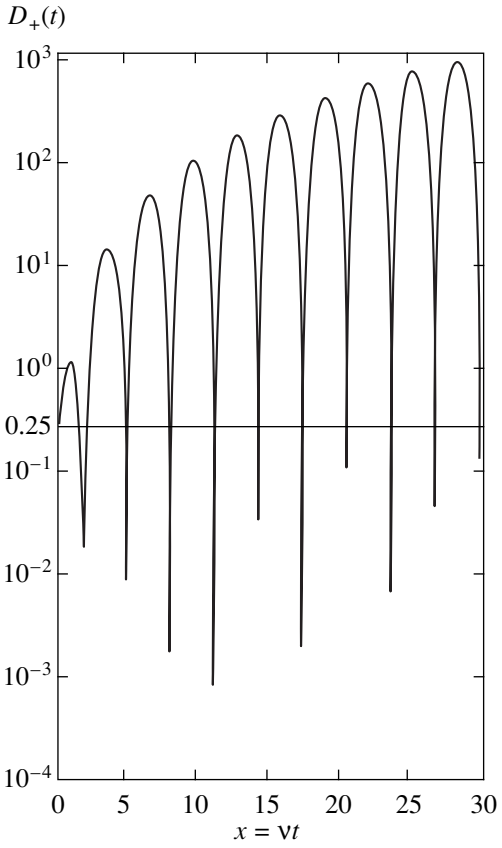


Fig. 2. The dynamics of the quadrature component $D_+(t)$ for the case $\nu = 2\varepsilon\rho$, $\varepsilon = \pm 1$; the parameter $\zeta = 0.50$.

3. COLLECTIVE PHENOMENA

3.1. Basic Notations

As we showed in the preceding section, in the case of simultaneous interaction of an atom with quantized and classical fields, the appropriated modes exchange photons, while the atom plays the role of a certain intermediary of photons. In this case, there exists a range of parameters where the number of photons in the quantized modes grows exponentially. As we noted above, condition (B) under which this phenomenon occurs imposes constraints during practical implementation. In this section, we consider a situation when N identical two-level atoms simultaneously interact with a quantized mode and a classical field; here, a two-photon resonance takes place with the quantized mode, whereas the classical field is quasi-resonant with the atomic transition. We assume that the atoms are situated at small distances as compared with the wavelengths of our modes, so that all atoms identically interact with optical fields. Here, we are interested in cooperative phenomena in this system. We will demonstrate that condition (B) in this case can be substantially weakened.

Thus, in this case, the wave function takes values in the space

$$L = F \otimes C^2 \otimes C^2 \ldots \otimes C^2,$$

where the Fock space F describes the states of the quantized mode and N copies of the space C^2 describe the states of the atoms. The space L can be represented as a linear envelope of vectors

$$n = f|e_{k_1}, e_{k_2}, \ldots, e_{k_N}\rangle,$$

where e_{k_n} are the eigenvectors of the matrix $\mathbf{U}\mathbf{J}_0\mathbf{U}^{-1}$ described above and $k_m = 1, 2$.

Define the operators $\mathbf{J}_0^{(m)}$, $\mathbf{J}_-^{(m)}$, $\mathbf{J}_+^{(m)}$, and $\Xi^{(m)}(t)$, $1 \le m \le N$ as follows: they act on the mth component of the vector $|\nu_1, \nu_2, \ldots, \nu_N\rangle$ as operators $\mathbf{J}_0$, $\mathbf{J}_-$, $\mathbf{J}_+$, and $\Xi(t)$ respectively, and do not act on other components. Then, under our assumptions (within the rotating wave

approximation), the Hamiltonian of the system describing the interaction of N atoms with quantized and classical fields can be expressed as follows:

$$\mathbf{H}_N = \omega \mathbf{a}^+ \mathbf{a} + \kappa \sum_{m=1}^{N} \mathbf{J}_0^{(m)} + \zeta \sum_{m=1}^{N} \{(\mathbf{a}^+)^2 \mathbf{J}_-^{(m)} + \mathbf{a}^2 \mathbf{J}_+^{(m)}\} + \mu \sum_{m=1}^{N} [\mathbf{J}_-^{(m)} \exp(i\Omega t) + \mathbf{J}_+^{(m)} \exp(-i\Omega t)]. \quad (26)$$

We will focus on the evolution of the wave function $\Psi(t)$, a solution to the appropriate Schrödinger equation. Let us separate the optical frequency using an analogue of relation (5):

$$\Psi(t) = \exp\left[-i\omega t\left(\mathbf{a}^+ \mathbf{a} + \sum_{m=1}^{N} \mathbf{J}_0^{(m)}\right)\right]\Phi(t).$$

For the function $\Phi(t)$, we obtain the equation

$$i\frac{\partial \Phi(t)}{\partial t} = \left\{(\kappa - \omega)\sum_{m=1}^{N} \mathbf{J}_0^{(m)} + \zeta \sum_{m=1}^{N} [(\mathbf{a}^+)^2 \mathbf{J}_-^{(m)} + \mathbf{a}^2 \mathbf{J}_+^{(m)}] + \mu \sum_{m=1}^{N} [\mathbf{J}_-^{(m)} \exp(i(\Omega - 2\omega)t) + \mathbf{J}_+^{(m)} \exp(i(2\omega - \Omega)t)]\right\}\Phi(t).$$

The matrices $\Xi^{(m)}(t)$, $m = 1, 2, \ldots, N$, commute. Introduce the matrix

$$\Xi_N(t) = \prod_{m=1}^{N} \Xi^{(m)}(t).$$

Here, $\Xi_N(0)$ is the identity matrix in

$$C^2 \otimes C^2 \ldots \otimes C^2,$$

and $\Xi_N(t)$ is a solution to the equation

$$i\frac{\partial}{\partial t}\Xi_N(t) = \left\{(\kappa - \omega)\sum_{m=1}^{N} \mathbf{J}_0^{(m)} + \mu \sum_{m=1}^{N} [\mathbf{J}_-^{(m)} \exp(i(\Omega - 2\omega)t) + \mathbf{J}_+^{(m)} \exp(i(2\omega - \Omega)t)]\right\}\Xi_N(t).$$

Substituting $\Phi(t) = \Xi_N(t)\varphi(t)$, we arrive at the equation for the function $\varphi(t)$:

$$i\frac{\partial \varphi(t)}{\partial t} = \zeta \Xi_N^{-1}(t) \times \sum_{m=1}^{N} [(\mathbf{a}^+)^2 \mathbf{J}_-^{(m)} + \mathbf{a}^2 \mathbf{J}_+^{(m)}]\Xi_N(t)\varphi(t). \quad (27)$$

Here, $\varphi(0) = \Phi(0)$. The operator

$$\exp\left[-i\omega t\left(\mathbf{a}^+ \mathbf{a} + \sum_{m=1}^{N} \mathbf{J}_0^{(m)}\right)\right]\Xi_N(t)$$

is unitary and commutes with the operator $\mathbf{a}^+\mathbf{a}$. Thus, for the purely Fock operator $G = (\mathbf{a}^+\mathbf{a})^m$, we obtain

$$\langle G \rangle = \langle \overline{\Psi(t)}, G\Psi(t) \rangle = \langle \overline{\varphi(t)}, G\varphi(t) \rangle.$$

3.2. Slow Dynamics

As in the case of a single atom, the right-hand side of Eq. (27) contains fast and slow oscillations. Deleting fast oscillations, we obtain an equation describing the slow evolution of the system during the interaction of atoms with a quantized field:

$$i\frac{\partial \varphi(t)}{\partial t} = \zeta \left\langle\!\!\left\langle \Xi_N^{-1}(t) \sum_{m=1}^{N} [(\mathbf{a}^+)^2 \mathbf{J}_-^{(m)} + \mathbf{a}^2 \mathbf{J}_+^{(m)}]\Xi_N(t)\varphi(t) \right\rangle\!\!\right\rangle.$$

As before, we discuss here the case

$$|2\omega - \Omega| = 2|\nu| \ll R,$$

where R is the Rabi parameter of the classical field. This condition imposes a constraint on the difference between the frequencies of the classical and quantized modes. In this case, the averaging procedure yields a nontrivial (i.e., nonzero in the leading asymptotic order) contribution; thus, we obtain

$$\mathbf{H}_{av} = \rho[(\mathbf{a}^+)^2 \exp(2i\nu t) + \mathbf{a}^2 \exp(-2i\nu t)] \times \sum_{m=1}^{N} \mathbf{U}_m \mathbf{J}_0^{(m)} \mathbf{U}_m^{-1},$$

where the matrix $\mathbf{U}_m$ acts as matrix $\mathbf{U}$ on the mth component of the wave function and does not act on other components. This Hamiltonian governs the evolution of a system consisting of atoms dressed by the field and a quantized component. As for a single atom, this Hamiltonian is a product of the Fock operator

$$(\mathbf{a}^+)^2 \exp(2i\nu t) + \mathbf{a}^2 \exp(-2i\nu t)$$

and a matrix operator. Thus, after passing to the expansion over the vector basis $|\mathbf{e}_{k_1}, \mathbf{e}_{k_2}, \ldots, \mathbf{e}_{k_N}\rangle$, which are the eigenvectors of the matrix

$$\sum_{m=1}^{N} \mathbf{U}_m \mathbf{J}_0^{(m)} \mathbf{U}_m^{-1},$$

the averaged Hamiltonian splits into a set of one-dimensional Hamiltonians. Accordingly, we represent the required wave function as follows:

$$\varphi(t) = \sum_{\sigma} \eta_\sigma(t) |\mathbf{e}_{k_1}, \mathbf{e}_{k_2}, \ldots, \mathbf{e}_{k_N}\rangle,$$

where the functions $\eta_\sigma(t)$ take values in a Fock space, σ denotes a set of N numbers $k_1, k_2, \ldots, k_N$ each of which assumes the values 1 or 2, and the summation is over all such sets σ. For each $\eta_\sigma(t)$, we obtain

$$i\frac{\partial \eta_\sigma(t)}{\partial t} = \rho S_\sigma \times [(\mathbf{a}^+)^2 \exp(2i\nu t) + \mathbf{a}^2 \exp(-2i\nu t)]\eta_\sigma(t), \tag{28}$$

where

$$S_\sigma = \sum_{m=1}^{N} (-1)^{k_m + 1},$$

here, the summation is performed over all number k_m that form a set σ. Thus, S_σ denotes the summation over m eigenvalues of the operators $\mathbf{U}_m \mathbf{J}_0^{(m)} \mathbf{U}_m^{-1}$, these eigenvalues being equal to ± 1. Note that only coefficient S_σ in Eq. (28) depends on σ. A solution to the initial problem for Eq. (28) with $\eta_\sigma(z, 0) = Q_\sigma(z)$ is described by an obvious modification of formulas (16) and (17). Thus, under our assumptions, the initial problem can be solved in analytic terms for any initial data in the case of N atoms as well.

3.3. Cooperative Behavior

Let us discuss the following initial problem: at the initial moment, a quantized mode is in the vacuum state, and all atoms are in the same initial state $v \in C^2$ such that

$$v = \mathbf{e}_1 \cos\chi + \mathbf{e}_2 \sin\chi.$$

Expanding the wave function of the system over the above basis, we arrive at the following initials for the individual components of this expansion:

$$\eta_\sigma(z, 0) = (\cos\chi)^{N-k} (\sin\chi)^k$$

(here, k denotes the number of twos in the set σ). For a given set σ, we denote this number by $|\sigma|$. Then,

$$\eta_\sigma(z, t) = (\cos\chi)^{N-k} (\sin\chi)^k \varphi_\sigma(z, t),$$

where the function $\varphi_\sigma(z, t)$ is defined by (18) with regard to the substitution $\rho \longrightarrow \rho S_\sigma$. Note that, if $|\sigma| = k$, then $S_\sigma = N - 2k$ and, for a given k, there exist $\binom{N}{k}$ different sets σ with fixed $|\sigma|$. According to the previous results, we obtain the following expression for the Fock operator $G = (\mathbf{a}^+\mathbf{a})^m$ within the Fock–Bargmann representation:

$$\begin{aligned}
\langle G \rangle &= \langle \overline{\varphi(z, t)}, G\varphi(z, t) \rangle \\
&= \sum_{\sigma} \int dz d\bar{z} \exp(-z\bar{z}) \overline{\eta_\sigma(z, t)} G \eta_\sigma(z, t) \\
&= \sum_{k=0}^{N} \sum_{|\sigma| = k} \int dz d\bar{z} \exp(-z\bar{z}) \overline{\eta_\sigma(z, t)} G \eta_\sigma(z, t) \\
&= \sum_{k=0}^{N} \binom{N}{k} (\cos\chi)^{2N-2k} (\sin\chi)^{2k} \\
&\quad \times \int dz d\bar{z} \exp(-z\bar{z}) \overline{\varphi_\sigma(z, t)} G \varphi_\sigma(z, t)|_{|\sigma| = k}.
\end{aligned}$$

Here, we have availed ourselves of the fact that the coefficients $\eta_\sigma(z, t)$ with identical S_σ coincide.

Using our results, let us determine the number of photons in the quantized mode. The appropriate integrals have already been calculated earlier, and we obtain

$$\begin{aligned}
\langle n(t) \rangle &= \sum_{k=0}^{N} \binom{N}{k} (\cos\chi)^{2N-2k} (\sin\chi)^{2k} \\
&\quad \times \frac{4\rho^2 (N-2k)^2}{4\rho^2 (N-2k)^2 - \nu^2} \\
&\quad \times [\sinh(\sqrt{4\rho^2 (N-2k)^2 - \nu^2}\, t)]^2.
\end{aligned} \tag{29}$$

Note that we used formula (23), which corresponds to the case

$$4\rho^2 (N-2k)^2 - \nu^2 > 0.$$

As follows from our results, other situations can be considered with the use of a procedure of analytic continuation of this relation.

The analytical investigation of the sum on the right-hand side of the last formula presents a difficult problem. However, this problem can be simplified when the number N is large. In this case, we can rewrite the sum as an integral after the substitutions

$$k = N\left(\frac{1-\xi}{2}\right), \quad d\xi \approx \frac{1}{N}, \quad \nu = 2N\rho\theta.$$

This transformation preserves the leading term of the asymptotic for large N. Using the Stirling formula for a gamma function, for $N \gg 1$ we obtain

$$\langle n(t)\rangle = \frac{1}{4}\sqrt{\frac{2N}{\pi}}|\sin\chi\cos\chi|^N \int_{-1}^{1} \frac{\xi^2 d\xi}{(\xi^2-\theta^2)\sqrt{1-\xi^2}} \times \{\exp[N\mu_+(\xi,t)] + \exp[N\mu_-(\xi,t)] - 2\exp[N\mu_-(\xi,0)]\}, \quad (30)$$

$$\mu_\pm(\xi,t) = -\frac{1}{2}\left[(1+\xi)\ln\left(\frac{1+\xi}{2}\right) + (1-\xi)\ln\left(\frac{1-\xi}{2}\right)\right] + \eta\xi \pm 4\rho t\sqrt{\xi^2-\theta^2},$$

where $\eta = \ln|\cot\chi|$. Thus, we have to calculate three integrals with a large parameter in the argument of the exponential function in the expression under the integral sign. To this end, we can apply the Laplace method. In this paper we restrict the analysis to the situation when the parameter θ is small. Recall that, for a single atom, it is a similar situation where we observed exponentially growing populations of the quantized mode.

Differentiating the function $\mu_\pm(\xi, t)$ with respect ξ, we obtain the following equations for the critical points $\xi_\pm(t)$:

$$\frac{d}{d\xi}\mu_\pm(\xi_\pm,t) = -\frac{1}{2}\ln\left(\frac{1+\xi_\pm}{1-\xi_\pm}\right) \pm \frac{4t\rho\xi_\pm}{\sqrt{\xi_\pm^2-\theta^2}} + \eta = 0. \quad (31)$$

These are transcendental equations that cannot be solved explicitly in the general case. However, we can analyze their solutions for small values of the parameter θ. For definiteness, let us set $\eta > 0$ and consider first the equation for $\xi_+(t)$. Neglecting the terms of order θ^2, we obtain

$$\xi_+(t) \approx \frac{\exp(2\eta + 8t\rho) - 1}{\exp(2\eta + 8t\rho) + 1}.$$

As t increases, $\xi_+(t)$ increases and tends to 1. Note that

$$\frac{d^2}{d\xi^2}\mu_+(\xi_+,t) \approx \frac{-1}{1-\xi_+^2} < 0,$$

so that the point ξ_+ maximizes the function $\mu_+(\xi, t)$ and, according to the Laplace method, determines the asymptotics in N of the corresponding integral.

Next, let us consider the behavior of the solution $\xi_-(t)$ as t increases. The calculations show that this quantity decreases starting from the value $\xi_-(0) = \eta$, becomes complex at $t \sim \eta/4\rho$, and tends to zero as t increases further. For complex values of $\xi_-(t)$, one should apply the saddle-point method instead of the Laplace method. However, even these qualitative considerations show that the main contribution to (30) is made by the first integral. Neglecting the terms of order θ^2, we can assume that

$$\mu_\pm(\xi_\pm,t) \approx -\frac{1}{2}\ln\left(\frac{1-\xi_\pm^2}{4}\right) = \ln[\exp(\eta \pm 4t\rho) + \exp(-\eta \mp 4t\rho)]$$

to a sufficient degree of accuracy. As a result, we obtain

$$\langle n(t)\rangle \approx \frac{1}{2}\left\{[\exp(4t\rho)\cos^2\chi + \exp(-4t\rho)\sin^2\chi]^N \times \frac{\xi_+^2(t)}{\xi_+^2(t)-\theta^2} + [\exp(-4t\rho)\cos^2\chi + \exp(4t\rho)\sin^2\chi]^N \times \left|\frac{\xi_-^2(t)}{\xi_-^2(t)-\theta^2}\right| - \frac{2\xi_-^2(0)}{\xi_-^2(0)-\theta^2}\right\}. \quad (32)$$

Recall that this result is obtained for $\eta > 0$, i.e., for

$$\cos^2\chi > \sin^2\chi.$$

The opposite case is analyzed similarly. The parameter χ is defined by the initial states of the atoms and the parameters of the classical field—its amplitude and the frequency detuning from the atomic transition frequency. Thus, the quantity χ can easily be set and controlled in the experiment.

It follows from (32) that the number of photons in the quantized mode exponentially depends on the number of two-level atoms N and behaves as $\exp(4N\rho t)$ for sufficiently large t; i.e., we observe exponential superradiation. Naturally, our results in this case are valid only on the time intervals where the basic relation (4) is satisfied, i.e., under the assumption that the classical field is much more intense than the quantized field.

The origin of this phenomenon is rather clear on the formal level. One can easily see that the parameter ρ enters in the argument of sine function in (21) and in its analogs for a single atom. With regard to the substitution $\rho \longrightarrow \rho S_\sigma$ (which corresponds to the transition form the case of a single atom to N atoms), this fact gives rise to the parameter N in the argument of sine in the expression (29) for the population of the quantized mode in the case of N atoms. This fact contrasts with the situation of a one-photon interaction of atoms with a quantized mode and classical fields, which was discussed in [7, 8], where only the coefficient of the oscillating harmonic depended on the parameter N.

From the physical viewpoint, the origin of this phenomenon lies in the cooperative interaction of atoms dressed by the field, which start to jointly transfer photons from the classical to the quantized mode. As we have shown above, when $\theta \ll 1$, this process becomes exponential; actually, it is sufficient to consider a con-

dition of the type $\theta \sim 0.1$. Recall that $\theta = \nu(2N\rho)^{-1}$, so that the latter condition implies

$$|\nu| = |\omega - \Omega/2| \le 0.2N\rho. \quad (33)$$

It is obvious that this constraint (for large N) is considerably weaker than the condition $|\nu| < 2\rho$ under which the population of the quantized mode exponentially grows in the case of a single atom. Recall that this condition describes the fine tuning of the frequency of the fields, while the coarse tuning is described by (4).

In addition, for real system, constraint (4) is necessary only for time intervals comparable with the relaxation time τ_{rel} of our physical system, which is determined by the lifetime of a photon in the resonator and the decay time of the atomic states involved. If this condition is not satisfied during τ_{rel}, we can observe only the initial stage of the exponential growth of the population of the quantized mode until this constraint is violated (under the assumption that the observation time is greater than τ_{rel}).

4. CONCLUSION

Let us summarize our results. We have discussed the dynamics of a system containing one or many ($N \gg 1$) two-level atoms that simultaneously interact with a quasi-resonant classical field and are in the two-photon resonance with a quantized mode. Here, we assume that the classical field is much more intense than the quantized one. We investigated the situation of open resonances when the photon exchange between the classical and quantized modes occurs without preservation of the number of excitations.

We have demonstrated that, for the interaction of a single atom with quantized and classical fields, one can develop conditions for a squeezed radiation in the quantized mode, where the squeezing can be made arbitrarily close to unity. Here, the amplitude of the quantized mode can be (relatively) large. Moreover, there are situations when the amplitude of the quantized mode grows exponentially with time. However, for a single atom, this situation is associated with a sufficiently fine condition for the tuning of the frequency of optical fields. For the case of $N \gg 1$ identical two-level atoms that simultaneously interact with quantized and classical fields, situations are also possible when the amplitude of the quantized mode exponentially grows with time. In this case, the exponent is a linear function of the number N, and the condition (33) for the frequencies of the fields is less restrictive by a factor of N. This fact gives us grounds to refer to this phenomenon as exponential superradiation. Recall that, for an ordinary superradiation, the number of photons of the excited field is proportional to N^2 (see [21] and the discussion of the one-photon variant of our problem in [7, 8]).

In conclusion, note that the tensor structure of the wave function and quantum interference, which actually underlie the phenomenon of exponential superradiation described above, also play a decisive role in exponentially fast algorithms for quantum computations [22].

REFERENCES

1. C. K. Law and J. H. Eberly, Phys. Rev. A **43**, 6337 (1991).
2. P. Alsing, D. S. Guo, and H. J. Carmichael, Phys. Rev. A **45**, 5135 (1992).
3. I. V. Jyotsna and G. S. Agarwal, Opt. Commun. **99**, 344 (1993).
4. M. Z. Smirnov, Zh. Éksp. Teor. Fiz. **112**, 818 (1997) [JETP **85**, 441 (1997)].
5. M. Z. Smirnov, Quantum Semiclassic. Opt. **10**, 765 (1998).
6. A. Ya. Kazakov, Quantum Semiclassic. Opt. **10**, 753 (1998).
7. R. A. Ismailov and A. Ya. Kazakov, Zh. Éksp. Teor. Fiz. **116**, 858 (1999) [JETP **89**, 454 (1999)].
8. R. A. Ismailov and A. Ya. Kazakov, Zh. Éksp. Teor. Fiz. **118**, 798 (2000) [JETP **91**, 691 (2000)].
9. F.-L. Li and S.-Y. Gao, Phys. Rev. A **62**, 043809 (2000).
10. A. Joshi, Phys. Rev. A **62**, 043812 (2000).
11. A. Ya. Kazakov, J. Opt. B **3**, 97 (2001).
12. A. Ya. Kazakov, Submitted to Int. J. Theor. Phys. Groups Theory Nonlinear Opt. (2001).
13. B. Buck and C. V. Sukumar, Phys. Lett. A **81**, 132 (1981).
14. Fam Le Kien, M. Kozierowski, and Trang Quang, Phys. Rev. A **38**, 263 (1988).
15. S. J. Phoenix and P. L. Knight, J. Opt. Soc. Am. B **7**, 116 (1990).
16. T. Nasreen and M. S. K. Razmi, J. Opt. Soc. Am. B **10**, 1292 (1993).
17. M. Kozierowski, Phys. Rev. A **47**, 723 (1993).
18. T. Nasreen and K. Zaheer, Phys. Rev. A **47**, 616 (1994).
19. A. M. Perelomov, *Generalized Coherent States and Their Applications* (Nauka, Moscow, 1987; Springer-Verlag, New York, 1986).
20. E. I. Aliskenderov, A. S. Shumovskiĭ, and Ho Trung Dung, Fiz. Élem. Chastits At. Yadra **24**, 409 (1993) [Phys. Part. Nucl. **24**, 177 (1993)].
21. A. V. Andreev, V. I. Emel'yanov, and Yu. A. Il'inskiĭ, *Cooperative Phenomena in Optics* (Nauka, Moscow, 1988).
22. P. W. Shor, SIAM J. Comput. **26**, 1484 (1997).

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 93, No. 5, 2001, pp. 1028–1034.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 5, 2001, pp. 1185–1193.

ATOMS, SPECTRA, RADIATION

Population Inversion on Transitions to the Ground State of Atoms upon Nonresonance Absorption of Laser Radiation

R. V. Markov, A. I. Plekhanov, and A. M. Shalagin*

Institute of Automation and Electrometry, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia
**e-mail: shalagin@iae.nsk.su*

Received June 27, 2001

Abstract—Lasing at the resonance transitions (D_1- and D_2-lines) of sodium was observed in the superradiance mode upon nonresonance optical excitation in the presence of a buffer gas. The dependences of the lasing intensity on the exciting radiation intensity and on the detuning of its frequency from the frequencies of resonance transitions were studied. It is found that, under specific conditions of the experiment (high pressure of a buffer gas and a rather high radiation intensity), in the case of a large positive detuning of the exciting radiation frequency from the resonance ("working") transition frequency, the population inversion is produced at the "working" transition, which results in lasing. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Collision relaxation plays an important role in the spectral redistribution of quasi-resonance radiation interacting with gaseous media. The study of such processes is important both from the fundamental and practical points of view. In particular, it is connected with attempts to obtain new methods for lasing and to extend its spectral region. Of special interest is the possibility of lasing at the transitions to the ground state of atoms or molecules or to the lowest energy levels that are maximally populated in the absence of perturbation. It is with this situation that the hope to obtain an extremely short-wavelength is associated.

Lasing on the D_1-line of alkali metals such as potassium [1], sodium [2], and rubidium [3] in vapors mixed with a buffer gas has already been demonstrated. The lasing was observed upon resonance excitation of the D_2-line and was achieved due to frequent collisions with buffer-gas particles, which managed to establish the Boltzmann population distribution among the fine-structure components ($P_{3/2}$ and $P_{1/2}$) during the exciting radiation pulse and the spontaneous relaxation time. In this case, the population of the $P_{1/2}$ level proves to be somewhat higher (by the Boltzmann factor) than that of the $P_{3/2}$ level. The exciting-radiation intensity was so high that it provided the equalization of populations of the ground $S_{1/2}$ and excited $P_{3/2}$ levels. Under these conditions, the population of the $P_{1/2}$ level is higher (by the same Boltzmann factor) than populations of both the $P_{3/2}$ and $S_{1/2}$ levels. Therefore, the population inversion is produced at the transition from the $P_{1/2}$ level to the ground $S_{1/2}$ state, resulting in lasing on the D_1-line. Under real experimental conditions [1, 2], lasing occurred in the superradiance mode.

In the next paper [4], we obtained lasing at resonance lines by choosing the appropriate polarization of pumping and using collisions with a buffer gas. In these studies, we used a more powerful laser setup, which allowed us to vary experimental conditions to a much greater extent. We have found unexpectedly in our experiments that coherent radiation at the D_1-line could be obtained not only upon tuning the pump frequency in resonance with the D_2-line but also upon a rather large (up to 250 cm^{-1}) detuning to blue. In this paper, we studied this effect experimentally and interpreted it theoretically.

2. EXPERIMENTAL SETUP

Figure 1 shows the scheme of the experimental setup. A tunable Rhodamine 6G laser was excited by a second harmonic from a Nd^{3+} : YAG laser.

A grazing-incidence diffraction grating was used as a dispersion element in the dye laser. The dye laser had a pulse power up to 15 kW, 5-ns pulse duration, a pulse repetition rate of 10 Hz, and emitted linearly polarized radiation tunable from 578 to 593 nm. The emission spectrum of the dye laser consisted of a powerful narrow line of width 0.3 cm^{-1} and a weak broad luminescence band of Rhodamine 6G. The spectral density of the broadband luminescence was three orders of magnitude lower than that of the narrow line. The radiation from the dye laser was focused by a lens with the focal distance 55 cm to the center of the heated region of a working cell. The radiation intensity at the focus was 10 MW/cm^2 and could be varied with the help of neutral filters. Sodium vapor was produced inside a glass cell of diameter 1.2 cm with a heated region of length 5 cm. The sodium vapor density ($N \sim 10^{14}$ cm^{-3}) was controlled by the vapor temperature, which was mea-

sured with a thermocouple. The pressure of a buffer gas (helium) was varied from 10 to 800 Torr. The radiation emerging from the cell was focused on the entrance slit of a Ramanor HG.2S monochromator (the width of the instrumental function was 0.5 cm^{-1}) and detected with a photomultiplier connected with an amplifier and an integrator. The output signals from the photomultiplier were fed to a computer, where they were accumulated and averaged over a large number of measurements. The spectrum of laser emission was detected point by point with a step of 0.5 cm^{-1}. The monochromator was calibrated using a sodium lamp.

3. EXPERIMENTAL RESULTS

The initial aim of our experiments was to study in detail coherent radiation at the D_1-line of sodium (the $3P_{1/2}$–$3S_{1/2}$ transition) in the forward and backward directions relative to the exciting radiation whose frequency was tuned in the vicinity of the D-lines of Na.

We found that such radiation did not appear in the absence of a buffer gas irrespective of the experimental conditions.

When the intensity of laser radiation was $I_L \approx 1$–2 MW/cm^2 and higher, the pressure p_{He} of helium exceeded 200 Torr, and the concentration of sodium vapors was $N \sim 10^{14}$ cm^{-3}, coherent radiation at the D_1-line was observed together with transmitted laser radiation whose frequency was varied in the vicinity of the D_2-line. The divergence of coherent radiation did not exceed that of the pump radiation. Coherent radiation at the D_1-line was detected not only in the direction of the pump radiation but in the opposite direction as well.

Under experimental conditions used in paper [2], we obtained the same results as in [2]. Namely, the threshold value of the helium pressure was 200 Torr, and the intensity of lasing at the D_1-line increased monotonically with increasing pressure to the atmospheric value. The lasing intensity was maximal when the pump frequency ω was exactly tuned to the frequency (ω_{D_2}) of the $3P_{3/2}$–$3S_{1/2}$ transition. The optimum concentration of sodium vapors was approximately 2×10^{14} cm^{-3} for a pump intensity equal to 3 MW/cm^2.

Under optimal conditions, the radiation power in the D_1-line was approximately 3.5% of the pump power (only forward radiation was taken into account). We estimated the gain at the $3P_{1/2}$–$3S_{1/2}$ transition at the length $L \approx 5$ cm of the heated region as $\alpha \approx 3$–3.5 cm^{-1}. In this case, the parameter αL amounts to 15–18, and we can speak already about superradiance if we take into account that it is induced by luminescence of the dye (broadband component of the exciting radiation). The observation of narrow-band backward radiation confirms the existence of superradiance. The intensity

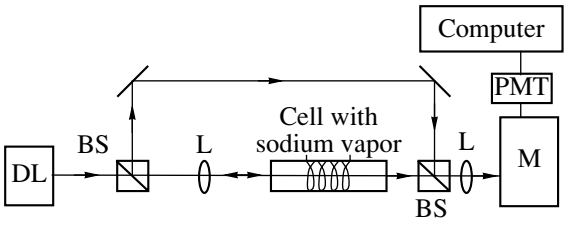


Fig. 1. Scheme of the experimental setup: (DL) pulsed dye laser; (BS) beamsplitter; (L) lens; (M) monochromator; (PMT) photomultiplier.

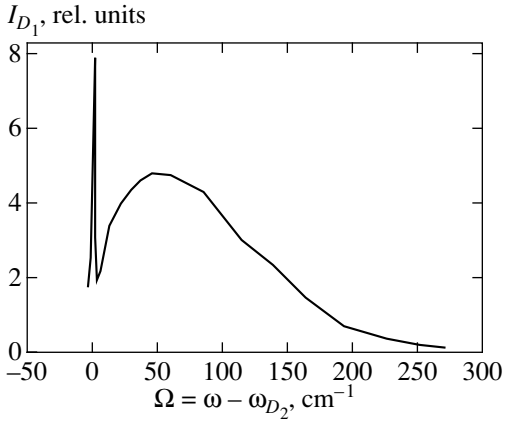


Fig. 2. Dependence of the radiation intensity I_{D_1} at the D_1-line on the frequency detuning $\Omega = \omega - \omega_{D_2}$. $I_L \approx 10$ MW/cm^2, p_{He}= 600 Torr, $N \approx 2 \times 10^{14}$ cm^{-3}.

of backward radiation is several tens times lower due to the absence of external seed radiation in this direction.

Upon excitation at higher powers (up to $I_L \sim$ 10 MW/cm^2), the following picture was observed (Fig. 2). Near zero detunings $\Omega = \omega - \omega_{D_2} \approx 0$, a sharp peak was observed in the dependence $I_{D_1}(\Omega)$. As the detuning Ω increases, the intensity I_{D_1} first gradually decreases and then suddenly increases again up to $\Omega \approx 60$ cm^{-1}. Then, the value of I_{D_1} decreases with increasing Ω; however, the signal of amplified coherent emission at the D_1-line can be reliably detected up to $\Omega \approx 250$ cm^{-1}. In addition, coherent emission also appears at the D_2-line at large detunings. Note that emission at the D_1- and D_2-lines is observed both in the forward and backward directions relative to the pump radiation. We specially measured the degree of spatial coherence of this emission and found that its divergence does not exceed that of the pump radiation. The efficiency of excitation of this emission at large detunings strongly depended on the intensity I_L. As I_L decreased by half, the region of detunings Ω where the effect was observed reduced by half, while the maximum intensity of coherent emission

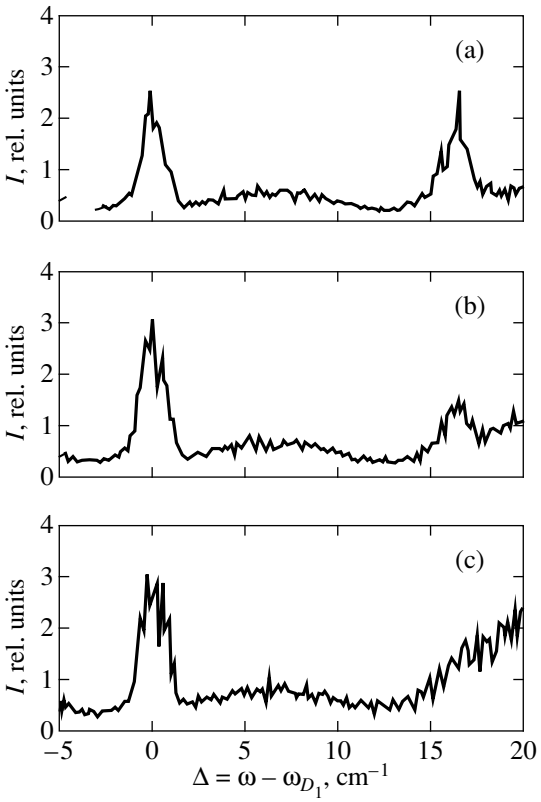


Fig. 3. Spectrograms of generated radiation $I(\Delta)$ for different detunings of the pump frequency $\Omega = \omega - \omega_{D_2} \approx 60$ (a), 33.5 (b), and 20 cm^{-1} (c); $I_L \approx 10$ MW/cm^2, $p_{He} = 600$ Torr, $N \approx 2 \times 10^{14}$ cm^{-3}.

decreased by a factor of five. The effect virtually disappeared at $I_L \approx 3$ MW/cm^2. Note an important fact that when the dye laser frequency was tuned in the region below the frequency of the D_1-line, no coherent emission was observed at resonance lines. We observed neither Raman scattering nor three-photon lines.

Figure 3 shows typical spectrograms of generated emission, which were obtained for some characteristic pump frequencies. On the abscissa is plotted the frequency of the detected emission. Zero corresponds to the position of the D_1-line of sodium. Upon the detuning $\Omega = 60$ cm^{-1} (Fig. 3a), coherent emission was observed both at the D_1-line and the D_2-line of approximately the same intensity (two peaks at $\Delta = 0$ and $\Delta = 17.2$ cm^{-1}). The nonzero signal between the peaks is caused by a broadband luminescence of Rhodamine 6G. When the pump frequency approached the D_2-line frequency, coherent emission at this frequency disappeared and only coherent emission at the D_1-line remained (Fig. 3c). As the pump frequency was further decreased, the D_1-line emission also disappeared at $\Omega = -5.5$ cm^{-1}. The increase in the signal at the right of Fig. 3c is caused by pump radiation.

4. POPULATION INVERSION IN A TWO-LEVEL SYSTEM UPON CW LASER EXCITATION

A new experimental result that we obtained, namely, the observation of coherent emission at resonance lines of sodium for large detunings of the pump frequency from the resonance, is not consistent with theoretical models available and requires special treatment.

The fact that nonresonance pumping produces amplified coherent emission both at the $3P_{1/2}$–$3S_{1/2}$ and $3P_{3/2}$–$3S_{1/2}$ transitions suggests that this effect can exist for a simple two-level system as well. In addition, at buffer-gas pressures used in our experiments, there is no essential difference between pulsed and cw pumping because the population distribution that is established during the pump pulse coincides with that produced upon cw pumping.

Therefore, the problem can be formulated as follows. Consider a gas of two-level atoms in the cw laser radiation field with frequency far from the resonance frequency. Consider also a buffer gas at a high pressure. Collisions with buffer-gas particles cause no inelastic transitions between the combining levels 1 and 2 of absorbing atoms. We should find out whether emission can be amplified at the transition frequency ω_{21} under these conditions. According to the standard theory (see, for example, [5]), the population difference under the above conditions is described by expressions (the Doppler broadening at high pressures of the buffer gas is negligible)

$$N_1 - N_2 = \frac{N}{1+\kappa},$$
$$\Omega = \omega - \omega_{21}, \quad \kappa = \frac{4|G|^2\Gamma}{\Omega^2\Gamma_2}, \quad G \equiv \frac{d_{21}\mathscr{E}_0}{2\hbar}, \tag{1}$$

where Γ_2 is the excited-state radiative decay; Γ is the collision half-width of an absorption line; N is the concentration of absorbing atoms; N_1 and N_2 are the populations of levels 1 and 2; $\mathscr{E}_0$ is the amplitude of the radiation electric field; and d_{21} is the matrix element of the dipole moment of the 2–1 transition. In (1), the condition $|\Omega| \gg \Gamma$ was used. According to expressions (1) and adopted concepts, continuous radiation cannot produce the population inversion, being capable of equalizing the level populations at most. In principle, the situations are possible (see, for example, [6]) in which spectral intervals exist due to the so-called nonlinear interference effects where a weak emission can be amplified even in the absence of the population inversion. However, this effect is not observed under the conditions discussed above. Expression (1) for populations proves to be no longer valid under our specific experimental conditions (high-pressure buffer gas, high-intensity emission, and nonresonance conditions) and should be corrected.

Note first of all that the effective saturation parameter k in (1) at a high pressure of the buffer gas can be

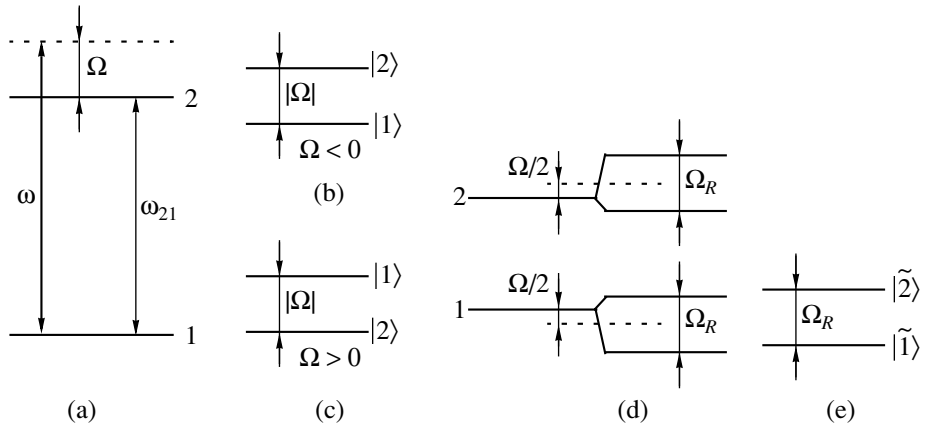


Fig. 4. Energy level diagrams for different basis wave functions.

large under the condition $|\Omega| \gg \Gamma, |G|$ as well due to the large ratio Γ/Γ_2 (collisions do not quench level 2 but increase the value of Γ). The absorption line experience a large field broadening although the value of $|G|$ can be not large compared to Γ.

Consider the effect of "inelastic" (with respect to the optical 2–1 transition) collisions on optical transitions and the population of levels 1 and 2 under the above conditions. The necessity of refinement of the standard model is clearly demonstrated by analyzing the problem based on the concept of compound systems [7]. Therefore, we will follow the approach developed in [7].

Let us introduce the Hamiltonian

$$\mathbf{H} = \mathbf{H}_{\mathscr{A}} + \mathbf{H}_{\mathscr{E}} + \mathbf{V}_{\mathscr{A}\mathscr{E}} + \mathbf{U}(t). \quad (2)$$

Here, $\mathbf{H}_{\mathscr{A}}$ is the Hamiltonian of particles interacting with radiation; $\mathbf{H}_{\mathscr{E}}$ is the Hamiltonian of the radiation field; and $\mathbf{V}_{\mathscr{A}\mathscr{E}}$ and $\mathbf{U}(t)$ are operators of the interaction of particles $\mathscr{A}$ with the field and buffer particle, respectively. The operator $\mathbf{U}(t)$ represents separate pulses, which are randomly distributed on the time scale, but the time dependence is specified inside each of the pulses. Such a representation of the operator $\mathbf{U}(t)$ corresponds to classical scattering.

In the resonance (two-level) approximation, the operator $\mathbf{H}_{\mathscr{A}}$ corresponds to two energy atomic states 1 and 2 (Fig. 4a). The operator $\mathbf{H}_{\mathscr{A}} + \mathbf{H}_{\mathscr{E}}$ has the states (Figs. 4b, 4c)

$$|1\rangle \equiv |1, n\rangle, \quad |2\rangle \equiv |2, n-1\rangle, \quad (3)$$

where n is the number of photons in the radiation field. These states belong to the atom + field compound system in the absence interaction the atom and the field. The mutual arrangement of the energy levels of these states depends on the sign of Ω: when Ω is negative, the state $|1\rangle$ is located below the state $|2\rangle$, and vice versa when Ω is positive. The operator $\mathbf{H}_{\mathscr{A}} + \mathbf{V}_{\mathscr{A}\mathscr{E}}$ has no stationary eigenstates; however, its states (quasi-energy levels) are often used to illustrate the field splitting of the atomic energy levels (the so-called dynamic Stark effect or the Authler–Townes effect). This is illustrated in Fig. 4d. Finally, the operator $\mathbf{H}_{\mathscr{A}} + \mathbf{H}_{\mathscr{E}} + \mathbf{V}_{\mathscr{A}\mathscr{E}}$ corresponds to the states $|\tilde{1}\rangle$ and $|\tilde{2}\rangle$ (Fig. 4e). In the absence of other interactions, these states are stationary and they are sometimes called "clothed"-atom states. A wave function, in the presence of any other interactions, can be represented in the basis of any of these pairs of states. In particular, in the basis of states $|1\rangle$ and $|2\rangle$ an arbitrary wave function has the form

$$\Psi = \exp\left(-i\frac{E_1}{\hbar} - in\omega t\right)[a_1|1\rangle + a_2|2\rangle\exp(i\Omega t)], \quad (4)$$

where a_1 and a_2 are the corresponding amplitudes of the states. From the Schrödinger equation with the Hamiltonian (2), we obtain for these amplitudes

$$\begin{aligned} \frac{d}{dt}a_1 &= iG^* e^{i\Omega t}a_2 - \frac{i}{\hbar}U_{11}a_1, \\ \frac{d}{dt}a_2 &= iGe^{-i\Omega t}a_1 - \frac{i}{\hbar}U_{22}a_2, \\ U_{ii} &= \langle i|\mathbf{U}|i\rangle. \end{aligned} \quad (5)$$

Here, only the diagonal elements of the operator of interaction with buffer-gas particles are nonzero. In the absence of the interaction **U, t**he solutions of Eqs. (5) are well known:

$$\begin{aligned} a_1^0 &= \exp\left(i\frac{\Omega}{2}t\right) \\ &\times\left[A_1\exp\left(i\frac{\Omega_R}{2}t\right) + A_2\exp\left(-i\frac{\Omega_R}{2}t\right)\right], \\ a_2^0 &= \exp\left(-i\frac{\Omega}{2}t\right)\left[A_1\frac{2G}{\Omega_R - \Omega}\exp\left(i\frac{\Omega_R}{2}t\right)\right. \end{aligned} \quad (6)$$

$$- A_2\frac{2G}{\Omega_R+\Omega}\exp\left(-i\frac{\Omega_R}{2}t\right)\Big],$$

$$\Omega_R \equiv \sqrt{\Omega^2+4|G|^2}.$$

Here, A_1 and A_2 are arbitrary coefficients determined by the initial conditions. The corresponding wave function Ψ^0 can be represented in the form

$$\Psi^0 = \exp\left(-i\frac{E_1}{\hbar}t - in\omega t + i\frac{\Omega}{2}t\right)$$

$$\times\left[\tilde{a}_1^0|\tilde{1}\rangle\exp\left(i\frac{\Omega_R}{2}t\right)+\tilde{a}_2^0|\tilde{2}\rangle\exp\left(-i\frac{\Omega_R}{2}t\right)\right], \qquad (7)$$

$$\tilde{a}_1^0 = \frac{A_1}{b_1},\quad \tilde{a}_2^0 = \frac{A_2}{b_2^*},$$

$$b_1 = \frac{1}{\sqrt{2}}\sqrt{1-\frac{\Omega}{\Omega_R}},\quad b_2 = \frac{G}{|G|}\frac{1}{\sqrt{2}}\sqrt{1+\frac{\Omega}{\Omega_R}}.$$

Here, we introduced new wave functions $|\tilde{1}\rangle$ and $|\tilde{2}\rangle$, which form the orthonormal basis and are related to functions $|1\rangle$ and $|2\rangle$ by the expressions

$$|\tilde{1}\rangle = b_1|1\rangle + b_2|2\rangle,\quad |\tilde{2}\rangle = b_2^*|1\rangle - b_1|2\rangle, \qquad (8)$$
$$|b_1\rangle^2 + |b_2|^2 = 1.$$

The inverse matrix coincides with the matrix of transformation (8). The states $|\tilde{1}\rangle$ and $|\tilde{2}\rangle$ are stationary according to (7), and their energy levels are shown in Fig. 4e. Therefore, these states are the stationary states of the compound system with the Hamiltonian $\mathbf{H}_{\mathscr{A}} + \mathbf{H}_{\mathscr{E}} + \mathbf{V}_{\mathscr{A}\mathscr{E}}$. Independently of the sign of Ω, the state $|\tilde{1}\rangle$ is located below the $|\tilde{2}\rangle$ state. The wave function corresponding to an arbitrary Hamiltonian also can expanded in the wave functions of the given compound system. Let us introduce the probability amplitudes $\tilde{a}_1$ and $\tilde{a}_2$ instead of $\tilde{a}_1^0$ and $\tilde{a}_2^0$ and write for them equations based on the Schrödinger equation with the full Hamiltonian (2)

$$i\hbar\frac{d}{dt}\tilde{a}_1 = \tilde{U}_{12}\tilde{a}_2 e^{-i\Omega_R t} + \tilde{U}_{11}\tilde{a}_1,$$

$$i\hbar\frac{d}{dt}\tilde{a}_2 = \tilde{U}_{21}\tilde{a}_1 e^{i\Omega_R t} + \tilde{U}_{22}\tilde{a}_2, \qquad (9)$$

$$\tilde{U}_{ij} = \langle\tilde{i}|\mathbf{U}|\tilde{j}\rangle.$$

Taking into account relations (8) for matrix elements $\tilde{U}_{ij}$ we obtain the following expressions

$$\tilde{U}_{21} = \frac{G}{\Omega_R}\Delta U,\quad \tilde{U}_{12} = \tilde{U}_{21}^*;$$

$$\tilde{U}_{11} = \overline{U} - \frac{\Omega}{2\Omega_R}\Delta U,\quad \tilde{U}_{22} = \overline{U} + \frac{\Omega}{2\Omega_R}\Delta U, \qquad (10)$$

$$\overline{U} = \frac{U_{11}+U_{22}}{2},\quad \Delta U = U_{11} - U_{22}.$$

We interpret the obtained result in the case when the operator $\mathbf{U}$ corresponds to the interaction in collisions with buffer particles in the following way. In the basis of the unperturbed states of the atom $\mathscr{A}$ (Fig. 4a) in the absence of emission, collisions do not induce transitions between the states 1 and 2, being "elastic" in this sense. When the radiation intensity tends to zero, collisions remain elastic in the basis of the states $|\tilde{1}\rangle$ and $|\tilde{2}\rangle$ as well (see (9), (10)). The nonzero radiation intensity results in transitions between the states $|\tilde{1}\rangle$ and $|\tilde{2}\rangle$ ($\tilde{U}_{21} \neq 0$); i.e., collisions acquire an inelastic channel with the energy gap equal to $\hbar\Omega_R$. The elastic scattering channel also changes. Note that the matrix element $\tilde{U}_{21}$ of the interaction potential, which causes inelastic transitions, contains both the parameters of the initial interaction operator for colliding particles and the radiation parameters. This means that a photon of the radiation field is involved in the corresponding collision event. Collisions of this type are called optical collisions [7]. According to [7], optical collisions are the main channel of the variation in the radiation energy under the condition $|\Omega| \gg \Gamma$, in contrast to the region $|\Omega| \leqslant \Gamma$, where the radiation energy changes mainly during the free transit time. In other words, radiative processes that change an atomic state under the condition $|\Omega| \gg \Gamma$ occur predominantly during collision events. This is a very important point for our subsequent conclusions.

In weak fields $|G| \ll |\Omega|$, radiative processes for $|\Omega| \gg \Gamma$ can be clearly interpreted in the basis of the unperturbed states of an atom (Fig. 5a). If the atom is at level 1, then it can absorb a photon $\hbar\omega$ in a collision event and transfer to level 2. In this case, the energy excess (or deficit) $\hbar\Omega$ is compensated by the translational energy of collision partners (Fig. 5a). The same is true for stimulated emission if the atom was at the level 2 before the collision (Fig. 5b).

In a strong field, the states $|\tilde{1}\rangle$ and $|\tilde{2}\rangle$ of the compound system are the real stationary states. Collisions and accompanying radiative process for these states are described by the interaction operator $\mathbf{U}$ with matrix elements (10). For such a clothed atom, we can obtain in a standard way the kinetic equations for the density matrix with a collision integral in which characteristics of an elementary scattering event will be determined by the operator $\mathbf{U}$ containing both elastic and inelastic scattering channels. Of course, all the known general results of the kinetic theory remain valid in this case as

well. In particular, the general thermodynamic result that collisions should lead to the equilibrium (Boltzmann) distribution over the energy levels will be also valid. In our case, this means that the population of the state $|\tilde{2}\rangle$ will be lower with time than that of the $|\tilde{1}\rangle$ state, in accordance with the Boltzmann factor $\exp(-\hbar\Omega_R/k_BT)$.

Consider now expression (8) relating the states $|\tilde{i}\rangle$ and $|i\rangle$. In the case of the positive detuning and under the condition $4|G|^2 \ll \Omega^2$, we have

$$|\tilde{1}\rangle \approx |2\rangle, \quad |\tilde{2}\rangle \approx \frac{G}{|G|}|1\rangle; \tag{11}$$

i.e., the upper level of the compound system corresponds to the lower level of the atom and vice versa. We conclude that in the case of a large positive detuning ($|\Omega| \gg |G|, \Gamma$), the population of the upper atomic state $|2\rangle$ caused by optical collisions proves to be higher than that of the lower state $|1\rangle$ by the Boltzmann factor $\exp(\hbar\Omega/k_BT)$:

$$|a_2|^2 = \exp(\hbar\Omega/k_BT)|a_1|^2. \tag{12}$$

In the opposite situation ($\Omega < 0$, $|\Omega| \gg |G|, \Gamma$), we have

$$|\tilde{1}\rangle \approx |1\rangle, \quad |\tilde{2}\rangle \approx -|2\rangle. \tag{13}$$

In this case, the lower atomic level is populated greater than the upper one. Expression (12) remains valid taking the sign of Ω into account. If the gas (thermostat) temperature is high, then the populations of the states $|1\rangle$ and $|2\rangle$ are equalized. The results of the adopted standard model prove to be valid only under the condition

$$\hbar|\Omega| \ll k_BT. \tag{14}$$

The results of paper [7] and of previous papers on optical and radiative collisions are in fact also restricted by the condition (14).

The result obtained above can be clearly interpreted in the basis of the states of an unperturbed atom (Fig. 5). In the case of a large positive detuning Ω of the pump frequency, the energy excess of the photon $\hbar\Omega$ is transferred to a medium in the absorption event, whereas in the stimulated emission event, the atom should take the energy from the medium to undergo the transition from the level 2 to the level 1. The number of buffer-gas particles capable of stimulated emission is lower by a factor of $\exp(-\hbar\Omega/k_BT)$ than the number of particles capable of absorption. As a result, the population of the state 2 becomes greater than that of the state 1; i.e., the population inversion appears.

5. DISCUSSION

We performed our analysis by neglecting relaxation processes from levels 1 and 2 in order to demonstrate the appearance of the population inversion in a pure form. In the presence of relaxation, its rate should provide the relaxation time $1/\Gamma_2$ (we assume that level 1 is the ground level) sufficient for the accumulation of the effect. The rate of optical transitions in an atom caused by optical collisions is described by the expression $2|G|^2\Gamma_{oc}/\Omega^2$ [7], where Γ_{oc} is the phase relaxation time in optical collisions. For estimates, we can assume that $\Gamma_{oc} \approx \Gamma$, where Γ is the collision half-width of the line. Therefore, the effect will be distinctly observed when

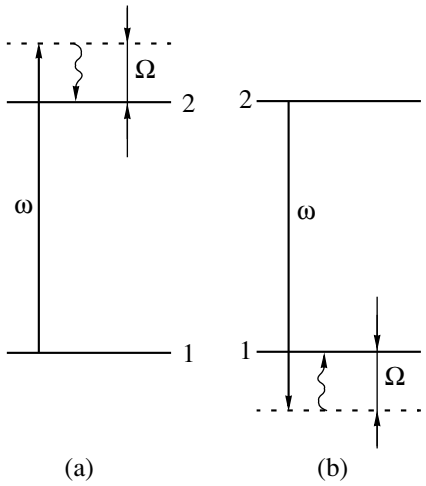


Fig. 5. Illustration of optical collision events in the basis of unperturbed atomic states: (a) absorption of photon $\hbar\omega$, (b) emission of a photon.

$$\frac{2|G|^2\Gamma}{\Omega^2\Gamma_2} \gg 1, \tag{15}$$

which means that the effective saturation parameter in (1) should be large.

In our experiments at the buffer-gas pressure p_{He} = 600 Torr, based on the data on broadening of the D_1-line of sodium by helium [8], we have $\Gamma \approx 10^4$ MHz. The value of Γ_2 coincides with the first Einstein coefficient for the $P_{1/2}$–$S_{1/2}$ transition and, according to the data available (see, for example, [9]), Γ_2 = 10 MHz. Therefore, under our experimental conditions, $\Gamma/\Gamma_2 \approx 10^3$. To estimate the value of $|G|$, we will use the expression from [5] for the $P_{1/2}$–$S_{1/2}$ transition in the case of linearly polarized radiation

$$|G| = 0.334 \times 10^{-2}\sqrt{\frac{1}{3}\bar{\lambda}\bar{S}f}\ \text{cm}^{-1},$$

where $f = 0.33$ is the oscillator strength for the $S_{1/2}$–$P_{1/2}$ transition; $\bar{\lambda}$ is the radiation wavelength in microns; and $\bar{S}$ is the radiation energy flux density (the Poynting vector) in W/cm^2. For $\bar{S} = 10^7$, we have $|G| \approx 3$ cm^{-1}. As a result, we obtain the value of the effective saturation parameter in (1) for Ω = 60 cm^{-1} (the detuning at which the effect is maximum) $\kappa \approx 10$; i.e., the experimental

conditions correspond to the manifestation of the population inversion effect. Taking also into account the fact that no amplified emission was observed at resonance transitions in the case of the negative detuning Ω, we can certainly assert that we experimentally observed the effect discussed in the previous section.

6. CONCLUSIONS

We have presented a theoretical explanation and experimentally observed a new effect—the appearance of the population inversion in a two-level system upon nonresonance absorption of continuous laser radiation. The effect is observed at high pressures of a buffer gas and is caused by the so-called optical collisions. This effect demonstrates the limitedness of the adopted physical concepts based on the widely used quantum kinetic equations for the density matrix (see, for example, [10–12]). These equations were derived assuming that the radiation field is not involved in collision events and all radiative processes occur during the free transit time. This assumption is valid in the region of frequency detuning that is not much wider than the collision linewidth. Otherwise, one should take into account that the balance between radiation and a quantum system is established for a time of the order of $|\Omega|^{-1}$ and particles are not in the pure atomic states at the collision event. This fact was taken into account in the theory of optical and radiative collisions (see [7] and references therein). However, the authors of these papers also did not pay attention to a rather obvious inevitability of the establishment of the Boltzmann distribution over the levels of a compound system, which can result in the population inversion. The inclusion of the radiation field in collision processes should result in the correction of conventional kinetic equations for the density matrix defined in the basis of unperturbed atomic states, which requires additional studies. The first steps in this direction were in fact already initiated by Pestov and Rautian [12] (see also [5]). Here, the required equations were obtained in the operator form. These equations should be modified for solving specific physical problems. It is not inconceivable, however, that for some problems the kinetic equations derived for the density matrix defined in the basis of the states of a compound system prove to be more convenient.

In our opinion, the effect found by us is useful from the point of view of the extension of possibilities for generating coherent radiation in new spectral regions. By using stepwise excitation of a quantum system to higher energy states, lasing can be obtained in the short-wavelength spectral region employing the population inversion at the transitions to the ground state.

ACKNOWLEDGMENTS

The authors thank M.G. Stepanov and E.V. Podivilov for useful discussions. This work was supported by the Russian Foundation for Basic Research (project no. 01-02-17433) and by the program “Physics of Quantum and Wave Processes” of the Ministry of Industry and Science.

REFERENCES

1. B. A. Glushko, M. E. Movsesyan, and T. O. Ovakimyan, Opt. Spektrosk. **52**, 762 (1982) [Opt. Spectrosc. **52**, 458 (1982)].
2. S. N. Atutov, A. I. Plekhanov, and A. M. Shalagin, Opt. Spektrosk. **56**, 215 (1984) [Opt. Spectrosc. **56**, 134 (1984)].
3. Z. Konefal, Opt. Commun. **164**, 95 (1999).
4. A. A. Apolonsky, S. A. Babin, S. I. Kablukov, *et al.*, Phys. Rev. A **61**, 33 408 (2000).
5. S. G. Rautian, G. I. Smirnov, and A. M. Shalagin, *Nonlinear Resonances in Atomic and Molecular Spectra* (Nauka, Novosibirsk, 1979).
6. S. G. Rautian and I. I. Sobel’man, Zh. Éksp. Teor. Fiz. **41**, 456 (1961) [Sov. Phys. JETP **14**, 328 (1961)].
7. S. I. Yakovlenko, Usp. Fiz. Nauk **136**, 593 (1982) [Sov. Phys. Usp. **25**, 216 (1982)].
8. I. I. Sobelman, *Atomic Spectra and Radiative Transitions* (Nauka, Moscow, 1977; Springer-Verlag, Berlin, 1979).
9. J. F. Kielkopf, J. Phys. B **13**, 3813 (1980).
10. R. F. Snider, J. Chem. Phys. **32**, 1051 (1960).
11. V. A. Alekseev, T. L. Andreeva, and I. I. Sobel’man, Zh. Éksp. Teor. Fiz. **62**, 614 (1972) [Sov. Phys. JETP **35**, 325 (1972)].
12. É. G. Pestov and S. G. Rautian, Zh. Éksp. Teor. Fiz. **64**, 2032 (1973) [Sov. Phys. JETP **37**, 1025 (1973)].

Translated by M. Sapozhnikov

Journal of Experimental and Theoretical Physics, Vol. 93, No. 5, 2001, pp. 1035–1051.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 5, 2001, pp. 1194–1212.

PLASMA, GASES

Dynamic Polarizability and the Theory of the Ion-Acoustic Turbulence in a Plasma Containing Ions of Two Species

I. V. Kuzora, V. P. Silin*, and S. A. Uryupin

Lebedev Institute of Physics, Russian Academy of Sciences, Moscow, 119991 Russia
**e-mail: silin@sci.lebedev.ru*

Received April 18, 2001

Abstract—A damping rate of the induced scattering of fast ion-acoustic waves on ions, the explicit form of which depends on the plasma polarizability at the frequency of beats of the interacting waves, was determined for a nonisothermal plasma containing hot electrons and cold ions of two species. In the case of a plasma containing mobile light ions and slow heavy ions, a new phenomenon of significant decrease in the probability of induced scattering was established. This effect is related to the fact that a contribution to the dielectric function of slow ions, determining the scattering amplitude, depends on both spatial and frequency dispersion. It is shown that this decrease in the induced scattering probability leads to a growth of the turbulent noise level and to a change of the anomalous transport coefficients in the limit of large turbulent Knudsen numbers. The same effect is responsible for a relative decrease in the runaway electron production. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The properties of plasmas featuring the ion-acoustic turbulence (IAT) have drawn the attention of researchers since the early 1960s (see, e.g., reviews [1, 2]). This interest is related to the fact that IAT is involved in such remarkable phenomena as the anomalous resistance and turbulent heating of a direct-discharge plasma. Already the first steps in development of the IAT theory showed that one of the main physical processes responsible for the formation of a quasistationary turbulence spectrum is the induced scattering of ion-acoustic waves on ions.

The first notions about the turbulence spectrum and the role of induced scattering were formulated by Kadomtsev [3] and Petviashvili [4] within the framework of a theory of simplest plasmas containing ions of a single species. During a long period of time, the concepts of Kadomtsev and Petviashvili concerning the induced scattering of ion-acoustic waves on ions were not subject to any revision. It was not until the beginning of the 1990s when the development of the IAT theory for the plasmas containing ions of two species revealed the effect of a significant increase in intensity of the induced scattering of ion-acoustic waves on ions, which was related to the ion charge separation [5, 6]. As a result of this increase in the scattering intensity, a quasistationary IAT level in the plasmas with two species of ions is significantly lower as compared to that in the simplest case of a plasma containing ions of a single species (see, e.g., reviews [7, 8]).

Recently [9], we have generalized the new conclusions derived in [5, 6] to the case of plasmas containing ions of more than two different species. It should be noted that the influence of the collective properties of a plasma on the induced scattering was described in [5, 6, 9] in the approximation of a static ion polarizability in the plasma (originally used in the pioneering works of Kadomtsev and Petviashvili). This approximation ignores the frequency (temporal) dispersion of the plasma polarizability and assumes that the entire collective effect of the plasma on the induced scattering is determined by the spatial dispersion manifested by a dependence of the polarizability on the wavevector. In a certain sense, this model is analogous to the approximation of the Debye screening of the plasma charge, the only significant difference being that, in the context of induced scattering, we are speaking of the frequency and wavevector of the beats of interacting waves leading to the manifestation of the collective properties of plasmas in the phenomenon of induced scattering.

On the other hand, it is the frequency dispersion of the plasma polarizability that determines the spectrum of collective oscillations and the decay interaction of such oscillations [10, 11]. However, the frequency dispersion of the ion polarizability determining the spectrum of ion-acoustic waves is still not studied as a factor influencing the process (very important in the IAT theory) of the induced scattering of ion-acoustic waves on ions. In this paper, we draw the attention of researchers for the first time to the necessity of taking into account the frequency dispersion of the ion polarizability in a plasma in descriptions of the induced scattering of ion-acoustic waves. This is necessary for the plasmas containing ions of several species characterized by significantly different thermal velocities. Under these conditions, the phase velocities of beats of the interacting waves are (i) comparable with the thermal velocities of

light ions and (ii) exceed the thermal velocities of slow, relatively heavy ions. It is the latter circumstance that makes it necessary to take into account the influence of the dynamic plasma polarizability on the induced scattering of ion-acoustic waves on ions.

Below we present the IAT theory for a plasma containing ions of two species in which the probability of the induced scattering of ion-acoustic waves on ions is determined with an allowance for the influence of a dynamic polarizability of the plasma. Section 2 indicates the conditions under which there is no need in considering the decay of ion-acoustic waves. In Section 3, the principles of the theory of induced scattering are formulated so as to include the effect of the dynamic plasma polarizability. An integral operator of the ion-acoustic wave damping rate related to the induced scattering on ions is reduced to a differential form in Section 4. A new spectrum of the turbulent ion-acoustic wave fluctuations is determined in Section 5. Some results of the theory of anomalous transport are discussed in Section 6, with the main attention paid to conditions under which the influence of the plasma polarizability is significant for both the turbulent noise level and the transport coefficients. It is shown that the dynamic polarizability of the given plasma at the frequency of beats of the interacting waves leads to an increase in the level of turbulent noise in the plasma with two (light and heavy) ion species, which is manifested by a corresponding decrease in the factor of limitation of the thermal flux of electrons (6.7) and an increase in the effective turbulent frequency of electron collisions (6.11), which is an analogue of the turbulent frequency (6.8) proposed by Sagdeev [12] for description of the anomalous resistance in a plasma containing ions of a single species. In Section 7, we determine the flux characteristic of the runaway electron production in a plasma with two species of ions. It is shown that a decrease in the induced scattering, caused by the dynamic polarizability of the plasma, leads to suppression of the runaway electron production. The possible manifestations of the relative decrease in the probability of induced scattering, related to an allowance for the dynamic plasma polarizability, are illustrated in Section 8 by examples of hydrocarbon (C_xH), lead oxide (Pb_xO), HXe_x, and $LiAu_x$ plasmas.

2. LOW-FREQUENCY WAVES AND THE POSSIBILITY OF DECAY IN A PLASMA WITH IONS OF TWO SPECIES

The main purpose of this section is to establish the conditions under which the decay of ion-acoustic waves in a plasma is impossible. This aim will be achieved by method of indirect demonstration: we will find the conditions necessary for the existence of decay, thus outlining the conditions when this process does not take place.

The first necessary condition for decay is the presence of weakly damped waves. For this reason, we will first recall the known data on the ion-acoustic waves in a fully ionized nonisothermal plasma containing relatively hot electrons and two species of relatively cold ions. First, let us consider the fast ion-acoustic waves possessing a group velocity that is lower than the thermal velocity of electrons v_T, but higher than the thermal velocities $v_{T\alpha}$ ($\alpha = 1, 2$) of the ions of both species:

$$v_T \gg \frac{\omega}{k} \gg v_{T\alpha}. \tag{2.1}$$

Under these conditions, the frequency of a fast ion-acoustic wave is related to the wavevector by a relationship of the following type:

$$\omega^2 = \frac{k^2 v_s^2}{1 + k^2 r_{De}^2} + \frac{3}{\omega_L^2} k^2 (\omega_{L1}^2 v_{T1}^2 + \omega_{L2}^2 v_{T2}^2), \tag{2.2}$$

where $v_s = r_{De}\omega_L$ is the ion-acoustic wave velocity; $r_{De} = v_T/\omega_{Le}$ and ω_{Le} are the Debye radius and Langmuir frequency of electrons, respectively; $\omega_L^2 = \omega_{L1}^2 + \omega_{L2}^2$; and $\omega_{L\alpha}$ is the Langmuir frequency of ions of the α species. In the region of wavelengths lower than (or comparable with) the reciprocal Debye radius of electrons (i.e., for $kr_{De} \lesssim 1$), a correction to the ion-acoustic wave velocity in (2.2) due to the thermal motion of ions is relatively small. This correction is necessary for description of the Cerenkov wave damping on ions, while an analysis of the wave frequency can be performed using a simpler expression of the type

$$\omega \approx k v_s / \sqrt{1 + k^2 r_{De}^2}. \tag{2.3}$$

The fast ion-acoustic waves characterized by the dispersion laws (2.2) or (2.3) are weakly damped as a result of the Cerenkov interaction with electrons and ions. Assuming that the damping rate $\gamma(k)$ is small as compared to the fast ion-acoustic wave frequency and using the dispersion law, we obtain

$$\gamma(k) = \gamma_s(k) + \gamma_1(k) + \gamma_2(k), \tag{2.4}$$

where $\gamma_s(k)$ is the damping rate on electrons expressed as

$$\gamma_s(k) = \sqrt{\frac{\pi}{8}}\omega\frac{\omega_L}{\omega_{Le}}\left(\frac{\omega}{k v_s}\right)^3 \ll \omega, \tag{2.5}$$

and $\gamma_\alpha(k)$ are the damping rates on ions of the α species. In the region of relatively large velocities, $v \geq \omega/k$, where the distribution of ions of the α species is described by the Maxwell function, the latter quantity can be expressed as

$$\gamma_\alpha(k) = \sqrt{\frac{\pi}{8}}\omega\frac{\omega_{L\alpha}^2}{\omega_L^2}\left(\frac{\omega}{k v_{T\alpha}}\right)^3 \exp\left[-\frac{1}{2}\left(\frac{\omega}{k v_{T\alpha}}\right)^2\right]. \tag{2.6}$$

The requirement of small contributions to the wave damping on ions of each species reduces to the condition $\gamma_\alpha(k)/\omega \ll 1$. Using this condition and expression (2.6), we obtain

$$\left(\frac{\omega}{kv_{T\alpha}}\right)^2 \gg \ln\left[\frac{\pi}{8}\frac{\omega_{L\alpha}^4}{\omega_L^4}\left(\frac{\omega}{kv_{T\alpha}}\right)^6\right]. \quad (2.7)$$

For the waves with $kr_{De} \lesssim 1$, Eq. (2.3) can be replaced by an approximate relation $\omega \approx kv_s$ and Eq. (2.6) can be rewritten in the form of inequality for the $v_s/v_{T\alpha}$ ratio:

$$\left(\frac{v_s}{v_{T\alpha}}\right)^2 \gg \ln\left[\frac{\pi}{8}\frac{\omega_{L\alpha}^4}{\omega_L^4}\left(\frac{v_s}{v_{T\alpha}}\right)^6\right]. \quad (2.8)$$

Let us consider inequality (2.8) for ions of the first species. If these ions are not treated as an impurity, we may put $\omega_L \approx \omega_{L1}$. Then (2.8) can be written approximately as

$$\frac{v_s^2}{v_{T1}^2} = \left(\frac{\omega_L r_{De}}{\omega_{L1} r_{D1}}\right)^2 \approx \frac{r_{De}^2}{r_{D1}^2} \gg 3. \quad (2.9)$$

This inequality presents the necessary condition for a small damping of the fast ion-acoustic wave on ions of the first species.

Now let us briefly consider the fast ion-acoustic wave damping on ions of the second species, for which we assume that

$$v_{T2} < v_{T1}. \quad (2.10)$$

If these ions are also not treated as an impurity, we may put $\omega_L \approx \omega_{L2}$ in (2.8). Then inequalities (2.8) and (2.10) show that small damping on ions of the first species ensures smallness of the damping on ions of the second species. If ions of the second species are treated as an impurity, we are to put $\omega_{L2} \ll \omega_L$, which makes the fast ion-acoustic wave damping on ions of the second species negligibly small. Thus, condition (2.9) provides for a small damping of the fast ion-acoustic waves.

Now we can proceed to analysis of the central problem concerning a slow ion-acoustic wave, by which we imply plasma oscillations with a phase velocity satisfying the conditions

$$v_{T1} \gg \frac{\omega}{k} \gg v_{T2}. \quad (2.11)$$

Below we will consider a situation in which, in addition to the slow acoustic wave obeying (2.11), a weakly damped fast acoustic wave exists as well (see (2.9)). Taking into account the conditions (2.11) and the fact that the ratio r_{D1}^2/r_{De}^2 is small, we obtain an expression for the spectrum of slow ion-acoustic wave:

$$\omega = \frac{kv_l}{\sqrt{1 + k^2 r_{D1}^2}}, \quad (2.12)$$

where $v_l = \omega_{L2} r_{D1}$ is the slow ion-acoustic wave velocity. Below we consider the case of $kr_{De} \lesssim 1$. For such wavevectors, formula (2.12) can be rewritten with an allowance for (2.9) as

$$\omega_l = kv_l. \quad (2.13)$$

The conditions (2.11) necessary for the existence of a slow acoustic wave are reduced to the following:

$$\frac{v_{T1}}{v_{T2}} \gg r \gg 1, \quad (2.14)$$

where

$$r = \frac{r_{D1}}{r_{D2}}. \quad (2.15)$$

Let us consider the damping rate of slow ion-acoustic waves. The damping rate on electrons is small because $r_{D1}^2 \ll r_{De}^2$ and $v_l \ll v_T$. For the Cerenkov damping on ions, the damping rate can be approximately expressed as

$$\gamma_l(k) = \gamma_{l1}(k) + \gamma_{l2}(k), \quad (2.16)$$

$$\gamma_{l1}(k) = \sqrt{\frac{\pi}{8}}\omega_l \frac{\omega_{L2}}{\omega_{L1}} \ll \omega_l, \quad (2.17)$$

$$\gamma_{l2}(k) = \sqrt{\frac{\pi}{8}}\omega_l \left(\frac{\omega_l}{kv_{T2}}\right)^3 \exp\left[-\frac{1}{2}\left(\frac{\omega_l}{kv_{T2}}\right)^2\right]. \quad (2.18)$$

The condition of small damping rate on ions of the first species is expressed by the left-hand inequality in (2.14). The damping on ions of the second species is treated by analogy with the fast ion-acoustic wave damping on ions of the first species considered above. Using the approximate relationship (2.13), notation (2.15), and formula (2.18), it is easy to show that the requirement for a small ratio $\gamma_{l2}/\omega_l \ll 1$ reduces to the following inequality:

$$3 < \ln\left[\frac{\pi}{8} r^6\right] \ll r^2. \quad (2.19)$$

The system of inequalities (2.9), (2.14), and (2.19) determines the conditions under which weakly damped fast and slow ion-acoustic waves may coexist.

For a decay process to be possible, it is necessary that (i) the ion-acoustic waves would exist and (ii) the maximum damping rate of the interacting waves would be smaller than the minimum wave frequency. For this reason, in addition to conditions (2.9), (2.14), and (2.19), it is necessary to ensure that the fast ion-acoustic wave damping rate $\gamma_1(k)$ (2.6) would be small as compared to the slow wave frequency ω_l (2.13). For the waves under consideration ($kr_{De} \lesssim 1$), this leads to the condition

$$\left(\frac{v_s}{v_{T1}}\right)^2 \gg \ln\left[\frac{\pi}{8}\frac{\omega_{L1}^2}{\omega_{L2}^2}\frac{\omega_{L1}^4}{\omega_L^4}\left(\frac{v_s}{v_{T1}}\right)^8\right]. \quad (2.20)$$

In the case of $\omega_{L2} \sim \omega_{L1}$, inequality (2.20) is essentially equivalent to (2.9), although the former condition poses a stronger restriction on the v_s/v_{T1} ratio: $(v_s/v_{T1})^2 \gg 4$. When $4\omega_{L2} < \omega_{L1}$, the limitation of the parameter v_s/v_{T1} becomes still more rigid and can be expressed (to within a logarithmic term) as

$$\left(\frac{v_s}{v_{T1}}\right)^2 \approx \frac{r_{De}^2}{r_{D1}^2} \gg \ln\left\{\frac{\pi}{8}\frac{\omega_{L1}^2}{\omega_{L2}^2}\ln^4\left[\frac{\pi}{8}\left(\frac{\omega_{L1}}{\omega_{L2}}\right)^2\right]\right\} > 4. \quad (2.21)$$

Thus, in the case of $4\omega_{L2} < \omega_{L1}$, the possibility of decay requires that the r_{De}/r_{D1} ratio (characterizing the degree of nonisothermicity of the plasma) would increase with decreasing ω_{L2}/ω_{L1}. This property becomes evident if we recall that the conditions (2.21) also provide for the slow acoustic wave frequency (proportional to ω_{L2}) being not small as compared to the fast acoustic wave damping rate γ (which decreases with growing ratio $v_s/v_{T1} \approx r_{De}/r_{D1}$).

Using inequalities (2.9), (2.14), (2.19), and (2.21), we may find conditions under which, on the one hand, weakly damped fast ion-acoustic waves may exist and, on the other hand, the fast waves cannot take part in decay interactions with the slow wave. Now we will indicate these conditions for a plasma with $\omega_{L1} \gtrsim \omega_{L2}$. The first is the so-called nonisothermicity condition (2.9). The second condition,

$$\left(\frac{v_l}{v_{T2}}\right)^2 = r^2 < 3, \quad (2.22)$$

is in fact an inverse condition with respect to (2.19) and represents a situation when the slow acoustic wave cannot exist because of strong damping. The third condition is

$$3 \ll \left(\frac{v_s}{v_{T1}}\right)^2 < \ln\left[\frac{\pi}{8}\frac{\omega_{L1}^2}{\omega_{L2}^2}\frac{\omega_{L1}^4}{\omega_L^4}\left(\frac{v_s}{v_{T1}}\right)^8\right], \quad (2.23)$$

which ensures, according to (2.20), that the damping rate of the fast wave would be greater than the slow wave frequency at a not too large degree of nonisothermicity, even if the inequality (2.22) is not satisfied. In (2.23), the left-hand inequality corresponds to condition (2.9). According to this, the decay interaction will be forbidden (if the condition (2.22) is not satisfied) only in a comparatively narrow interval of nonisothermicity. For $4\omega_{L2} < \omega_{L1}$, (2.23) is replaced by the condition

$$3 \ll \left(\frac{v_s}{v_{T1}}\right)^2 \approx \frac{r_{De}^2}{r_{D1}^2} < \ln\left\{\frac{\pi}{8}\frac{\omega_{L1}^2}{\omega_{L2}^2}\ln^4\left[\frac{\pi}{8}\frac{\omega_{L1}^2}{\omega_{L2}^2}\right]\right\}. \quad (2.24)$$

According to (2.24), the range of the degree of nonisothermicity in which the decay interaction in the plasma becomes possible logarithmically weakly expands with decreasing ω_{L2}.

In the case when a difference between v_{T1} and v_{T2} is not very large, the dispersion equation for fast and slow ion-acoustic waves can be studied by numerical methods (see, e.g., [13–15]). In this context, let us consider the results of such a numerical investigation reported by Williams *et al.* [15] for a fully ionized CH plasma with equal concentrations and temperatures of hydrogen and carbon ions, for which $v_{T1}^2 = 12\,v_{T2}^2$ and $r_{D1}^2 = 36 r_{D2}^2$ (accordingly, $r^2 = 36$). As was demonstrated in [15], no fast acoustic wave exists for $T_e/T_i \lesssim 5$; this corresponds to violation of the inequality (2.9). In contrast, for $T_e/T_i \gtrsim 5$, the fast acoustic wave exists and the slow acoustic wave is strongly damped. This case, not treated in the analytical considerations presented above, corresponds to violation of the left-hand inequality (2.11), which is obviously not satisfied for the CH plasma studied in [15]. Thus, one can see that our analytical treatment provides for a simple interpretation of the results of numerical calculations. Since the CH plasma studied in [15] cannot feature coexisting fast and slow ion-acoustic waves, this plasma will not exhibit decay interactions for any degree of nonisothermicity.

In this context, we would like to mention the results reported by Vu *et al.* [16] for the ion-acoustic waves in a plasma with ions of two species. According to these data, a necessary condition for the existence of a weakly damped fast wave for $kr_{De} < 1$ is

$$\frac{r_{De}^2}{r_{D1}^2} > \left(1 + \frac{\omega_{L2}^2}{\omega_{L1}^2}\right)^{-1}, \quad (2.25)$$

while a weakly damped slow wave can exist provided that

$$\frac{\omega_{L1}^2}{\omega_{L2}^2} > \frac{r_{De}^2}{r_{De}^2 + r_{D1}^2} > \frac{r_{D2}^2}{r_{D1}^2}. \quad (2.26)$$

When only one of these conditions is satisfied, only the corresponding weakly damped ion-acoustic wave exists. Both conditions fulfilled simultaneously provide for the coexistence of two weakly damped waves. In particular, for a plasma with $\omega_{L1}^2/\omega_{L2}^2 > 1$, conditions (2.25) and (2.26) can be expressed as

$$\frac{r_{De}^2}{r_{D1}^2} > 1, \quad (2.27)$$

$$\frac{\omega_{L1}^2}{\omega_{L2}^2} > 1 > \frac{r_{D2}^2}{r_{D1}^2}. \quad (2.28)$$

Inequalities (2.25)–(2.28) follow from the conditions (2.1) as (2.11) relating the phase velocities of ion-acoustic waves to the thermal velocities of ions. According to an analysis performed in this section, inequalities (2.27) and (2.28) follow from (2.9) and (2.19), respectively.

Therefore, the conditions (2.25)–(2.28) established in [16] follow from the inequalities derived above (so as to provide for a small level of the Cerenkov damping of both fast and slow waves) and impose no additional limitations on the necessary conditions formulated above for the existence of decay processes.

Taking into account the above results, we will consider plasmas with strongly different thermal velocities of the ions of two species assuming that the decay interaction can be ignored provided that either condition (2.22) or the inequalities (2.23) or (2.24) are satisfied. It should be emphasized that decays are impossible in the plasmas similar to that studied in [15].

3. INDUCED SCATTERING OF THE FAST ION-ACOUSTIC WAVES ON IONS

Below we will consider plasmas in which the fast ion-acoustic waves can exist in the absence of decay interactions (since some necessary conditions for the latter are not satisfied). In these plasmas, the main nonlinear process for IAT is the induced scattering of fast ion-acoustic waves on ions. According to the general concepts of the nonlinear plasma theory, the damping rate of the induced scattering of ion-acoustic waves on ions is as follows [17]:

$$\gamma_{nl}(\mathbf{k}) = \sum_{\alpha = 1, 2} \int \frac{d\mathbf{v}}{2m_\alpha} \times \int \frac{d\mathbf{k}'}{(2\pi)^3} N(\mathbf{k}') W_\alpha(\mathbf{k}, \mathbf{k}', \mathbf{v}) \left(\mathbf{k}'' \frac{\partial f_\alpha}{\partial \mathbf{v}}\right), \tag{3.1}$$

where m_α is the mass of ions of the α species ($\alpha = 1, 2$), $N(\mathbf{k})$ is the number density of the ion-acoustic waves, $\mathbf{k}'' = \mathbf{k} - \mathbf{k}'$, $\omega'' = \omega - \omega'$, and $f_\alpha = f_\alpha(\mathbf{v})$ is the distribution function of ions of the α species, and $W_\alpha(\mathbf{k}, \mathbf{k}', \mathbf{v})$ is the probability of induced scattering. The latter quantity is determined by the formula

$$W_\alpha(\mathbf{k}, \mathbf{k}', \mathbf{v}) = \frac{4(2\pi)^9 |\Lambda_\alpha(\mathbf{k}, \mathbf{k}', \mathbf{v})|^2}{\omega^2 \left[\frac{\partial}{\partial \omega} \varepsilon(\omega, \mathbf{k})\right] \left[\frac{\partial}{\partial \omega'} \varepsilon(\omega', \mathbf{k}')\right]} \times \delta(\omega'' - \mathbf{k}'' \cdot \mathbf{v}). \tag{3.2}$$

Here, $\varepsilon(\omega, \mathbf{k})$ is the longitudinal dielectric function of the plasma, $\Lambda_\alpha(\mathbf{k}, \mathbf{k}', \mathbf{v})$ is the scattering probability amplitude defined as

$$\Lambda_\alpha(\mathbf{k}, \mathbf{k}', \mathbf{v}) = \frac{\omega}{(2\pi)^3} \left\{ \frac{e_\alpha^2}{m_\alpha} \frac{\mathbf{k} \cdot \mathbf{k}'}{kk'} \frac{1}{(\omega - \mathbf{k} \cdot \mathbf{v})(\omega' - \mathbf{k}' \cdot \mathbf{v})} + \frac{4\pi e_\alpha}{kk'} \frac{1}{k''^2 \varepsilon(\omega'', \mathbf{k}'')} \sum_{\beta = e, 1, 2} \frac{e_\beta^3}{m_\beta^2} \int \frac{d\mathbf{v}'}{\omega - \mathbf{k} \cdot \mathbf{v}'} \times \left[\left(\mathbf{k}' \frac{\partial}{\partial \mathbf{v}'}\right) \frac{1}{\omega'' - \mathbf{k}'' \cdot \mathbf{v}'} \left(\mathbf{k}'' \frac{\partial}{\partial \mathbf{v}'}\right) f_\beta(\mathbf{v}') + \left(\mathbf{k}'' \frac{\partial}{\partial \mathbf{v}'}\right) \frac{1}{\omega' - \mathbf{k}' \cdot \mathbf{v}'} \left(\mathbf{k}' \frac{\partial}{\partial \mathbf{v}'}\right) f_\beta(\mathbf{v}') \right] \right\}, \tag{3.3}$$

and e_α is the charge of ions of the α species. Expression (3.3) for the scattering probability amplitude can be simplified using the fact that the phase velocity of a fast ion-acoustic wave is small as compared to the thermal velocity of electrons but large in comparison to the thermal velocity of ions (for more detail, see [6]).

In what follows, we take into account that most significant oscillations in the IAT theory are the charge density oscillations with wavenumbers on the order of the reciprocal Debye radius for electrons ($k \sim r_{De}^{-1}$). In addition, we assume that ions of at least one species obey the relationships

$$r_{De} \gg r_{D\alpha}, \quad Z_\alpha T_e \gg T_\alpha, \tag{3.4}$$

where T_e is the electron temperature, Z_α and T_α are the ionization multiplicity and the temperature of ions of the α species, respectively. Under these conditions, expression (3.3) yields an approximate relationship [5, 6]

$$\Lambda_\alpha(\mathbf{k}, \mathbf{k}', \mathbf{v}) \approx \frac{e_\alpha^2}{(2\pi)^3 m_\alpha \omega'} \frac{\mathbf{k} \cdot \mathbf{k}'}{kk'} \times \left[\frac{\mathbf{k} \cdot \mathbf{v}}{\omega} + \frac{\mathbf{k}' \cdot \mathbf{v}}{\omega'} + L_\alpha(\omega'', \mathbf{k}'') \right], \tag{3.5}$$

where

$$L_\alpha(\omega'', \mathbf{k}'') = \left(1 - \frac{e_\beta m_\alpha}{m_\beta e_\alpha}\right) \frac{\delta\varepsilon_\beta(\omega'', \mathbf{k}'')}{\delta\varepsilon_1(\omega'', \mathbf{k}'') + \delta\varepsilon_2(\omega'', \mathbf{k}'')}, \quad \alpha \neq \beta, \tag{3.6}$$

and $\delta\varepsilon_\beta(\omega'', \mathbf{k}'')$ is the partial contribution to a longitudinal dielectric function of ions of the β species. Taking into account the approximate relationship (3.5) for the scattering amplitude and using expressions (3.1) and (3.2), we obtain

$$\gamma_{nl}(\mathbf{k}) = \int d\mathbf{k}' N(\mathbf{k}') K(\mathbf{k}, \mathbf{k}'), \tag{3.7}$$

where

$$K(\mathbf{k}, \mathbf{k}') = \frac{\omega\omega'}{2\omega_L^4} \left(\frac{\mathbf{k} \cdot \mathbf{k}'}{kk'}\right)^2 \sum_{\alpha = 1, 2} \frac{e_\alpha^4}{m_\alpha^3} \int d\mathbf{v} \delta(\omega'' - \mathbf{k}'' \cdot \mathbf{v}) \times \left(\mathbf{k}'' \frac{\partial f_\alpha}{\partial \mathbf{v}}\right) \left\{ \left(\frac{\mathbf{k} \cdot \mathbf{v}}{\omega} + \frac{\mathbf{k}' \cdot \mathbf{v}}{\omega'}\right)^2 + 2\left(\frac{\mathbf{k} \cdot \mathbf{v}}{\omega} + \frac{\mathbf{k}' \cdot \mathbf{v}}{\omega'}\right) \times \mathrm{Re}[L_\alpha(\omega'', \mathbf{k}'')] + |L_\alpha(\omega'', \mathbf{k}'')|^2 \right\}. \tag{3.8}$$

In order to simplify expression (3.8), we assume that the distribution of ions with respect to velocity is

described by the Maxwell function. Then, integrating over the ion velocities, the kernel (3.8) can be presented in the following form:

$$K(\mathbf{k}, \mathbf{k}') = -\frac{\omega\omega'}{8\pi\sqrt{2\pi}\omega_L^4}\left(\frac{\mathbf{k}\cdot\mathbf{k}'}{kk'}\right)^2 \sum_{\alpha=1,2}\left(\frac{\omega''}{k''v_{T\alpha}}\right) \times\left(\frac{e_\alpha}{m_\alpha r_{D\alpha}}\right)^2 \exp\left(-\frac{\omega''^2}{2k''^2 v_{T\alpha}^2}\right) \times [K_{\alpha 0}(\mathbf{k}, \mathbf{k}') + K_{\alpha m}(\mathbf{k}, \mathbf{k}') + K_{\alpha t}(\mathbf{k}, \mathbf{k}')]. \quad (3.9)$$

Here, the terms involving the kernel

$$K_{\alpha 0}(\mathbf{k}, \mathbf{k}') = 2\left[\frac{kk'v_{T\alpha}(\omega+\omega')}{k''\omega\omega'}\right]^2\left(\frac{[\mathbf{k}\times\mathbf{k}']}{kk'}\right)^2 \quad (3.10)$$

are similar to those appearing in the theory of plasma containing ions of a single species. The terms proportional to the kernel,

$$K_{\alpha m}(\mathbf{k}, \mathbf{k}') = \frac{2kk'}{\omega\omega'}\left(\frac{\omega''}{k''}\right)^2 \mathrm{Re}[L_\alpha(\omega'', k'')] \times\left[\frac{\omega}{\omega-\omega'}\left(1-\frac{k'}{k}\right)\left(1+\frac{k}{k'}\right)-\frac{k}{k'}+\frac{\mathbf{k}\cdot\mathbf{k}'}{kk'}\right], \quad (3.11)$$

are due to the mixing of the induced scattering probability contributions determining kernels (3.10) and (3.12). Terms containing the latter kernel $K_{\alpha t}(\mathbf{k}, \mathbf{k}')$ are related to a nontrivial distinction between plasmas containing ions of two species and those with ions of a single species. These terms, as well as the terms with a kernel of the (3.11) type, are nonzero only when ions of the two species possess significantly different charge-to-mass ratios ($e_1/m_1 \neq e_2/m_2$). For this reason, the potential field of interacting ion-acoustic waves acts differently upon the unlike ions. The resulting effect of the dynamic ion charge separation leads to an increase in the induced scattering probability as compared to the case of a simple plasma containing ions of a single species. The kernel

$$K_{\alpha t}(\mathbf{k}, \mathbf{k}') = |L_\alpha(\omega'', k'')|^2 \quad (3.12)$$

is determined by the function L_α defined as (3.6), which depends on the susceptibilities $\delta\varepsilon_1(\omega'', k'')$ and $\delta\varepsilon_2(\omega'', k'')$ corresponding to the Maxwell distribution of ions. According to [18], the latter quantities can be expressed as

$$\delta\varepsilon_\alpha(\omega'', k'') = \frac{1}{(k''r_{D\alpha})^2}\left[1 - J_+\left(\frac{\omega''}{k''v_{T\alpha}}\right)\right], \quad \alpha = 1, 2. \quad (3.13)$$

Expressions (3.7)–(3.13) serve as a basis for the following analysis of induced scattering of the ion-acoustic waves with the dispersion law (2.3) in a plasma containing ions of two species.

In order to use expression (3.13), it will be necessary to obtain an explicit expression for the function $J_+(x)$ and its asymptotics as functions of the real variable x. These expressions are as follows [18]:

$$J_+(x) = J'_+(x) + iJ''_+(x), \quad (3.14)$$

$$J'_+(x) = xe^{-x^2/2}\int_0^x dt e^{t^2/2}, \quad (3.15)$$

$$J''_+(x) = -\sqrt{\frac{\pi}{2}}xe^{-x^2/2}. \quad (3.16)$$

Note that the function $1 - J'_+(x)$ changes sign at $x \approx 1.3$, where $-J''_+(x)$ is in the order of unity. The imaginary part of $-J''_+(x)$ reaches a maximum value ~0.76 at $x = 1$. Far from $x \sim 1$, the asymptotic behavior of $J''_+(x)$ is as follows. For $x \ll 1$,

$$1 - J'_+(x) \approx 1 - x^2, \quad -J''_+(x) \approx x\sqrt{\pi/2} \ll 1, \quad (3.17)$$

while for $x \gg 1$, the function $-J''_+(x)$ is exponentially small according to (3.16) and

$$1 - J'_+(x) \approx -x^{-2} - 3x^{-4}. \quad (3.18)$$

4. REDUCTION OF THE INDUCED SCATTERING DAMPING RATE FROM INTEGRAL TO DIFFERENTIAL

The nonlinear damping rate of the ion-acoustic waves $\gamma_{nl}(\mathbf{k})$ plays a key role in determining the IAT spectrum. Owing to smallness of the momentum transported from waves to ions during the induced scattering, $\gamma_{nl}(\mathbf{k})$ usually can be presented in the form of a differential operator acting upon the modulus of the wavevector of the ion-acoustic wave. In the case under consideration, this representation is possible as well.

The kernel of the integral operator in formula (3.7) has an exponent with the parameter

$$\frac{\omega''}{k''v_{T\alpha}} = \frac{v_s}{v_{T\alpha}}\frac{\omega''}{k''v_s}, \quad (4.1)$$

where the ratio of the acoustic wave velocity to the thermal velocity of ions is much greater than unity ($v_s/v_{T\alpha} \gg 1$) by virtue of the conditions (2.1).

Figure 1 shows the plots of $\omega''/k''v_s$ versus k'/k calculated for $kr_{De} = 1$, $\mathbf{k}\cdot\mathbf{k}'/kk' = 0.3$ and 0.9 and for $kr_{De} = 0.3$, $\mathbf{k}\cdot\mathbf{k}'/kk' = 0.9$. Together with the condition $v_s/v_{T\alpha} \gg 1$, the data in Fig. 1 indicate that, in the range of wavenumbers $kr_{De} \leqslant 1$ (of most interest for the theory of plasmas with IAT), only the values of k'/k around unity are significant: in this region, the kernel (3.9) is not exponentially small. It should be noted that, for $kr_{De} \leqslant 1$, small values of the parameter $\omega''/k''v_{T\alpha}$ are

also possible for $k' > \omega_L/v_{T\alpha}$. However, according to condition (2.1), there are no waves proper in this region of wavelengths. Below we will employ the proximity of k'/k to unity, which corresponds to small variations of the absolute value of the wave momentum, in reducing expression (3.7) to a differential form.

First, consider the quantity $\gamma_0(\mathbf{k})$ representing a contribution to the damping rate proportional to the $K_{\alpha 0}$ kernel defined in (3.10). Introducing the value

$$\Delta = 1 - \frac{k'}{k} \approx \frac{\omega''/k''}{d\omega/dk}\sqrt{2\left(1 - \frac{\mathbf{k}\cdot\mathbf{k}'}{kk'}\right)}, \tag{4.2}$$

representing a small parameter ($|\Delta| \ll 1$), we can write the following relationships:

$$k''^2 \approx 2k^2\left(1 - \frac{\mathbf{k}\cdot\mathbf{k}'}{kk'}\right)(1-\Delta), \tag{4.3}$$

$$\omega'' \approx \Delta k\frac{d\omega}{dk} - \frac{1}{2}\Delta^2 k^2\frac{d^2\omega}{dk^2}, \tag{4.4}$$

$$dx \equiv d\left(\frac{\omega''}{k''v_{T\alpha}}\right) \approx -\frac{dk'}{kv_{T\alpha}} \times \frac{d\omega/dk}{\sqrt{2(1-\mathbf{k}\cdot\mathbf{k}'/kk')}}\left[1 + \Delta - \Delta k\frac{d^2\omega/dk^2}{d\omega/dk}\right], \tag{4.5}$$

where x is a variable related to the parameter Δ as

$$\Delta = x\frac{v_{T\alpha}dk}{d\omega}\sqrt{2\left(1 - \frac{\mathbf{k}\cdot\mathbf{k}'}{kk'}\right)}. \tag{4.6}$$

Since $v_s \gg v_{T\alpha}$, the limits of integration with respect to x, corresponding to small and large k' values in Eq. (3.7), can be extended to $+\infty$ and $-\infty$, respectively. Retaining only the terms linear in Δ and integrating with respect to x, we obtain from (3.7)–(3.10) a contribution to the damping rate due to only the kernel (3.10):

$$\gamma_0(\mathbf{k}) = \frac{1}{2\pi}\left[\sum_{\alpha=1,2}\left(\frac{e_\alpha}{m_\alpha}v_{T\alpha}\frac{\omega_{L\alpha}}{\omega_L^2}\right)^2\right]\frac{k^2dk}{d\omega} \times\frac{\partial}{\partial k}\left\{\frac{k^4dk}{d\omega}\int do_{k'}\left(\frac{\mathbf{k}\cdot\mathbf{k}'}{kk'}\right)^2\left(\frac{[\mathbf{k}\times\mathbf{k}']}{kk'}\right)^2 N\left(k, \frac{\mathbf{k}'}{k'}\right)\right\}, \tag{4.7}$$

where $do_{k'}$ is the solid angle element of the vector $\mathbf{k}'$. A difference of Eq. (4.7) from an analogous expression in the theory of plasmas containing ions of a single species is determined by the term in square brackets. In the latter theory, formula (4.7) transforms into a well-known expression for the damping rate of induced scattering on thermal ions [8]. Accordingly, expression (4.7) converts into the Kadomtsev–Petviashvili operator [3, 4] for $kr_{De} \ll 1$ or into the Galeev–Sagdeev operator [7] for $kr_{De} \gg 1$.

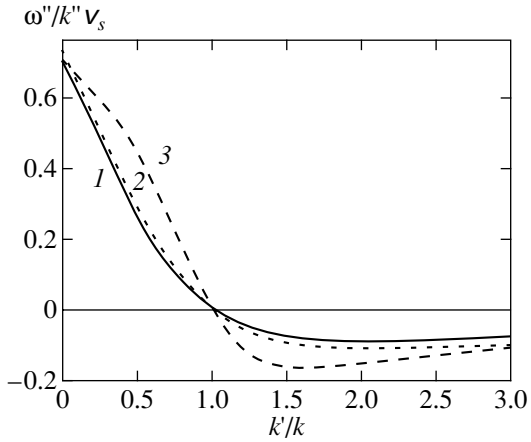


Fig. 1. The plots of $\omega''/k''v_s$ versus k'/k calculated for the parameters $kr_{De} = 1$ (*1*, *3*), 0.3 (*2*) and $\mathbf{kk}'/kk' = 0.3$ (*1*), 0.9 (*2*, *3*).

In the region of wavenumbers $k' \sim k$, the kernels (3.10) and (3,11) can be estimated as

$$K_{\alpha 0}(\mathbf{k}, \mathbf{k}') \sim 8\left(\frac{v_{T\alpha}}{v_s}\right)^2\left(\frac{[\mathbf{k}\times\mathbf{k}']}{kk'}\right)^2, \tag{4.8}$$

$$K_{\alpha m}(\mathbf{k}, \mathbf{k}') \sim 2x^2\left(\frac{v_{T\alpha}}{v_s}\right)^2 \mathrm{Re}[L_\alpha(\omega'', k'')], \tag{4.9}$$

where $x^2 \leqslant 1$. Comparing these expressions to the kernel $K_{\alpha t}(\mathbf{k}, \mathbf{k}')$ defined in (3.12), we can readily see that $K_{\alpha m}$ is always smaller than $\max[K_{\alpha 0}, K_{\alpha t}]$. Therefore, we may ignore the contribution due to the mixed terms (3.11) in the analysis of the induced scattering in plasmas containing ions of two species.

By analogy with the derivation of formula (4.7) for the damping rate $\gamma_0(\mathbf{k})$ we can obtain an expression for the differential contribution to the induced scattering damping rate related to terms specific of the plasma containing two species of ions. This contribution, related to the $K_{\alpha t}$ kernel in (3.9), will be denoted by $\gamma_t(\mathbf{k})$. Using relationships (4.3)–(4.6), we derive the following expression for this damping rate from Eqs. (3.7), (3.9), (3.12), and (3.13):

$$\gamma_t(\mathbf{k}) = \frac{1}{4\pi}\left(\frac{e_1}{m_1} - \frac{e_2}{m_2}\right)^2\sum_{\alpha=1,2}\frac{\omega_{L\alpha}^2}{\omega_L^4}|L_\alpha(0)|^2 D_\alpha\frac{k^2}{d\ln\omega/d\ln k} \times\frac{\partial}{\partial k}\left\{\frac{k^4}{d\ln\omega/d\ln k}\int do_{k'}\left(\frac{\mathbf{k}\cdot\mathbf{k}'}{kk'}\right)^2\left(1 - \frac{\mathbf{k}\cdot\mathbf{k}'}{kk'}\right)N\left(k, \frac{\mathbf{k}'}{k'}\right)\right\}, \tag{4.10}$$

where

$$D_\alpha(u, r) = \frac{1}{\sqrt{2\pi}|L_\alpha(0)|^2}\int_{-\infty}^{\infty} dx x^2 e^{-x^2/2}|L_\alpha(x)|^2, \quad \alpha = 1, 2, \tag{4.11}$$

$$L_1(x) = \left[1 - J_+\left(\frac{x}{u}\right)\right] \times \left\{1 - J_+\left(\frac{x}{u}\right) + r^{-2}[1 - J_+(x)]\right\}^{-1}, \quad (4.12)$$

$$L_2(x) = [1 - J_+(xu)] \times \{1 - J_+(xu) + r^2[1 - J_+(x)]\}^{-1}, \quad (4.13)$$

$L_\alpha(0) = r_{D\alpha}^2/(r_{D1}^2 + r_{D2}^2)$, $u = v_{T2}/v_{T1}$, and $r = r_{D1}/r_{D2}$. In cases when we can neglect the difference between $L_\alpha(x)$ and $L_\alpha(0)$ in (4.11), we obtain $D_1 = D_2 = 1$. In such cases, the damping rate (4.10) coincides with that obtained in [5], where the partial ion contribution to the longitudinal permittivity of plasma was described by an expression corresponding to the static limit: $\delta\varepsilon_\alpha(0, k) = (kr_{D\alpha})^{-2}$. Using the new, more general expression for the ion contributions to the induced scattering damping rate for the plasma with ions of two species, we can describe the induced scattering on ions under the conditions where the effects of dynamic polarizability of the plasma cannot be ignored.

5. ION-ACOUSTIC TURBULENCE SPECTRUM

Using the above expressions for additive contributions to the damping rate of induced scattering for the plasmas with ions of two species, it is possible to determine conditions under which the contribution $\gamma_t(\mathbf{k})$ would prevail. A comparison of expressions (4.7) and (4.10) for $kr_{De} \leqslant 1$ shows that the latter contribution dominates when

$$\frac{r_{De}^2}{r_{D1}^2} \gg D^{-1}\left(\frac{e_1}{m_1} - \frac{e_2}{m_2}\right)^{-2} \frac{(r^2+1)^2}{r^2(r^2+u^2)} \times \left[\left(\frac{e_1\omega_{L1}}{m_1\omega_L}\right)^2 + \left(\frac{e_2\omega_{L2}}{m_2\omega_L}\right)^2\right], \quad (5.1)$$

where

$$D = \frac{D_2u^2 + D_1r^2}{u^2 + r^2}. \quad (5.2)$$

If the inequality (5.1) is satisfied, relationship (4.10) for the damping rate of induced scattering can be written as

$$\gamma_{nl}(\mathbf{k}) = D\gamma_{\text{stat}}(\mathbf{k}), \quad (5.3)$$

where $\gamma_{\text{stat}}(\mathbf{k})$ describes the induced scattering in a plasma with two species of ions, provided that $\delta\varepsilon_\alpha(\omega, k)$ can be replaced by $\delta\varepsilon_\alpha(0, k)$ in the expression for the scattering probability amplitude [6]

$$\gamma_{\text{stat}}(\mathbf{k}) = \frac{1}{4\pi\omega_L^4}\left(\frac{e_1}{m_1} - \frac{e_2}{m_2}\right)^2 (r_{D1}^2 + r_{D2}^2)^{-2} \times \left(\sum_{\alpha=1,2} v_{T\alpha}^2 r_{D\alpha}^2\right) \frac{k^2}{d\ln\omega/d\ln k} \frac{\partial}{\partial k}\left\{\frac{k^4}{d\ln\omega/d\ln k} \times \int do_{k'}\left(\frac{\mathbf{k}\cdot\mathbf{k}'}{kk'}\right)^2\left(1 - \frac{\mathbf{k}\cdot\mathbf{k}'}{kk'}\right)N\left(k, \frac{\mathbf{k}'}{k'}\right)\right\}. \quad (5.4)$$

The quantity D in relationship (5.3) characterizes a difference between $\gamma_{nl}(\mathbf{k})$ and $\gamma_{\text{stat}}(\mathbf{k})$ caused by the dynamic polarizability effects. The function D determined by formulas (5.2), (4.11)– (4.13) depends on two parameters: $r = r_{D1}/r_{D2}$ and $u = v_{T2}/v_{T1}$. As was noted at the end of Section 4, putting zero argument ($x = 0$) in formulas (4.12) and (4.13) for the functions J_+ corresponds to neglect of the deviation of the ion polarizability from that in the static case. Then $D = 1$ and $\gamma_{nl}(\mathbf{k})$ coincides with $\gamma_{\text{stat}}(\mathbf{k})$. However, expressions (4.11)–(4.13) show that, in the general case, we cannot assume that the J_+ value is much smaller than unity. This implies that a consistent consideration of the induced scattering must take into account that the ion polarizability of a plasma differs from that in the static case.

However, according to Eqs. (5.2) and (4.11)–(4.13), there are still conditions under which the coefficient D is close to unity and, hence, the induced scattering can be considered in the static approximation. Such conditions exist, in particular, in the plasmas characterized by equal thermal velocities of ions. In this case, the contributions due to ions of different species to the permittivity are proportional to a function of the same argument $1 - J_+(x)$ and, hence, $D = 1$. Should the thermal velocities of ions differ but little, so that $v_{T1} - v_{T2} \ll v_{T2} < v_{T1}$, the D value can be approximately expressed as (see Appendix, Eqs. (A.4) and (A.5))

$$D \approx 1 - 4(1-u)^2r^2(1+r^2)^{-2}. \quad (5.5)$$

Note that formula (5.5) is symmetric with respect to the substitution of $1/r$ for r. This approximate expression exhibits a minimum $D \approx 1 - (1-u)^2$ for $r = 1$, in which case D is close to unity for $1 - u \ll 1$.

An analysis of formula (5.5) reveals a trend whereby the interval of velocities in which D is close to unity increases when r deviates from $r = 1$. The same trend can be followed in the exact expressions (5.2) and (4.11)–(4.13). For example, when $r \longrightarrow 0$ in formula (5.2), it is sufficient to retain only the coefficient D_2 which, according to (4.11) and (4.13), is close to unity for any finite u. In the case of $r \longrightarrow \infty$, the principal quantity in (5.2) is the coefficient D_1 that is also close to unity according (4.11) and (4.12). Such a behavior implies that, in the plasmas with strongly differing Debye radii of ions, the domain of applicability of the static approximation in the theory of induced scattering

is much greater than in a plasma with $r_{D1} \approx r_{D2}$. This conclusion follows from approximate expressions for the function D (see Appendix, Eqs. (A.6) and (A.8)). In contrast to the approximate formula (5.5), which is valid for $(1-u)^2 \ll 1$, approximate formulas (A.6) and (A.8) for $u^2 \ll 1$ yield

$$D \approx 1 - r^2/u^2, \quad r^2 < u^2 \ll 1,$$
$$D \approx 1 - (ru)^{-2}, \quad r^2u^2 > 1, \quad u^2 \ll 1. \tag{5.6}$$

According to these formulas, D is close to unity when $r < u$ or $r > 1/u$. In view of the latter condition, it should be recalled that the decay interaction of ion-acoustic waves in the plasma with $u \ll 1$ can be ignored for $r^2 < 3$ (see inequality (2.22)). Thus, relationships (5.6) show that plasmas with the r values far from unity may possess the D function close to unity even for significantly different thermal velocities of light and heavy ions. It is also seen that, as the u value approaches r or $1/r$, the D function becomes small as compared to unity. Therefore, D deviates most significantly from unity in the interval

$$u < r < 1/u. \tag{5.7}$$

These inequalities determine a relationship between the plasma parameters for which the probability of induced scattering drops most significantly due to a difference of the ion polarizability from the static value.

For arbitrary values of the parameters r and u, the degree of proximity of the D function to (or difference from) unity is illustrated in Fig. 2, where the D value is plotted versus the ion thermal velocity ratio v_{T2}/v_{T1} for various values of the parameter $r = r_{D1}/r_{D2}$. According to these data, the D value tends to unity together with the velocity ratio. The region in which D is close to unity increases when the r_{D1}/r_{D2} ratio deviates from unity. For D close to unity, $\gamma_{nl}(\mathbf{k})$ is close to $\gamma_{\text{stat}}(\mathbf{k})$ and a difference of the ion susceptibility from the static value is not manifested in the magnitude of induced scattering. In contrast, for $v_{T2}/v_{T1} \ll 1$, the situation is qualitatively different as seen from formulas (5.6) and (5.7) and illustrated in Fig. 2. In this case, a consistent consideration of the dynamic polarizability in a plasma containing two species of ions with strongly different thermal velocities leads to a decrease in the damping rate of induced scattering of ion-acoustic waves on ions as compared to the value for a plasma with insignificantly different thermal velocities of ions. The smaller the v_{T2}/v_{T1} ratio, the lower the D value. At the same time the D value cannot be arbitrarily small. The domain of applicability of the relationship (5.3), as well as the range of D, is limited by the condition (5.1). If this inequality is violated, $\gamma_{nl}(\mathbf{k})$ is determined by a contribution $\gamma_0(\mathbf{k})$ (4.7) to the damping rate inherent in the plasma with a single ion species. The above considerations show that the effect of dynamic polarizability is significant for the plasmas with strongly different ion velocities and may lead to a considerable decrease

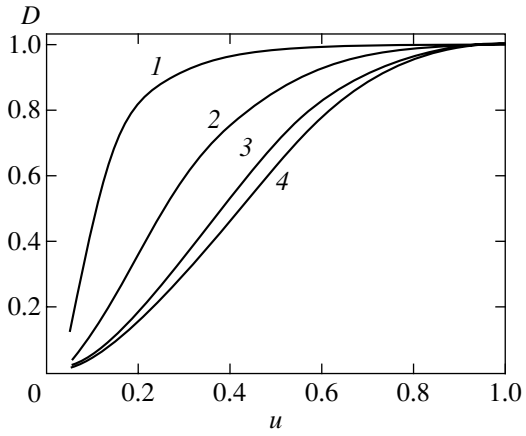


Fig. 2. Plots of the coefficient D versus the ion thermal velocity ratio $u = v_{T2}/v_{T1}$ for various values of the parameter $r = r_{D1}/r_{D2}$: (*1*) 0.1; (*2*) 0.3 and 10/3; (*3*) 0.6 and 10/6; (*4*) 1.

(down to the $\gamma_0(\mathbf{k})$ value) of the anomalously large damping rate $\gamma_{\text{stat}}(\mathbf{k})$ observed in a plasma with two species of ions. The decrease of $\gamma_{nl}(\mathbf{k})$ with decreasing D in formula (5.1) is possible as long as the condition (5.3) is fulfilled. When this inequality is not valid, the $\gamma_{nl}(\mathbf{k})$ value is determined by the $\gamma_0(\mathbf{k})$ contribution.

Now, when a differential form of the nonlinear damping rate is established, we can proceed to analysis of the ion-acoustic wave distribution in a plasma containing ions of two species. In order to establish the new regularities most clearly, we will restrict the consideration to a plasma with the quasistationary electric field $\mathbf{E} = (0, 0, E)$ and the gradients of the pressure $p = n_e\kappa T_e$, where n_e is the electron density and κ is the Boltzmann constant, and the electron temperature T_e are oriented along the z axis. In this case, the electron increment of the ion-acoustic instability $\gamma_e(\mathbf{k})$ is determined by the effective force density $\mathbf{R} = (0, 0, R)$ with

$$R = en_eE - \partial p/\partial z > 0, \tag{5.8}$$

where e is the electron charge. A quasistationary distribution of the ion-acoustic wave number density $N(\mathbf{k})$ is established due to a competition between the Cerenkov wave generation by electrons and the damping of these waves as a result of the induced scattering and the Cerenkov interaction with hot resonance ions. By the resonance ions we imply those possessing velocities above the phase velocity of the ion-acoustic wave ($v > \omega/k \sim v_s$). Since a nonisothermal plasma is characterized by $v_s \gg v_{T\alpha}$, the resonance ions occur at the tail of the distribution function. The relative fraction $n_{h\alpha}$ of such ions is small and their number density is much smaller than that of the major ion fraction n_α. At the same time, this small group of ions is effectively heated by the ion-acoustic waves due to the Cerenkov interaction. The Cerenkov absorption leads to a raid heating of the resonance ions. Eventually, the velocity distribution of the

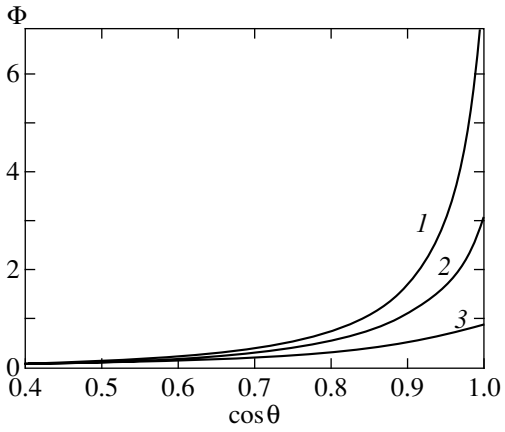


Fig. 3. Plots of the angular term in the ion-acoustic wave number density versus wavevector angle θ in the limit of small turbulent Knudsen values for $K_{stat} = 5$, $\delta = 5$, and three values of the parameter $D = 0.3$ (*1*), 0.5 (*2*), 1.0 (*3*).

heated resonance ions is characterized by an effective thermal velocity $v_{h\alpha}$ exceeding the acoustic wave velocity v_s, for which reason such ions are usually called the hot resonance ions (for more detail, see [8]). Under these conditions, the $N(\mathbf{k})$ is determined by the following equation:

$$\gamma_e(\mathbf{k}) + \gamma_1(\mathbf{k}) + \gamma_2(\mathbf{k}) + \gamma_{nl}(\mathbf{k}) = 0, \qquad (5.9)$$

where $\gamma_\alpha(\mathbf{k}) = \delta_\alpha\gamma_s(k)$ and $\gamma_s(k)$ is the damping rate of the Cerenkov damping on electrons determined by formula (2.5), $\delta_\alpha = (n_{h\alpha}v_T^3/n_e v_{h\alpha}^3)(e_\alpha^2 m/e^2 m_\alpha)$ is the ratio of the damping rate of wave damping on hot resonance ions to $\gamma_s(k)$, $n_{h\alpha}$ and $v_{h\alpha}$ are the density and thermal velocity of hot ions of the α species, respectively, and m is the electron mass. Equation (5.9) differs from an analogous relationship studied in [6] only by the presence of an additional factor D in $\gamma_{nl}(\mathbf{k})$ determined by (5.3), which allows the results obtained previously [5, 6] to be used here.

Following [5, 6], we obtain the following relationship from (5.9):

$$N(k, \cos\theta) = \Phi(\cos\theta)\sqrt{\frac{\pi}{2}}\frac{\omega_L^6 r_{De}}{\omega_{Le}}\frac{(r_{D1}^2 + r_{D2}^2)^2}{v_{T1}^2 r_{D1}^2 + v_{T2}^2 r_{D2}^2}$$

$$\times D^{-1}\left(\frac{e_1}{m_1} - \frac{e_2}{m_2}\right)^{-2}\frac{1}{k^4(1 + k^2 r_{De}^2)}\left[\ln\frac{1 + \sqrt{1 + k^2 r_{De}^2}}{k r_{De}}\right. \qquad (5.10)$$

$$\left. - \frac{1}{\sqrt{1 + k^2 r_{De}^2}} - \frac{1}{3(1 + k^2 r_{De}^2)^{3/2}}\right],$$

$$\cos\theta \geq 0,$$

where θ is the angle between vectors **k** and **R**. In this relationship, the explicit form of the function $\Phi(\cos\theta)$ depends on the turbulent Knudsen number. This quantity is proportional to the parameter D given in (5.2) and is determined as

$$\mathrm{Kn} = DK_{stat} \equiv \frac{6\pi^2}{\lambda}\frac{R\omega_{Le}^2}{v_s\omega_L^7}$$

$$\times \frac{(v_{T1}^2 r_{D1}^2 + v_{T2}^2 r_{D2}^2)}{(r_{D1}^2 + r_{D2}^2)^2} D\left(\frac{e_1}{m_1} - \frac{e_2}{m_2}\right)^2, \qquad (5.11)$$

where $\lambda \approx 0.55$. In the limit of small Knudsen numbers, $\mathrm{Kn} < (1 + \delta)^2$ (where $\delta = \delta_1 + \delta_2$), we obtain

$$\Phi(\cos\theta) = \frac{4\mathrm{Kn}}{3\pi(1 + \delta)\cos\theta}\frac{d}{d(\cos\theta)} \qquad (5.12)$$

$$\times [\cos^4\theta(1 + \epsilon - \cos\theta)^{\alpha_\epsilon - 1}].$$

Here ϵ and α_ϵ are small parameters: $\alpha_\epsilon \approx \ln 2/\ln[(1 + \delta)^2/\mathrm{Kn}] \ll 1$ and $\epsilon \approx 2\mathrm{Kn}/3\pi(1 + \delta)^2\alpha_\epsilon \ll 1$.

In the opposite limit of $\mathrm{Kn} \gg (1 + \delta)^2$, the angular distribution is as follows:

$$\Phi(\cos\theta) = \frac{2}{\pi\cos^2\theta}\sqrt{\mathrm{Kn}}\frac{d}{d(\cos\theta)}\int_0^{\cos\theta}\frac{t^5 dt}{\sqrt{\cos^2\theta - t^2}}$$

$$\times\left[0.51 + 0.08t^2 - 0.33t^4\right. \qquad (5.13)$$

$$\left. - 0.92t^2\sqrt{1 - t^2}\ln\left(\frac{1 + \sqrt{1 - t^2}}{t}\right)\right]^{-1}.$$

Equations (5.10)–(5.13) completely determine the distribution of the ion-acoustic wave number density with respect to the wavevector **k**. The shape of the distribution with respect to the wavenumber modulus and, hence, the frequency, is independent of the parameter D. On the contrary, the shape of the angular distribution of the wavevector **k** is extremely sensitive to the parameter D, because various D values may correspond to different turbulent Knudsen numbers. As a result, a decrease in D is accompanied by relative reduction of the region of large turbulent Knudsen numbers in which distributions of the type (5.13) take place. Corresponding to the asymptotic limit of large turbulent Knudsen numbers, this form of the distribution is independent of D. On the contrary, in the limit of small turbulent Knudsen numbers, a decrease in D (accompanied by a decrease in Kn), together with small values of the parameters ϵ and α_ϵ, leads to "sharpening" of the distribution of the ion-acoustic wave number density in the direction of the force density **R** generating the instability. Figure 3 shows the angular distribution function $\Phi(\cos\theta)$ of the type (5.12) for $K_{stat} = 5$, $\delta = 5$, and various $D = 1$, 0.5, and 0.3.

6. ANOMALOUS TRANSPORT

Based on the above-described variations in the IAT spectrum, we may consider the corresponding changes of the values characterizing some electromagnetic and kinetic phenomena in a nonisothermal plasma with significantly different thermal velocities of ions. The variations in the IAT spectrum are most significantly pronounced by changes in the transport coefficients manifested in the limit of large Knudsen numbers. A description of the transport process in a plasma with developed IAT is usually based on a kinetic equation for the electron distribution function f. In the case when a quasistationary electric field $\mathbf{E}$ in the plasma and the gradient of the electron distribution function are directed along the same axis (z axis), we obtain the following equation for the limit of large Knudsen numbers:

$$\frac{\partial f}{\partial t} + v\frac{\partial f}{\partial z}\cos\theta_v + \frac{e}{m}E\left(\frac{\partial f}{\partial v}\cos\theta_v - \frac{\sin\theta_v}{v}\frac{\partial f}{\partial\theta_v}\right) = \frac{\partial}{\sin\theta_v\partial\theta_v}\left[\frac{D_{\theta\theta}}{v^2}\frac{\partial f}{\partial\theta_v}\sin\theta_v\right], \quad (6.1)$$

where $f = f(v, \theta_v, z, t)$, θ_v is the angle between the velocity vector $\mathbf{v}$ and axis z, and $D_{\theta\theta}$ is the component of the quasilinear diffusion tensor describing relaxation of the electron momentum upon scattering on the ion-acoustic oscillations of the charge density. For an IAT spectrum described by Eqs. (5.10) and (5.13), the $D_{\theta\theta}$ tensor component can be expressed as

$$D_{\theta\theta} \approx \frac{e^2}{\pi v m^2\omega_L^2}\int_0^\infty dk\omega^3 k \int_0^{\sin\theta_v} d(\cos\theta)\left(\frac{\cos\theta}{\sin\theta_v}\right)^2 \times \frac{N(k, \cos\theta)}{\sqrt{\sin^2\theta_v - \cos^2\theta}} = \frac{\nu}{\mathrm{Kn}}\frac{v_T^3}{v} \times \int_0^{\sin\theta_v} d(\cos\theta)\left(\frac{\cos\theta}{\sin\theta_v}\right)^2 \frac{\Phi(\cos\theta)}{\sqrt{\sin^2\theta_v - \cos^2\theta}}, \quad (6.2)$$

where ν is the frequency characterizing relaxation of the momentum of thermal electrons in the turbulent plasma:

$$\nu = \sqrt{\frac{9\pi}{8}}\frac{R}{n_e m v_s}. \quad (6.3)$$

In deriving Eq. (6.2), we took into account that the phase velocity of the ion-acoustic waves is small as compared to the electron velocity. In addition, the integration with respect to the wavenumber modulus was performed from zero to infinity. This approximation is justified by the fact that the main contribution to the integral with respect to k is observed in the region of $k \sim 1/r_{De}$. Let us assume that the field $\mathbf{E}$ and the electron distribution inhomogeneity lead to a small deviation $\delta f = f - f_0$ of the electron distribution function from the isotropic value $f_0 = f_0(v, z, t)$ in the space of velocities. Then, using (6.1), we obtain the following relationship for the times greater than the momentum relaxation time for the major electron fraction:

$$-\frac{\partial\delta f}{\partial(\cos\theta_v)} = \frac{v^2}{2D_{\theta\theta}}\left(\frac{e}{m}E\frac{\partial f_0}{\partial v} + v\frac{\partial f_0}{\partial z}\right). \quad (6.4)$$

Using Eq. (6.4) it is possible to calculate the electron fluxes along the z axis. Assuming that the function f_0 for the major electron fraction is close to the Maxwell distribution f_m and using relationships (5.13) and (6.2)–(6.4), we obtain an expression for the current density:

$$j = e\int d\mathbf{v} f v\cos\theta_v = -e\pi\int_0^\infty dv\, v^3\int_0^\pi d\theta_v\frac{\partial\delta f}{\partial\theta_v}\sin^2\theta_v = \frac{8}{\pi}\beta e n_e v_s\sqrt{\mathrm{Kn}}\left(1 - \frac{3p}{2R}\frac{\partial}{\partial z}\ln T_e\right), \quad (6.5)$$

where $\beta \approx 0.45$. By the same token, we calculate the thermal electron flux:

$$q = \frac{m}{2}\int d\mathbf{v} f v^3\cos\theta_v = -\frac{\pi}{2}m\int_0^\infty dv\, v^5\int_0^\pi d\theta_v\frac{\partial\delta f}{\partial\theta_v}\sin^2\theta_v = \frac{32}{\pi}\beta p v_s\sqrt{\mathrm{Kn}}\left(1 - \frac{5p}{2R}\frac{\partial}{\partial z}\ln T_e\right). \quad (6.6)$$

Under conditions frequently encountered in the case of laser plasmas, the current density is zero ($j = 0$). In the zero-current plasma, it is interesting to determine the thermal flux limiting factor $f_q = -q/pv_s$. Using the condition $j = 0$, we obtain from (6.5) and (6.6) an expression

$$f_q = \frac{64}{3\pi}\beta\frac{\omega_L}{\omega_{Le}}\sqrt{\mathrm{Kn}} \approx 3\frac{\omega_L}{\omega_{Le}}\sqrt{DK_{\mathrm{stat}}} < 3\frac{\omega_L}{\omega_{Le}}\sqrt{K_{\mathrm{stat}}}. \quad (6.7)$$

According to this, the thermal flux limiting factor is $\sqrt{D}$ times smaller than that appearing in the theory neglecting the frequency dispersion of the plasma polarizability in the amplitude of probability of the induced scattering of ion-acoustic waves on ions.

In the case of a plasma characterized by a homogeneous electron temperature, relationship (6.5) allows the effective frequency of electron collisions to estimated as $\nu_{\mathrm{eff}} = R/n_e mu$, where $u = j/en_e$ is the electron drift velocity.

The first estimate of ν_{eff} obtained by Sagdeev [12] referred to a turbulent plasma containing ions of a single species:

$$\nu_{\mathrm{eff}}^{(s)} = \rho_s\omega_{Li}\frac{u}{v_s}\frac{ZT_e}{T_i}, \quad (6.8)$$

where Z and ω_{Li} are the ionization multiplicity and the Langmuir frequency of ions, respectively, $v_s = \omega_{Li}r_{De}$,

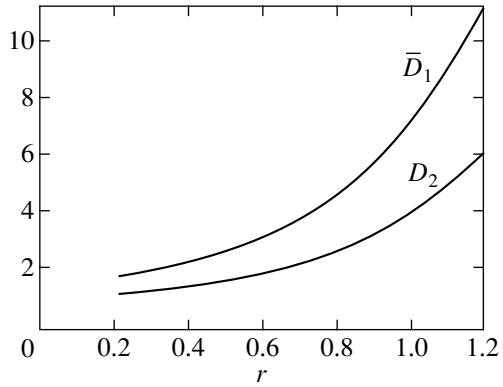


Fig. 4. The plots of coefficients $\bar{D}_1$ and D_2 versus r for $r \le 1.2$.

and $\rho_s \approx 10^{-2}$ [12]. An expression of the (6.8) type also appears in the theory of IAT for a plasma containing ions of a single species, but with a slightly different coefficient $\rho_s \approx 2.5 \times 10^{-2}$ (for more detail, see [8]).

If a plasma contains the ions of two species, while the effect of dynamic charge separation can be ignored (so that an inequality opposite to (5.1) is valid), the IAT theory yields

$$\nu_{\text{eff}} = \rho_s u v_s \frac{\omega_L^7}{\omega_{Le}^4}\left[\sum_{\alpha=1,2}\left(\frac{e_\alpha m}{m_\alpha e} v_{T\alpha}\omega_{L\alpha}\right)^2\right]^{-1}, \quad (6.9)$$

where $\rho_s = \lambda\pi/786\beta_1^2 = 0.025$ and $\beta_1 = 0.28$. Relationship (6.9) generalizes the Sagdeev formula to the case of plasma containing ions of two species. This generalization corresponds to conditions when the nonlinear damping rate of induced scattering on ions γ_{nl} can be calculated by adding the two contributions (due to ions of different species) analogous to those determined in a theory of plasma with ions of a single species.

A different generalization of the Sagdeev formula for a plasma containing ions of two species corresponds to conditions when ν_{stat} takes into account an increase in the probability of induced scattering related to the dynamic charge separation for ions. Taking $D = 1$, which correspond to the static polarizability of ions at the frequency of beats of the interacting waves, and assuming that inequality (5.1) is valid, we arrive at the expression obtained in [6]:

$$\nu_{\text{stat}} = \rho\frac{u}{v_s}\frac{\omega_L^7}{\omega_{Le}^4} \times \frac{(r_{D1}^2 + r_{D2}^2)^2}{(\omega_{L1}^2 r_{D1}^4 + \omega_{L2}^2 r_{D2}^4)m^2}\frac{e^2}{}\left(\frac{e_1}{m_1} - \frac{e_2}{m_2}\right)^{-2}, \quad (6.10)$$

where $\rho = \pi\lambda/96\beta^2 \approx 0.1$. Under the conditions (5.1) with $D = 1$, the effective frequency ν_{stat} given by (6.10) is smaller than the value determined using formula (6.9), which contains a relatively large parameter—the ratio of the square acoustic wave velocity (v_s^2) to the squares of ion thermal velocities ($v_{T\alpha}^2$)—and depends in a different way on the plasma parameters. The lower ν_{eff} value obtained in [6] is explained by a relative increase in the probability of induced scattering due to the dynamic charge separation for ions.

Still another generalization of the Sagdeev formula is obtained if we take into account both the effect of the dynamic charge separation and the difference between the dynamic and static polarizabilities of the plasma. This very case is studied in our theory, where conditions (5.1) are assumed to be valid. Using (5.11) and (6.5), we obtain

$$\nu_{\text{eff}} = \nu_{\text{stat}}/D. \quad (6.11)$$

Thus, the effective frequency ν_{eff} (6.11) differs by the factor $1/D \ge 1$ from the value of ν_{stat} given by (6.10). For D approaching unity, ν_{eff} is close to ν_{stat}. On the contrary, the case of $D \ll 1$ corresponds to a considerable increase of ν_{eff} as compared to ν_{stat} in the theory ignoring the difference between dynamic and static plasma polarizabilities. However, this relative increase in ν_{eff} (6.11) with decreasing D is observed as long as inequality (5.1) is valid. When D becomes so small that inequality (5.1) is no longer satisfied, expression (6.9) instead of (6.11) should be used for ν_{eff}.

Now we will discuss in more detail the form of the parameter D for a plasma with significantly different thermal velocities of ions:

$$v_{T1} \gg v_{T2}. \quad (6.12)$$

As can be seen from Fig. 2, this case corresponds to the most pronounced manifestation of the dynamic polarizability effect. A minimum D value in this system is achieved for $r \approx 1$. In view of this, let us consider a restricted interval of r values $r \le 1.2$. This would allow us, on the one hand, to study the region of minimum D values and, on the other hand, to derive a simpler expression for D (see formula (6.15) below).

Under condition (6.12), formulas (4.11)–(4.13) yield the approximation

$$D_1 \approx \left(\frac{v_{T2}}{v_{T1}}\right)^3 \frac{(1+r^2)^2}{\sqrt{2\pi}} \times \int_{-\infty}^{\infty} dx x^2 \left|\frac{1 - J_+(x)}{1 + r^2[1 - J_+(x)]}\right|^2 \equiv \bar{D}_1\left(\frac{v_{T2}}{v_{T1}}\right)^3, \quad (6.13)$$

$$D_2 \approx \frac{(1+r^2)^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dx x^2 \frac{\exp(-x^2/2)}{\left|1 + r^2[1 - J_+(x)]\right|^2}. \quad (6.14)$$

The plots of $\bar{D}_1$ and D_2 are presented in Fig. 4. As can be seen, the D_2 value is close to unity for $r \ll 1$ and increases approximately six times on approaching

$r = 1.2$. In the same interval of r values, the coefficient $\bar{D}_1$ increases from ~1.6 to ~11. In the whole interval, the D_2 and $\bar{D}_1$ values differ no more than by a factor of 2. Taking into account this behavior of the coefficients $\bar{D}_1$ and D_2 for $r \le 1.2$, expression (5.2) for the parameter D can be simplified to yield

$$D = \frac{D_2 + \bar{D}_1 r^2 v_{T2}/v_{T1}}{1 + r^2 (v_{T1}/v_{T2})^2} \approx \frac{D_2}{1 + (r_{D1} v_{T1}/r_{D2} v_{T2})^2}. \quad (6.15)$$

From this it follows that, according to (5.7), the parameter D is significantly smaller than unity provided that

$$\frac{v_{T1}}{v_{T2}} \gg \frac{r_{D2}}{r_{D1}} \gtrsim 1. \quad (6.16)$$

In the plasmas satisfying conditions (6.16), an allowance for the difference of the polarizability of slow ions from the static value leads to a significant modification of both the turbulence spectrum and to a considerable change of the transport coefficients in the limit of large turbulent Knudsen numbers. It should be recalled that, for the manifestation of the dynamic polarizability effect, it is also necessary to satisfy, in addition to conditions (6.16), inequality (5.1) ensuring smallness of the contribution to the induced scattering damping rate (inherent in the plasmas with ions of a single species). Under the conditions studied, inequality (5.1) takes the form

$$\left(1 - \frac{e_2 m_1}{e_1 m_2}\right)^2 \gg \frac{v_{T1}^2}{v_s^2} \frac{\omega_{L1}^2}{\omega_{L2}^2}, \quad (6.17)$$

which can be satisfied in the plasmas with sufficiently hot electrons, when the acoustic wave velocity is much greater than the thermal velocity of light ions. In addition, the charge-to-mass ratios of the ions of two species must be not close to each other.

7. GENERATION OF RUNAWAY ELECTRONS

In this section, we will consider the formation of runaway electrons in a homogeneous plasma containing ions of two species, exposed to a sufficiently strong quasistationary electric field. The system is studied under the same conditions as in Section 6—for a sufficiently large turbulent Knudsen number $\mathrm{Kn} \gg (1 + \delta)^2$—when the effect of the dynamic ion polarizability at the frequency of beats of the interacting ion-acoustic waves is most pronounced.

In order to determine the flux of runaway electrons, we will use a kinetic equation for the large isotropic part of the electron distribution function f_0. This equation is derived from (6.1) by averaging over the velocity vector directions and using relationship (6.4). Ignoring the inhomogeneity of the electron distribution in (6.1) and (6.4) we obtain

$$\frac{\partial f_0}{\partial t} = \frac{e^2 E^2}{8 m^2 v^2} \frac{\partial}{\partial v} \times \left[\int_{-1}^{1} d(\cos\theta_v) \frac{\sin^2\theta_v}{D_{\theta\theta}} v^2 \frac{\partial f_0}{\partial v} \right] + \mathrm{St}(f_0), \quad (7.1)$$

where $\mathrm{St}(f_0)$ is the electron–electron collision integral. The presence of this integral implies that we consider the conditions under which, on the one hand, the effect of electron–electron collisions on the anomalous transport is negligibly small and the results of Sections 5 and 6 are applicable; on the other hand, it is assumed that the electron–electron collisions are sufficient to provide that the isotropic part of the distribution function of the major fraction of electrons is close to the Maxwell distribution. When the distribution function of the thermal electrons is close to the Maxwell distribution, the kinetics of nonthermal electrons with large velocities ($v \gg v_T$) can be described using an approximate expression for the electron–electron collision integral of the type

$$\mathrm{St}(f_0) = v \nu_{ee}(v) \frac{\partial}{\partial v} \left[\frac{v_T^2}{v} \frac{\partial f_0}{\partial v} + f_0 \right], \quad (7.2)$$

where $\nu_{ee}(v)$ is the frequency of the electron–electron collisions determined by the formula

$$\nu_{ee}(v) = 4\pi e^4 n_e \Lambda m^{-2} v^{-3} \quad (7.3)$$

and Λ is the Coulomb integral.

Let us consider a quasistationary solution to Eq. (7.1) under the conditions of a relatively small diffusion flux of electrons toward the region of high velocities. Taking into account relationships (5.13), (6.2), and (7.2), the result of integration of Eq. (7.1) can be presented in the following form:

$$\frac{v_T^2}{v} \left[1 + \left(\frac{v}{v_m} \right)^6 \right] \frac{d f_0}{d v} + f_0 = -\frac{S(v)}{4\pi v_T^3 \nu_{ee}}, \quad (7.4)$$

where $S(v)$ is the electron flux in the space of velocities defined as

$$S(v) = -4\pi \frac{d}{dt} \int_0^v dv\, v^2 f_0 \quad (7.5)$$

(since there is no source of electrons for $v = 0$, we take $S(0) = 0$). In Eq. (7.4), $\nu_{ee} = \nu_{ee}(v_T)$ and the quantity v_m denotes the characteristic electron velocity depending on the turbulent Knudsen number:

$$v_m = v_T \left(\frac{9\pi}{\beta} \frac{\nu_{ee}}{\nu_E} \frac{v_T^2}{v_s^2 \sqrt{\mathrm{Kn}}} \right)^{1/6} \gg v_T, \quad (7.6)$$

where $\nu_E = (9\pi/8)^{1/2}|e|E/mv_s > 0$. Following the theory of the formation of runaway electrons [19], we will consider the solution to Eq. (7.4) in the approximation of a stationary electron flux in the space of velocities. In this approximation, Eq. (7.4) yields

$$f_0 \approx -\frac{S}{4\pi v_T^3 \nu_{ee}} + \frac{n_e}{(2\pi)^{3/2} v_T^3} \times \exp\left[-\int_0^v \frac{v'dv'}{v_T^2(1 + v'^6/v_m^6)}\right]. \quad (7.7)$$

Using the condition that the velocity-independent part of the distribution function (7.7) is zero for $v \gg v_m$, we obtain the following expression for the flux of runaway electrons:

$$S = \sqrt{\frac{2}{\pi}} n_e \nu_{ee} \exp\left(-\frac{\pi}{3\sqrt{3}}\frac{v_m^2}{v_T^2}\right) = \sqrt{\frac{2}{\pi}} n_e \nu_{ee} \exp\left(-\sqrt{\frac{E_c}{E}}\right), \quad (7.8)$$

where E_c is the critical electric field strength

$$E_c = E_{Dr}\sqrt{\frac{8}{\pi}}\left(\frac{\pi}{3}\right)^{8/3}\left(\frac{\omega_{Le}}{\omega_L}\frac{\nu_N}{\beta^2\nu_{ee}}\right)^{1/3} D^{-1/3} \gg E_{Dr}, \quad (7.9)$$

$E_{Dr} = m\nu_{ee}v_T|e|^{-1}$ is the Drycer field strength, and ν_n is the characteristic turbulent frequency

$$\nu_N = \frac{\lambda}{\sqrt{2\pi}}\frac{\omega_L^7}{\omega_{Le}^4}\frac{(r_{D1}^2 + r_{D2}^2)^2}{(v_{T1}^2 r_{D1}^2 + v_{T2}^2 r_{D2}^2)} \times \frac{e^2}{m^2}\left(\frac{e_1}{m_1} - \frac{e_2}{m_2}\right)^{-2}. \quad (7.10)$$

According to formula (7.9), the critical electric field strength E_c in the turbulent plasma is much greater than the Drycer field strength determining the runaway electron flux in a laminar plasma. The effect of the dynamic ion polarizability on the critical electric field strength E_c in a plasma containing ions of two species is manifested in (7.9) by the additional factor $D^{-1/3} > 1$. This relative increase in the critical electric field strength implies that the dynamic polarizability leads to a relative suppression of the runaway electron formation in a turbulent plasma containing ions of two species.

8. DISCUSSION

The IAT theory developed above refers to a plasma with ions of two species and takes into account the difference of the ion polarizability from the static value at a frequency of beats of the interacting ion-acoustic waves. This theory allowed us to establish the conditions for which the previously used static approximation is applicable to description of the induced scattering of ion-acoustic waves on ions and indicate the conditions under which the difference between the dynamic and static polarizability leads to a significant decrease in the probability of induced scattering. According to an analysis of the D function (Section 5), the static approximation can be used for the description of plasmas with slightly different thermal velocities of ions, for which the D value is close to unity. This is related to the fact that the frequency dependences of the contributions of different ions to the dynamic polarizability of the plasma are described by comparable functions.

The possibility of using the static polarizability approximation may also take place for a plasma with strongly different thermal velocities of ions of the two species, provided that the ratio of the Debye radii of these ions satisfies the conditions (5.6): $r_{D1}/r_{D2} < v_{T2}/v_{T1} \ll 1$ or $r_{D1}/r_{D2} > v_{T1}/v_{T2} \gg 1$. Under these conditions, the function D is determined for the most part by a contribution of ions of the second or first species, respectively, while the corresponding coefficients D_2 and D_1 are close to unity by themselves. Note that, in the case of $r_{D1}/r_{D2} < v_{T2}/v_{T1} \ll 1$, the parameter r must be not too small. The limitation of r from below follows from the condition (5.1). When this condition is not satisfied, the damping rate of the induced scattering on ions is determined by a contribution of the type (4.7) characteristic of a plasma with ions of a single species and the IAT theory is analogous to the theory of such plasmas.

On the contrary, if the ratio of the Debye radii of ions satisfies the conditions $v_{T2}/v_{T1} < r_{D1}/r_{D2} < v_{T1}/v_{T2}$ in plasmas with considerably different thermal velocities of ions, the D value can be significantly smaller than unity (see Fig. 2 and Eq. (5.7)). For these plasma parameters, the difference of the ion polarizability from the static value leads to a significant decrease in the probability of induced scattering.

In order to elucidate the conditions favoring experimental manifestations of the dynamic polarizability effect studied above, let us consider two examples of plasmas containing ions of two species with not strongly different thermal velocities. The ion temperatures for both species will be assumed equal: $T_1 = T_2 = T_i$. We will use the conventional notations of the mass number $A_\alpha = m_\alpha/m_p$ and the ionization multiplicity $Z_\alpha = e_\alpha|e|^{-1}$, where $|e|$ and m_p are the proton charge and mass, respectively. In these notations, the plasma parameters determining the main coefficient $D = D(u, r)$ (5.2) in the IAT theory developed in this study are as follows:

$$u = \frac{v_{T2}}{v_{T1}} = \sqrt{\frac{A_1}{A_2}}, \quad r = \frac{r_{D1}}{r_{D2}} = \frac{Z_2}{Z_1}\sqrt{\frac{n_2}{n_1}}, \quad (8.1)$$

where n_α is the number density of ions of the α species ($\alpha = 1, 2$). In these terms, the two principal conditions (2.9)

and (5.1) for applicability of the static theory acquire the following form

$$\frac{T_e}{T_i} \gg F_1\left(\frac{Z_1}{Z_2}, u, r\right) = \frac{3}{Z_1}\left(1 + r^2\frac{Z_1}{Z_2}\right)(1 + r^2u^2)^{-1}, \quad (8.2)$$

$$\frac{T_e}{T_i} \gg F_2\left(\frac{Z_1}{Z_2}, u, r\right) = D^{-1}(u, r)\left(1 - \frac{Z_2}{Z_1}u^2\right)^{-2} Z_1^{-1} \times \frac{(1 + r^2)^2(1 + r^2Z_1/Z_2)}{r^2(r^2 + u^2)(1 + r^2u^2)}\left[1 + r^2u^8\left(\frac{Z_2}{Z_1}\right)^2\right], \quad (8.3)$$

representing requirements to the degree of plasma nonisothermicity.

First, let us consider the coefficient D determined by relationships (5.2) and the conditions (8.2) and (8.3) for a fully ionized C_xH plasma, where $x = n_2/n_1$ is the ratio of the atomic number densities of carbon and hydrogen. For this plasma, $Z_1/Z_2 = 1/6$ and $u^2 = A_1/A_2 = 1/12$. Figure 5 shows plots of the functions D (5.2), F_1 (8.2), and F_2 (8.3) versus the parameter r (8.1) (hence, versus the n_2/n_1 ratio). As can be seen from this figure, D exhibits a minimum ($D \approx 0.28$) for $r \approx 1$ ($n_2/n_1 \approx 1/36$). At this point, according to the condition (8.3), the plasma must be significantly nonisothermal: $T_e/T_i > 50$. As the carbon density increases, the limitation from below on the degree of nonisothermicity becomes not as strong. For example, at $r \approx 3$ ($n_2/n_1 \approx 1/4$), we obtain $T_e/T_i > 10$. In this case, however, the effect of decreasing induced scattering is characterized by $D \sim 0.5$.

Another example refers to a lead oxide Pb_xO plasma, where $x = n_2/n_1$ is the ratio of the atomic number densities of lead and oxygen. The estimates were obtained for $Z_1/Z_2 = 8/25$, which corresponds to the full ionization of oxygen ($Z_1 = 8$) and a partial ionization of lead ($Z_2 = 25$); $A_1/A_2 = 16/207$. The corresponding plots of the functions D (5.2), F_1 (8.2), and F_2 (8.3) versus the parameter r (8.1) are depicted by dashed curves in Fig. 5. Here, the coefficient D is also minimum ($D \approx 0.28$) for $r \approx 1$ ($n_2/n_1 \approx 64/625 \approx 0.1$). At this point, limitations from below on the plasma nonisothermicity are not significant: $T_e/T_i > 4$. For $r \approx 3$ ($n_2/n_1 \approx 0.9$), we obtain $D \approx 0.5$ and there is virtually no limitation from below on the degree of nonisothermicity: $T_e/T_i > 1$.

Now we will consider two more examples referring to the plasmas with relatively small ratios of the thermal velocities of ions. The first is an HXe_x plasma with the ion mass ratio $u^2 = A_1/A_2 = 1/131$. Let us assume that the hydrogen component is fully ionized ($Z_1 = 1$) and the xenon component is ionized to $Z_2 = 2$. Figure 6 shows the plots of $0.001F_2$ (8.3) and $10D$ (5.2) for $r = 2\sqrt{n_2/n_1}$ varied from 0.3 to 3. In this range, F_1 (8.2) is much smaller than F_2 (8.3) and the plot of the former value is omitted. According to these data, D has a minimum at $r = 1$ (or $n_2 = 0.25n_1$); the minimum value of D is much smaller then unity: $D_{\min} \approx 0.03$. Such a small value can be realized only for $T_e/T_i > 200$. The limitation from below on the nonisothermicity decreases with increasing atomic number density of xenon: for $n_2 \approx n_1$ ($r = 2$), $T_e/T_i > 100$ (in this case D only slightly increases up to ~0.045). Such a relatively high degree of nonisothermicity of this plasma at a small degree of xenon ionization can be reached in quasistationary current discharge regimes at a small plasma density. For example, this is possible in an HXe_x plasma with $T_e \sim 20$ eV, $T_i \sim 0.2$ eV, and $n_1 \sim n_2 \sim 10^{13}$ cm^{-3}. From the data presented in Fig. 6, we may conclude that a decrease in the induced scattering owing to the dynamic polarizability effect in the HXe_x plasma can be very large. The corresponding small D values are realized only provided a considerable nonisothermicity of the plasma.

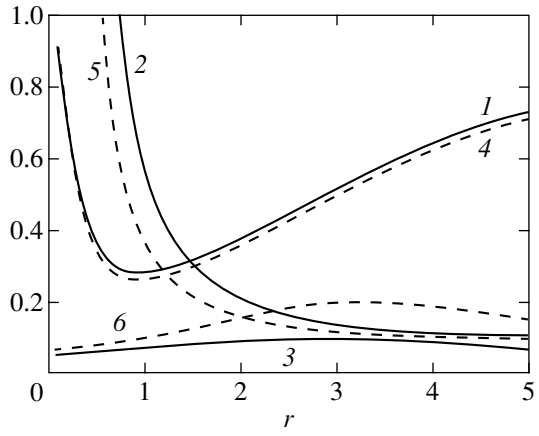


Fig. 5. Plots of the coefficient D (determining a decrease in the induced scattering probability) and the functions F_1 and F_2 (characterizing limitations of the degree of nonisothermicity) versus the parameter $r = \sqrt{x}Z_2/Z_1$: solid curves *1–3* show D, $0.01F_2$, and $0.01F_1$ values for a fully ionized C_xH plasma with $Z_1 = 1$, $Z_2 = 6$, $x = n_2/n_1$; dashed curves *4–6* show D, $0.1F_2$, and $0.1F_1$ values for a partly ionized Pb_xO plasma with $Z_1 = 8$, $Z_2 = 25$.

Similar behavior is observed in a $LiAu_x$ plasma containing the ions of lithium and gold with $A_1/A_2 = 7/197$. Assuming $Z_1 = 3$ and $Z_2 = 12$, we obtain $r = 4\sqrt{n_2/n_1}$. The plots of $0.01F_2$ (8.3) and D (5.2) in the range $0.3 \leq r \leq 3$ are depicted by dashed curves in Fig. 6. In the entire interval, F_1 is significantly smaller than F_2 and the former value is not plotted in Fig. 6. As can be seen, the D value is minimum ($D_{\min} \approx 0.13$) for $r = 1$ ($n_2 = n_1/16$). Here, the required degree of nonisothermicity is not too large: $T_e/T_i > 16$. For $n_2 = n_1/4$) ($r = 2$), we obtain $D \approx 0.19$ and $T_e/T_i > 6$. Thus, a decrease in the induced scattering is large at a relative weak nonisothermicity of

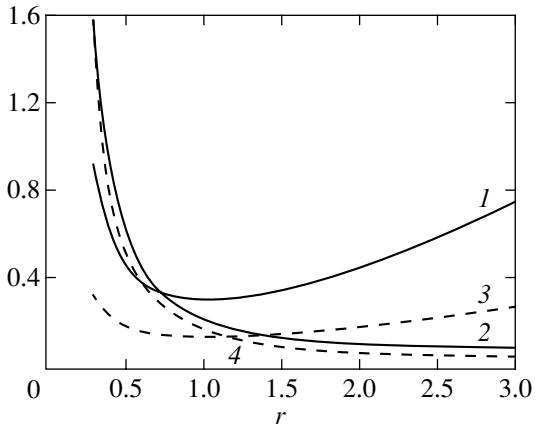


Fig. 6. Solid curves: the plots of (*1*) $10D$ and (*2*) $0.001F_2$ versus $r = \sqrt{x}Z_2/Z_1$ for a HXe_x plasma with $Z_1 = 1$, $Z_2 = 2$, $A_1 = 1$, and $A_2 = 131$; dashed curves: the plots of (*3*) D and (*4*) $0.01F_2$ versus $r = \sqrt{x}Z_2/Z_1$ for a $LiAu_x$ plasma with $Z_1 = 3$, $Z_2 = 12$, $A_1 = 7$, and $A_2 = 197$.

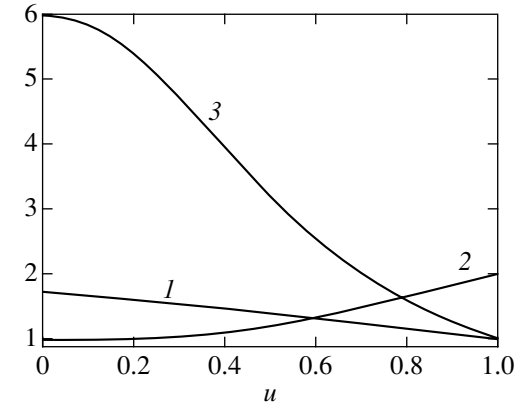


Fig. 7. The plots of (*1*) $a(u)$, (*2*) $b(u)$, and (*3*) $c(u)$ functions.

the plasma. It should be emphasized that the conditions (2.22) are valid for both of the last examples.

9. CONCLUSION

The above considerations and examples show that a significant decrease in the induced scattering probability may take place for a comparatively readily attainable plasma parameters. A relative decrease in the nonlinear damping rate of ion-acoustic waves is accompanied by the corresponding growth in the level of turbulent pulsations of the charge density. This leads to a relative increase in the effective frequency of electron collisions, a decrease in the factor of limitation of the thermal flux of electrons, and a reduction in the runaway electron production. The results presented above are of importance for description of the turbulent state of objects such as laser plasmas and current discharges in gases containing ions of two species.

ACKNOWLEDGMENTS

This study was supported in parts by the Russian Foundation for Basic Research (project no. 99-02-18075), Civilian Research and Development Foundation (CRDF grant no. RP1-2268), International Scientific-Technological Center (ISTC grant no. 1253), and the Federal Program of Support for Leading Scientific Schools (project no. 00-15-96720).

APPENDIX

Consider the coefficient D (5.2) for a plasma with close thermal velocities of ions, for which

$$\Delta \equiv 1 - u \ll 1. \tag{A.1}$$

Using Eqs. (5.2) and (4.11)–(4.13) and the following expansions,

$$J_+(xu) \approx J_+(x) - x\Delta\frac{d}{dx}J_+(x) + \frac{1}{2}x^2\Delta^2\frac{d^2}{dx^2}J_+(x), \tag{A.2}$$

$$\begin{aligned} J_+(x/u) \approx J_+(x) + x\Delta(1+\Delta)\frac{d}{dx}J_+(x) \\ + \frac{1}{2}x^2\Delta^2\frac{d^2}{dx^2}J_+(x), \end{aligned} \tag{A.3}$$

we obtain to within the terms quadratic in Δ:

$$D \approx 1 + Ar^2\Delta^2(1+r^2)^{-2}, \tag{A.4}$$

where A is a numerical coefficient:

$$\begin{aligned} A = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{dx x^3\exp(-x^2/2)}{|1-J_+(x)|^2} \\ \times\left\{\frac{d}{dx}|1-J_+(x)|^2\left[3 - 2x\frac{d}{dx}\ln|1-J_+(x)|\right.\right. \\ \left.\left. + x\left|\frac{d}{dx}[1-J_+(x)]\right|^2 + \frac{d^2}{dx^2}|1-J_+(x)|^2\right]\right\} \approx -4. \end{aligned} \tag{A.5}$$

The expression for the function D can be simplified in the case of small and large r values. For $r^2 \ll 1$, we may neglect a weak difference of D_2 from unity and the dependence of D_1 on r in (5.2) and (4.11)–(4.13) to obtain the following approximate relationships:

$$D \approx 1 - \frac{r^2}{u^2 + r^2}[1 - a(u)u^3], \quad r^2 \ll 1, \tag{A.6}$$

$$a(u) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dx x^2 \exp\left(-\frac{u^2x^2}{2}\right)\left|\frac{1-J_+(x)}{1-J_+(xu)}\right|^2. \quad \text{(A.7)}$$

The value of $a(u)$ is comparable with unity in the entire range $0 < u \le 1$ (see Fig. 7) and $a(1) = 1$.

If the parameter r is very large, so that $r^2u^2 \gg 1$, we may neglect the dependence of D_2 on r. Retaining the terms D_1 proportional to r^{-2} in (5.2) and (4.11)–(4.13), we obtain the following approximate relationships:

$$D \approx 1 - \frac{u^2}{r^2}[1-c(u)] - \frac{1}{r^2u^2}[b(u) - 2u^2], \quad r^2u^2 \ll 1, \quad \text{(A.8)}$$

$$b(u) = u^2\sqrt{\frac{2}{\pi}}\int_{-\infty}^{\infty} dx x^2 \exp\left(-\frac{x^2}{2}\right)\mathrm{Re}\left[\frac{1-J_+(x)}{1-J_+(x/u)}\right], \quad \text{(A.9)}$$

$$c(u) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dx x^2 \exp\left(-\frac{x^2}{2}\right)\left|\frac{1-J_+(xu)}{1-J_+(x)}\right|^2. \quad \text{(A.10)}$$

Here, $b(0) = 1$, $b(1) = 2$, $c(0) = 6$, and $c(1) = 1$. Plots of the functions $b(u)$ and $c(u)$ in the interval $0 < u \le 1$ are presented in Fig. 7.

REFERENCES

1. E. K. Zavoiskii and L. I. Rudakov, At. Énerg. **23**, 417 (1967).
2. E. D. Volkov, N. F. Perepelkin, V. A. Suprunenko, and E. A. Sukhomlin, *Collective Phenomena in Current-Carrying Plasma* (Naukova Dumka, Kiev, 1979).
3. B. B. Kadomtsev, in *Reviews of Plasma Physics*, Ed. by M. A. Leontovich (Atomizdat, Moscow, 1964; Consultants Bureau, New York, 1968), Vol. 4, p. 258.
4. V. I. Petviashvili, Dokl. Akad. Nauk SSSR **153**, 1295 (1963) [Sov. Phys. Dokl. **8**, 1218 (1964)].
5. V. P. Silin and S. A. Uryupin, Zh. Éksp. Teor. Fiz. **102**, 78 (1992) [Sov. Phys. JETP **75**, 41 (1992)].
6. V. P. Silin and S. A. Uryupin, Fiz. Plazmy **19**, 894 (1993) [Plasma Phys. Rep. **19**, 464 (1993)].
7. A. A. Galeev and R. Z. Sagdeev, in *Reviews of Plasma Physics*, Ed. by M. A. Leontovich (Atomizdat, Moscow, 1973; Consultants Bureau, New York, 1979), Vol. 7, p. 5.
8. V. Yu. Bychenkov, V. P. Silin, and S. A. Uryupin, Phys. Rep. **164**, 119 (1988).
9. I. V. Kuzora, V. P. Silin, and S. A. Uryupin, Phys. Lett. A **258**, 329 (1999).
10. V. N. Oraevskiĭ and R. Z. Sagdeev, Zh. Tekh. Fiz. **32**, 1291 (1962) [Sov. Phys. Tech. Phys. **7**, 955 (1963)].
11. V. V. Pustovalov and V. P. Silin, Tr. Fiz. Inst. Akad. Nauk SSSR **61**, 42 (1972).
12. R. Z. Sagdeev, in *Magneto-Fluid and Plasma Dynamics: Proceedings of the 18th Symposium on Applied Mathematics, New York, 1965* (American Mathematical Society, Providence, 1967), p. 281.
13. B. D. Fried, R. B. White, and T. K. Samec, Phys. Fluids **13**, 1290 (1970).
14. I. M. A. Gledhill and M. A. Hellberg, J. Plasma Phys. **36**, 75 (1986).
15. E. A. Williams, R. L. Berger, R. P. Drake, *et al.*, Phys. Plasmas **2**, 129 (1995).
16. H. X. Vu, J. M. Wallace, and B. Bezzerides, Phys. Plasmas **1**, 3542 (1994).
17. V. N. Tsytovich, in *Theory of Turbulent Plasma* (Atomizdat, Moscow, 1971; Plenum, New York, 1974), p. 93.
18. V. P. Silin and A. A. Rukhadze, in *Electromagnetic Properties of Plasma and Plasmalike Media* (Atomizdat, Moscow, 1961), p. 91.
19. A. V. Gurevich and Yu. N. Zhivlyuk, Zh. Éksp. Teor. Fiz. **49**, 214 (1965) [Sov. Phys. JETP **22**, 153 (1966)].

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 93, No. 5, 2001, pp. 1052–1064.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 4, 2001, pp. 1213–1226.

PLASMA, GASES

Weak Magnetohydrodynamic Turbulence of a Magnetized Plasma

E. A. Kuznetsov

Landau Institute of Theoretical Physics, Russian Academy of Sciences, Moscow, 117334 Russia
e-mail: kuznetso@itp.ac.ru

Received June 5, 2001

Abstract—A weak turbulence of the magnetohydrodynamic waves in a strongly magnetized plasma was studied in the case when the plasma pressure is small as compared to the magnetic field pressure. In this case, the principal nonlinear mechanism is the resonance scattering of fast magnetoacoustic and Alfvén waves on slow magnetoacoustic waves. Since the former waves are high-frequency (HF) with respect to the latter, the total number of HF waves in the system is conserved (adiabatic invariant). In the weak turbulence regime, this integral of motion generates a Kolmogorov spectrum with a constant flux of the number of HF waves toward the longwave region. The shortwave region features a Kolmogorov spectrum with a constant energy flux. An exact angular dependence of the turbulence spectra is determined for the wave propagation angles close to the average magnetic field direction. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The central place in the theory of turbulence belongs to the concept of a turbulence spectrum representing the energy distribution over scales. Determining the turbulence spectrum is a difficult problem that still remains unsolved. Important results in this field were obtained by Kolmogorov [1] and Obukhov [2], which showed an self-similar character of the spectrum of developed hydrodynamic turbulence.

In the 1970s, the ideas of Kolmogorov and Obukhov were fruitfully developed and applied, mostly due to the effort of Zakharov, in the theory of weak wave turbulence (see monograph [3] and the first original papers [4–6]). The wave turbulence has proved to be, in a certain sense, somewhat simpler than the hydrodynamic turbulence. The presence of a wave dispersion results in that there exists a wave intensity region where the interaction between waves can be considered as weak. If the initial phase distribution of the waves is random, the weak nonlinear interaction provides for a small correlation between phases of the interacting waves. For this reason, the waves can be described in terms of the pair correlation functions with the Fourier images coinciding (to within a factor) with the number of waves n_k (occupation number) possessing a given wavevector $\mathbf{k}$. In turn, the occupation numbers n_k obey the kinetic wave equations. In this theory, the Kolmogorov spectra appear in the form of stationary scale-invariant solutions to the kinetic equations, annulating the collisional term. These spectra, in contrast to the thermodynamically equilibrium ones, refer to solutions of the flux type realizing a constant flux of some integral of motion (energy, number of particles, etc.) over scales. It is important to note that the concept of the inertial interval (a region where the pumping and damping effects can be ignored), which is formulated as an assumption (a hypothesis of the locality of interaction) in the case of a developed hydrodynamic turbulence, is explicitly established as the locality of spectra for the weak wave turbulence.

Most of the investigations devoted to the Kolmogorov spectra of weak turbulence refer to isotropic media (for a complete bibliography, see [3]). The effect of anisotropy, for example, of the magnetic field in a plasma, was studied to a smaller extent. The first example of determining the Kolmogorov spectra in anisotropic media for a weak turbulence of magnetized ion-sound waves was reported by the author in 1972 [7]. It was found that the collisional term in the kinetic wave equations is invariant with respect to stretching in the two independent directions (along and across the magnetic field), which allowed the anisotropic Kolmogorov spectra to be constructed with a power dependence on both longitudinal (k_z) and transverse ($k_\perp$) components of the wavevector. This, in turn, made it possible to determine (with the aid of generalized Zakharov transformations) the Kolmogorov indices and find the exact angular dependence of the Kolmogorov spectra. Later, the ideas of that study were used in determining the turbulence spectra of the drift waves and the Rossby waves (see, e.g., [8, 9]).

This paper is devoted to the study of a weak turbulence of the magnetohydrodynamic (MHD) waves in a strongly magnetized plasma in the case when the plasma (thermal) pressure nT is small as compared to

the magnetic field pressure $H^2/8\pi$:

$$\beta = \frac{8\pi nT}{H^2} \ll 1.$$

Under these conditions, the turbulence spectra are determined (unlike the cases studied previously [4, 7]) by solving three linked kinetic equations for the Alfvén waves and the fast and slow magnetoacoustic waves.

For $\beta \ll 1$, the main nonlinear interaction of MHD waves is the scattering of fast magnetoacoustic and Alfvén waves on slow magnetoacoustic waves (Section 2). In these processes (involving the decay of one wave into two, as well as the reverse process of merging), the fast magnetoacoustic and Alfvén waves act as high-frequency (HF) with respect to the slow magnetoacoustic waves. In every scattering event, a change in the frequency of the former waves (referred to below as *A*-waves) is relatively small (due to the small β value), which makes this process analogous to the Mandelstamm–Brillouin scattering of electromagnetic waves on acoustic phonons. As a result of this time scale separation, whereby the waves are divided into HF and low-frequency (LF) components, the wave-decay interaction retains, in addition to the energy, an adiabatic invariant—the total number of HF waves. This, however, does not exhaust the analogy with the Mandelstamm–Brillouin scattering. It is established that the matrix element of this interaction is maximum for a maximum value of the longitudinal momentum component transferred from *A*-waves to slow magnetoacoustic waves. This result can be derived, in particular, from an expression derived by Galeev and Oraevskii [10] for the growth rate of the decay instability of a monochromatic Alfvén wave. It should be recalled that the matrix element for the Mandelstamm–Brillouin scattering is proportional to the square root of the transmitted momentum, which accounts for the maximum backscattering of electromagnetic waves. In view of this behavior of the *A*-wave scattering amplitude, it is naturally assumed that a stationary angular distribution of these waves must be strongly anisotropic and concentrated along the magnetic field direction in the *k*-space. Under these assumptions, the kinetic equations acquire additional symmetry and become invariant with respect to stretching in the two independent directions (along and across the magnetic field), which allows the transformations developed previously [7] to be used in this case as well.

Owing to these two features of the kinetic equations in the transparency range, it was possible to find two scale-invariant (with respect to longitudinal and transverse wavevectors) Kolmogorov spectra corresponding to a constant energy flux directed toward the shortwave region of scales (direct cascade) and a constant flux of the number of *A*-waves toward the region of small *k* (inverse cascade). This study is based on the results reported by the author long ago in the form of a preprint in Russian [11] and remained, for this reason, unavailable abroad. Moreover, it turned out that the work was also little known in Russia: despite an almost three-decade history, the results have still not been repeated. Recently, however, the question of MHD turbulence spectra was studied in the other limiting case ($\beta \gg 1$) by Galtier *et al.* [12]. This limit significantly differs from that considered below. First, a plasma with $\beta \gg 1$ can be treated as an incompressible fluid. Second, this limit introduces no significant difference between the Alfvén waves and slow magnetoacoustic waves: the waves of both types exhibit the same dispersion law and differ only by polarization. Such a degeneracy significantly changes the character of nonlinear interactions. Nevertheless, this case also admits two types of Kolmogorov spectra featuring dependence on the wavenumber analogous to that reported below. However, a physical explanation of the two spectra existing in the case of $\beta \gg 1$ is different from the interpretation given in this paper for $\beta \ll 1$.

The material is arranged as follows. Section 2 provides for a canonical description of the ideal MHD wave turbulence following the original work of Zakharov and the author [13] and the recent review [14]. Using the Hamiltonian approach, in Section 3 we derive averaged equations describing the interaction of *A*-waves with slow magnetoacoustic waves. It is shown that the *A*-waves represent an HF force acting upon the slow magnetoacoustic waves. Since the potential of this force is negative (in contrast to the potential of interaction between the Langmuir waves and the ion-sound waves) [15], the plasma is drawn into the regions of *A*-wave localization to form the density "humps." Stability of a monochromatic *A*-wave is also studied in Section 3. Section 4 is devoted to the Kolmogorov spectra of a weak MHD turbulence.

2. VARIATION PRINCIPLE AND NORMAL VARIABLES

Let us consider the ideal MHD equations for a barotropic flow in a plasma, the internal energy ε of which can be considered as dependent only on the plasma density ρ:

$$\frac{\partial \rho}{\partial t} + \mathrm{div}\rho\mathbf{v} = 0, \tag{1}$$

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v}\cdot\nabla)\mathbf{v} = -\nabla w + \frac{1}{4\pi\rho}[\mathrm{curl}\mathbf{H}\times\mathbf{H}], \tag{2}$$

$$\frac{\partial \mathbf{H}}{\partial t} = \mathrm{curl}[\mathbf{v}\times\mathbf{H}]. \tag{3}$$

Here, $\mathbf{v}$ is the plasma velocity and w is the enthalpy related to the pressure $p = p(\rho)$ and the internal energy ε by the relationships

$$dw = \frac{dp}{\rho}, \quad w = \frac{\partial}{\partial\rho}\varepsilon(\rho).$$

A variational principle for this system of equations can be formulated as follows. First, it can be seen from

Eqs. (1)–(3) that the vector $\mathbf{H}/\rho$ moves with the "fluid current" line; in other words, each field line moves with the particles occurring on this line, which corresponds to a well-known concept of the "frozen-in" magnetic field (see, e.g., [16]). This circumstance allows the magnetic field $\mathbf{H}$ and the plasma density ρ to be considered as generalized coordinates.

To formulate the variational principle, we will use the well-known expression for the Lagrangian of the electromagnetic field containing particles of a fluid [17]. We will write an expression for the Lagrangian L in the MHD approximation with neglect of a contribution due to the electric field relative to that due to the magnetic field, since $E \sim (v/c)H \ll H$. Taking into account relationships expressed by Eqs. (1) and (3) and the condition $\mathrm{div}\mathbf{H} = 0$ as constraints, we can write

$$L = \frac{\rho\mathbf{v}^2}{2} - \varepsilon(\rho) - \frac{H^2}{8\pi} + \mathbf{S}\cdot\left(\frac{\partial\mathbf{H}}{\partial t} - \mathrm{curl}[\mathbf{v}\times\mathbf{H}]\right) + \Phi\left(\frac{\partial\rho}{\partial t} + \mathrm{div}\rho\mathbf{v}\right) + \psi\mathrm{div}\mathbf{H},$$

where $\mathbf{S}$, Φ, and ψ are the Lagrange multipliers. Now we can use the so determined Lagrangian to introduce the functional of action

$$I = \int L dt d\mathbf{r},$$

the variation of which with respect to variables $\mathbf{v}$, ρ, and $\mathbf{H}$ leads to the following set of equations:

$$\rho\mathbf{v} = \mathbf{H}\times\mathrm{curl}\mathbf{S} + \rho\nabla\Phi, \tag{4}$$

$$\frac{\partial\Phi}{\partial t} + (\mathbf{v}\cdot\nabla)\Phi - \frac{v^2}{2} + w(\rho) = 0, \tag{5}$$

$$\frac{\partial\mathbf{S}}{\partial t} + \frac{\mathbf{H}}{4\pi} - \mathbf{v}\times\mathrm{curl}\mathbf{S} + \nabla\psi = 0. \tag{6}$$

The first equation suggests the change of variables whereby the velocity $\mathbf{v}$ is expressed in terms of the new variables $\mathbf{S}$ and Φ. It must be emphasized that this change is ambiguous, since we may add a vector $\mathbf{S}_0$ to $\mathbf{S}$ and a scalar Φ_0 to Φ such that

$$\mathbf{H}\times\mathrm{curl}\mathbf{S}_0 + \rho\nabla\Phi_0 = 0.$$

The two other equations, (5) and (6), represent the Bernoulli equation for the potential Φ and the equation of motion for the new vector $\mathbf{S}$ with an unknown potential ψ. The latter potential is set by fixing the gauge of vector $\mathbf{S}$. For example, the Coulomb gauge ($\mathrm{div}\mathbf{S} = 0$) determined ψ to within an arbitrary solution ψ_0 of the Laplace equation $\Delta\psi_0 = 0$:

$$\psi = \frac{1}{\Delta}\mathrm{div}[\mathbf{v}\times\mathrm{curl}\mathbf{S}] + \psi_0.$$

In particular, if $\mathbf{v} \longrightarrow 0$, $\mathbf{H} \longrightarrow \mathbf{H}_0$, and $\rho \longrightarrow \rho_0$ for $r \longrightarrow \infty$, the term ψ_0 is conveniently chosen so that $S \longrightarrow 0$ for $r \longrightarrow \infty$. Then

$$\psi_0 = -\frac{\mathbf{H}_0\cdot\mathbf{r}}{4\pi}.$$

Now we have to check that system (4)–(6) does not contradict the set of MHD equations. Substituting (4) into the equation of motion (2) and making simple transformations, we obtain

$$\nabla\left(\frac{\partial\Phi}{\partial t} + (\mathbf{v}\cdot V)\Phi - \frac{\mathbf{v}^2}{2} + w(\rho)\right) + \left[\frac{\mathbf{H}}{\rho}\times\mathrm{curl}\left\{\frac{\partial S}{\partial t} + \frac{\mathbf{H}}{4\pi} - \mathbf{v}\times\mathrm{curl}\mathbf{S}\right\}\right] = 0.$$

By virtue of Eqs. (5) and (6), this equation turns into an identity. Thus, we have proved that the new system of Eqs. (1), (3), (5), and (6) is equivalent to the set of MHD equations. Indeed, any solution of this system generates, by virtue of (4), a solution to the MHD equations. If we assume uniqueness of the Cauchy problem for systems (1)–(4) and (1), (3), (5) and (6), the inverse statement is also valid: for any solution to Eqs. (1)–(4) we can find a certain class of solutions to system (1), (3), (5), and (6). Indeed, this is achieved by constructing all possible sets of $\mathbf{S}$ and Φ satisfying Eq. (4) for a given set of $\mathbf{v}$, $\mathbf{H}$, ρ at a given time instant t_0 and taking these $\mathbf{S}$ and Φ values as the initial conditions for system (1), (3), (5), and (6).

Once the Lagrange function is known, we can define the generalized momenta and construct a system Hamiltonian:

$$\mathcal{H} = \int(\mathbf{S}\cdot\mathbf{H}_t + \Phi\rho_t - L)d\mathbf{r} = \int\left\{\frac{\rho\mathbf{v}^2}{2} + \varepsilon(\rho) + \frac{\mathbf{H}^2}{8\pi} - \psi\mathrm{div}\mathbf{H}\right\}d\mathbf{r},$$

which coincides in magnitude with the total energy. The equations of motion (1), (3), (5), and (6) represent the Hamilton equations

$$\frac{\partial\rho}{\partial t} = \frac{\delta\mathcal{H}}{\delta\Phi}, \quad \frac{\partial\Phi}{\partial t} = -\frac{\delta\mathcal{H}}{\delta\rho}, \quad \frac{\partial\mathbf{H}}{\partial t} = \frac{\delta\mathcal{H}}{\delta\mathbf{S}}, \quad \frac{\partial\mathbf{S}}{\partial t} = -\frac{\delta\mathcal{H}}{\delta\mathbf{H}}, \tag{7}$$

where the variables (ρ, Φ) and $(\mathbf{H}, \mathbf{S})$ represent the pairs of canonically conjugated values.

The change of variables defined by Eq. (4) and the canonical description (7) were originally introduced for the magnetic hydrodynamics in [13]. The transformation (4) is an analog for the Clebsch representation in the ideal hydrodynamics; accordingly, the variables $\mathbf{H}$ and $\mathbf{S}$ in Eq. (4) play the role of the Clebsch variables

(for the latter, see [18, 19] and a recent review [14]). Later, the same substitution was employed by Frenkel *et al.* in [20], where the velocity vector and the magnetic field were expressed through the scalar Clebsch variables; using simple transformations, this reduces to Eq. (4).

The MHD flows described by Eq (4), as well as the flows in the ideal fluid parametrized by the Clebsch variables, represent a partial flow type. For such MHD flows, the topological invariant of linkage of the magnetic field and vorticity lines [21]

$$I = \int (\mathbf{v} \cdot \mathbf{H}) d\mathbf{r}$$

is identically equal to zero.

Vladimirov and Moffatt [22] suggested an analogue of the Weber transform for the ideal MHD flows:

$$\mathbf{v} = u_{0k}(\mathbf{a}) \nabla a_k + \nabla \Phi + \frac{1}{\rho} \mathbf{H} \times \operatorname{curl} \mathbf{S}, \qquad (8)$$

where $\mathbf{a} = \mathbf{a}(\mathbf{r}, t)$ are the Lagrange markers of fluid particles (this is an inverse transformation with respect to $\mathbf{r} = \mathbf{r}(\mathbf{a}, t)$ determining the trajectory of a particle with the marker $\mathbf{a}$) and $\mathbf{u}_0(\mathbf{a})$ is a new Lagrange invariant.

The Weber transform (8) is a transformation of the general type containing the substitution (4) in a particular case of $\mathbf{u}_0 = 0$, which was not taken into account in [22]. The equations of motion for the potentials Φ and $\mathbf{S}$ have the same form as Eqs. (5) and (6). If $\Phi = 0$ and $\mathbf{S} = 0$ at $t = 0$, then $\mathbf{u}_0(\mathbf{a})$ is the initial velocity. It should be noted that it is the first term in (8) that ensures a nonzero value of the topological invariant I (this term is nonlinear if (8) is expanded in powers of small amplitude). Recently, Ruban [23] (see also [24]) elucidated a physical meaning of the new vector field $\mathbf{S}$. According to this, the quantity $\operatorname{curl} \mathbf{S}$ can be expressed through the displacement $\mathbf{d}$ between an electron and ion (considered as fluid particles) at a point $\mathbf{r}$ at the time instant t (the initial coordinates are assumed to coincide):

$$\operatorname{curl} \mathbf{S} = \frac{e}{Mc} \mathbf{d} \frac{\rho(\mathbf{r}, t)}{\rho_0(\mathbf{a})}.$$

Here M and e are the ion mass and charge, respectively, and $\rho_0(\mathbf{a})$ is the initial distribution of the plasma density.

Introduction of the canonical variables allows us to classify and study all nonlinear processes in a conventional manner, using the perturbation theory with respect to small wave amplitudes. For this purpose, it is necessary to expand the velocity and internal energy in Eq. (8) in powers of the canonical variables. If the plasma is placed into a homogeneous external magnetic field $\mathbf{H}_0$, the approximation linear in the wave amplitude must retain the terms linear in Φ and $\mathbf{S}$ and ignore the first (nonlinear) term in (8). As a result, the velocity expansion can be written as

$$\mathbf{v} = \mathbf{v}_0 + \mathbf{v}_1 + \dots, \qquad (9)$$

where the first-order term is

$$\mathbf{v}_0 = \frac{1}{\rho_0} \mathbf{H}_0 \times \operatorname{curl} \mathbf{S} + \nabla \Phi.$$

Three independent pairs ($\operatorname{div} \mathbf{H} = \operatorname{div} \mathbf{S} = 0$) of the canonically conjugated quantities correspond to the waves of three types. In the linear approximation, these waves do not interact with each other. The dispersion and polarization laws can be determined from an analysis of the quadratic (in powers of the canonical variables) Hamiltonian $\mathcal{H}_0$. The three-wave interaction corresponds to a cubic term, the magnitude of which is determined by a quadratic (in the wave amplitude) correction to the velocity:

$$\mathbf{v}_1 = \frac{\rho_1}{\rho_0^2} \mathbf{H}_0 \times \operatorname{curl} \mathbf{S} + \frac{1}{\rho_0} \mathbf{h} \times \operatorname{curl} \mathbf{S},$$

which takes into account only the "wave" degrees of freedom and neglects the first term in (8). Here, $\mathbf{h}$ and ρ_1 are the deviations of the magnetic field strength and the plasma density from the corresponding equilibrium values $\mathbf{H}_0$ and ρ_0. As a result, the Hamiltonian of the medium can be also written as an expansion in powers of the wave amplitude

$$\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_3 + \dots, \qquad (10)$$

with the quadratic Hamiltonian

$$\mathcal{H}_0 = \int \left\{ \frac{\rho_0 \mathbf{v}_0^2}{2} + \frac{\mathbf{h}^2}{8\pi} + c_s^2 \frac{\rho_1^2}{2\rho_0} \right\} d\mathbf{r},$$

and the cubic Hamiltonian

$$\mathcal{H}_3 = \int \left\{ \rho_0 (\mathbf{v}_0 \cdot \mathbf{v}_1) + \frac{\rho_1}{2} v_0^2 + q c_s^2 \frac{\rho_1^3}{2\rho_0^2} \right\} dr.$$

In these expressions, the squared sound velocity c_s^2 and the dimensionless coefficient q appeared as a result of expansion of the internal energy ε in powers of ρ_1:

$$\Delta\varepsilon(\rho) = \frac{\rho_0 c_s^2}{2} \left\{ \left(\frac{\rho_1}{\rho_0} \right)^2 + q \left(\frac{\rho_1}{\rho_0} \right)^3 + \dots \right\}.$$

Now let us accomplish the Fourier transform with respect to coordinates and pass to the new variables $a_j(k)$ ($j = 1, 2, 3$), which yields

$$\mathbf{h}(k) = \mathbf{e}_1(k) \sqrt{2\pi\omega_1} (a_1(k) + a_1^*(-k))$$

$$+ \mathbf{e}_2(k) \sum_{l = 2, 3} \lambda_l \sqrt{2\pi\omega_l} (a_l(k) + a_1^*(-k)),$$

$$\mathbf{S}(k) = -i\mathbf{e}_1(k)\frac{1}{\sqrt{8\pi\omega_1}}(a_1(k) - a_1^*(-k))$$
$$- i\mathbf{e}_2(k)\sum_{l=2,3}\lambda_l\frac{1}{\sqrt{8\pi\omega_l}}(a_l(k) - a_l^*(-k)), \quad (11)$$
$$\rho_1(k) = \sum_{l=2,3}\left(\frac{\rho_0\omega_l}{2c_s^2}\right)^{1/2}\mu_l(a_k(l) + a_{-k}^*(l)),$$
$$\Phi(k) = -i\sum_{l=2,3}\left(\frac{c_s^2}{2\rho_0\omega_l}\right)^{1/2}\mu_l(a_l(k) - a_l^*(-k)).$$

Here

$$\omega_1(k) = |\mathbf{k}\cdot\mathbf{V}_A|,$$
$$\omega_{2,3}(k) = \frac{1}{2}\Big|\sqrt{k^2V_A^2 + k^2c_s^2 + 2(\mathbf{k}\cdot\mathbf{V}_A)kc_s}$$
$$\pm\sqrt{k^2V_A^2 + k^2c_s^2 - 2(\mathbf{k}\cdot\mathbf{V}_A)kc_s}\Big|$$

are the dispersion laws for the Alfvén waves ($j = 1$) and the fast ($j = 2$) and slow ($j = 3$) magnetoacoustic waves; the corresponding unit polarization vectors are

$$\mathbf{e}_1(k) = \frac{[\mathbf{k}\times\mathbf{n}_0](\mathbf{k}\cdot\mathbf{n}_0)}{|\mathbf{k}\times\mathbf{n}_0||\mathbf{k}\cdot\mathbf{n}_0|}, \quad \mathbf{e}_2(k) = \frac{\mathbf{k}\times[\mathbf{k}\times\mathbf{n}_0]}{k|\mathbf{k}\times\mathbf{n}_0|}$$

($\mathbf{n}_0 = \mathbf{H}_0/H_0$ is the unit vector of the average magnetic field);

$$\mathbf{V}_A = \frac{\mathbf{H}_0}{(4\pi\rho_0)^{1/2}}$$

is the Alfvén velocity; and

$$\lambda_2 = -\mu_3 = -\left(1 - \frac{\omega_3^2 - k^2c_s^2}{\omega_2^2 - k^2c_s^2}\right)^{1/2},$$
$$\lambda_3 = \mu_2 = \left(1 - \frac{\omega_2^2 - k^2c_s^2}{\omega_3^2 - k^2c_s^2}\right)^{-1/2}.$$

The change of variables $a_k(j)$ represents a canonical U–V transform diagonalizing the Hamiltonian $\mathcal{H}_0$:

$$\mathcal{H}_0 = \sum_j\int\omega_j(k)a_j(k)a_j^*(k)d\mathbf{k}.$$

The amplitudes $a_j(k)$ play the role of normal variables, for which the equations of motion acquire the canonical form

$$\frac{\partial a_j(k)}{\partial t} = -i\frac{\delta\mathcal{H}}{\delta a_j^*(k)}.$$

In the linear approximation, the quantities $a_j(k)$ obey the following equations:

$$\frac{\partial a_j(k)}{\partial t} + i\omega_j(k)a_j(k) = 0,$$

which imply that the amplitude modulus $|a_j(k)|$ does not change with the time t, while the phase grows linearly with t.

In order to find an expression for the interaction Hamiltonian in terms of the variables $a_j(k)$, it is necessary to substitute transform (11) into expansion (10). As a result, the Hamiltonian of the wave interaction has the form of an integro-power series with respect to $a_j(k)$. In the lowest order with respect to the wave amplitude, the principal nonlinear process is the three-wave resonance interaction corresponding to the Hamiltonian

$$\mathcal{H}_{\text{int}} = \frac{1}{2}\int\sum_{lmn}[V_{kk_1k_2}^{lmn}a_l^*(k)a_m(k_1)a_n(k_2) + \text{c.c.}] \quad (12)$$
$$\times\delta_{k-k_1-k_2}dkdk_1dk_2.$$

This Hamiltonian is obtained by substituting transform (11) into the cubic Hamiltonian $\mathcal{H}_3$ and separating the corresponding resonance terms. The remaining terms in $\mathcal{H}_3$ are small and can be excluded with the aid of a canonical transformation (for more detail, see [14]). Note that calculation of the matrix elements $V_{kk_1k_2}^{ije}$ in this scheme is a purely algebraic procedure involving the Fourier transform in the integrals, substitution of (11), and symmetrization with respect to variables $a_k(i)$ [for example, with respect to (k_1, m) and (k_2, n) in Eq. (12)].

3. AVERAGED EQUATIONS

Expressions for the dispersion laws and the matrix elements of interaction can be significantly simplified in the case of a plasma with small $\beta = 8\pi nT/H^2$ (representing the ratio of the thermal plasma pressure nT to the magnetic field pressure $H^2/8\pi$). The condition $\beta \ll 1$ implies that $V_A \gg c_s$. In this limit, the fast magnetoacoustic waves possess an isotropic dispersion law $\omega_2 = kV_A$ and their phase (and group) velocity coincides with the group velocity of the Alfvén waves. In this linear approximation, the velocity of plasma in the Alfvén waves and the fast magnetoacoustic waves is determined by the following formula:

$$\mathbf{v}_{HF} = \frac{1}{\rho_0}\mathbf{H}_0\times\operatorname{curl}\mathbf{S}.$$

The potential part of the plasma velocity $\nabla\Phi$ is a small quantity with respect to the parameter β. In contrast, the main contribution to the velocity of slow magnetoacoustic

waves is due to the potential part. For this reason, this velocity is directed along the magnetic field $\mathbf{H}_0$,

$$\mathbf{v}_s = \mathbf{n}_0 \frac{\partial \Phi}{\partial z}, \tag{13}$$

and the dispersion of the slow magnetoacoustic waves becomes strongly anisotropic:

$$\omega_3 \equiv \Omega_s = |k_z| c_s. \tag{14}$$

The transverse velocity components in these waves $[\mathbf{H}_0 \times \operatorname{curl} \mathbf{S}]/\rho_0$ are compensated by the term $\nabla_\perp \Phi$.

If the plasma is collisionless and strongly nonisothermal ($T_e \gg T_i$), the slow magnetoacoustic waves represent the magnetized ion-sound waves (considered in more detail in [7]). In this case, the sound velocity in Eq. (14) can be expressed as

$$c_s = \sqrt{T_e/M}.$$

As for a nonlinear interaction of the MHD waves, the principal nonlinear mechanism is the resonance scattering of fast magnetoacoustic and Alfvén waves on the slow magnetoacoustic waves. This is clearly seen from a comparison of the calculated matrix elements V^{lmn} in the Hamiltonian (12). In this process, the former waves (called *A*-waves) appear as high-frequency (HF) with respect to the latter (simply referred to below as the acoustic waves, or S-waves). This conclusion follows from an analysis of the resonance conditions for this decay process:

$$\omega_A(k) = \omega_A(k_1) + \Omega_s(k_2), \quad \mathbf{k} = \mathbf{k}_1 + \mathbf{k}_2. \tag{15}$$

It is quite easy to see qualitatively how this interaction proceeds in the system studied. As an *A*-wave packet propagates in the plasma, the average characteristics (plasma density and velocity) slowly vary under the action of these waves. Owing to this, the average Alfvén velocity differs from a local value by the quantity

$$\Delta V_A = -V_A \rho_{1s}/2\rho_0,$$

where ρ_{1s} is the low-frequency (LF) density variation. As a result, frequencies of the *A*-wave acquire an increment $\Delta\omega_\rho \sim k\Delta V_A$. Due to a slow motion with the drift velocity v_D, the *A*-wave frequency changes by $\Delta\omega_D \sim kv_D$ (Doppler effect). The ratio of the two frequencies ($\Delta\omega_D$ and $\Delta\omega_\rho$) is a small quantity with respect to the parameter c_s/V_A. Therefore, a principal interaction is the scattering on the LF density fluctuations. Note that the LF characteristics of the plasma change under the HF force action of the *A*-waves.

An expression for the HF force is most simply derived by averaging the Hamiltonian over the HF oscillations. Upon this averaging, the Hamiltonian acquires the following form:

$$\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_{\text{int}}, \tag{16}$$

where

$$\mathcal{H}_0 = \int \left\{ \frac{1}{2\rho_0} \langle [\mathbf{H}_0 \times \operatorname{curl} \mathbf{S}]^2 \rangle + \frac{\langle \mathbf{h}^2 \rangle}{8\pi} \right\} d\mathbf{r}$$

$$+ \int \left\{ \frac{\rho_0 \Phi_z^2}{2} + c_s^2 \frac{\rho_{1s}^2}{2\rho_0} \right\} d\mathbf{r},$$

$$\mathcal{H}_{\text{int}} = -\int \frac{\rho_{1s}}{2\rho_0^2} \langle [\mathbf{H}_0 \times \operatorname{curl} \mathbf{S}]^2 \rangle d\mathbf{r}$$

(the angle brackets denote the averaging over high frequencies). The first integral in $\mathcal{H}_0$ corresponds to the *A*-waves, while the second represents the acoustic oscillations of the magnetized plasma. Variation of the interaction Hamiltonian with respect to ρ_{1s} yields the following expression for the HF interaction potential:

$$U \equiv M \frac{\delta \mathcal{H}_{\text{int}}}{\delta \rho_{1s}} = -\frac{1}{2Mn_0^2} \langle [\mathbf{H}_0 \times \operatorname{curl} \mathbf{S}]^2 \rangle. \tag{17}$$

According to this, the equation of motion for the potential Φ_s is as follows:

$$\frac{\partial \Phi_s}{\partial t} + c_s^2 \frac{\rho_{1s}}{\rho_0} = \frac{\langle [\mathbf{H}_0 \times \operatorname{curl} \mathbf{S}]^2 \rangle}{2\rho_0^2}. \tag{18}$$

It is important to note that the HF potential (17) is negative, which implies that the HF force acting in the region of localization of the *A*-waves leads to the appearance of the density "humps" (instead of the "dips" observed for the interaction of ion-sound waves with the Langmuir waves [25]).

The equations of motion for the slow wave component are closed by the equation of continuity for ρ_{1s} According to (13), this equation can be written as

$$\frac{\partial \rho_{1s}}{\partial t} + \rho_0 \frac{\partial^2 \Phi_s}{\partial z^2} = 0. \tag{19}$$

Using Eq. (18) and (19), we obtain

$$\frac{\partial^2 \rho_{1s}}{\partial t^2} - c_s^2 \frac{\partial^2 \rho_{1s}}{\partial z^2} = -\frac{1}{2\rho_0} \frac{\partial^2}{\partial z^2} \langle [\mathbf{H}_0 \times \operatorname{curl} \mathbf{S}]^2 \rangle. \tag{20}$$

In order to derive equations for the *A*-waves, it is necessary to perform averaging of the interaction Hamiltonian $\mathcal{H}_{\text{int}}$. This is achieved by retaining terms of the type $a_\lambda^* a_{\lambda_1}$ ($\lambda = 1, 2$ is the HF wave number):

$$H_{\text{int}} = -\int \frac{\rho_{1s}(k)}{2\rho_0} \sum_{\lambda\lambda_1} F_{kk_1}^{\lambda\lambda_1} a_\lambda^*(k) a_{\lambda_1}(k_1) \delta_{k-k_1-\kappa} d\mathbf{k} d\kappa,$$

where

$$F_{kk_1}^{\lambda\lambda_1} = (\omega_\lambda(k)\omega_{\lambda_1}(k_1))^{1/2} \mathbf{n}_\lambda(k) \cdot \mathbf{n}_{\lambda_1}(k_1),$$

$$\mathbf{n}_2 = \frac{\mathbf{k}_\perp}{k_\perp}, \quad \mathbf{n}_1 = -\mathbf{n}_2 \times \mathbf{n}_0.$$

As a result, the *A*-wave equations acquire the following form:

$$\frac{\partial a_\lambda(k)}{\partial t} + i\omega_\lambda(k)a_\lambda(k) = -i\frac{\delta H_{\rm int}}{\delta a_\lambda^*(k)}, \quad \lambda = 1, 2. \quad (21)$$

A collisionless isothermal ($T_e \approx T_i$) plasma features no slow magnetoacoustic oscillations as a result of the strong Landau damping on ions. Accordingly, the *A*-wave decay interaction [15] changes to the induced scattering of *A*-waves on ions. In this case, Eqs. (20) have to be replaced by a system of the kinetic drift equation [26] for a slow variation of the ion distribution function f_i (cf. [27]),

$$\frac{\partial f_i}{\partial t} + v_z\frac{\partial f_i}{\partial z} - \frac{1}{M}\frac{\partial}{\partial z}(e\tilde{\varphi} + U)\frac{\partial f_0}{\partial v_z} = 0, \quad (22)$$

and the condition of quasineutrality for the slow motions ($\Omega_k = k_z c_s \ll \omega_{pi}$),

$$\delta n_i = \int f_i d\mathbf{v} = \frac{n_0}{T_e}e\tilde{\varphi} = \frac{\rho_{1s}}{M}, \quad (23)$$

where f_0 is the equilibrium ion distribution function and $\tilde{\varphi}$ is the LF electrostatic potential. The equations of motion for the *A*-waves retain the form of Eq. (21) and the plasma density is linearly expressed via the HF potential in terms of Green's function for the system of Eqs. (22) and (23):

$$G_{\kappa\Omega} \equiv \frac{\rho_{1s}(\kappa, \Omega)}{U_{\kappa\Omega}} = -\frac{n_0\kappa^2}{\omega_{pi}^2}\frac{\epsilon_e\epsilon_i}{\epsilon_e + \epsilon_i}. \quad (24)$$

Here, $\rho_{1s}(\kappa, \Omega)$ and $U_{\kappa\Omega}$ are the Fourier images of the LF density and HF potential, respectively, and $\epsilon_{e,i}$ are the partial permittivities of electrons and ions:

$$\epsilon_e = \frac{1}{\kappa^2 r_d^2},$$

$$\epsilon_i = \frac{4\pi e^2}{M\kappa^2}\int\frac{\kappa_z(\partial f_0/\partial v_z)}{\Omega - k_z v_z}d\mathbf{v},$$

($r_d^2 = T_e/4\pi n_0 e^2$ is the squared Debye radius). In a strongly nonisothermal plasma ($T_e \gg T_i$), Green's function (24) converts into

$$G_{\kappa\Omega} = \frac{n_0\kappa_z^2}{\Omega^2 - \kappa_z^2 c_s^2},$$

which coincides with the expression for the function determined by Eq. (20).

The system of equations (22)–(24) completely describes the interaction of *A*-waves in a strongly magnetized plasma with an arbitrary ratio of the electron and ion temperatures. However, because of the Landau damping on ions, the Hamiltonian $H_0 + H_{\rm int}$ is no longer conserved.

4. INSTABILITY OF A MONOCHROMATIC WAVE

Now we proceed to analysis of the equations derived in the preceding section. First, let us consider the behavior of a narrow *A*-wave packet. A qualitative pattern of this process can be outlined by studying the stability of a monochromatic *A*-wave. For simplicity, we will restrict the consideration to the stability of an Alfvén wave in a hydrodynamic limit. For a collisionless plasma, this implies that a phase velocity of Ω/k_z of beats in the *A*-wave exceeds the thermal ion velocity v_{Ti}. Under these conditions, we may neglect the Landau damping on ions for the slow acoustic oscillations and use Eqs. (20) or (24). It should be born in mind that the acoustic waves in a strongly nonisothermal plasma represent intrinsic oscillations, whereas the sound generated in an isothermal plasma ($T_e \approx T_i$) represents induced oscillations in the plasma density. However, under the condition $\Omega/k_z \gg v_{Ti}$, the hydrodynamic description is applicable in both cases.

It is convenient to express ρ_{1s} through the normal variables $a_3(k) \equiv b_k$:

$$\rho_{1s}(k) = \left(\frac{\rho_0\Omega_k}{2c_s}\right)^{1/2}(b_k + b_{-k}^*).$$

Equations for the new variables $b(k)$ are obtained by variation of the total Hamiltonian $H_0 + H_{\rm int}$:

$$\frac{\partial b_k}{\partial t} + i\Omega(k)b_k = -i\frac{\delta H_{\rm int}}{\delta b_k^*}. \quad (25)$$

A monochromatic Alfvén wave corresponds to the following equation of Eqs. (21) and (25):

$$a_\lambda(k) = \frac{A}{\omega_0^{1/2}}\delta_{\lambda 1}\exp(-i\omega_0 t)\delta_{k-k_0}, \quad b_k = 0,$$

$$\omega_0 = \omega_1(k_0).$$

Here, the Alfvén wave amplitude is selected so that the value $W = |A|^2$ would coincide with the energy density of the oscillations.

Upon linearizing Eqs. (22)–(24) relative to the exact solution and taking perturbations in the form of

$$\delta a_\lambda(k) \propto \exp(-i(\Omega + \omega_0)t)\delta_{k-k_0-\kappa},$$

$$\delta a_\lambda^*(k) \propto \exp(-i(\Omega - \omega_0)t)\delta_{k-k_0+\kappa},$$

we obtain the following dispersion relationship for Ω:

$$\frac{WG}{4Mn_0^2\omega_0}\sum_\lambda\left\{\frac{\left|F^{1\lambda}_{k_0, k_0+\kappa}\right|^2}{\Omega + \omega_0 - \omega_\lambda(k_0 + \kappa)} + \frac{\left|F^{1\lambda}_{k_0, k_0-\kappa}\right|^2}{-\Omega + \omega_0 - \omega_\lambda(k_0 - \kappa)}\right\} = 1. \quad (26)$$

Now we will present the results of investigation of the dispersion relationship (26) in various special cases depending on the oscillation energy density W and the ratio of electron and ion temperatures in the plasma.

For $T_e \gg T_i$ and sufficiently small oscillation amplitudes, the A-wave exhibits decay instability with the ion-sound excitation [10]. For this instability, the frequency Ω can be expressed through the matrix element of the decay interaction

$$V_{kk_1k_2}^{\lambda\lambda_1} = \left(\frac{\Omega(k_2)}{8\rho_0 c_s^2}\right)^{1/2} F_{kk_1}^{\lambda\lambda_1} \tag{27}$$

and the energy density W as

$$\Omega = \frac{1}{2}[\omega_0 - \omega_\lambda(k_0 - \kappa) + \Omega(\kappa)] \pm \left\{\frac{1}{4}[\omega_0 - \omega_\lambda(k_0 - \kappa) - \Omega(\kappa)]^2 - \frac{W}{\omega_0}\left|V_{k_0, k_0-\kappa, \kappa}^{1\lambda}\right|^2\right\}^{1/2}. \tag{28}$$

From this expression, it follows that the instability takes place in the vicinity of the resonance surface

$$\omega_0 = \omega_\lambda(k_0 - \kappa) + \Omega(\kappa) \tag{29}$$

with a maximum growth rate

$$\Gamma = \left[\frac{W}{8nT}\frac{\Omega_\kappa}{\omega_0}\left|F_{k_0, k_0-\kappa}^{1\lambda}\right|^2\right]^{1/2}. \tag{30}$$

The growth rate width with respect to the frequency is on the order of the maximum value (30).

Since the matrix element is proportional to the square root of the slow sound frequency, the maximum growth rate on the resonance surface (29) is reached for the maximum value of $|\kappa_z|$. Upon decay into the Alfvén wave and the slow acoustic wave,

$$\max|\kappa_z| \approx 2|k_{0z}|,$$

which implies that the secondary Alfvén wave propagates in the direction opposite to that of the primary (exciting) Alfvén wave. The character of the decay instability is typical of the Mandelstamm–Brillouin scattering, the matrix element for which is proportional to the square root of the momentum of light transmitted to the acoustic waves as a result of scattering. This accounts for the maximum backscattering of light.

Now we can readily investigate the decay instability in all other channels of the decay process $A \longrightarrow A + S$. In all these cases, the growth rates are on the same order of magnitude as the growth rate determined by formula (30):

$$\Gamma \sim (\omega_0 \Omega_s W/nT)^{1/2}.$$

This instability takes place for $W/nT < \beta^{1/2}$. As the W/nT ratio increases, the decay instability is modified. For $W/nT > \beta^{1/2}$, we may neglect Ω_s^2 in comparison to Ω^2 in relationship (26). Then, the unstable wavevectors are lying on the surface,

$$\omega_1(k_0) = \omega_\lambda(k_0 - \kappa).$$

This instability is referred to as the modified decay instability [15, 28]. In the case of interaction between the Alfvén waves and the slow acoustic waves, this instability has a growth rate reaching maximum at $\kappa_z = 2k_{0z}$:

$$\Gamma \approx \frac{\sqrt{3}}{2}\omega_0\left(\frac{W}{\rho_0 V_A^2}\right)^{1/3}. \tag{31}$$

Since this value is independent of the temperature, the same instability may take place for $W/nT > 1$ as well (up to $W/nT \sim 1/\beta$, when the main assumption of adiabaticity, $\Gamma \sim \omega_0$, fails to be valid).

In the other channels, the behavior of instability with increasing parameter W/nT exhibits the same pattern: for $W/nT > \beta^{1/2}$, the growth rate reaches maximum at $\kappa \sim k_0$ and coincides in the order of magnitude with the value given by formula (31).

As can be readily seen, a decay instability with the growth rate (30) for an arbitrary channel $A \longrightarrow A + S$ belongs to instability of the convective type. According to relationship (28), the group velocities of the excited waves are significantly different from the group velocity of the primary (exciting) wave. Therefore, for the wave packet with a length L, the instability will be significant only provided L is sufficiently large so that the gain G would exceed the Coulomb logarithm Λ:

$$G = \Gamma L/V_A \approx \Lambda.$$

For smaller L values, the decay instability will be not manifested since the perturbation amplitude acquires only a small increment during the time required for the perturbation to travel through the entire packet length. In this case, the wave packet dynamics is determined by slower processes. Among these, the most important are related to the unstable perturbations propagating together with the wave packet. Should it be a decay instability, this instability must be absolute (in the coordinate system moving with the wave packet). This is one of the possible factors for a collapse of the fast magnetoacoustic waves producing a special effect on the structure of collisionless shock waves in a plasma [29, 30]. The collapse of fast magnetoacoustic waves arises as a result of a three-wave interaction involving only the fast magnetoacoustic waves.

5. THE KOLMOGOROV SPECTRA

In the preceding section, we have considered the instability of a wave packet narrow in the k-space. Upon the decay of a monochromatic wave obeying the resonance conditions (29) (i.e., reaching the maximum growth rate given by formula (30)), the sum of phases

of the excited waves ($\phi_A + \phi_s$) is strictly related to the pumping wave phase (ϕ_0):

$$\phi_0 + \pi/2 = \phi_A + \phi_s.$$

The phase difference in the pair of excited waves with a fixed wavevector κ is arbitrary. As can be readily checked, the above phase correlation is lost on deviating from the resonance conditions (29). Both these factors introduce an element of stochasticity into the system of interacting triads related to the pumping wave. Thus, each triad is characterized by a single random phase. In the next stage (secondary cascade), new random phases are added and the memory of a coherent pumping wave is lost. Upon numerous repeats of this process, the system must pass to a turbulent state in which the wave phases can be considered random. Therefore, the randomization time must be equal to several times the reciprocal growth rate given by formula (30). This scenario of transition to a turbulent state seems to be quite realistic. A series of recent numerical experiments were aimed at the verification of this hypothesis (see, e.g., [31, 32].

Based on the above considerations, it is clear that a regime of developed turbulence can be expected to possess a wide spectrum of waves. This spectrum can be statistically described in terms of correlation functions. For the waves of small intensity, it is sufficient to restrict the consideration to pair correlation functions obeying the kinetic equations. This regime is referred to as weakly turbulent.

In the case of a weak MHD turbulence ($\beta \ll 1$), we obtain three pair correlation functions determined by the formulas

$$\langle a_\lambda(k) a^*_{\lambda_1}(k_1)\rangle = N^\lambda_k \delta_{\lambda\lambda_1}\delta_{k-k_1},\ \langle b_k b^*_{k_1}\rangle = n_k \delta_{k-k_1}.$$

Here, the coefficients N^λ_k and n_k having the sense of occupation numbers obey the following system of equations:

$$\begin{aligned}\dot n_k &= 2\pi\int |V_{k_1k_2k}|^2 (N_{k_1}N_{k_2} - n_k N_{k_1} + n_k N_{k_2}) \\ &\quad\times \delta_{k+k_1-k_2}\delta_{\Omega+\omega_1-\omega_2}dk_1dk_2,\end{aligned} \tag{32}$$

$$\begin{aligned}\dot N_k &= 2\pi\int |V_{kk_1k_2}|^2 (N_{k_1}n_{k_2} - N_k n_{k_2} - N_k N_{k_1}) \\ &\quad\times \delta_{k-k_1-k_2}\delta_{\omega-\omega_1-\Omega_2}dk_1dk_2 \\ &\quad - 2\pi\int |V_{k_1kk_2}|^2 (N_k n_{k_2} - N_{k_1}n_{k_2} - N_k N_{k_1}) \\ &\quad\times \delta_{k_1-k-k_2}\delta_{\omega_1-\omega-\Omega_2}dk_1dk_2,\end{aligned} \tag{33}$$

where $\omega \equiv \omega(k)$, $\omega_1 \equiv \omega(k_1)$, etc. In these equations (and below) we omitted the summation with respect to λ, which can be restored by substituting

$$dk_1 \longrightarrow \sum_n dk_1, \quad N_k \longrightarrow N^\lambda_k,$$

$$\omega_k \longrightarrow \omega_{k\lambda}, \quad V_{kk_1k_2} \longrightarrow V^{\lambda\lambda_1}_{kk_1k_2},$$

etc.

Equations (32) and (33) assume that the nonlinear wave interaction is weak. In this particular case, the most significant condition is

$$\Omega_s \gg 1/\tau,$$

where τ is the characteristic nonlinear time determined from the kinetic equations (32) and (33). In order to estimate τ, it is necessary to take into account that, in every wave decay event and the reverse (wave merging) process, the *A*-wave frequencies change by a small increment $\Delta\omega_A = \Omega_s \ll \omega_A$ so that the *A*-wave energy is redistributed over the spectrum in a diffusion manner. Taking his into account, we obtain the following estimate for τ:

$$\frac{1}{\tau} \sim \omega_A \frac{W}{\rho V_A^2}.$$

Note that this τ value is significantly greater than the characteristic time of randomization determined as the reciprocal growth rate Γ^{-1} determined by formula (30). Finally, the criterion can be written in the following form:

$$\frac{W}{\rho V_A^2} \ll \beta^{1/2}.$$

Now let us include the sources of turbulence and damping into Eqs. (32) and (33). For this purpose, we introduce the terms $\Gamma_k n_k$ and $\gamma_{k\lambda}N_{k\lambda}$ into the left-hand parts of these equations, respectively. It will be assumed that the domains of pumping (Γ_k, $\gamma_{k\lambda} > 0$) and damping (Γ_k, $\gamma_{k\lambda} < 0$) are separated in the *k*-space by an intermediate region (inertial interval) in which the turbulence dynamics is determined only by the wave interaction. If we can neglect the pumping and damping effects in the inertial interval (which has to be proved), the n_k and $N_{k\lambda}$ distributions would be independent of the particular form of γ_k and Γ_k.

It should be recalled that determination of the turbulence spectrum (describing the distribution of fluctuations over scales) in the theory of hydrodynamic turbulence is based on the two Kolmogorov hypotheses [1]. According to the first hypothesis, concerning the self-similar character of the turbulence spectrum, the spectrum in the inertial interval is determined by a single quantity *P* representing a constant energy flux along scales. The second hypothesis stipulates that the interaction of fluctuations with different *k* values has a local character.

Applying these hypotheses to our situation, the turbulence spectra can be determined proceeding from the dimensionality considerations. In this case, the kinetic equations (32) and (33) have two conservation laws—for the energy and the number of HF waves. Each of

these laws must correspond to a Kolmogorov spectrum of its own. Indeed, a constant flux of the number of HF waves (N_k^λ)

$$P_N = \frac{\partial}{\partial t}\sum_\lambda \int N_{k\lambda} dk,$$

corresponds to the spectrum

$$N_{k\lambda} \propto P_N^{1/2} k^{-4}, \quad n_k \propto P_N^{1/2} k^{-4}, \tag{34}$$

while the constant energy flux

$$P_\varepsilon = \frac{\partial}{\partial t}\int\left(\omega_k n_k + \sum_\lambda \omega_{k\lambda} N_{k\lambda}\right) dk$$

corresponds to

$$N_{k\lambda} \propto P_\varepsilon^{1/2} k^{-3/2}, \quad n_k \propto P_\varepsilon^{1/2} k^{-3/2}. \tag{35}$$

Based on the conservation of the total number of HF waves and the energy in the inertial interval, one may readily infer that the flux of the particle number N is directed toward small k, whereas the energy flux P_ε is directed toward large k. The rough approximation of the turbulence spectra given by formulas (34) and (35) may only provide for a correct dependences on the wavenumber and the fluxes but ignore the diffusion character of the wave decay process. It should be also recalled that the above conclusions are based on the hypothesis of a local character of the wave interaction.

The spectra (34) and (35) fail to describe fine details of the distribution functions (such as the angular dependence). Therefore, these functions are determined to within an arbitrary function of the angles. The angular dependence can be described by solving the exact equations (32) and (33). The solutions can be obtained, in particular, for the interaction of Alfvén waves with acoustic waves ($N_2 \equiv 0$). For this purpose, Eqs. (32) and (33) can be conveniently rewritten as

$$\dot{n}_k = -\int U_{k_2|kk_1} T_{k_2|kk_1} dk_1 dk_2, \tag{36}$$

$$\dot{N}_k = \int (U_{k|k_1k_2} T_{k|k_1k_2} - U_{k_1|kk_2} T_{k_1|kk_2}) dk_1 dk_2, \tag{37}$$

where

$$U_{k|k_1k_2} = 2\pi \left|V^{11}_{k'k'_1k_2}\right|^2 \delta_{k-k_1-k_2}\delta_{\omega-\omega_1-\omega_2},$$

$$T_{k|k_1k_2} = N_{k_1}n_{k_2} - N_k n_{k_2} - N_k N_{k_1}.$$

As can be readily checked, Eqs. (36) and (37) possess thermodynamically equilibrium solutions

$$N_k = \frac{N}{\omega_k + \mu}, \quad n_k = \frac{T}{\Omega_k},$$

representing the Rayleigh–Jeans distributions for which the collisional term is zero.

In order to determine the nonequilibrium distributions, note that the function U (having the sense of the decay probability) possesses the following properties: U is a bihomogeneous function of variables k_z and $k_\perp$, the degree of homogeneity being +1 for all k_z and –2 for $k_\perp$. In addition, U is invariant with respect to rotation relative to the z axis coinciding with the direction of the magnetic field $\mathbf{H}_0$.

For these reasons, it is natural to seek for solutions in the form of

$$n_k = Ak_z^\alpha k_\perp^\beta, \quad N_k = Bk_z^\alpha k_\perp^\beta. \tag{38}$$

Consider stationary solutions to Eq. (37):

$$\int (U_{k|k_1k_2} T_{k|k_1k_2} - U_{k_1|kk_2} T_{k_1|kk_2}) = 0. \tag{39}$$

Let us map the integration domain (determined by the conservation laws) of the second integral (39) onto that of the first integral. This is conveniently performed by introducing the complex quantity $\zeta = k_x + ik_y$. Using this value, the conservation laws determining the integration domain of the second integral can be written as:

$$k_{z1} - k_z - k_{z2} = 0,$$

$$\zeta_1 - \zeta - \zeta_2 = 0,$$

$$\omega_1 - \omega - \Omega_2 = 0.$$

This domain is mapped onto that of the first integral (39) by the following conformal mapping with respect to variables k_z and ζ

$$k_z = k'_z \frac{k_z}{k'_z}, \quad \zeta = \zeta' \frac{\zeta}{\zeta'},$$

$$k_{z1} = k_z \frac{k_z}{k'_z}, \quad \zeta_1 = \zeta \frac{\zeta}{\zeta'}, \tag{40}$$

$$k_{z2} = k''_z \frac{k_z}{k'_z}, \quad \zeta_2 = \zeta'' \frac{\zeta}{\zeta'}\frac{\zeta}{\zeta'}.$$

Here, each separate transformation is an inversion: relative to the point k_z for the z-components of the wave vectors and relative to the circumference of radius $|k_\perp|$ for the transverse components. As a result, the vector $\mathbf{k}$ transforms into $\mathbf{k}_1$, $\mathbf{k}_1$ into $\mathbf{k}$, and $\mathbf{k}_2$ into $\mathbf{k}_2$. Simultaneously, the z-components are stretched by the factor $|k_z/k_{z1}|$ and the transverse components, by the factor $|k_\perp/k_{1\perp}|$; this is complemented by rotation about the z axis through the angle $\arg(\zeta/\zeta_1)$.

As a result (with allowance for the properties of U and T values), the integrand in the second integral (39) converts into the integrand of the first integrand multiplied by a power factor:

$$\int U_{k|k_1k_2} T_{k|k_1k_2}\left[1 - \left(\frac{k_z}{k_{z1}}\right)^{2\alpha+4}\left(\frac{k_\perp}{k_{\perp 1}}\right)^{2\beta+4}\right] dk_1 dk_2 = 0.$$

From this we may infer that, besides thermodynamically equilibrium solutions (for which T becomes zero), there exist nonequilibrium solutions

$$n_k = Ak_z^{-2}k_\perp^{-2}, \quad N_k = Bk_z^{-2}k_\perp^{-2}, \tag{41}$$

which correspond to the solutions obtained previously from the dimensionality considerations for the constant particle number flux P_N. A relationship between the coefficients A and B in (41) is determined from the stationary equation (32) ($\partial/\partial t = 0$). From this we readily obtain an estimate $c_s A \sim V_A B$, which implies that the energies of the acoustic and Alfvén waves in the stationary case are on the same order of magnitude.

Note that the set of all mappings of the type (40) forms a group G, which is a direct product

$$G = G(1) \times G(2)$$

of groups $G(1)$ and $G(2)$ acting in one- and two-dimensional spaces, respectively. For power-type spectra, mappings of this type lead to factorization of the collisional term. The one-dimensional transformations (in the frequency space) for isotropic spectra were found by Zakharov [5, 6, 25]. Generalizations of these transformations to the two- and tree-dimensional cases were formulated by Kats and Kontorovich [33]. Mappings (40) represent a partial case of the quasiconformal mapping.

In order to determine the second nonequilibrium solution (35), we introduce a quantity

$$\varepsilon_k = \omega_k N_k + \Omega_k n_k$$

representing the energy density in the k-space. According to (36) and (37), ε_k obeys the equation

$$\frac{\partial \varepsilon_k}{\partial t} = \int \{\omega_k U_{k|k_1k_2} T_{k|k_1k_2} - \omega_k U_{k_1|kk_2} T_{k_1|kk_2} - \Omega_k U_{k_2|k_1k} T_{k_2|k_1k}\} dk_1 dk_2. \tag{42}$$

Let us find the stationary solutions to this equation in the same form (38) as above. To this end, we also map the integration domain of the second and third integrals (42) onto that of the first integral. For the second integral, the mapping coincides with (40); for the third integral, the required mapping is

$$k_z = k_z'' \frac{k_z}{k_z''}, \quad \zeta = \zeta'' \frac{\zeta}{\zeta''},$$

$$k_{z_1} = k_z' \frac{k_z}{k_z''}, \quad \zeta_1 = \zeta' \frac{\zeta}{\zeta''}, \tag{43}$$

$$k_{z_2} = k_z \frac{k_z}{k_z''}, \quad \zeta_2 = \zeta \frac{\zeta}{\zeta''}.$$

Substituting (40) and (43) into the stationary equation (42), we obtain

$$0 = \int |V_{kk_1k_2}|^2 \delta_{k-k_1-k_2} \delta_{\omega-\omega_1-\Omega_2} T_{k|k_1k_2} dk_1 dk_2 \times \left\{ \omega(k) - \omega(k_1) \left(\frac{k_z}{k_{z_1}}\right)^{2\alpha+5} \left(\frac{k_\perp}{k_{\perp_1}}\right)^{2\beta+4} - \Omega(k_2) \left(\frac{k_z}{k_{z_2}}\right)^{2\alpha+5} \left(\frac{k_\perp}{k_{\perp_2}}\right)^{2\beta+4} \right\}.$$

This equation indicates that the expression in braces turns zero for $\alpha = -5/2$ and $\beta = -2$, so that the required solutions have the following form:

$$n_k = Ak_z^{-5/2}k_\perp^{-2}, \quad N_k = Bk_z^{-5/2}k_\perp^{-2}. \tag{44}$$

These solutions correspond to the spectra with a constant energy flux P_ε. As above, a relationship between the coefficients A and B is determined from the stationary equations (32). These equations lead to the same estimate:

$$c_s A \sim V_A B.$$

The Kolmogorov type solutions obtained above refer only to the channel of interaction between the Alfvén waves and slow magnetoacoustic waves, which markedly reduces the significance of these results. It should be recalled that processes involving the fast magnetoacoustic waves are on the same order of magnitude and, hence, cannot be ignored. However, this channel can be successfully included into the above scheme without need in significant improvements. As was noted in the preceding section, a maximum scattering of the A-waves (i.e., of the Alfvén waves and fast magnetoacoustic waves) takes place for the maximum z-projection of the momentum transferred to the slow acoustic waves. It is naturally assumed that this behavior of the A-wave scattering amplitude would lead to strongly anisotropic distributions of the waves concentrated in the k-space within a narrow-angle cone in the magnetic field direction: $k_z \gg k_\perp$. Under these conditions, the fast sound frequency coincides with that of the Alfvén waves:

$$\omega_2 \approx |k_z| V_A.$$

Another important circumstance in this case is that the matrix element of the interaction between A-waves and the slow magnetoacoustic waves is diagonal with respect to the polarization λ:

$$V^{\lambda\lambda_1}_{k'k_1'k_2} \approx \delta_{\lambda\lambda_1} V^{11}_{k'k_1'k_2}.$$

Thus, in the case of an almost longitudinal (i.e., extended in the magnetic field direction) distribution, we observe no difference between the Alfvén waves and fast magnetoacoustic waves. Moreover, there is even no energy exchange between these waves. This

implies that (in the given angular range) the Kolmogorov spectra for the fast magnetoacoustic waves will exhibit the same shape as the spectra (41) and (44) obtained above. In these expressions, we have to replace N_k and B by N_k^λ and B_λ and determine the coefficient A as

$$A \sim \beta^{-1/2} \frac{\sum B_\lambda^2}{\sum B_\lambda}.$$

The Kolmogorov spectra (41) and (44) possess a physical meaning provided that the condition of local turbulence is fulfilled, according to which the contribution to the wave interaction due to the regions of pumping and damping must be small. This condition reduces to the requirement of convergence of the integrals in Eqs. (36) and (37).

The convergence of integrals with respect to k_z in the kinetic equations is ensured by the presence of two δ-functions of k_z. The integrals over transverse wavevectors exhibit logarithmic divergence. In our opinion, the logarithmic divergence is not as dangerous as the power-type behavior, since it falls on the boundary of locality. The appearance of this divergence is related to the aforementioned bihomogeneity of the probability U. If the medium is isotropic and the matrix elements of V and the frequencies possess the same degree of homogeneity as that in the MHD case of $\beta \ll 1$ (this situation takes place for the Mandelstamm–Brillouin scattering), the condition of locality would be satisfied (see [34]). The condition of bihomogeneity for the interaction of Alfvén waves and slow magnetoacoustic waves is violated in the case of a nearly transverse wave propagation:

$$k_\perp/k_z \sim \beta^{-1/2},$$

and for the interaction of fast and slow magnetoacoustic waves, in the region of angles

$$k_\perp/k_z \sim \beta^{1/2}.$$

For this reason, the integrals in the kinetic equations should be truncated at smaller angles ($\lesssim \beta^{1/2}$). The logarithmic divergence can be eliminated by introducing powers for logarithms of the transverse momenta $k_\perp$ in the spectra (41) and (44). However, this procedure does not lead to determining these powers (while ensuring the convergence of integrals). Finally, it should be noted that both spectra (41) and (44) are characterized by the same dependence on the transverse momenta:

$$n_k, N_k \propto k_\perp^{-2},$$

which coincides with the degree of homogeneity of $\delta(\mathbf{k}_\perp)$. This implies that, besides the anisotropic Kolmogorov spectra (41) and (44), there may exist singular Kolmogorov spectra of the type

$$n_k = Ak_z^{-2}\delta(\mathbf{k}_\perp), \quad N_k = Bk_z^{-2}\delta(\mathbf{k}_\perp)$$

and

$$n_k = Ak_z^{-5/2}\delta(\mathbf{k}_\perp), \quad N_k = Bk_z^{-5/2}\delta(\mathbf{k}_\perp).$$

A rigorous answer to the question as to which stationary spectra are in fact realized can be obtained from the investigation of stability of the obtained solutions or (on a qualitative level) from the results of numerical modeling. Both approaches require special consideration.

ACKNOWLEDGMENTS

The author is grateful to V.E. Zakharov for fruitful discussions. Many thanks to the Nice Observatory, where this paper was completed within the Landau–CNRS agreement.

This study was supported by the Russian Foundation for Basic Research (project no. 00-01-00929), the Federal Program of Support for Leading Scientific Schools (project no. 00-15-96007), and the INTAS foundation (project no. 00-100292).

REFERENCES

1. A. N. Kolmogorov, Dokl. Akad. Nauk SSSR **30**, 299 (1941).
2. A. M. Obukhov, Izv. Akad. Nauk SSSR, Ser. Geogr. Geofiz. **5**, 453 (1941).
3. V. E. Zakharov, V. S. L'vov, and G. E. Falkovich, *Kolmogorov Spectra of Turbulence I* (Springer-Verlag, Berlin, 1992).
4. V. E. Zakharov, Zh. Éksp. Teor. Fiz. **51**, 688 (1966) [Sov. Phys. JETP **24**, 455 (1967)].
5. V. E. Zakharov, Prikl. Mekh. Tekh. Fiz. **4**, 34 (1965).
6. V. E. Zakharov and N. N. Filonenko, Prikl. Mekh. Tekh. Fiz. **5**, 62 (1967).
7. E. A. Kuznetsov, Zh. Éksp. Teor. Fiz. **62**, 584 (1972) [Sov. Phys. JETP **35**, 310 (1972)].
8. S. V. Nazarenko, A. B. Mikhailovski, *et al.*, Phys. Lett. A **133**, 407 (1988).
9. A. M. Balk, V. E. Zakharov, and S. V. Nazarenko, Zh. Éksp. Teor. Fiz. **98**, 446 (1990) [Sov. Phys. JETP **71**, 249 (1990)].
10. A. A. Galeev and V. N. Oraevskii, Dokl. Akad. Nauk SSSR **147**, 71 (1962) [Sov. Phys. Dokl. **7**, 988 (1963)].
11. E. A. Kuznetsov, Preprint No. 81-73, IYaF SO AN SSSR (Inst. of Nuclear Physics, Siberian Division, Academy of Sciences of USSR, Novosibirsk, 1973).
12. S. Galtier, S. V. Nazarenko, A. C. Newell, and A. Pouquet, J. Plasma Phys. **63**, 447 (2000); in *Nonlinear MHD Waves and Turbulence*, Ed. by T. Passot and P.-L. Sulem (Springer-Verlag, New York, 1999).
13. V. E. Zakharov and E. A. Kuznetsov, Dokl. Akad. Nauk SSSR **194**, 1288 (1970) [Sov. Phys. Dokl. **15**, 913 (1971)].
14. V. E. Zakharov and E. A. Kuznetsov, Usp. Fiz. Nauk **167**, 1137 (1997) [Phys. Usp. **40**, 1087 (1997)].
15. V. E. Zakharov, Zh. Éksp. Teor. Fiz. **62**, 1745 (1972) [Sov. Phys. JETP **35**, 908 (1972)].

16. L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Gostekhizdat, Moscow, 1959).
17. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Nauka, Moscow, 1967; Pergamon, Oxford, 1975).
18. H. Lamb, *Hydrodynamics* (Cambridge Univ. Press, Cambridge, 1932; Gostekhizdat, Moscow, 1947).
19. B. I. Davydov, Dokl. Akad. Nauk SSSR **69**, 165 (1949).
20. A. Frenkel, E. Levich, and L. Shtilman, Phys. Lett. A **88**, 461 (1982).
21. H. K. Moffatt, J. Fluid Mech. **35**, 117 (1969).
22. V. A. Vladimirov and H. K. Moffatt, J. Fluid Mech. **283**, 125 (1995).
23. V. P. Ruban, Zh. Éksp. Teor. Fiz. **116**, 563 (1999) [JETP **89**, 299 (1999)].
24. E. A. Kuznetsov and V. P. Ruban, Phys. Rev. E **61**, 831 (2000).
25. V. E. Zakharov, Zh. Éksp. Teor. Fiz. **60**, 1714 (1971) [Sov. Phys. JETP **33**, 927 (1971)].
26. A. A. Vedenov, E. P. Velikhov, and R. Z. Sagdeev, Usp. Fiz. Nauk **73**, 132 (1961).
27. E. A. Kuznetsov, Fiz. Plazmy **2**, 327 (1976) [Sov. J. Plasma Phys. **2**, 178 (1976)].
28. V. E. Zakharov and A. M. Rubenchik, Prikl. Mekh. Tekh. Fiz. **5**, 84 (1972).
29. E. A. Kuznetsov, S. L. Musher, and A. V. Shafarenko, Pis'ma Zh. Éksp. Teor. Fiz. **37**, 204 (1983) [JETP Lett. **37**, 241 (1983)].
30. E. A. Kuznetsov and S. L. Musher, Zh. Éksp. Teor. Fiz. **91**, 1605 (1986) [Sov. Phys. JETP **64**, 947 (1986)].
31. A. Pushkarev and V. Zakharov, Phys. Rev. Lett. **76**, 3320 (1996).
32. V. E. Zakharov, A. I. D'yachenko, and O. A. Vasil'ev, Pis'ma Zh. Éksp. Teor. Fiz. **73**, 68 (2001) [JETP Lett. **73**, 63 (2001)].
33. A. V. Kats and V. M. Kontorovich, Pis'ma Zh. Éksp. Teor. Fiz. **14**, 392 (1971) [JETP Lett. **14**, 265 (1971)].
34. V. E. Zakharov and E. A. Kuznetsov, Zh. Éksp. Teor. Fiz. **75**, 904 (1978) [Sov. Phys. JETP **48**, 458 (1978)].

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 93, No. 5, 2001, pp. 1065–1073.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 5, 2001, pp. 1227–1236.

PLASMA, GASES

Mechanism of Ion Flow Generation in Vacuum Arcs

S. A. Barengolts[a,*], G. A. Mesyats[b], and D. L. Shmelev[b,]**

[a]*Center for Natural Science Research, Institute of General Physics, Russian Academy of Sciences, Moscow, 119991 Russia*
[b]*Institute of Electrophysics, Ural Division, Russian Academy of Sciences, ul. Komsomol'skaya 34, Yekaterinburg, 620016 Russia*
**e-mail: sb@nsc.gpi.ru*
***e-mail: shmelev@iep.uran.ru*
Received April 14, 2001

Abstract—The ecton model of the cathode spot is used to analyze the main parameters of ion flow in vacuum arcs (ion erosion, mean charge, and velocity). It is shown that the arc plasma is formed as a result of microexplosions at the cathode surface, induced by the Joule heating by the high-density current of explosive electron emission. Ionization processes are localized in a narrow region of the order of a micrometer near the cathode and the ionization composition of the plasma subsequently remains unchanged. Under the action of the electron pressure gradient, ions acquire directional velocities of the order of 10^6 cm/s even over small distances of the order of several micrometers. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

High-velocity particle flows emitted from the region of the cathode spot of a vacuum arc were discovered by Tanberg [1]. It was assumed that such a flow is a metal vapor jet in which the velocity of particles is equal to 1.6×10^6 cm/s for a copper cathode. Plyutto *et al.* [2] established that this flow is formed not by vapor but by a plasma flow in which the ionization rate is 50–100% for metals with a relatively high boiling point (Cu, Ag, Mg); it also contained doubly and triply charged ions, and the velocity of plasma jets was of the order of 10^6 cm/s. For readily evaporating metals (Zn, Cd), the velocity was found to be slightly lower and the degree of ionization of metal vapor was 10–25%. The energies of ions were distributed between 0 and 50–70 eV. The presence of high-energy ions in plasma jets led Plyutto *et al.* [2] to the conclusion that a potential hump exists in the region of cathode-spot plasma.

Mesyats *et al.* [3] proved that similar plasma jets are also present at the spark stage. For example, in the case of a copper cathode, the velocity of such jets was 2×10^6 cm/s and was determined by microexplosions on the surface of the cathode, which were accompanied by explosive electron emission. It was shown [4] that such a value of the plasma jet velocity can be obtained proceeding from the assumption of explosive gasdynamic expansion of plasma [5] for specific energies corresponding to the electric explosion of the metal. It was established later [6] that the functioning of cathode spots of a vacuum spark and arc are controlled by the same physical processes which govern microexplosions at the cathode accompanied by the formation of ectons (individual portions of electrons) in the course of explosive emission of electrons [7].

A considerable advance has been made in recent years in the investigation of ion parameters of the arc plasma. This is associated with the creation of vacuum-arc ion sources which were used for obtaining distribution over the charge states of virtually all conducting materials [8]. According to the results obtained in these investigations, the charge composition of the vacuum-arc plasma remains unchanged upon an increase in the arc current from 50 to 500 A. An important result of these investigations was the experimentally established fact of equality of the velocities of ions having different charges [9]. These results unambiguously support the gasdynamic mechanism of acceleration of ions.

In the present work, the parameters of the ion flow (charge composition, mean velocity, and ion erosion) are considered from the positions of the ecton model of the cathode spot of a vacuum arc. According to this model, the cathode spot of the vacuum arc consists of individual cells emitting ectons [7]. The main experimental regularities established for ion flows emitted from the cathode spot of the vacuum arc form the basis of the ecton model. For example, an increase in current increases the number of cells, and the main ion parameters are formed as a result of functioning of a unit cell of the spot during explosive erosion of a cathode region by Joule heating. The substance of the cathode is gradually transformed into a series of states (condensed state and the states of nonideal and ideal plasma). During this transformation, the charge composition of the plasma is formed and ions are accelerated by the pressure gradient of the plasma, whose concentration decreases by several orders of magnitude over distances of ten micrometers from the surface of the cathode.

1063-7761/01/9305-1065$21.00

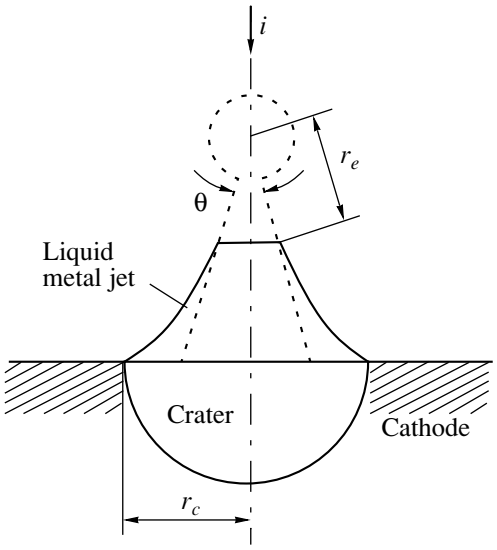


Fig. 1. Model of a liquid-metal jet in which an ecton exists (r_c is the radius of the melt at the cathode and r_e is the radius of the ecton zone).

It should be noted that the physical processes leading to the formation of the charge composition and acceleration of ions of the vacuum arc are considered more than once (see, for example, [10]). Among recent publications in this field, we can mention the theoretical investigations by Anders [11], who obtained the charge distribution for ions of an arc plasma. However, a common drawback of these publications is that the mechanism of ion generation was not determined in them. For this reason, the ion flow parameters were considered without taking into account their interrelation, and the main parameters used in calculations were chosen quite arbitrarily without necessary theoretical substantiation.

2. ECTON PROCESSES AT CATHODE OF A THE VACUUM ARC

2.1. Parameters of Ecton Processes

The study of explosive electron emission [3, 4, 6, 12] made it possible to substantiate the nature of physical processes in the cathode spot of a vacuum arc. It was found that the plasma of a cathode flame in the case of explosive electron emission (the spark stage) is characterized by the same parameters (expansion rate and charge composition) as an arc plasma. The traces of cathode damage in both cases turned out to be identical; the expansion rate of the liquid-metal fraction, erosion rate, the cathode voltage drop, and the threshold currents also coincided with the results of measurements in the arc. This led to the conclusion that in the case of vacuum arc, we are also dealing with explosive electron emission.

The basic principles of the ecton model of the cathode spot of a vacuum arc, which is based of the explosion-emission processes, were formulated in [7]. An ecton is initiated during the interaction between a liquid-metal jet (see Fig. 1) formed as a result of the expulsion of the liquid-metal fraction from the cathode spot region and the cathode plasma.

Explosive electron emission as well as electric explosion of conductors is associated with the Joule heating under the action of the passing current [6]. Using the analogy with electric explosion of a conductor, we can estimate the current density required for initiating an ecton from the formula for the specific action $\bar{h}$:

$$J = (\bar{h}/t_z)^{1/2}, \tag{1}$$

where t_z is the time delay of the explosion. Since $\bar{h} \sim 10^9$ A^2 s cm^{-4} for most metals, the current density amounts to approximately 10^9 A/cm^2 for $t_z \approx 10^{-9}$ s. The high current density leads to rapid heating of a microscopic region of the cathode and its explosion accompanied by explosive electron emission. During the explosion period, the emission zone increases, the current density decreases, and the heat withdrawal due to thermal conduction and the energy removal due to the expulsion of the plasma and heated liquid metal become significant. For this reason, the current of explosive emission terminates, giving rise to a short-lived portion of electrons (ecton).

The ecton lifetime can be estimated as follows [7]:

$$t_e = \frac{I^2}{\pi^2 a^2 \bar{h} \theta^4}, \tag{2}$$

where a is the thermal diffusivity of the cathode material and I is the ecton current. While writing formula (2), we assumed that the ecton appears as a result of explosion of the liquid-metal tip of conic geometry with a small cone angle θ (Fig. 1). The mass carried away from the cathode during the time t_e is given by

$$M = \frac{2}{3\pi^2} I^3 \frac{\rho}{(a\bar{h})^{3/2}\theta^4}, \tag{3}$$

where ρ is the density of the cathode material. The total electron charge flowing during the ecton lifetime is given by

$$q_e = \frac{I^3}{\pi^2 a^2 \bar{h} \theta^4}. \tag{4}$$

Since the arc discharge is self-sustained due to the explosion of liquid-metal inhomogeneities, we must use in formulas (1)–(4) the values of a, ρ, and $\bar{h}$ for the liquid state.

The finite lifetime of an ecton is responsible for the cyclic nature of the processes in the cathode spot. Such a cycle has two stages, viz., the first stage lasting the time t_e during which the ecton actually exists and the

second stage that last for time t_i during which a new ecton is initiated by the ion current from the cathode plasma. The time of the cycle estimated from the experimental data on voltage oscillations in the region of threshold currents of arc glow and noise voltage has the value $t_c \approx 30$ ns for copper and tungsten cathodes, while the relative fraction of the ion stage of the cycle is

$$\alpha = \frac{t_i}{t_e + t_i} \approx 0.2$$

(see [7, 13]).

Another important property of the cathode spot is the existence of the internal structure, which is manifested in presence of individual cells or the fragments of the cathode spot. According to the results obtained by Kesaev, each cell of the cathode spot carries a current equal to the doubled threshold current I_p of the arc glow [14]. In the framework of the ecton model, a cell of the spot is an explosion-emission center emitting an ecton.

One of the main arguments in favor of the ecton model is the presence of microcraters having a radius r_c of the order of 10^{-4} cm in the trace of the cathode spot [4, 14, 15]. This is not only a qualitative, but also a quantitative evidence of ecton processes in the cathode spot. Indeed, if a microcrater is formed due to thermal conduction, the velocity of propagation of the thermal boundary amounts to 10^4 cm/s and, hence, the time of the formation of such a crater is $\tau_c \sim 10^{-8}$ s. If the current flowing through a cell is $I = 2I_p$ and $I_p \sim 1$ A [14], the density of the current through the cell is $J_c \sim 10^8$ A/cm^2. Such values of the parameters r_c, τ_c, and J_c are due to the ecton processes occurring in the cathode spot of a vacuum arc.

2.2. Ion Erosion

An analysis of the energy distribution for ions of an arc plasma proved that they escape from the region of the cathode spot with mean velocities of the order of 10^6 cm/s. These ions create a current opposite to the arc current since they move from the cathode to the anode. Important measurements in this field were made by Kimblin [16], who established that the ion current collected at a cylindrical screen attains a limiting value which is independent of the geometrical parameters, is approximately proportional to the total arc current, and depends weakly on the cathode material. The proportionality factor between the ion current and the total current amounts to approximately 0.1.

The ion current from the cathode leads to a loss of the cathode mass in the form of ions. The existence of the vacuum discharge is determined precisely by the generation of a conducting medium in the electrode gap. The measure of ion erosion is the erosion rate γ_i, equal to the ratio of the mass carried away in the form of ions to the passing charge $q = I\Delta t$. The experimental results on the cathode erosion are rather contradictory since, along with ion erosion, the cathode material is carried away during the discharge in the form of macroparticles, drops, and neutral vapor. The conventional methods of investigation of erosion (such as weighing and the estimating the erosion structures from the change in the geometrical parameters) were determined to a considerable extent by the arc current, the duration of the arc glow, and the cathode geometry. Daalder [17] carried out a series of experiments with copper electrodes of diameters of 25 and 10 mm. The current varied from 33 to 200 A. It turned out that upon a decrease of the value of q to 0.1 C, different dependences of erosion rate obtained for different currents give the same value ~40 μg/C, which is precisely the ion erosion rate. The ion erosion can also be determined from the measurements of the mean ion charge with the help of the obvious formula [17]

$$\gamma_i = \frac{I_i}{I}\frac{m}{Ze}, \tag{5}$$

where I_i is the charge carried by ions from the cathode per unit time, Ze is the mean ion charge of the arc plasma, and m is the mass of an ion.

Let us analyze the characteristics of the ion flow emitted by the cathode sport from the positions of the ecton model. The mass carried away during the time t_e is defined by formula (3), while during the time t_i, the ion current mainly flows to the cathode; according to the results obtained in [13], this current amounts to 0.1 of the arc current. Consequently, the total mass loss from the cathode during a cycle is $M(1 - 2\alpha)$. Taking into account relation (4), we can write the following expression for ion erosion:

$$\gamma_i = \frac{2}{3}\rho\left(\frac{a}{\bar{h}}\right)^{1/2}(1 - 2\alpha). \tag{6}$$

Using formulas (5) and (6), we can determine the mean ion charge of the plasma formed as a result of functioning of the ecton:

$$Z = \frac{3I_i}{2I}\frac{m}{e\rho(1-2\alpha)}\left(\frac{\bar{h}}{a}\right)^{1/2}. \tag{7}$$

Considering that the ratio $I_i/I \approx 0.1$ for all materials [16], we find that the values of γ_i and Z are independent of current and are determined only by the parameters of the cathode material, which is in accord with the experimental results [7, 8, 17]. The mean values of charges and ion erosion rate obtained in accordance with formulas (6) and (7) for a number of metals for which the values of specific action are known [7] are presented in the table. The values of thermophysical coefficients for metals are borrowed from [18]. In analogy with W and Cu, the value of α was put equal to 0.2 for all metals.

The observed agreement with the experimental results is even more astonishing if we take into account the approximate nature of thermophysical parameters of the materials under investigation.

Ion erosion and mean charge of arc plasma for various materials of the cathode

Cathode material	ρ, g/cm^3	a, cm^2/s	$\bar{h} \times 10^{-9}$, A^2 s cm^{-4}	γ_i, μg/C	Z	Z [8, 19]
Cu	8.0	0.42	3.1	37.2	1.76	1.7–2.0
Au	17.2	0.40	1.3	120.6	1.69	1.6–2.0
Al	2.3	0.40	1.4	15.5	1.80	1.5–1.7
Ag	9.3	0.56	2.0	62.2	1.77	1.8–2.1
W	17.0	0.14	1.5	65.7	2.90	3.0–3.1

3. PROCESSES IN THE CATHODE PLASMA OF A VACUUM ARC

3.1. General System of Equations

The simulation of the initial stage of explosion of microtips in the case of explosive electron emission [20] proved that the density of the cathode material changes during the functioning of an ecton over short distances (of the order of a few micrometers) from the solid-state value to that corresponding to the state of ideal plasma. The maximum energy contribution takes place upon a transition to the plasma state for a plasma concentration $n \leq 10^{21}$. It is important to note that, according to calculations, the value of the specific action integral

$$\int_0^t I^2 dt$$

remains unchanged by the time of explosion upon a change in current, justifying the analogy with an electric explosion of a conductor used in the previous section in the analysis of ecton processes. The expansion of the plasma formed as a result of the explosion of a microtip at the cathode is of the gasdynamic type and may be described by a system of equations of 2D two-temperature magnetic gas dynamics of completely ionized plasma taking into account the thermal conductivity of electrons in the cylindrical system of coordinates:

$$\frac{\partial n_i}{\partial t} + \frac{\partial n_i u_z}{\partial z} + \frac{1}{r}\frac{\partial r n_i u_r}{\partial r} = 0, \tag{8}$$

$$\frac{\partial c_j}{\partial t} + u_z\frac{\partial c_j}{\partial z} + u_r\frac{\partial c_j}{\partial r} = n_e(\alpha_{j-1,j}c_{j-1} - \beta_{j,j-1}c_j n_e - \alpha_{j,j+1}c_j + \beta_{j+1,j}c_{j+1}n_e),$$

$$\alpha_{j-1,j} = C\left(\frac{I_H}{I_j}\right)^3 T_e\sqrt{\frac{8T_e}{\pi m_e}}\left(\frac{I_j}{T_e}+2\right)\exp\left(-\frac{I_j}{T_e}\right),$$

$$\beta_{j,j-1} = C\left(\frac{I_H}{I_j}\right)^3\left(\frac{I_j}{T_e}+2\right)\frac{h^3}{2\pi e m_e^2}, \tag{9}$$

$$j = 1, \ldots, 4,$$

$$n_e = n_i\sum_{j=1}^{4} jc_j = Zn_i, \tag{10}$$

$$\frac{\partial n_i u_z}{\partial t} + \frac{\partial u_z n_i u_z}{\partial z} + \frac{1}{r}\frac{\partial r u_r n_i u_z}{\partial r} + \frac{1}{m}\frac{J_r B}{c} + \frac{1}{m}\frac{\partial(n_i T_i + n_e T_e)}{\partial z} = 0,$$

$$\frac{\partial n_i u_r}{\partial t} + \frac{\partial u_z n_i u_r}{\partial z} + \frac{1}{r}\frac{\partial r u_r n_i u_r}{\partial r} - \frac{1}{m}\frac{J_z B}{c} + \frac{1}{m}\frac{\partial(n_i T_i + n_e T_e)}{\partial r} = 0, \tag{11}$$

$$\frac{3}{2}\left(\frac{\partial n_i T_i}{\partial t} + \frac{\partial u_z n_i T_i}{\partial z} + \frac{1}{r}\frac{\partial r u_r n_i T_i}{\partial r}\right) + n_i T_i\left(\frac{\partial u_z}{\partial z} + \frac{1}{r}\frac{\partial r u_r}{\partial r}\right) = 3\frac{m_e}{m}\frac{n_e}{\tau_{ei}}(T_e - T_0), \tag{12}$$

$$\mathbf{v} = \frac{\mathbf{J}}{eZn_i} + \mathbf{u}, \quad \mathbf{J} = \frac{c}{4\pi}\mathrm{curl}\,\mathbf{B}, \tag{13}$$

$$\frac{3}{2}\left(\frac{\partial n_e T_e}{\partial t} + \frac{\partial v_z n_e T_e}{\partial z} + \frac{1}{r}\frac{\partial r v_r n_e T_e}{\partial r}\right) + n_e T_e\left(\frac{\partial v_z}{\partial z} + \frac{1}{r}\frac{\partial r v_r}{\partial r}\right) + \left(\frac{\partial Q_z}{\partial z} + \frac{1}{r}\frac{\partial r Q_r}{\partial r}\right) = -3\frac{m_e}{m}\frac{n_e}{\tau_{ei}}(T_e - T_0) + \frac{m_e n_e}{\tau_{ei}}(\mathbf{v}-\mathbf{u})\cdot\left(\mathbf{v}-\mathbf{u}-\frac{3\mathbf{Q}}{5n_e T_e}\right) - E_i, \tag{14}$$

$$\mathbf{Q} = \beta_0 n_e T_e(\mathbf{v}-\mathbf{u}) - \gamma_0\frac{n_e T_e \tau_{ei}}{m_e}\mathrm{grad}\,T_e, \tag{15}$$

$$\frac{\partial B}{\partial t} + \frac{\partial v_z B}{\partial z} + \frac{\partial v_r B}{\partial r} = \frac{c^2}{4\pi}\left(\frac{\partial}{\partial r}\left(\frac{1}{r\sigma}\frac{\partial r B}{\partial r}\right) + \frac{\partial}{\partial z}\left(\frac{1}{\sigma}\frac{\partial B}{\partial z}\right)\right), \tag{16}$$

$$\sigma = \frac{n_e e^2 \tau_{ei}}{m_e} \frac{13 + 4\sqrt{2}/Z}{4 + 4\sqrt{2}/Z}, \tag{17}$$

$$\tau_{ei} = \frac{3}{4} \frac{\sqrt{m_e} T_e^{3/2}}{\sqrt{2\pi} e^4 Z^2 n_i \Lambda}, \tag{18}$$

$$\Lambda = \max\left\{1, \ln\left(\frac{3T_e}{Ze^2}\sqrt{\frac{1}{4\pi e^2 n}\frac{T_e T_i}{T_e + T_i}}\right)\right\},$$

$$\nabla\phi = \frac{1}{\sigma}\mathbf{J} + \frac{1}{ecn_e}\mathbf{J}\times\mathbf{B} + \frac{\nabla(n_e T_e)}{en_e} + \frac{\gamma}{e}\nabla T_e. \tag{19}$$

Here, u is the hydrodynamic velocity of ions, m_e and m are the masses of electrons and ions, respectively, T_e and T_i, n_e and n_i are their temperatures and concentrations, v is the hydrodynamic (or current) velocity of electrons, J is the current density, Q is the electron thermal flux, E_i is the ionization energy, λ is the interpolation function depending on concentration, c_j is the fraction of ions having the jth charge in n_i, τ_{ei} is the electron–ion relaxation time, B is the magnetic field, φ is the potential, β_0, γ_0, and γ are Braginskii's coefficients [21], $\alpha_{j-1,j}$ and $\beta_{j,j-1}$ are the ionization and recombination rates between transitions $j-1$ and j, and I_j is the jth ionization potential. The initial expressions for $\alpha_{j-1,j}$ and $\beta_{j,j-1}$ are taken from [5], where they are written for a plasma in the semiclassical approximation.

The continuity equation (8) is written for the heavy component, and the electron concentration was determined with the help of formula (10) taking into account the quasineutrality condition. The ionization kinetics equations (9) are written for ions with charges from +1 to +4. The momentum conservation equations (11) take into account the effect of the intrinsic magnetic field. The inertia of electrons was disregarded since the current velocity of electrons is much lower than their sound velocity. The hydrodynamic velocity of electrons in this case can be determined using formula (13).

The balance equation for the internal energy is split into two parts: the equation for the heavy plasma component and the equation for the electron energy. Ions and electrons exchange energy, which is described by the first term on the right-hand sides of Eqs. (12) and (14). We assumed that the Joule energy (the second term on the right-hand side of Eq. (14)) is supplied directly to the electron subsystem and then to the ion subsystem through the exchange term. We also assumed that the ionization energy is supplied to the electron component (the third term on the right-hand side of Eq. (14)). In the balance equation for the electron energy, the electron thermal conductivity (15) plays a significant role. Equation (16) is written for the ϕth component of the magnetic field B, which is the only nonzero component in the cylindrical system of coordinates. We also took into account diffusion (right-hand side) and magnetic field transport by electrons (the second and third terms on the left-hand side).

The system of equations (8)–(19) was solved numerically for a copper cathode for a given current of 3.2 A, which, according to [14], corresponds to the current of an individual cell of the cathode spot. In our calculations, we assumed that an ion flow having a velocity smaller than the sound velocity and an electric current of a preset value pass into the rated region through a cylindrical tube of radius $r_0 = 1$ μm and height $h = 0.5$ μm (analogue of a crater). At the lower end of the tube (i.e., for $z = 0$), the flow has the given mean charge $Z = 1$; the temperatures $T_e = T_i = 1.5$ eV and the plasma concentration is of the order of 10^{21} cm^{-3}.

In the numerical solution, we neglected the backward thermal electron current to the plane $z = h$, but the heat flux of these electrons was taken into account in the boundary condition for Eq. (14) in the form

$$\gamma_0 \frac{en_e T_e \tau_{ei}}{m_e} \frac{\partial T_e}{\partial z} = j_r^t (2T_e + \phi),$$
$$j_r^t = \frac{1}{4} en_e \sqrt{\frac{8T_e}{\pi m_e}} \exp\left(-\frac{\phi}{T_e}\right), \tag{20}$$

where ϕ is the plasma boundary potential determined by integrating Eq. (12) and j_r^t is the thermal current of electrons from the plasma to the cathode through the potential barrier ϕ (backward current).

The results of numerical simulation of the basic parameters of an arc plasma are presented in Fig. 2. It can be seen that our results are in good agreement with the experimental data obtained from the investigations of the ion parameters, presented in [7–9, 19]. The mean charge of the plasma coincides with its values presented in the table. The ion velocity slightly exceeds the value 1.28×10^6 cm/s presented in [9]; however, we can assume that the agreement with the experiment is satisfactory on account of the wide spread in the experimental data obtained by different methods of measurements and the conditions of the vacuum [7, 9]. It should be noted that the plasma potential and temperature are also in accord with the experimental results [10].

Thus, according to the results of numerical simulation, the formulation of the problem in the form (8)–(19) using the typical parameters of ecton processes (the current through a cell of the spot and the size of the plasma source) makes it possible to obtain detailed information on the physical processes occurring in the arc plasma near the cathode.

According to the results of numerical simulation, the geometry of the cathode plasma jet is close to spherical. Test calculations taking into account and disregarding the intrinsic magnetic field of the jet proved that it does not affect the jet formation significantly. The spherically symmetric expansion of the cathode plasma considerably simplifies the system of equations (8)–(19) and makes it possible to obtain the analytic estimates of a number of plasma parameters.

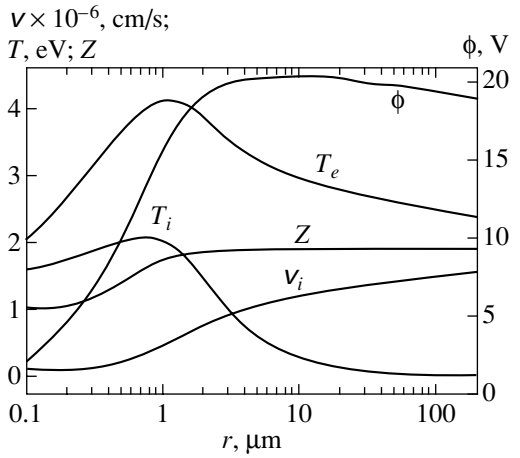


Fig. 2. Distribution of plasma parameters along the jet axis (2D calculations).

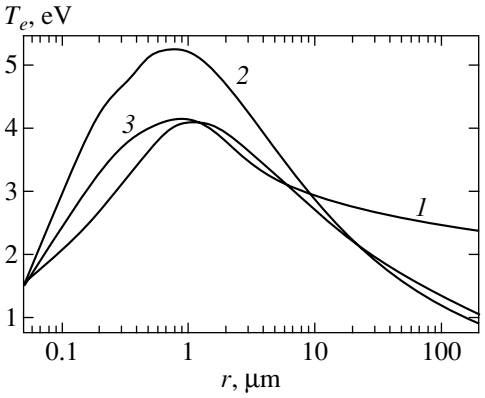


Fig. 3. Electron temperature distribution: 2D calculations (*1*), formula (23) (*2*), formula (26) (*3*)

3.2. Electron Temperature

The current density in the immediate vicinity of the cathode is quite high; consequently, the electron temperature increases due to Joule heating. The intensity of heating decreases in proportion to r^{-4}. For this reason, the electron temperature decreases upon an increase of the distance to the cathode due to the work done by pressure forces, forming a characteristic "hump" at distances of the order of a micrometer from the cathode surface. In the immediate vicinity of the base of the jet, the electron temperature deviates from the ion temperature (see Fig. 2). The difference increases rapidly upon expansion. This is due to an abrupt decrease in the exchange term (right-hand side of Eq. (12)) proportional to concentration and $T_e^{-3/2}$, which leads to the effect of "runaway" of the electron temperature.

At the base of the plasma jet, the plasma concentration gradient is high and, hence, the processes in this region can be described to a high degree of accuracy under the assumption of the spherically symmetric expansion of the plasma. Taking this into account, we can write Eq. (14) for a steady-state spherically symmetric plasma jet disregarding ionization losses and the energy exchange with ions:

$$v\frac{\partial T_e}{\partial r} + \frac{2}{3}\frac{T_e}{r^2}\frac{\partial r^2 v}{\partial r} - \frac{2}{3}\left(1 - \frac{3}{5}\beta_0\right)\frac{m_e}{\tau_{ei}}v^2 = \frac{2}{3n_e}\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\gamma_0\frac{n_e T_e \tau_{ei}}{m_e}\frac{\partial T_e}{\partial r}\right). \quad (21)$$

Disregarding the coordinate derivative of velocity and assuming that $n \propto r^{-2}$, $j \propto r^{-2}$, and $nur^2 = \text{const}$, we obtain

$$\frac{\partial T_e}{\partial r} + \frac{4}{3}\frac{T_e}{r} - \frac{H\Lambda Z}{T_e^{3/2}}\frac{I}{r^2} = \frac{G}{IZ}\frac{\partial}{\partial r}\left(r^2 T_e^{5/2}\frac{\partial T_e}{\partial r}\right), \quad (22)$$

where H is a constant approximately equal to 5, G is a constant approximately equal to 0.03 (if r is measured in micrometers and temperature in electronvolts), and I is the total current in amperes. For small r, we can also neglect the term on the right-hand side of Eq. (22) (thermal conductivity). We also assume that the Coulomb logarithm Λ is a constant equal to 5; in this case, the equation can easily be solved and its solution has the form

$$T_e(r) = \left(\frac{15H\Lambda ZI(r^{7/3} - r_0^{7/3}) + 14r_0^{9/3}T_0^{5/2}}{14r^{10/3}}\right)^{2/5}. \quad (23)$$

For large values of r, the $T_e(r)$ dependence behaves as $r^{-2/5}$:

$$T_e(r) \approx \left(\frac{15}{14}\frac{H\Lambda ZI}{r}\right)^{2/5}, \quad (24)$$

and $T_e(r)$ from Eq. (23) is virtually independent of the initial condition even for $r > 10$ μm. On the other hand, for large r, thermal conductivity becomes the principal term in Eq. (22). In this case, $T_e(r)$ behaves as $r^{-2/7}$; a similar behavior of temperature was also obtained as a result of the numerical solution of the system (8)–(19) for $r \geq 30$ μm. It can be seen from Fig. 3 that formula (23) for $I = 3.2$ A and $Z = 1.8$ gives an exaggerated result for small r, which is due to the disregard of the energy exchange with ions and the averaging of the Coulomb logarithm. Let us take into account the effect of variation of the Coulomb logarithm. We can write this quantity in the form [21]

$$\Lambda = 23.4 - 1.15\log n + 3.45\log T_e. \quad (25)$$

In this expression, we disregard the variation of T_e (of the order of 4 eV) as compared to n. Instead of expression (23),

we then obtain

$$T_e(r) = r^{-4/3}\left\{r_0^{10/3}T_0^{5/2} + HIZ\left(3.06(r^{7/3} - r_0^{7/3}) - 0.535\left[r^{7/3}\ln\left(\frac{IZ}{r^2}\right) - r_0^{7/3}\ln\left(\frac{IZ}{r_0^2}\right)\right]\right)\right\}^{2/5}. \quad (26)$$

It can be seen from Fig. 3 that this expression describes the temperature for $r \le 10$ μm more correctly.

3.3. Acceleration of Ions

Let us write the equations (11) of momentum conservation in the form of the equation of motion for a spherically symmetric plasma jet:

$$\frac{1}{u}\frac{\partial u}{\partial r} = \frac{\frac{1}{S}\frac{\partial S}{\partial r} - \frac{1}{T'}\frac{\partial T'}{\partial r}}{\frac{mu^2}{T'} - 1}, \quad T' = \frac{P}{n} = T_i + ZT_e, \quad (27)$$

where S *i*s the cross-sectional area of the jet.

Let us use this formula for estimating the behavior of the ion velocity for large values of r. In this case, we can neglect unity in the denominator and the ion temperature in the expression for T', taking T_e in the form (23). The result of integration of Eq. (27) in this case is presented in Fig. 4. It can be seen that the results are in satisfactory agreement with the results of 2D numerical calculation. The solution of Eq. (27) with temperature (23) for large r can be approximately written as

$$u(r) \approx 3.5\sqrt{(H\Lambda I)^{2/5}Z^{7/5}m^{-1}\left(1 - \frac{1}{r^{2/5}}\right)}. \quad (28)$$

For $I = 3.2$ A, $Z = 1.8$, $r = 200$ μm, and $\Lambda = 5$, the velocity is 1.46×10^6 cm/s, while the value obtained as a result of numerical calculations is 1.55×10^6 cm/s. The obtained expression for velocity for $r \longrightarrow \infty$ has a finite limiting value equal to 1.56×10^6 cm/s. Thus, the main contribution to the acceleration of ions comes from the electron pressure gradient.

3.4. Mean Ion Charge

The dependence of the mean charge of the plasma on the distance is depicted in Fig. 2. It can be seen that the mean charge of ions increases monotonically and the main ionization processes are localized at a distance as small as 2 μm. At a distance exceeding 5 μm, reactions cease altogether and a "freezing" of ion composition, which remains unchanged upon the further expansion, takes place. This is due to the fact that, in accordance with Eq. (9), the reaction rates are proportional to the concentration to the first (ionization) and second (recombination) power. Upon the expansion of the plasma, its concentration decreases rapidly (in proportion to r^{-3} at the initial stage); consequently, the reaction rates also decrease, the recombination being "cut off" more effectively. The ions which acquire a large charge at the jet base, where ionization can be regarded as virtually in equilibrium, enter the region of slow reactions at a high velocity. However, the time of charge equalization to the equilibrium value in this region is longer than the time of flight of ions. Thus, the mean charge of the plasma jet differs significantly from the equilibrium charge, depends on the acceleration, and cannot unambiguously reflect the plasma parameters in the region of the cathode spot, as was assumed in [11]. In order to prove this, we consider a steady-state spherically expanding plasma in which reactions with singly, doubly, and triply charged ions take place. In this case, Eqs. (9) assume the form

$$u\frac{\partial c_1}{\partial r} = -\alpha_{12}c_1n_e + \beta_{21}c_2n_e^2,$$
$$u\frac{\partial c_3}{\partial r} = \alpha_{23}c_2n_e - \beta_{32}c_3n_e^2, \quad (29)$$
$$c_2 = 1 - c_1 - c_3, \quad n_e = n_0\frac{r_0^2}{r^2}(2 + c_3 - c_1).$$

Since the main reactions occur in the immediate vicinity of the cathode, we take the electron temperature in the form (26). We assume that $n_0 = 10^{20}$ cm^{-3}, $r_0 = 1$ μm and vary velocity u. The result is presented in Fig. 5. It can be seen that the mean charge of the plasma jet differs considerably from the equilibrium value and depends on the velocity of ions.

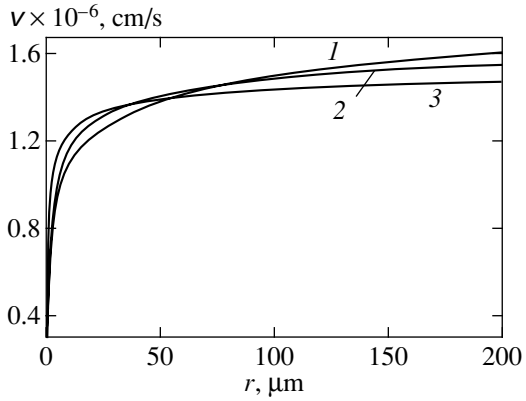


Fig. 4. Ion velocity distribution: formula (28) for T_e = 4.5 eV (*1*), 2D calculations (*2*), and formula (28) with T_e defined by formula (23) (*3*).

3.5. Potential Distribution

It was stated in [22] that the cathode plasma cannot transmit a high-density current (~10^8 A/cm^2) for typical values of the cathode potential drop at a level of 10–30 V. This conclusion was drawn only on the basis

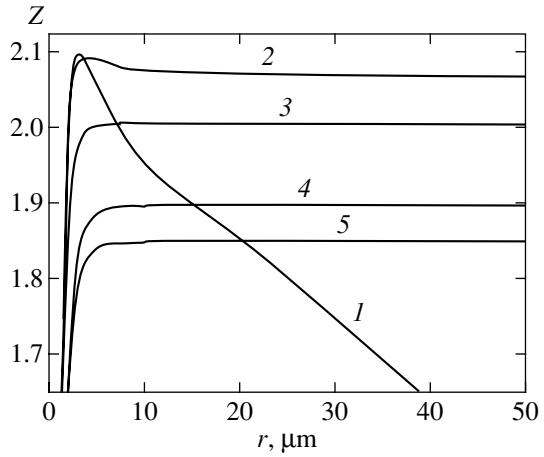


Fig. 5. Mean ion charge distribution along the jet, obtained by integrating system (29) for different values of velocity u: $u = 0$, equilibrium ionization (*1*), $u = 10^5$ cm/s (*2*), $u = 5 \times 10^5$ cm/s (*3*), $u = 10^6$ cm/s (*4*), and formula (27) with T_e from Eq. (26) (*5*).

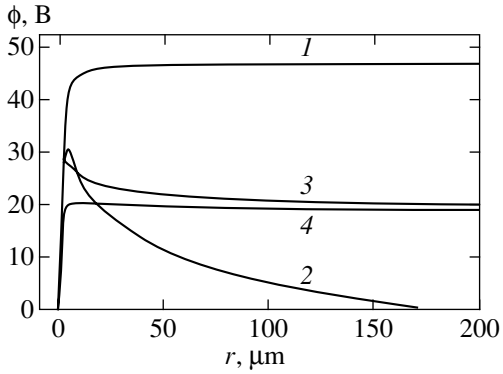


Fig. 6. Potential distribution along the jet: formula (31) for $T_e = 4.5$ eV disregarding the gradient term (*1*), formula (31) for $T_e = 4.5$ eV taking into account the gradient term (*2*), the result of integration of Eq. (30) with T_e from (23) (*3*), and 2D calculations (*4*).

of an analysis of the ohmic term in the generalized Ohm's law (19), which does not reflect the actual situation in which the plasma pressure gradient plays a significant role.

The potential distribution obtained from the numerical solution of the system of equations (8)–(19) is shown in Fig. 2. The main feature of the behavior of the potential is the existence of the peak at a distance of about 20 μm from the cathode. It should be noted that it is not a potential hump having a height of several tens of volts, whose existence was presumed by Plyutto *et al.* [2]. The potential peak height in our case is of the order of several volts. Let us prove that its emergence is due to the counteraction of the ohmic and gradient terms in the generalized Ohm law (19). Let us write Eq. (19) in the spherically symmetric case, retaining only these terms:

$$\frac{\partial \phi}{\partial r} = \frac{Z}{\sigma_0 T_e^{3/2}} \frac{I}{r^2} + r^2 \frac{\partial r^{-2} T_e}{\partial r}. \quad (30)$$

Under the assumption concerning the constancy of T_e, the integral of (30) is equal to

$$\phi = \frac{ZI}{\sigma_0 T_e^{3/2}} \left(\frac{1}{r_0} - \frac{1}{r} \right) - 2T_e \ln \frac{r}{r_0}. \quad (31)$$

The first (ohmic) term is bounded for $r \longrightarrow \infty$, while the second (gradient) term increases indefinitely, which inevitably leads to the formation of a hump. Physically, this means that the electron current induced by the pressure gradient exceeds the preset current of the jet, leading to the emergence of an electric field decelerating electrons in the plasma. The potential distribution calculated using formula (31) for $T_e = 4.5$ eV, $Z = 1.8$, and $I = 3.2$ A is presented in Fig. 6 (curve *2*). It can be seen that behind the inflection, the potential decreases indefinitely in accordance with Eq. (31). It can easily be shown that the $T_e(r)$ dependence of the form r^k for $0 > k > -2/3$ leads to potential stabilization. The actual behavior of temperature corresponds to this law. Curve *3* in Fig. 6 shows the potential distribution obtained as a result of the substitution of temperature in the form (23) into Eq. (30). It is close to the dependence obtained with the help of 2D calculations (curve *4* in Fig. 6). The magnitude of potential for $r \longrightarrow \infty$ is finite and equal to 17.4 V, which is close to the value of the cathode potential drop (16 V) given in [14].

4. CONCLUSIONS

An analysis of the results of the theoretical investigation of ecton processes in vacuum arcs leads to the following mechanism of the formation of the ion flow parameters in vacuum arcs. The concentration of energy due to Joule heating in microscopic volumes at the surface of the cathode leads to their explosive degradation. The cathode material consecutively passes during short time periods (~1 ns) through the condensed state and the stages of nonideal and ideal plasma. The current density at the base of the plasma jet is rather high and the cathode material continues being heated after the transition to the plasma state also. The Joule energy liberated at this stage is the kinetic energy of electrons, acquired in the electric field. A part of this energy is spent in ionization and heating of ions. Ionization processes occur in a narrow region of the order of a micrometer in the vicinity of the cathode, and the ionization composition of the plasma does not change subsequently. Under the action of the electron pressure gradient, ions acquire directional velocities of the order of 10^6 cm/s even at distances of a few micrometer.

ACKNOWLEDGMENTS

This work was financed by the RFBR-INTAS (grant no. 97-0663) and by the Russian Foundation for Basic Research (grant no. 99-02-18234).

REFERENCES

1. R. Tanberg, Phys. Rev. **35**, 1080 (1930).
2. A. A. Plyutto, V. V. Ryzhkov, and A. T. Kapin, Zh. Éksp. Teor. Fiz. **47**, 494 (1964) [Sov. Phys. JETP **20**, 328 (1964)].
3. G. M. Kassirov and G. A. Mesyats, Zh. Tekh. Fiz. **34**, 1476 (1964) [Sov. Phys. Tech. Phys. **9**, 1141 (1965)].
4. G. A. Mesyats, in *Proceedings of the X International Conference on Phenomena in Ionized Gases, Oxford, 1971*, Part 2, p. 333.
5. Ya. B. Zel'dovich and Yu. P. Raizer, *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena*, 2nd ed. (Nauka, Moscow, 1966; Academic, New York, 1966).
6. G. A. Mesyats and D. I. Proskurovsky, *Pulsed Electrical Discharge in Vacuum* (Springer-Verlag, Berlin, 1989).
7. G. A. Mesyats, *Ectons in Vacuum Discharge: Breakdown, Spark, Arc* (Nauka, Moscow, 2000).
8. I. G. Brown, B. Feinberg, and J. E. Galvin, J. Appl. Phys. **63**, 4889 (1988).
9. G. Yu. Yushkov, E. M. Oks, A. Anders, and I. G. Brown, J. Appl. Phys. **88**, 5618 (2000).
10. *Handbook of Vacuum Arc: Science and Technology*, Ed. by R. L. Boxman, P. J. Martin, and D. M. Sanders (Noyes Publ., Park Ridge, 1995).
11. A. Anders, Phys. Rev. E **55**, 969 (1997).
12. E. A. Litvinov, G. A. Mesyats, and D. I. Proskurovsky, Usp. Fiz. Nauk **139**, 265 (1983) [Sov. Phys. Usp. **26**, 138 (1983)].
13. V. F. Puchkarev, D. I. Proskurovsky, and A. M. Murzakaev, Zh. Tekh. Fiz. **57**, 1224 (1987) [Sov. Phys. Tech. Phys. **32**, 1405 (1987)]; Zh. Tekh. Fiz. **58**, 88 (1988) [Sov. Phys. Tech. Phys. **33**, 51 (1988)].
14. I. G. Kesaev, *Cathodic Processes in Electric Ars* (Nauka, Moscow, 1968).
15. B. Juttner, Beitr. Plasma Phys. **19**, 25 (1979).
16. C. W. Kimblin, J. Appl. Phys. **44**, 3074 (1973).
17. J. E. Daalder, J. Phys. D **8**, 1647 (1975).
18. V. E. Zinov'ev, *Thermophysical Properties of Metals at High Temperature: Handbook* (Metallurgiya, Moscow, 1989).
19. I. G. Brown, Rev. Sci. Instrum. **65**, 3061 (1994).
20. V. V. Loskutov, A. V. Luchinskiĭ, and G. A. Mesyats, Dokl. Akad. Nauk SSSR **271**, 1120 (1983) [Sov. Phys. Dokl. **28**, 654 (1983)].
21. S. I. Braginskii, in *Reviews of Plasma Physics*, Ed. by M. A. Leontovich (Gosatomizdat, Moscow, 1963; Consultants Bureau, New York, 1963), Vol. 1.
22. M. P. Zektser and V. I. Rakhovskiĭ, Dokl. Akad. Nauk SSSR **276**, 86 (1984).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 93, No. 5, 2001, pp. 1074–1081.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 5, 2001, pp. 1237–1245.

SOLIDS
Structure

The Solution of the Kinetic Equation for Phonon Heat Conductivity by the Method of Momenta and the Influence of Isotopic Disorder on Phonon Heat Conductivity of Germanium and Silicon Crystals at $T = 300$ K

A. P. Zhernov[†]

Institute of Superconductivity and Solid State Physics, Russian Research Centre Kurchatov Institute, pl. Kurchatova 1, Moscow, 123182 Russia

Received January 18, 2001

Abstract—The problem of solving the kinetic equation for the heat conductivities of dielectric substances and semiconductors by the method of momenta is discussed. Microscopic models are used to estimate the effect of isotopic disorder on the thermal resistance of germanium crystals in the multimomentum approximation. The contributions of acoustic and optical phonons are taken into account. Excess thermal resistance ΔW of samples with a natural isotopic composition compared with highly enriched samples is calculated. For germanium, the theoretical and experimental ΔW values are in close agreement with each other. For silicon, the theoretical ΔW value is much smaller than its experimental excess thermal resistance. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In the past decade, virtually chemically pure and highly enriched isotopically diamond, germanium, and silicon massive single crystals were synthesized thanks to progress in the technology of producing high-purity materials. Single crystals with intermediate contents of isotopes of various kinds were also prepared. Diamond single crystals with different isotopic compositions were grown at the General Electric (USA) Laboratory, and growth of germanium single crystals was due to joint efforts of the Institute of Molecular Physics (Russian Research Centre Kurchatov Institute) and Lawrence National Laboratory at Berkeley (USA). Recently, isotopically enriched silicon single crystals were prepared thanks to joint efforts of Russian, German, and Japanese scientists. These single crystals were used to study the influence of the ratio between the concentrations of isotopes on various physical properties. In particular, the behavior of heat conductivity $\kappa(T)$ was investigated. Heat transfer in diamond was studied in [1–4], and the behavior of $\kappa(T)$ of germanium and silicon, in [5, 6] and [7–9], respectively.

Note that in 2000, perfect and highly enriched ^{28}Si (99.9%) single crystals were grown at the laboratory headed by G.G. Devyatykh (Institute of Chemistry of High-Purity Substances, Nizhni Novgorod). These samples were used to measure their thermal resistance W. The measurements were performed by M. Caradona's team (Stuttgart) at temperatures from helium to room [8]. A.V. Gusev *et al*. (Nizhni Novgorod) studied the temperature range ~70–300 K [10]. The W values for ^{28}Si were compared with those obtained for silicon samples with a natural isotopic composition (^{nat}Si). Note that the data on excess thermal resistance $\Delta W = W(^{nat}\text{Si}) - W(^{28}\text{Si})$ caused by isotopic disorder obtained in [8, 11] at room temperature $T = 300$ K differed several times.

The general problems of the theory of heat conductivity of dielectric substances and semiconductors were discussed in several book-length monographs (e.g., see [11, 12]).

Heat conductivity is usually considered on the assumption that various nonequilibrium phonon scattering processes, namely, boundary scattering on sample walls, elastic scattering caused by isotopic disorder and impurities, and inelastic anharmonic collisions, are independent. It is well known that normal (N) anharmonic phonon scattering processes, that is, scattering with the conservation of the quasi-momentum, do not cause a finite thermal resistance by themselves. At the same time, in the temperature range in which resistance processes of phonon scattering on phonons, that is, processes accompanied by the loss of quasi-momentum (U-processes), are frozen, N-processes can determine the structure of a stationary nonequilibrium distribution of phonons. As a result, N-processes play a fairly important role. For this reason, the results obtained in [13] (also see [14]) are extensively used in analyzing various relaxation mechanisms. In [13], the special role played by N-processes was qualitatively taken into account, and a comparatively simple equation for the

[†]Deceased.

heat conductivity coefficient was derived. These results were obtained based on several simplifying assumptions. Namely, the Debye approximation was used to describe phonon modes, and no distinction was drawn between longitudinal and transverse modes. In addition, the frequency and temperature dependences of the rates of anharmonic relaxation of phonon modes were described by comparatively simple power laws suggested in [15]. The equations for the relaxation rates of *N*- and *U*-processes contained parameters found by adjustment to experimental results. It was also assumed that the rate of relaxation caused by isotopic disorder was proportional to ω^4 (as concerns boundary scattering, e.g., see [16]).

Note that the theory suggested in [13] was generalized to strongly anisotropic phonon spectra characteristic of silicon and germanium because of the presence of soft transverse modes (see [6, 17]).

The author knows of only two works [18, 19] concerned with Group IV semiconductors in which consistent calculations within the framework of the kinetic equation were performed based on microscopic models for describing phonon modes and anharmonic interaction. In [18], the behavior of the heat conductivity of germanium was analyzed in a wide temperature range (from helium to room temperatures). The roles played by various anharmonic *N*- and *U*-processes ($t + t \longrightarrow l$, $t + l \longrightarrow l$, and $l + l \longrightarrow l$) were considered in detail. Simultaneously, boundary scattering and scattering related to isotopic disorder were taken into account. Note also that phonons and anharmonic interactions were analytically described in [18] based on the isotropic continual model. The kinetic equation was solved in the multimomentum approximation, and the nonequilibrium correction to the equilibrium distribution function of phonons was selected in the form

$$\Phi_z(\mathbf{q}, j) = -q_z \sum_{r'} \eta^j_{r'} \left(\frac{|\mathbf{q}|}{q_D}\right) r'.$$

In calculations, the expansion in momenta was performed taking into account the first several terms.

In work [19], published comparatively recently, heat transfer in Ge and Si crystals with different isotopic compositions was studied in detail. The authors developed a model of anisotropic pair interatomic interaction and used this model to describe photon spectra, thermal expansion, and anharmonic interaction of phonons. The contributions of acoustic and optical phonon modes were taken into account. As far as the kinetic equation is concerned, it was solved iteratively. Note that the theoretical results obtained in [19] for highly enriched ^{28}Si at a $\kappa(T)$ maximum were several times smaller than the experimental values [8]. This might be caused by the use of an iterative procedure, which might prove ineffective precisely near the $\kappa(T)$ maximum.

We see that, during the past years, much experimental data on the heat conductivities of C, Si, and Ge crystals highly enriched in certain isotopes have been collected. The theoretical calculations performed to date allow isotopic disorder effects on the behavior of $\kappa(T)$ both in the region of its maximum and at room temperature to be described only qualitatively. Substantial quantitative discrepancies between theory and experiment require using more elaborate procedures for theoretical calculations and, in particular, elucidating the role played by anisotropy of elastic scattering of phonons on isotopes.

In this work, a method for determining linearly independent symmetrized momenta present in the problem of lattice heat conductivity is described; the method is based on the theory of representations. Crystals with cubic structures and also with hexagonal and tetragonal structures (point symmetry groups D_{6h} and D_{4h}, respectively) are considered.

Next, calculations for Ge and Si crystals with natural isotopic compositions are performed. The thermal resistance part ΔW caused by isotopic disorder is determined. Our consideration will be confined to the practically important case of room temperatures, $T = 300$ K. Generally speaking, the kinetic equation was solved in the multimomentum approximation taking into account the contributions of the isotopic mechanism of scattering and anharmonic interaction of phonons to the scattering operator. The phonon spectrum was described by the many-particle model of charges on bonds [20, 21] and the model suggested in [19], which treated harmonic and anharmonic interatomic interactions in the pair approximation. The model of charges on bonds is more precise than the model suggested in [19], but setting anharmonicity in this model involves difficulties.

As a first approximation, ΔW can be determined solely based on the isotopic scattering mechanism with a trial phonon nonequilibrium distribution function. This trial function possesses necessary properties determined by crystal symmetry and allows a possible difference in group velocity sign between of acoustic and optical phonons to be taken into account; precisely for this purpose, we used the 3*C* model of charges on bonds.

A general solution to the kinetic equation in the multimomentum approximation, when both isotopic and anharmonic mechanisms of phonon scattering should be taken into account, was obtained using the model suggested in [19].

2. THE ZIMAN THEORY OF HEAT CONDUCTIVITY. VARIATIONAL PRINCIPLE

As mentioned, the rigorous theory of phonon heat conductivity was constructed by Ziman in terms of the kinetic equation based on the variational principle. This theory, which is described, for instance, in the well-known monograph [11], is still topical.

In the Ziman theory, the kinetic equation is linearized through replacing the $n(\mathbf{q}, j)$ occupation numbers

of phonon modes by their equilibrium values $n_0(\mathbf{q}, j)$ in the field term. Simultaneously, the $\Phi_\alpha(\mathbf{q}, j)$ function appears in the integral of collisions; this function characterizes small deviations from equilibrium. In the lowest order with respect to the temperature gradient, we have

$$n(\mathbf{q}, j) = n_0(\mathbf{q}, j) + \Phi_z(\mathbf{q}, j)n_0(\mathbf{q}, j)(n_0(\mathbf{q}, j) + 1)\frac{\partial T}{\partial z}. \quad (1)$$

The Φ_z function satisfies the equation

$$X_z(\mathbf{q}j) = -\sum_{j'}\int d\mathbf{q}' P(\mathbf{q}, j; \mathbf{q}', j')\Phi_z(\mathbf{q}', j'), \quad (2)$$

where

$$X_z(\mathbf{q}j) = \frac{\hbar\omega(\mathbf{q}, j)}{k_B T^2} n_0(\mathbf{q}, j)(n_0(\mathbf{q}, j) + 1)v_z(\mathbf{q}, j). \quad (3)$$

Here and throughout, $\omega(\mathbf{q}, j)$, $v_z(\mathbf{q}, j)$, and $\mathbf{e}(\mathbf{q}, j)$ are the frequency, the group velocity, and the polarization vector of the $\{\mathbf{q}, j\}$ phonon mode, respectively. The $P(\mathbf{q}, j; \mathbf{q}', j')$ value is the matrix element of the scattering operator between the $\mathbf{q}$, j and $\mathbf{q}'$, j' states. According to the microscopic reversibility principle, $P(\mathbf{q}, j; \mathbf{q}', j') = P(\mathbf{q}', j'; \mathbf{q}, j)$. In the problem under consideration, the P operator is the sum of three terms, which describe phonon scattering at sample boundaries, anharmonic interaction, and isotopic disorder effects.

Let us write the explicit expressions for the $P_{\text{iso}}(\mathbf{q}, j; \mathbf{q}', j')$ and $P_{\text{anh}}(\mathbf{q}, j; \mathbf{q}', j')$ terms, which describe isotopic and anharmonic scattering (e.g., see [11] and [12, 22], respectively). First, we have

$$P_{\text{iso}}(\mathbf{q}, j; \mathbf{q}', j') = \frac{V_0\xi^2}{8\pi^2}\omega(\mathbf{q}, j)\omega(\mathbf{q}', j') \times |\mathbf{e}(\mathbf{q}, j)\mathbf{e}(\mathbf{q}', j')|^2 N_{qq'}^{jj}\delta\{\omega(\mathbf{q}, j) - \omega(\mathbf{q}', j')\}. \quad (4)$$

In addition, it is assumed that $N_{qq'}^{jj} = n_0(\mathbf{q}, j)(n_0(\mathbf{q}', j') + 1)$. Here, $\xi^2 = (\langle M^2\rangle - \langle M\rangle^2)/\langle M\rangle^2$ is the parameter that describes isotopic disorder, and

$$\langle M\rangle = \sum_i c_i M_i,$$

where c_i is the concentration of atom–isotopes of kind i.

The equation for the anharmonic part of the operator of collisions is cumbersome. Many terms of this equation can, however, be combined by formally introducing negative polarization indices $-j \le j_s \le j$. Assume by definition that $\omega(\mathbf{q}, -j) = -\omega(\mathbf{q}, j)$. The square of the modulus of the anharmonic vertex will be denoted as

$$\left|\Phi_{jj'j''}^{\mathbf{q}\mathbf{q}'\mathbf{q}''}\right|^2 = D_{jj'j''}^{\mathbf{q}\mathbf{q}'\mathbf{q}''}.$$

We then have

$$D_{jj'j''}^{\mathbf{q}\mathbf{q}'\mathbf{q}''} = -D_{-jj'j''}^{\mathbf{q}\mathbf{q}'\mathbf{q}''}.$$

In addition, $\Phi_z(-\mathbf{q}, -j) = -\Phi_z(\mathbf{q}, j)$ This eventually gives

$$P_{\text{anh}}(\mathbf{q}, j_s; \mathbf{q}', j_{s'}') = \frac{\pi}{2k_B T^2\hbar^2} \times \sum_{\mathbf{q}'', j_{s''}''} \frac{D_{j_s j_{s'}' j_{s''}''}^{\mathbf{q}\mathbf{q}'\mathbf{q}''}\delta\{\omega(\mathbf{q}, j_s) + \omega(\mathbf{q}', j_{s'}') + \omega(\mathbf{q}'', j_{s''}'')\}}{\sinh\beta_{j_s}^{\mathbf{q}}\sinh\beta_{j_{s'}'}^{\mathbf{q}'}\sinh\beta_{j_{s''}''}^{\mathbf{q}''}}, \quad (5)$$

where $\beta_{j_s}^{\mathbf{q}} = \hbar\omega(\mathbf{q}, j_s)/2k_B T$. The heat conductivity coefficient is given by the equality

$$\kappa(T) = -\frac{k_B T^2}{8\pi^3}\sum_j\int d\mathbf{q}\chi(\mathbf{q}, j)\Phi(\mathbf{q}, j). \quad (6)$$

Following Ziman, let us write thermal resistance $W = 1/\kappa$ in the form

$$W = \frac{8\pi^3}{k_B T^2}\sum_{j, j'}\iint d\mathbf{q}d\mathbf{q}'\Phi(\mathbf{q}, j)P(\mathbf{q}, j; \mathbf{q}', j')\Phi(\mathbf{q}', j') \times\left(\sum_j\int d\mathbf{q}X(\mathbf{q}, j)\Phi(\mathbf{q}, j),\right)^{-2}. \quad (7)$$

Note that, according to the variational principle, the $\Phi_z(\mathbf{q}, j)$ function satisfying kinetic equation (2) minimizes (6).

A trial function is, as a rule, selected in the form of the expansion in some set of functions $\{\varphi_r\}$ with coefficients η_r to be found. The corresponding system of equations with respect to η_r has the form

$$X_r = \sum_{r'} P_{rr'}\eta_{r'},$$

where X_r and $P_{rr'}$ are the matrix elements of the X and P operators in the system of functions $\{\varphi_r\}$.

The selection of functions in the method of momenta is considered in the next section.

3. SELECTION OF MOMENTA

At low temperatures, phonon-phonon transitions are essentially inelastic. As only phonons with small quasimomenta are excited, umklapp transitions (U-processes) are possible for nonequilibrium modes situated only in some regions of the irreducible part of the Brillouin zone. The $P(\mathbf{q}, j; \mathbf{q}', j')$ transition probability in an ordered crystal should therefore depend both on the absolute values of the $\mathbf{q}$ and $\mathbf{q}'$ vectors and on their mutual orientation in the reciprocal lattice space. Normal processes ensure the existence of a finite number of nonequilibrium modes in the region of their most effective umklapp scattering. The appearance of isotopes of

different kinds in the crystal lattice, on which phonons are scattered elastically, generally makes the $\Phi(\mathbf{q}, j)$ nonequilibrium phonon distribution function more isotropic. This effect is observed at temperatures at which inelastic scattering by phonons and elastic scattering by impurities make contributions of the same order of magnitude. Under the conditions of a strong phonon spectrum anisotropy (this anisotropy may, for instance, be related to weak-dispersion transverse modes), elastic scattering, however, also becomes essentially anisotropic as a result of isotopic disorder. It follows that, at low temperatures, because of *U*-processes and phonon spectrum anisotropy, explicitly taking into account the nonstandard structure of the $\Phi(\mathbf{q}, j)$ distribution function is of fundamental importance. At temperatures $T \geq \theta_D/6$, virtually all groups of acoustic phonons are excited, and anharmonic scattering processes of various types become predominant. The role played by isotopic scattering is then insignificant in principle. In this situation, the distribution of the standard form

$$\Phi_z(\mathbf{q}j) \propto q_z\omega(\mathbf{q}, j)$$

should satisfactorily describe the real momentum distribution. We cannot, however, rule out the possibility of a substantial influence of strong phonon spectrum anisotropy on the distribution function Φ.

Taking into account the aforesaid, consider the problem of selecting momenta present in expansions of the $\Phi_\alpha(\mathbf{q}, j) = \omega(\mathbf{q}, j)\phi(\mathbf{q}, j)$ stationary distribution. By definition, $\Phi_\alpha(\mathbf{q}, j)$ is a piecewise continuous periodic function with the lattice period, and its point symmetry group coincides with that of the crystal. Therefore, $\phi(\mathbf{q}, j)$ can be expanded in a series in $g_\alpha^{(i_m)}(\mathbf{q}, j)$ and $f_j^{(r_l)}(\mathbf{q})$ functions. Momentum $g_\alpha^{(i_m)}$ transforms under the vector representation and characterizes the angular dependence of the distribution. Momenta of the type

$$f_j^{(r_l)}(\mathbf{q}) = (q/q_D)^l, \quad l = 0, n_r$$

(if $\mathbf{q}$ is given outside the first Brillouin zone, we must consider the $\mathbf{q} + \mathbf{B}$ value, where $\mathbf{B}$ is one of the reciprocal lattice vectors) take into account the dependence of the distribution on $|\mathbf{q}|$.

According to the Riesz–Fisher theorem, if $\{\varphi(r)\}$ is a complete orthonormalized set of functions and the sequence of η_r numbers is such that the $\sum_{r=1}^{\infty}\eta_r^2$ series converges, then $\eta_1\varphi^{(1)} + \eta_2\varphi^{(2)}\ldots$ converges in mean to Φ. According to the uniqueness theorem, the $\{\eta_r\}$ coefficients determine uniquely $\Phi(\mathbf{q})$ almost everywhere in the domain of definition. In particular, if $\Phi(\mathbf{q})$ is continuous, it is continuous everywhere.

Consider momenta of the $g_\alpha^{(i_m)}$ type in detail. Under the conditions of the problem, they possess the same properties as the components of the group velocity vector $v_\alpha(\mathbf{q}, j) = \partial\omega(\mathbf{q}, j)/\partial q_\alpha$. To make formulas more compact, the j index of mode polarization will further be omitted in this section.

Taking into account the aforesaid, the momenta describing the angular dependence $\Phi_\alpha(\mathbf{q})$ can conveniently be selected in the form

$$g_\alpha^{(i_m)}(\mathbf{q}) = v_{\alpha_1}^{(i_1)}\ldots v_{\alpha_j}^{(i_j)}, \quad i_1 + i_2 + \ldots + i_j = m, \quad (8)$$

where i_s are integers. In a cubic crystal under symmetry transformations $G \in O_h$, the $v_\alpha(\mathbf{k})$ component transforms as the basis function of the $\nu = F_{1u}$ irreducible representation. Momenta (8) at $n \neq 1$ form the basis of the $\Gamma_{n\nu}$ reducible representation, which is the direct product of n Γ_ν representations,

$$\Gamma_{n\nu} = \bigotimes_{i=1}^{n}\Gamma_\nu. \quad (9)$$

The Φ_α function transforms under the F_{1u} irreducible representation. We must therefore select functions of the F_{1u} symmetry type from set of momenta (8), which form bases of representations (9). The sought momenta can be determined using the projection operators

$$P^{(\nu)} = \frac{f_\nu}{g}\sum_G \chi_\nu(G)G, \quad (10)$$

where f_ν is the dimension of irreducible representation ν and $\chi_\nu(G)$is the character of irreducible representation ν for G. We have

$$g_\nu = P^{(\nu)}g_n, \quad (11)$$

where g_n is the basis function of the $\Gamma_{n\nu}$ reducible representation.

The group velocity components v_x, v_y, and v_z transform under the action of G as the x, y, and z coordinates. Taking this into account and directly using (10) and (11), we can find the g_ν functions.

The total number of independent functions of form (8) for some m is $(m + 1)(m + 2)/2$. The number of linearly independent momenta of the $\{g_\alpha^{(i_m)}\}$ type present in the expansion of $\Phi_\alpha(\mathbf{q})$ is substantially smaller. This number is determined by the multiplicity of the νth representation m_ν in the $\Gamma_{n\nu'}$ reducible representation. If characters $[\chi^n]G$ of the $\Gamma_{n\nu'}$ representation are known, then

$$m_\nu = \frac{1}{g}\sum_C g_C\chi_\nu(C)[\kappa^n](C),$$

where g_C and g are the numbers of the elements in the C class and in the group. By definition, the $\Gamma_{n\nu'}$ representation generated by the $v_{\alpha_1}^{(i_1)}\ldots v_{\alpha_j}^{(i_j)}$ functions is symmetrical with respect to all indices $\alpha_1\ldots\alpha_j$. The

Table 1. Characters of tensor representations $\Gamma_{nv} = \bigotimes_{i=1}^{n} \Gamma_{1u}$ and equivalent basis functions of the Γ_{1u} irreducible representation

n	Class					m	Linearly independent momenta
	E	$3C_4^2$	$6C_4$	$6C_2$	$8C_3$		
1	3	–3	1	–1	0	1	x
3	10	–2	0	–2	1	2	x^3, $x(y^2 + z^2)$
5	21	–3	1	–3	0	4	x^5, $x^3(y^2 + z^2)$, $x(y^4 + z^4)$, xy^2z^2
7	36	–4	0	–4	0	6	x^7, $x^5(y^2 + z^2)$, $x^3(y^4 + z^4)$, $x(y^6 + z^6)$, $x(y^4z^2 + y^2z^4)$, $x^3y^2z^2$
9	55	–5	1	–5	1	9	x^9, $x^7(y^2 + z^2)$, $x^5(y^4 + z^4)$, $x^3(y^6 + z^6)$, $x(y^8 + z^8)$, $x(y^2z^6 + y^6z^2)$, $x^3(y^2z^4 + y^4z^2)$, $x^5y^2z^2$, xy^4z^4

Note: m is the multiplicity of F_{1u} in Γ_n for the O_h group.

characters of such representations are found by the formula [23]

$$[\chi^n]G = \sum_{\{p\}} \frac{\chi^{p_1}(G)\ldots\chi^{p_n}(G)}{p_1!2p_2!\ldots np_n!}. \quad (12)$$

Here, p_i are integral positive numbers. The $\{\ldots\}$ symbol denotes the summation over all possible sets of p_i satisfying the condition

$$p_1 + 2p_2 + \ldots + np_n = n. \quad (13)$$

The characters of the $\Gamma_{nv} = \bigotimes_{i=1}^{n} \Gamma_{1u}$ tensor representations, the m values (m is the multiplicity of F_{1u} in Γ_{nv}), and the equivalent bases of the Γ_{1u} representation are listed in Table 1 for cubic crystals.

Next, consider crystals with hexagonal and tetragonal structures, whose point symmetry groups are D_{6h} and D_{4h}, respectively. The z coordinate along the fourfold symmetry axis for D_{4h} and the sixfold symmetry axis for D_{6h} transforms under the one-dimensional irreducible representation denoted by A_{2u} in both symmetry groups, and the x and y coordinates are basis functions of two-dimensional irreducible representations, E_{1u} in the D_{6h} group and E_u in the D_{4h} group. For this reason, polynomials of the form $x^iy^jz^k$ can be represented by products of functions from the basis sets of the $\Gamma_n = \bigotimes_{i=1}^{n} A_{2u}$ and $\Gamma_m = \bigotimes_{i=1}^{m} E_{1u}$ or $\Gamma_m = \bigotimes_{i=1}^{m} E_u$ tensor representations. The Γ_n representation is one-dimensional for all n, and the z^n function is a basis function for the A_{1g} or A_{2u} representation depending on whether n is even or odd, respectively. The characters of the Γ_m representations for the D_{6h} and D_{4h} symmetry groups and the equivalent basis sets for the A_{1g} and E_{1u} (or E_u) representations are listed in Table 2.

Note that the tables given in [24] contain explicit expressions for the simplest $\mathbf{A}(\mathbf{k})$ vector functions symmetry-compatible with crystals and textures; the tables include most of the crystallographic classes and textures. These tables allow the first two or three additional angular momenta to be determined. Lastly, if the distribution is selected in the form

$$\Phi_z(\mathbf{q}, j) = v_z(\mathbf{q}, j)\omega(\mathbf{q}, j)F_j(\mathbf{q}), \quad (14)$$

the angular dependence can be described by specially chosen harmonics. They are found in the form of some combinations of spherical harmonics and possess the symmetry properties of the crystal. Such harmonics for hexagonal, tetragonal, and trigonal crystals were determined in [25].

For cubic crystals in the Debye approximation $v_\alpha(\mathbf{q})$ is proportional to v_α, the $\{g^{(a, n)}\}$ momenta are polynomials of v_x, v_y, and v_z. The number of linearly independent $\{g^{(i_m)}\}$ functions then substantially decrease. We have

$$g_\alpha^{(1)} = g_s(\mathbf{n}) = \sum_\alpha v_\alpha n_\alpha,$$

$$g_\alpha^{(2)} = \sum_\alpha v_\alpha^3 n_\alpha, \quad g_\alpha^{(3)} = \sum_\alpha v_\alpha^5 n_\alpha,$$

$$g^{(3)} = (v_x^4 + v_y^4 + v_z^4)g_s(\mathbf{n}), \quad (15)$$

$$g_\alpha^{(5)} = \sum_\alpha v_\alpha^5 n_\alpha, \quad g^{(6)} = (v_x^6 + v_y^6 + v_z^6)g_s(\mathbf{n}),$$

$$g_\alpha^{(7)} = \sum_\alpha v_\alpha^9 n_\alpha, \quad g_\alpha^{(8)} = (v_x^8 + v_y^8 + v_z^8)g_s(\mathbf{n})$$

($\mathbf{n}$ is the unit vector in the external field direction).

Table 2. Characters of tensor representations $\Gamma_n = \otimes_{i=1}^{n} A_{2u}$ and $\Gamma_m = \otimes_{i=1}^{m} E_u$ and equivalent basis functions of the A_{1g} (even n) and E_u (odd n) irreducible representations for groups D_{4h} and D_{6h}

n	Class							Linearly independent momenta	
	$\frac{E^*}{E}$	$\frac{C_2^*}{C_2}$	$2C_4^*$	$2C_3$	$2C_6$	$\frac{2U_2^*}{3U_2}$	$\frac{2U_2^*}{3U_2}$	D_{4h}	D_{6h}
1	2	–2	–1	0	1	0	0	x	x
2	3	3	0	–1	0	1	1	$x^2 + y^2$	$x^2 + y^2$
3	4	–4	1	0	–1	0	0	xy^2, x^3	$x(x^2 + y^2)$
4	5	5	–1	1	–1	1	1	x^2y^2, $x^4 + y^4$	$(x^2 + y^2)^2$
5	6	–6	0	0	0	0	0	x^3y^2, xy^4, x^5	$x(xy + y^2)^2$, $xy^2(y^2 - 3x^2)$
6	7	7	1	–1	1	1	1	$x^2y^2(x^2 + y^2)$, $y^6 + z^6$	$(x^2 + y^2)^3$, $x_6 - 15x^4y^2 + 15x^2y^4 - y^6$
7	8	–8	–1	0	1	0	0	x^5y^2, x^3y^4, xy^6, x^7	$x(x^2 + y^2)^3$, $x^3(x^4 + 7y^4)$, $xy^2(y^2 - 3x^2)(x^2 + y^2)$

Note: Classes of the D_{4h} group elements are labeled by asterisks.

4. NUMERICAL CALCULATION RESULTS FOR GERMANIUM AND SILICON AT T = 300 K: THE CONTRIBUTION OF THE ISOTOPIC MECHANISM OF SCATTERING TO THERMAL RESISTANCE

We started by applying the variational procedure to analyze the heat conductivity data obtained for germanium in the region of comparatively high temperatures.

The contribution of isotopic scattering to thermal resistance was first estimated based on the standard relations of the Ziman theory [Eqs. (4), (5), (8)]. The phonon spectra were described using the model of charges on bonds. Earlier, we used this model to analyze the influence of the ratio between the amounts of isotopes on the lattice parameters and the linear thermal expansion coefficients of germanium and silicon [27]. The calculations were performed using representations of the form $\Phi_a \propto \omega(\mathbf{q}, j)v_z(\mathbf{q}, j)$ for the nonequilibrium part of the phonon distribution function with or without taking the contribution of optical modes into account.

Compare the trial function

$$\Phi_z(\mathbf{q}j) \sim \omega(\mathbf{q}j)v_z(qj) \tag{16}$$

with the standard function

$$\Phi_z(\mathbf{q}j) \propto q_z, \tag{17}$$

which describes the distribution of nonequilibrium phonons for an isotropic continuum. We see that, first, unlike (17), (16) can have an arbitrary sign because of the group velocity $v_z(\mathbf{q}, j)$ factor. This is an important circumstance because optical phonons with $v_z < 0$ may make a noticeable contribution to ΔW (as a rule, $v_z > 0$ for acoustic phonons). Secondly, function (16) has translational symmetry and, therefore, correctly describes contributions to ΔW of phonons with $\mathbf{q}_1$ and $\mathbf{q}_2$ vectors differing by one of the reciprocal lattice vectors (the corresponding contributions are equal). Function (17) has no such property.

According to the theoretical calculations performed for T = 300 K, the $\Delta W = W(^{\text{nat}}\text{Ge}) - W(^{28}\text{Ge})$ value for the natural isotopic composition ($\xi^2 = 5.89 \times 10^{-4}$) equals 0.20 W^{-1} cm K without the contribution of optical modes and 0.32 W^{-1} cm K if this contribution is taken into account. Note that the standard selection of $\Phi_a \propto q_a$ yields ΔW = 0.75 W^{-1} cm K. The experimental ΔW value is 0.26 W^{-1} cm K.

The experimental data on heat conductivity κ and the contribution of the isotopic mechanism of scattering to thermal resistance ΔW are listed in Table 3. Recall that isotopic disorder is characterized by the parameter $\xi^2 = (\langle M^2\rangle - \langle M\rangle^2)/\langle M\rangle^2$.

In addition, we estimated the role played by multimomentum corrections when the kinetic equation is solved within the framework of the model of dynamic pair interactions described in detail in [19]. This model takes into account both harmonic and anharmonic interaction parts. Unlike the model of charges on bonds [20, 21], the model suggested in [19] only includes pair interatomic interactions. The first three additional momenta from (15) and two momenta of the $(|\mathbf{q}|/q_D)^l$ type (l = 1, 2) were included in the calculations. The kinetic equation was solved numerically simultaneously for the anharmonic and isotopic phonon scattering mechanisms. It was found that the inclusion of

Table 3. Experimental heat conductivities κ and thermal resistances ΔW caused by isotopic scattering in germanium crystals [5]

ξ^2	8.16×10^{-8}	7.75×10^{-5}	
T, K	κ, W cm^{-1} K^{-1}	κ, W cm^{-1} K^{-1}	ΔW, W^{-1} cm K
300	0.697	0.695	0.004
250	0.862	0.844	0.029
200	1.14	1.105	0.028
150	1.70	1.62	0.030

ξ^2	5.89×10^{-4}		1.54×10^{-3}	
T, K	κ, W cm^{-1} K^{-1}	ΔW, W^{-1} cm K	κ, W cm^{-1} K^{-1}	ΔW, W^{-1} cm K
300	0.590	0.26	0.537	0.43
250	0.795	0.258	0.628	0.44
200	0.889	0.248	0.753	0.45
150	1.23	0.223	0.981	0.43

additional momenta decreased ΔW by approximately 20%. The ΔW value found by the model suggested in [19] was $\Delta W = 0.23$ W^{-1} cm K. (The influence of the momenta of the higher orders was insignificant.) It follows that, for germanium at $T = 300$ K, the contribution of the isotopic scattering mechanism to the total thermal resistance amounts to 15–20%, in close agreement with the experimental data.

An interesting result follows from the data obtained in [5] and listed in Table 3. The data on the impurity part of thermal resistance ΔW for ^{70}Ge (96.3%) samples and samples of a natural isotopic composition obtained in the temperature range $T = 150$–250 K fall on a straight line passing through the origin (within the error of measurements). In other words, the relation $\Delta W \propto \xi^2$ holds. The data on the $^{70/76}$Ge (50%/50%) composition, however, fall outside this straight line. The experimental ΔW values are substantially lower than those predicted theoretically. A similar situation arises when the spectra of inelastic neutron scattering are analyzed [28]. It may well be that the deviations of ΔW for $^{70/76}$Ge from the linear dependence are caused by local ordering of heavy and light isotopes.

We also estimated ΔW for natural silicon ($\xi^2 = 2.02 \times 10^{-4}$). At $T = 300$K, the same models, [20, 21] and [19], and approximations were used to solve the kinetic equation. This gave ΔW equal to 0.09 and 0.071 W^{-1} cm K. As mentioned, the experimental ΔW value is $\Delta W \approx 0.142$ W^{-1} cm K [7, 21], which amounts to about 60% of the total thermal resistance. (According to [21], $\Delta W/W \approx 0.1$.) As distinguished from germanium, the experimental and theoretical ΔW values for silicon were substantially different.

5. CONCLUSION

To summarize, we used the theory of representations to suggest a method for determining symmetrized linearly independent momenta present in the problem of lattice thermal conductivity. Momenta for crystals with cubic, hexagonal, and tetragonal structures were determined.

Next, we estimated the influence of isotopic disorder on thermal resistance of natural germanium and silicon. Microscopic dynamic interatomic interaction models capable of describing both harmonic and anharmonic contributions were used. The kinetic equation was solved in the multimomentum approximations. According to the calculations, the ΔW contribution of isotopic scattering to the total thermal resistance was about 15–20% for natural germanium at $T = 300$ K, in close agreement with the experimental data. The corresponding theoretical contribution for silicon did not exceed 35%, whereas according to the Stuttgart team data, this contribution was substantially larger, $\Delta W/W =$ 60%. Nor was the theoretical result in agreement with the data obtained in Nizhni Novgorod. Clearly, the experimental values for ^{28}Si should be refined and brought in consistency with each other.

Note that taking into account the fine structure of the nonequilibrium phonon distribution function for germanium results in severalfold changes in ΔW. As distinguished from the isotropic space where $\Phi_a^{(S)} \propto q_a$, agreement between theory and experiment was quantitative rather than qualitative. Calculations in the multimomentum approximation improve agreement with the experimental data.

ACKNOWLEDGMENTS

The author thanks L.A. Maksimov for valuable comments. This work was financially supported by the Russian Foundation for Basic Research (project no. 01-02-16508).

REFERENCES

1. T. R. Anthony, W. F. Banholzer, J. F. Fleischer, *et al.*, Phys. Rev. B **42**, 1104 (1990).
2. I. W. Bray and T. R. Anthony, Z. Phys. B **84**, 3764 (1991).
3. Lanhua Wei, P. K. Kuo, R. L. Tomas, *et al.*, Phys. Rev. Lett. **70**, 3764 (1993).
4. Lanhua Wei, P. K. Kuo, R. L. Tomas, *et al.*, J. Appl. Phys. **79**, 3764 (1996).
5. V. I. Ozhogin, A. V. Inyushkin, A. N. Toldenkov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **63**, 463 (1996) [JETP Lett. **63**, 490 (1996)].
6. M. A. Asen-Palmer, N. Bartcovsky, E. Gmelin, *et al.*, Phys. Rev. B **56**, 9431 (1997).
7. W. S. Capinski, H. J. Maris, E. Bauser, *et al.*, Appl. Phys. Lett. **71**, 2109 (1997).
8. T. Ruf, R. W. Henn, M. A. Asen-Palmer, *et al.*, Solid State Commun. **115**, 243 (2000).

9. W. S. Capinski, H. J. Maris, and S. Tamura, Phys. Rev. B **59**, 10 105 (1999).
10. A. V. Inyushkin, private communication based on report of A.V. Gusev on conference on super-pure materials, Nizhni Novgorod, 2001.
11. J. M. Ziman, *Electrons and Phonons* (Clarendon Press, Oxford, 1960; Inostrannaya Literatura, Moscow, 1962).
12. B. M. Mogilevskiĭ and A. F. Chudnovskiĭ, *Thermal Conductivity of Semiconductors* (Nauka, Moscow, 1972).
13. J. Callaway, Phys. Rev. **113**, 1046 (1959).
14. R. Berman, *Thermal Conduction in Solids* (Clarendon Press, Oxford, 1976; Mir, Moscow, 1979).
15. C. Herring, Phys. Rev. **95**, 954 (1954).
16. G. P. Srivastata and G. S. Verma, Can. J. Phys. **51**, 223 (1973).
17. A. P. Zhernov, Fiz. Tverd. Tela (St. Petersburg) **40**, 1185 (1999) [Phys. Solid State **40**, 1079 (1999)].
18. R. A. Hamilton and H. J. E. Parrott, Phys. Rev. **178**, 1284 (1969).
19. M. Omini and A. Sparavigna, Nuovo Cimento **19**, 1537 (1996).
20. W. Weber, Phys. Rev. B **15**, 4789 (1977).
21. R. Eriget and I. Herman, Phys. Rev. B **53**, 7775 (1996).
22. P. G. Klemens, Proc. Phys. Soc. London, Sect. A **268**, 1113 (1955).
23. M. I. Petrashen' and E. D. Trifonov, *Applications of Group Theory in Quantum Mechanics* (Butterworths, London, 1969; Nauka, Moscow, 1970).
24. Yu. I. Sirotkin and M. P. Shaskolskaya, *Fundamentals of Crystal Physics* (Nauka, Moscow, 1979; Mir, Moscow, 1982).
25. D. D. Betts, A. B. Bhatia, and G. K. Horton, Phys. Rev. **104**, 43 (1956).
26. A. P. Zhernov, Zh. Éksp. Teor. Fiz. **114**, 654 (1998) [JETP **87**, 357 (1998)].
27. A. P. Zhernov, Fiz. Nizk. Temp. **26**, 1226 (2000) [Low Temp. Phys. **26**, 908 (2000)].
28. A. Gobel, D. T. Wang, M. Cardona, *et al.*, Phys. Rev. B **58**, 10 510 (1998).

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 93, No. 5, 2001, pp. 1082–1090.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 5, 2001, pp. 1246–1255.

SOLIDS
Electronic Properties

Giant Electric Conductivity at the CuO–Cu Interface: HTSC-Like Temperature Variations

V. V. Osipov*, I. V. Kochev, and S. V. Naumov **

Institute of Metal Physics, Ural Division, Russian Academy of Sciences, Yekaterinburg, 620219 Russia
**e-mail: osipov@imp.uran.ru*
***e-mail: viglin@imp.uran.ru*

Received January 16, 2001

Abstract—The temperature dependence of electric conductivity and current–voltage characteristics were studied in CuO single crystals with Cu films deposited onto natural faces by thermal evaporation in vacuum or by electrolysis. After electric (resistive) or thermal annealing of the samples, the conductivity of Cu films in this system significantly increases (by a factor of up to 1.5×10^5 and above) as compared to that of the control Cu films on a glass-ceramic substrate. The effect is attributed to an interfacial layer formed between CuO and Cu, the high conductivity mechanism in which is unclear. It is suggested that the giant electric conductivity and its HTSC-like temperature dependence, as well as nonlinear current–voltage characteristics of the samples can be due to the formation of superconducting regions with the critical temperatures significantly higher than 400 K. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The problem of obtaining high-temperature superconductors (HTSCs) with increasing critical temperatures (T_c) is still among important directions of research in modern physics. Previously, various researchers reported on the HTSC-like behavior in CuO-based compounds, which was manifested by a sharp drop in resistance with decreasing temperature. Azzoni *et al.* [1] observed a decrease in the resistance of polycrystalline CuO_{1-x} samples in the temperature interval from 180 to 220 K. Schönberger *et al.* [2] reported on the HTSC-like drop of resistance in copper-rich Y–Ba–Cu–O films in the same temperature interval. One of the authors also previously observed various HTSC-like anomalies in the electrical, magnetic, and some other properties of polycrystalline CuO_{1-x} in a range of temperatures from 200 to 300 K [3]. However, a distinctive feature of all the aforementioned observations was a significant time instability and poor reproducibility of the results of measurements for various samples. In addition, the effects disappeared after several cycles of measurements with the temperature varied in the aforementioned interval. These phenomena are sometimes even referred to as "nonreproducible superconductivity" effects.

On the other hand, a series of investigations based on the results reported by Mitin *et al.* in [4] showed evidence of the well-reproducible effects related to certain features in the behavior of $R(T, H, I)$ and the derivatives dR/dI and dR/dH. Using the measurements of these dependences, evolution of the electron properties was followed in a series of cuprates $LnBa_2Cu_3O_{6+\delta}$ (Ln = La, Nd, Gd, Dy, Lu) on the passage from a dielectric state to the percolation superconductivity threshold with increasing δ. Upon an analysis of the features observed for $R(T, H, I)$ and the derivatives dR/dI and dR/dH, it was suggested [4] that the system studied features the formation of localized microfragments characterized by local superconductivity with a broad distribution of the local critical temperatures (up to 160–180 K). The estimates even showed that the local superconductivity may be characterized by critical temperatures as high as 400–800 K. The concepts formulated in [4] were developed in [5], where clearer scenarios were outlined involving the formation of heat-resistant filamentary superconducting regions with T_c = 800–1000 K in ferromagnetic ordered stripes.

In recent years, some carbon-based materials were reported to exhibit anomalous magnetization effects which can also be explained by manifestations of the local high-temperature superconductivity. Tsebro *et al.* [6] observed nondecaying currents induced by a magnetic field in the fragments of cathode carbon deposits with a high content of multilayer nanotubes, where the magnetic flux trapping took place similar to that in a multiply connected superconducting stricture. At liquid helium temperatures, no decrease in the trapped magnetic flux was detected over a time period of 20 h. At an intermediate (30 K) and room temperature, the trapped magnetic flux decayed with a characteristic time of 150 and 15 h, respectively.

Recently, Kopelevich *et al.* [7] reported on the simultaneous observation of hysteresis loops characteristic of both superconductors and ferromagnets (for various magnetic field orientations relative to the characteristic growth directions) in highly ordered pyrolytic graphite at temperatures both below and above room

temperature. The behavior of the diamagnetic signal was much like the hysteresis typical of the HTSC cuprates, and it was concluded that the results observed in the entire temperature range studied (up to 400 K) are most probably explained by local (interfacial) superconductivity, in particular, at the grain boundaries. We may note that the total relative volume of a superconducting phase in the materials studied in [6, 7] was small and could not lead to the percolation superconductivity or hyperconductivity.

The previous communication [8] reported on the first observation of a giant electric conductivity in the interface between CuO single crystals and Cu films deposited by thermal evaporation in vacuum. The giant electric conductivity effect is probably explained by the formation of superconducting regions at the interface, which shunt the metal film. The experimental results could only be interpreted assuming that the HTSC regions are characterized by the critical temperature T_c significantly above 400 K.

Below we report on the results obtained in continuation of the study of electric conductivity in the samples prepared by depositing Cu films onto natural mirror faces of CuO single crystals either electrolytically or by thermal evaporation in vacuum.

2. SAMPLE PREPARATION

We have prepared several series of samples with the CuO–Cu interfaces. The size of CuO single crystals was $(3–7) \times (0.5– 1.5) \times (0.5–1.5)$ mm^3. The as-grown single crystals were washed in ethanol and then in distilled water. Prior to thermal deposition, the crystal surface was etched with an ion beam. The thickness of Cu films was 4000 Å in the first and second series and 1000 Å in the third series. Analogous control Cu films were deposited onto sitall substrates (glass ceramics of the pyracetam type).

In the first and second series of samples, the metal was deposited in vacuum onto substrates heated to a temperature of T_d = 200 and 350°C, respectively, and the deposition time τ_d was several tens of seconds in both cases. In the third series, the substrate temperature was also T_d = 200°C, but the total deposition and annealing time was about 30 min. Finally, the current and potential indium contacts were attached to the samples by ultrasonic soldering. Since the melting temperature of indium is 426 K, the subsequent measurements were performed in the temperature range from 73 to 420 K. The flow cryostat, employing liquid nitrogen boiling at a reduced pressure allowed the temperature range to be extended below the boiling point of nitrogen at normal atmospheric pressure.

The fourth series of samples was prepared by electrolytically depositing copper from an aqueous $CuSO_4$ solution. The current density during this process was about 0.1 A/cm^2. Lower current densities did not provide for the continuous metal film formation on the entire sample surface, which can be explained by electrical inhomogeneity of the CuO single crystals and/or by variation of the electrical properties on the crystal surface. At a current density above 0.1 A/cm^2, the metal coated the entire single crystal face area exposed, but the deposit possessed a granular structure comprising closely packed grains of approximately spherical shape with a diameter of 0.01 mm and below. The "electrical" thickness of the deposit, estimated using the measured resistance values and assuming a plane-parallel geometry, was within 0.4–4 μm.

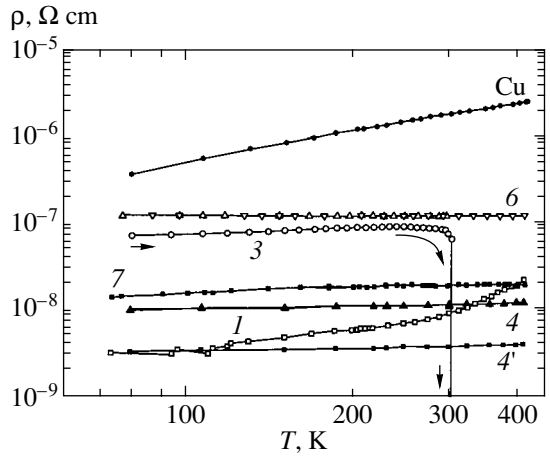


Fig. 1. The plots of ρ(*T*) for a control copper film on sitall (curve denoted by Cu) and for CuO–Cu samples nos. 1, 3, 4, 4' of the first series and 6, 7 of the second series (figures at the curves indicate the sample numbers).

3. ELECTRICAL MEASUREMENTS AND RESULTS

Resistivities of the control Cu films prepared by thermal deposition on sitall substrates for the first three series of samples were 20–25% greater than the reference value ($\rho = 1.55 \times 10^{-6}$ Ω cm) for pure copper. Figure 1 shows a dependence ρ(*T*) typical of such control Cu films. In the first series of CuO–Cu samples (with the Cu films deposited at T_d = 200°C for $\tau_d \approx 10$ s), the plots of ρ(*T*) in all cases (except for that referring to CuO–Cu sample no. 1) were quantitatively and qualitatively similar to the ρ(*T*) curves for Cu films on the ceramic substrate in the entire temperature range from 73 to 420 K. This result was quite expectable, since CuO is in fact an insulator in comparison with Cu. Indeed, the resistivity of a CuO single crystal amounts to $\rho_{300\,K} \approx 10^2$ Ω cm at 300 K, increases to $\rho_{100\,K} \approx 10^{10}$ Ω cm at 100 K, and becomes infinitely large (cannot be measured by conventional method) at 77 K [9]. As can be seen from the ρ(*T*) curve for CuO–Cu sample no. 1 depicted in Fig. 1, the room-temperature resistivity of this sample is lower by a factor of 199 as compared to the value for a control film. Moreover, the

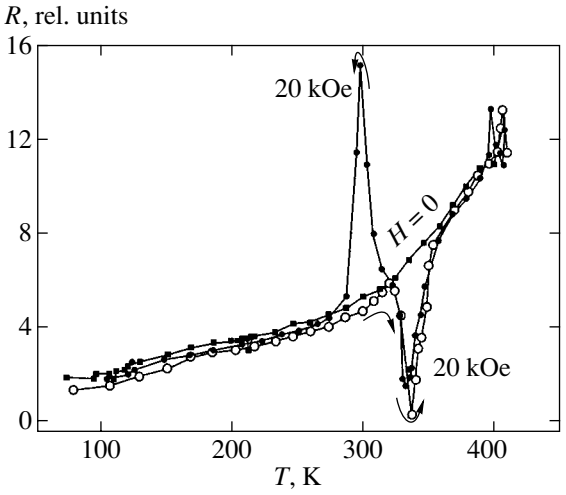


Fig. 2. The $R(T)$ curves for CuO–Cu sample no. 1 measured on heating and cooling with (H = 20 kOe) and without (H = 0) applied magnetic field.

shape of the $\rho(T)$ curve is qualitatively different from that of the control film.

All samples in the first series (except for no. 1) were subjected to electric (resistive) annealing, whereby a pulsed current (with an amplitude of up to a few amperes and a pulse duration of several tens of microseconds) or a constant current (up to several hundred milliamperes for 1–3 s) was passed via the current contacts. The annealing was performed in a nitrogen atmosphere in a cryostat maintained at room temperature. Judging by the $\rho(T)$ curves, the Cu films were heated up to approximately 300–400°C. The CuO single crystal, as well as the massive current and potential electrodes, were heated to much lower temperatures because of their considerable thermal inertia. Anyhow, the indium contacts were not melted. It should be noted that the current contacts (transverse strips across the whole film surface) were more massive than the potential contacts. The latter contacts were soldered either as spots or as thin strips across the sample.

In the course of annealing, all films showed a gradual decrease in resistance, which dropped several times or several dozen times after each annealing step. In attempts at decreasing the resistance of Cu films to a minimum possible level, most of the annealed CuO–Cu samples burned. Apparently, copper could be either evaporated as a result of overheating or oxidized to CuO or Cu_2O. In such cases, the voltage between potential contacts increased to a level corresponding to resistivity of the CuO single crystal measured.

The most interesting result was obtained for CuO–Cu sample no. 3. Before annealing, the voltage between potential electrodes on this sample was 15 mV at room temperature and a probing current of 50 mA. After the first three steps of annealing, the sample resistance decreased sequentially by a factor of 7, 2, and 1.4. As a result, the $\rho(T)$ curve for this sample was no longer analogous to that for Cu. Then, during the $\rho(T)$ measurement on heating from 77 K, the resistance of sample no. 3 smoothly (within a few seconds) dropped down to the instrumental zero level (Fig. 1, curve *3*). In this state, switching on and off the probing current (50 mA) did not affect the voltage between the potential contacts at a maximum voltmeter sensitivity of 0.1 μV. Eventually, the average resistivity of this film dropped by a factor exceeding 1.5×10^5 as compared to the initial value. Attempts to measure the critical current of sample no. 3 led to burning of one of the current contacts and a part of the Cu film at this contact.

Then we soldered a new indium contact between potential strips and the remaining current contact and repeated the four-point-probe $\rho(T)$ measurements for the intact sample area. The results of these measurements are represented by curve *4* in Fig. 1. The room-temperature resistivity decreased by a factor of approximately 160 as compared to that of the control Cu film, while a relative change in ρ in the temperature interval from 73 to 421 K amounted to about 25% (instead of 7.44 in the control). A difference between the ratios $\rho_{Cu}/\rho_3 > 1.5 \times 10^5$ and $\rho_{Cu}/\rho_4 \approx 160$ determined at T = 293 K (for curves *3* and *4* in Fig. 1, respectively) may be evidence of a partial failure of the highly conducting layer as a result of uncontrolled (above 400°C) heating of the sample. After several cycles of heating in the temperature interval from 73 to 420 K used for the $\rho(T)$ measurements, the ρ_{Cu}/ρ_4 ratio at T = 293 K additionally decreased by a factor of three. The plot of $\rho(T)$ for this case is given by curve *4'* in Fig. 1. It should be noted that all samples exhibited a tendency (especially pronounced at T = 373–420 K) to a further decrease in resistance. The resistance additionally decreased during storage for several months with periodic measurements.

All samples in the second series, where the films were deposited at T_d = 350°C for $\tau_d \approx 10$ s, initially possessed a lower resistivity as compared to that of the control films (Fig. 1, curves *6* and *7*). The electric annealing could also produce an additional, albeit small (severalfold), decrease in the sample resistance.

In some samples, the resistance was measured by both four- and two-point-probe techniques. In the latter case, additional potential contacts were soldered to the current contacts in order to eliminate the influence of connecting wires. The results of measurements using two methods coincided.

In all annealed samples of the first and second series except no. 1, an external magnetic field with a strength of $H \le 20$ kOe produced no effect on the electric conductivity. In sample no. 1, the conductivity was affected the magnetic field. However, this behavior was not described by a single-valued function of T and H and depended on the thermomagnetic prehistory. Figure 2 shows the plots of $R(T)$ for sample no. 1 measured at H = 0 and 20 kOe. As can be seen, there is a significant

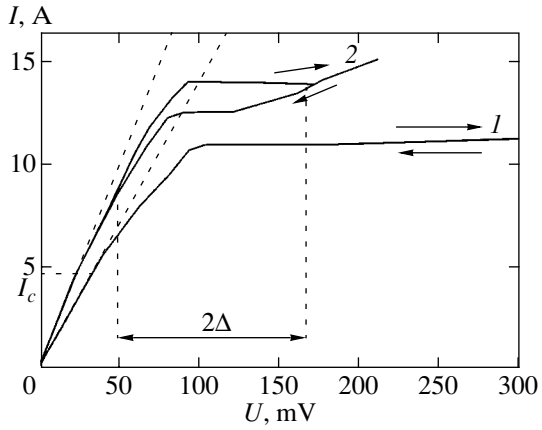


Fig. 3. The current–voltage characteristics of CuO–Cu sample no. 6 measured at $T = 300$ K (*1*) before and (*2*) after electric pulse annealing.

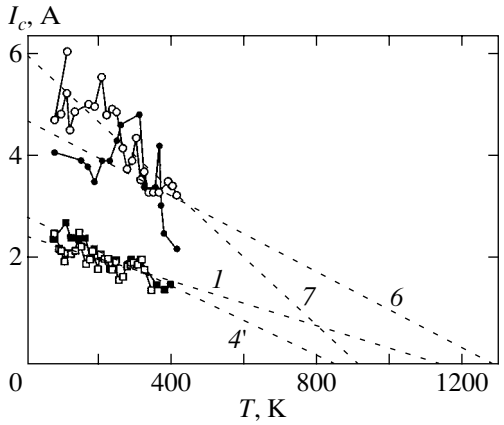


Fig. 4. The $I_c(T)$ curves for CuO–Cu samples nos. 1, 4', 6, and 7 (indicated by the same numbers). Dashed lines show extrapolation to the temperature axis.

difference between the $R(T)$ curves measured on heating and cooling the sample in the magnetic field. On heating in the field, the resistance at 340 K was 50 times lower than that in the absence of the field (or 10^4 lower than the resistance of a control Cu film on sitall). Note that the $R(T)$ curves exhibited no peaks if the temperature was changed with the field periodically switched on and off; in this case, the range of the field-induced resistance variation was narrower by several dozens of times. Therefore, the shape of the $R(T)$ curve depends on the thermomagnetic prehistory of a given sample. In order to check for the reliability of data presented in Fig. 2, we repeated the measurements several times. The results were completely reproducible.

The current–voltage (I–U) characteristics were measured using both a pulse technique (current generator mode) and a dc technique based on the conventional voltammeter measurements at small currents and voltages. The current pulses were produced by a thyristor generator, whereby one thyristor connected a storage capacitor to a sample via a low-ohmic load and another thyristor shunted the sample after a certain time period, determining the pulse duration. In our experiments, the pulse duration was varied from 20 to 200 μs and the current pulse amplitude reached 100 A and above. If necessary, a pulse taken from the potential contacts was amplified by a Unipan-232B broadband amplifier. The upper boundary of the amplification band was selected so as to suppress microsecond spikes and overshoots at the pulse fronts related to the current switching. These spikes corresponded to rapid changes in the magnetic field surrounding the sample ($H \propto I$) and, hence, induced electromagnetic signals between the potential contacts ($U \propto dI/dt$). These parasitic signals induced in the potential circuit were different in various samples and depended, in particular, on the mutual arrangement of the current and potential wires.

Upon detecting the potential and current pulses, synchronized with the aid of a two-channel lock-in gated device, the I–U curves were recorded on a two-coordinate recorder. Features of the current–voltage characteristics of sample no. 1 were described in detail elsewhere [8]. The I–U curves measured by a dc technique were linear within several tens of microvolts.

Figure 3 shows the current–voltage characteristics of CuO–Cu sample no. 6 from the second series. As can be seen, the curve deviates from the straight line (corresponding to the Ohm law) on reaching the critical current. Then the deviation from linearity increases and the I–U curve becomes a line almost parallel to the voltage axis (Fig. 3, curve *1*). It should be noted that such horizontal regions in the I–U characteristics were observed for all samples. For sufficiently small pulsed voltages applied to the samples (up to ten times the values indicated in Fig. 3), the current–voltage characteristics were reversible. A greater pulse amplitude led to irreversible annealing of the sample, after which the I–U curve exhibited a hysteresis behavior (Fig. 3, curve *2*) resembling the Josephson characteristic of two-particle tunneling in an individual tunnel junction. Based on this analogy, we separated a region with a width of 2Δ, Δ representing an effective bandgap. Similar features in the I–U curves were observed for other samples.

Figure 4 shows the plots of critical current I_c versus temperature for samples nos. 1, 4', 6, and 7. A linear extrapolation of these data to intersection with the temperature axis yields $T_c \approx 800$–1100 K. It is interesting to note that a linear extrapolation of the $\rho(T)$ curve plotted in the logarithmic coordinates for CuO–Cu sample no. 1 intersects with the analogous plot for the control Cu film on sitall at approximately the same temperatures.

In the CuO–Cu samples of the third series (T_d = 200°C, thickness 1000 Å), the annealing in vacuum during $\tau_a \approx 1800$ s led to a complete or partial oxidation (dissolution) of the Cu film on CuO. For example, the Cu film could react with CuO so as to form Cu_2O or some other copper–oxygen compound. The control

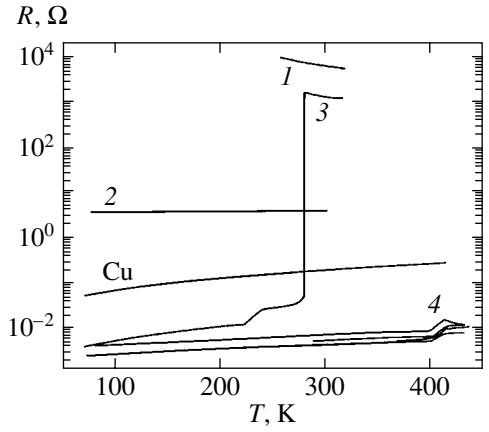


Fig. 5. Evolution of the temperature dependence of resistance for CuO–Cu sample no. 8 upon sequential electric pulse annealings (*1–4*). The curve denoted by Cu refers to a control copper film on sitall.

films on sitall treated under the same conditions remained unchanged. This experiment indicated that the characteristic length of the interpenetration of copper and oxygen atoms at the CuO–Cu interface for a diffusion time of 1800 s reaches approximately 1000 Å. In two of the five samples of this series subjected to electric pulse annealing, significant additional changes were observed in the electric conductivity and in the temperature dependence of resistance.

Figure 5 shows the evolution of the resistance of a CuO–Cu sample exhibiting the most pronounced changes in the electric conductivity upon electric pulse annealing. Before annealing, the $R(T)$ curve corresponded to that for a CuO single crystal with an activation energy of 0.3 eV (Fig. 5, curve *1*). After the first annealing, the sample was characterized by $R(T) = \text{const}$ (curve *2*), which was similar to the behavior of samples in the first and second series, except for the resistance being greater than that of a control Cu film with the same dimensions on sitall. Subsequent annealing led to an HTSC-like behavior of the $R(T)$ curve: a sharp drop by a factor of approximately 10^5 (Fig. 5, curve *3*) at a certain temperature, after which the sample resistance was 25 times smaller than that of the control Cu film.

Figure 6 shows a sequence of such changes in the resistance of sample no. 8 subjected to cyclic heating and cooling. As can be seen, the HTSC-like drops in the resistance take place in the temperature interval from 250 to 255 K. Subsequent electric pulse annealings led to an additional decrease in the sample resistance (as depicted by a bundle of curves *4* in Fig. 5). As a result, the room-temperature resistance of this sample decreased by a factor of about 50 as compared to the value for a control Cu film on sitall. The HTSC-like drop (approximately twofold) in resistance shifted to $T \approx 400$ K. The thermal cycling in the vicinity of $T \approx 400$ K led to a further decrease in the resistance of sample no. 8, which is illustrated in Fig. 7 for the series of curves *4* from Fig. 5.

Figure 8 demonstrates the effect of a magnetic field on the HTSC-like behavior of the resistance of sample no. 8 at $T \approx 400$ K. Since the R value at $T \approx 400$ K kept decreasing with time, the magnetic field effect was

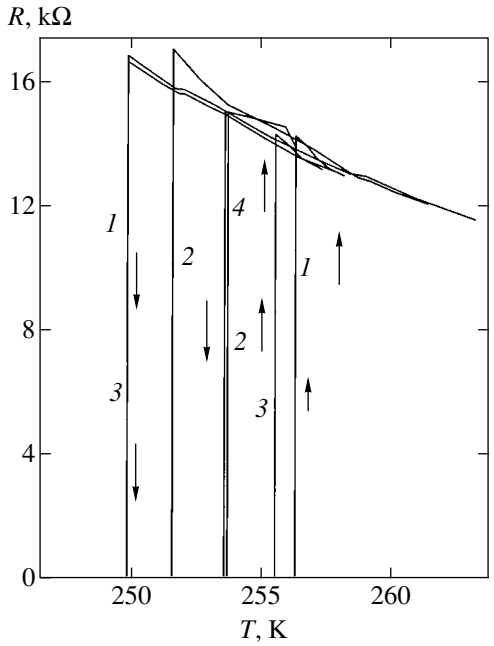


Fig. 6. A sequence of the HTSC-like transitions in CuO–Cu sample no. 8 at $T = 250$–255 K (corresponding to curve *3* in Fig. 5). Figures and arrows at the curves indicate the order of measurements and the direction of temperature variation.

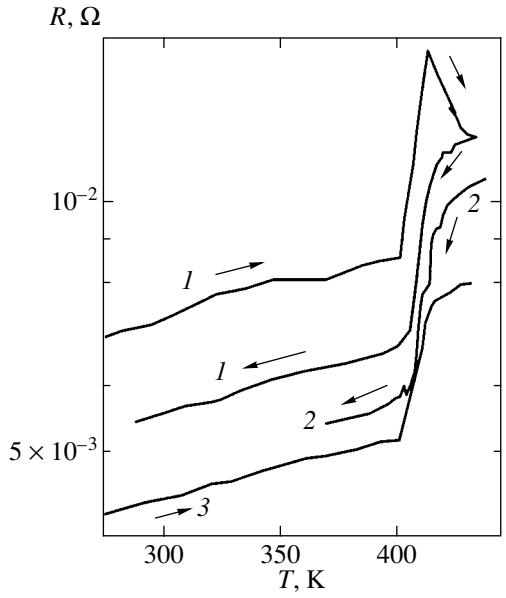


Fig. 7. A sequence of the HTSC-like transitions in CuO–Cu sample no. 8 at $T = 400$ K (corresponding to curves *4* in Fig. 5). Figures and arrows at the curves indicate the order of measurements and the direction of temperature variation.

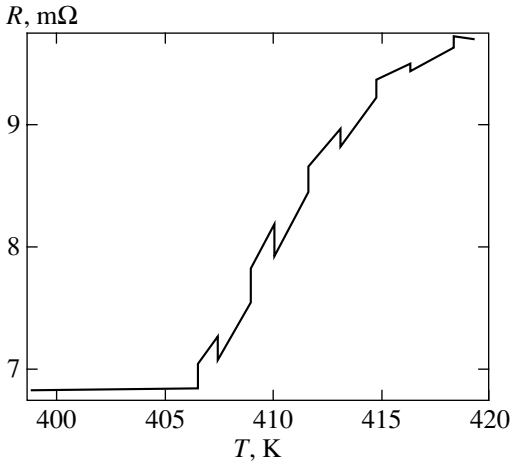


Fig. 8. The effect of a periodic magnetic field switching on (H = 20 kOe) and off upon the HTSC-like transitions in CuO–Cu sample no. 8 at T = 400 K.

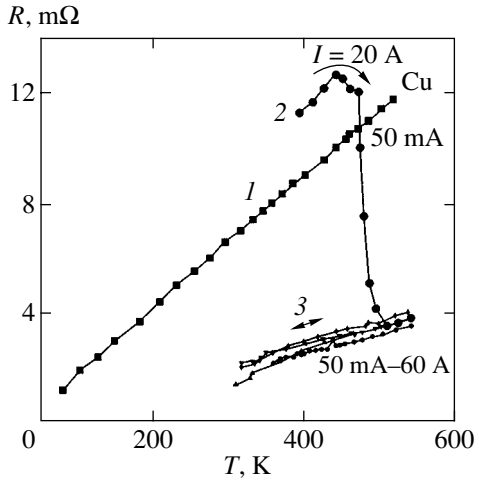


Fig. 9. Evolution of the temperature dependence of resistance of a CuO–Cu sample no. 1 with electrolytically deposited copper film under simultaneous action of temperature and high-power electric pulses.

studied by periodically switching the field as depicted in Fig. 8. As can be seen, application of the magnetic field (H = 20 kOe) decreases the HTSC-like transition temperature by approximately 0.7 K. Another electric pulse treatment of sample no. 8 restored the shape of the $R(T)$ curve characteristic of the CuO single crystal, which became exactly like curve *1* in Fig. 5. Subsequent annealings led to very small deviations from this curve.

It should be mentioned that a strong influence of current pulses (with sufficiently high repetition rate and duration) on the electrical properties of a polycrystalline HTSC of the $DyBa_2Cu_3O_{7-\delta}$ type, including the temperature dependence of resistance and the temperature of the superconducting transition, was reported by Mitin *et al.* [10]. In particular, the superconducting transition temperature could vary from 0 to ~90 K; these results were explained by redistribution of oxygen ions over phase-inhomogeneous material.

As noted above, the electrolytically deposited films possessed a granular structure comprising closely packed grains of approximately spherical shape with a diameter of 0.01 mm and below. Therefore, the system features no continuous CuO–Cu interface over the entire CuO single crystal surface. A highly-conducting interface may form only between the lower boundary of a Cu grain and the single crystal. Possessing this structure, a Cu film can be only partly shunted. Thus, the probing current used in the conductivity measurements passes through the interface and then through the Cu grains and intergranular contacts connected in series with the interface. Therefore, we may only expect a decrease in resistance of the granular copper film. However, the temperature dependence of the electric resistance of such films is analogous to that of copper. In order to provide for the measurements at higher temperatures, the contacts to electrolytically deposited copper films were soldered with a POS-30 alloy possessing a melting temperature of 530 K.

Figure 9 (curve *1*) shows a temperature dependence of the resistance of an electrolytically deposited CuO–Cu sample no. 1 measured in a helium atmosphere. The metal film thickness was about 4 μm. As can be seen, this resistance versus temperature curve measured using a small probing current (50 mA) exhibits a shape typical of copper. Measurement of the same dependence at $T > 430$ K using a pulsed 20-A current (Fig. 9, curve *2*) was accompanied by a fourfold drop in the resistance because of the combined effect of thermal and electric pulse annealing. This would correspond to a 40-fold drop in the resistance of the 4000-Å-thick films considered above. Upon this decrease, the subsequent temperature dependences of resistance are still characteristic of copper for all values of the probing current (Fig. 9, curves *3*).

Figure 10 shows the evolution of the temperature dependence of resistance of a CuO–Cu sample no. 2 with electrolytically deposited copper film in the course of sequential electric pulse annealings. Before annealing, the $R(T)$ curve is analogous to that for a Cu film with an "electrical" thickness of 0.4 μm (curve denoted by Cu). Upon the first pulse annealing stage, the sample heated to 500 K exhibited an HTSC-like drop in the resistance by 3.5 orders of magnitude at 350–400 K (Fig. 10, curves *1*). After the second annealing and heating to 500 K, the HTSC-like resistance drops are observed at T = 400–500 K (Fig. 10, curves *2*). The third annealing and heating to 500 K lead to a further decrease in the resistance by a factor of about 50 (curve *3*) relative to the initial sample resistance. Here, we may suggest that the HTSC-like transition has shifted toward still higher temperatures. It should be noted that sample no. 2 subjected to a combined action of high temperature (500 K) and electric pulse annealing exhib-

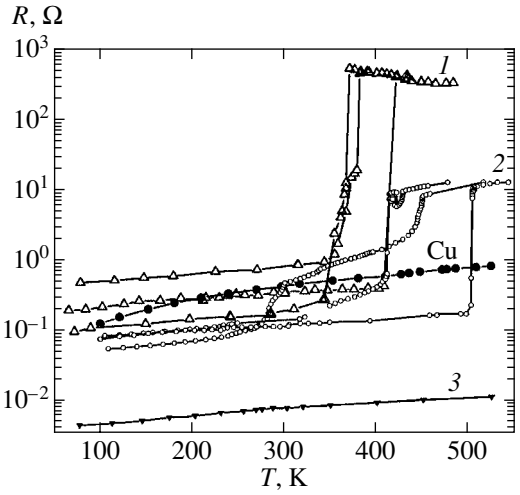


Fig. 10. Evolution of the temperature dependence of resistance of a CuO–Cu sample no. 2 with electrolytically deposited copper film in the course of sequential electric pulse annealings.

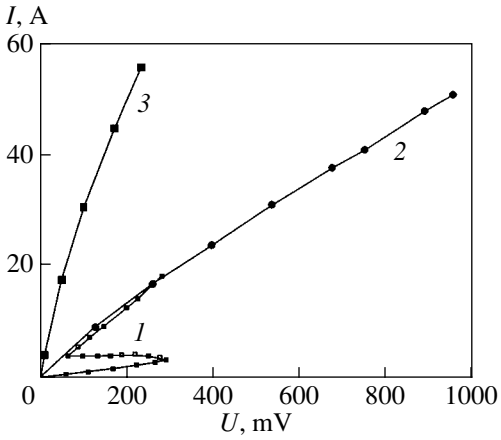


Fig. 11. The S-shaped current–voltage characteristic formation (curve *1*) followed by a decrease in resistance (curve *2*) of the electrolytically prepared CuO–Cu sample no. 3 under the action of high-power current pulses. For curve *1*, the resistance is 101 mΩ before and 15 mΩ after switching. In the same sample treated by subsequent high-power current pulses, the resistance dropped further to 3 mΩ (curve *3*).

ited a change in appearance of the copper film, which acquired a dark color showing evidence of the formation of some compound of the Cu_xO_y type (islands of metal copper were also distinguishable). We are planning to study various CuO–Cu interfaces and identify the new compounds by methods of X-ray diffraction and Auger electron spectroscopy.

Figure 11 presents data on the electric pulse annealing (in a helium atmosphere at room temperature) of a Cu film with a thickness of about 1 μm electrolytically deposited onto a CuO single crystal. As can be seen, the system acquires an S-shaped current–voltage characteristic (curve *1*) and then the resistance drops by a factor of more than six (curve *2*). Additional annealing by current pulses with an amplitude of up to 120 A leads to a further decrease in resistance (curve *3*), on the total by more than 30 times relative to the initial value. For a film thickness reduced to 4000 Å, this corresponds to a drop in the sample resistance by a factor exceeding 75.

We attempted to prepare highly conducting interfaces by oxidizing Cu films in air to CuO or by treating polycrystalline CuO samples. However, both thermal and electric pulse treatments of the CuO–Cu interfaces could decrease the sample resistance only by a few percent because of a significantly granular structure of the material and low quality of the grain surfaces. We may expect that preparing samples with a better quality of the grain surface and smaller grain spacing (i.e., with the grain surface quality approaching that of single crystals) will allow the effect to be significantly increased.

4. DISCUSSION OF RESULTS

Let us first consider the results of investigation of the CuO–Cu interfaces prepared by thermal deposition in vacuum (i.e., samples of the first and second series) in the temperature range from 73 to 420 K. Taking into account that these samples did not change in appearance upon the electric pulse annealing, we may suggest that a considerable growth in the electric conductivity of the CuO–Cu interface is due to the formation of a layer possessing high average conductivity, which shunts the copper film.

We may evaluate the thickness of this layer from the following experimental data. In some cases, a 1000-Å-thick Cu film deposited onto CuO single crystal and annealed for 30 min at 200°C (in the third series of samples) completely dissolved, for example, by reacting with CuO (to form CuO_2 or some other compounds of the Cu_xO_y type); on some other samples of this kind, the film was only partly retained. At the same time, the control films deposited onto sitall and treated under the same conditions remained unchanged. From these results, we may infer that the characteristic length of the interpenetration of copper and oxygen atoms at the CuO–Cu interface for a diffusion time of 1800 s reaches approximately 1000 Å. Since the duration of annealing using a constant current was 1–3 s and the time of annealing by the electric pulses was even shorter, estimates based on the diffusion equation indicate that the proposed layer thickness after the first annealing amounts to 25–40 Å. Therefore, the layer thickness after several annealing stages can be on the order of 100 Å. For sample no. 3, the ratio of the electric conductivity of this layer to that of copper at $T = 293$ K exceeds 6×10^6. The analogous ratio for other samples is somewhat lower.

The experimental results can probably be explained by the appearance of an unknown copper–oxygen compound possessing a high conductivity. A tempting

explanation is based on the appearance of HTSC phase regions shunting the copper film at the CuO–Cu interface. Evidence for this hypothesis is provided by a high average conductivity of the proposed shunting layer, which must be significantly higher than the conductivity of copper. The critical temperature of such an interfacial HTSC phase must be significantly above 400 K, as indicated by the extrapolation of $I_c(T)$ curves to zero. The critical current density at the interface may be as high as 10^7–10^8 A/cm^2.

Thus, according to the experimental data, the conductivity of the interfacial layer must (i) significantly exceed the conductivity of copper, (ii) be independent of the temperature, and (iii) vary within broad limits (see Fig. 1) depending on a particular sample and the degree of annealing. At present, we cannot describe a possible interface structure containing extended HTSC fragments, the conductivity of which satisfies all these requirements. Nevertheless, some details of this structure can be specified.

As can be seen from Fig. 3 (curve *2*), the current–voltage characteristic of sample no. 6 resembles (primarily, by the presence of a characteristic hysteresis) the Josephson characteristic for two-particle tunneling in an individual tunnel junction. We may suggest that, before an electric pulse annealing, the horizontal portion of the *I–U* curve in Fig. 3 (curve *1*) reflects the sequential switching of many tunneling junctions with supercritical current. In the course of electric annealing, a large number of junctions are formed with a broad distribution with respect to the critical current I_c. As a result (Fig. 3, curve *2*), a dominating single tunneling junction with a minimum switching current is separated, the individual current–voltage characteristic of which is observed. The effective bandgap width is about 100 mV. It is interesting to note that a close value of the pseudogap was estimated from various experimental data in [11] for an undoped HTSC cuprate phase with a minimum hole concentration (below 0.05 hole per CuO_2 unit cell), in which the percolation superconductivity (indicated by a drop down to the instrumental zero in the potential difference between contacts) is absent. It is possible that unusual behavior of the $R(T)$ curve observed for unannealed sample no. 1 exposed to a magnetic field is also determined by interference processes in a system of tunnel junctions. By now, the results of measurements using a SQUID magnetometer revealed no reliable nor reproducible diamagnetic signals from the CuO–Cu interface. Observation of the Meissner effect in this situation is probably complicated by several factors. First, it is necessary to distinguish the diamagnetic signal from a thin HTSC film on the background of a signal from the bulk of antiferromagnetic CuO. Second, an anomalously large London depth of the magnetic field penetration into CuO-based HTSC materials (reaching 0.1–1 μm) may significantly exceed the thickness of the highly conducting interfacial layer, which would decrease the diamagnetic signal intensity. Third, the diamagnetic signal may also significantly decrease if the highly conducting interfacial layer would split into separate HTSC regions.

Taking into account the above notions, it would also be interesting to consider the results reported in [1–3] for the "nonreproducible superconductivity" effects. It is possible that all these cases are random manifestations of the high conductivity (or HTSC phase formation) at the CuO–Cu grain boundaries similar to that observed in our experiments. Nonreproducibility of the previous results could be explained, for example, by mechanical displacements of the grains in the course of cyclic $R(T)$ measurements leading to degradation of the HTSC properties of the interface. It should be emphasized that, in contrast to the "nonreproducible superconductivity" effects, our results are characterized by a long-term stability and even by a tendency to decrease in the interface resistance with time.

5. CONCLUSIONS

(1) We have observed the phenomenon of giant electric conductivity of the CuO–Cu interface in the temperature interval from 77 to 400–500 K and above. The interfacial layer conductivity may significantly (by more than a factor of 6×10^6) exceed the conductivity of copper. In the case of Cu films deposited onto a CuO single crystal surface in vacuum, the interfacial layer conductivity was independent of the temperature. The thickness of a highly conducting interfacial layer is estimated at about 100 Å.

(2) Possessing a giant electric conductivity, the interfacial layer is capable of carrying a current density of up to 10^8 A/cm^2. As the current density increases above this level, the interfacial layer resistance sharply grows.

(3) The conductivity of the CuO–Cu interface is significantly affected by the electrical or thermal treatment conditions.

ACKNOWLEDGMENTS

This study was supported in parts by the Federal Program "Surface Atomic Structures" (project no. 2.4.99) and by the Russian Foundation for Basic Research (R-2001 Ural project no. 01-02-96429).

REFERENCES

1. C. B. Azzoni, G. B. Paravicini, G. Samoggia, *et al.*, Z. Naturforsch. A **45**, 790 (1990).
2. R. Schönberger, H. H. Otto, B. Brunner, *et al.*, Physica C (Amsterdam) **173**, 159 (1991).
3. A. A. Samokhvalov, T. I. Arbuzova, V. V. Osipov, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **38**, 3277 (1996) [Phys. Solid State **38**, 1788 (1996)].

4. A. V. Mitin, G. N. Kuzmicheva, V. V. Murashov, *et al.*, Zh. Éksp. Teor. Fiz. **107**, 1943 (1995) [JETP **80**, 1075 (1995)].

5. A. V. Mitin, in *Proceedings of the XIII Ural International Winter School on the Physics of Semiconductors, Yekaterinburg, 1999*, p. 120.

6. V. I. Tsebro, O. E. Omel'yanovskiĭ, and A. P. Moravskiĭ, Pis'ma Zh. Éksp. Teor. Fiz. **70**, 457 (1999) [JETP Lett. **70**, 462 (1999)].

7. Y. Kopelevich, P. Esquinazi, J. H. S. Torres, and S. Moehlecke, J. Low Temp. Phys. **119**, 691 (2000).

8. V. V. Osipov and A. A. Samokhvalov, Fiz. Met. Metalloved. **89** (5), 43 (2000) [Phys. Met. Metallogr. **89**, 463 (2000)].

9. A. A. Samokhvalov, N. A. Viglin, B. A. Gizhevskiĭ, *et al.*, Zh. Éksp. Teor. Fiz. **103**, 951 (1993) [JETP **76**, 463 (1993)].

10. A. V. Mitin, N. E. Alekseevskiĭ, and E. P. Khlybov, Sverkhprovodimost: Fiz., Khim., Tekh. **5**, 290 (1992).

11. B. Batlogg and Chandra M. Varma, http://Physics-Web.org/article/world/13/2/8 (2000).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 93, No. 5, 2001, pp. 1091–1100.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 5, 2001, pp. 1256–1267.

SOLIDS
Electronic Properties

"Magnetodipole" Self-Organization of Charge Carriers in High-T_c Superconductors and the Kinetics of Phase Transition

A. V. Voronov*, V. M. Petnikova, and V. V. Shuvalov

International Laser Center, Moscow State University, Vorob'evy gory, Moscow, 119899 Russia
**e-mail: avoronoff@mail.ru*

Received April 4, 2001

Abstract—A phenomenological model describing "magnetodipole" self-organization of charge carriers (the formation of so-called stripe-structures and the energy gap in the spectrum of states) was suggested to interpret the data of nonstationary nonlinear spectroscopy of high-T_c superconductors. It was shown that, after rapidly heating a superconducting sample, the kinetics of the succeeding phase transition depended on initial temperature T. At small "overheatings" $T^* < T < T_m \approx (1.4–1.5)T^*$ (T_c and $T^* \approx T_c$ are the temperatures of the transition to the superconducting state and the formation of stripe-structures) and the optimal level of doping, the decay of stripe-structures (and of the gap in the spectrum of states) occurred at a low rate (in times above to 10^{-9} s) in spite of the virtually instantaneous disappearance of superconductivity. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The discovery of high-temperature superconductivity in copper-oxide compounds [1] has led to explosive growth of scientific activity in this domain of science. An analysis of the data obtained [2–4] shows that these compounds are ionic semiconductors and their "metallic" conductivity is caused by deviations from stoichiometry (by a certain level of structure imperfection, or "doping"). The electronic structure of these compounds is determined by a complex interaction between "well" and "badly" localized electronic states sensitive to short-range order in the arrangement of atoms. Strong anisotropy results in a quasi-two-dimensional character of the Fermi surface.

The common property of copper-oxide compounds is ordering of Cu ion spins in so-called cuprate (CuO_2) planes. In stoichiometric compounds, these ions (usually denoted by Cu2) have "holes" with spin $s = 1/2$ in the $3d$ shell. Indirect (through oxygen ions) exchange establishes long-range antiferromagnetic order at comparatively high Neel temperatures $T_N = 300–500$ K [5]. The magnetic structure and spin correlations in the metallic phase were studied in detail by neutron scattering for La_2CuO_4 and $YBa_2Cu_3O_{7-\delta}$ single crystals [6]. It was shown that, in spite of the absence of long-range order, strong spin fluctuations with broad excitation spectra were preserved. The magnetic moments of Cu2 ions in CuO_2–Y–CuO_2 bilayers were antiferromagnetically ordered, and Cu1 ions that formed linear Cu–O chains had zero magnetic moments. Diffuse magnetic scattering was used to determine the dependence of magnetic correlation length ξ_m on the mean concentration of holes $\langle n \rangle$ in the cuprate plane [7]. Magnetic inelastic neutron scattering in $YBa_2Cu_3O_{7-\delta}$ was studied to determine exchange interaction constants between two nearest Cu2 ions ($2J \approx 0.170$ eV) in bilayers ($2J_b \approx 10^{-2} \times 2J$) and Cu2–O1 chains ($2J' \approx 10^{-5} \times 2J$) [8], which proved the quasi-two-dimensional character of the dynamics of spins in $YBa_2Cu_3O_{7-\delta}$. Similar data were obtained for $La_{2-\delta}Sr_\delta CuO_4$ [9].

The nature of the energy gap in the spectrum of states is one of the key problems in understanding the phenomenon of high-temperature superconductivity. In low-T_c superconductors, this ("superconducting") gap has so-called *s*-symmetry and is formed below the T_c temperature as a result of electron–phonon interactions in the transition to the superconducting phase [9]. The experimental data on high-T_c superconductors are evidence of *d*- or mixed (*s*-*d*)-symmetry of the gap [10]; at the optimal level of doping ($\delta = \delta_{opt}$), at which T_c is maximum, this gap has *d*-symmetry [11]. What is more, the so-called pseudogap was observed in the spectrum of states of high-T_c superconductors by angle-resolved photoemission spectroscopy (ARPES) [12]. It was shown that, at $\delta = \delta_{opt}$ and $\delta > \delta_{opt}$ (overdoped materials), the energy gap disappeared (closed) at $T = T_c$ at all Fermi surface points, as in low-T_c superconductors. At $\delta < \delta_{opt}$ (underdoped compounds), the superconducting gap transformed into a *d*-symmetry pseudogap under heating, which disappeared at $T = T^* \geq T_c$. The closure of the gap occurred at $T = T_c$ at the $(\pi/4, \pi/4)$ Brillouin zone point and at $T \approx 180$ K $\gg T_c$ at the $(\pi, 0)$ point. The nature of the pseudogap and its relation to the supercon-

ducting gap still raise heated debate. Although the most convincing ARPES experiments were performed for $Ba_2Sr_2CaCu_2O_{9-\delta}$ single crystals, it is believed that the observed behavior of the gap is common to all high-T_c superconductors. Note that two energy gaps with different kinetics have recently been observed in overdoped $YBa_2Cu_3O_{7-\delta}$ at $T < T_c$ in nonlinear nonstationary spectroscopy experiments [13]. One of these gaps was interpreted in [13] as the superconducting gap, and the other, as the pseudogap.

No less interesting data on lanthanum compounds were obtained in electron, neutron, and X-ray diffraction studies [14–16]. It was shown that a periodic sequence of alternating strongly extended regions, stripes, enriched in holes and virtually devoid of holes, was formed in the cuprate layers of these compounds at $T = T^*$. The correlation length (further, merely "length") of stripes oriented along the $(\pm 1, \pm 1, 0)$ crystallographic directions decreased as $\langle n \rangle$ increased. No such stripes were observed in similar experiments with bismuth, barium, and mercury compounds. At the same time, measurements of bismuth compound heat and electric conductivities, whose temperature dependences contained kinks, gave indirect evidence of the presence of stripe-structures [16]. Some authors hold the opinion that this is evidence of the dynamic character of the corresponding stripe-structures and smallness of their correlation lengths.

Two approaches to describing stripe-structures are most popular. The so-called *t–J* model is based on the Hamiltonian that takes into account kinetic energy and correlations on neighboring nodes [17]. The two main parameters of the model are energy *t* of hole "jumps" to the nearest nodes and energy *J* of exchange interaction with these nodes. Sometimes, terms describing hole jumps to remoter nodes are included. According to the *t–J* model, a sharp zone narrowing and the corresponding increase in the effective mass occur as a result of antiferromagnetic spin correlations, which suppress hole jumps between neighboring nodes. At $J \gg t$, spatial separation of phases (stratification), that is, the formation of real regions filled by holes and virtually devoid of holes, is energetically favorable. Numerical calculations based on the *t–J* model are laborious and are usually performed for small-sized two-dimensional clusters (typically, 8×8 clusters) [18]. These calculations show that the stratification at small $\langle n \rangle$ should be observed at $J/t \geq 1$ [19], whereas, in real compounds, the stratification occurs at $J/t \approx 0.35$ [20].

The other approach uses the Peierls–Hubbard model and is based on taking into account Jahn–Teller lattice distortions [21]. The model Hamiltonian describes hole jumps over copper sites, one-node Coulomb correlations, spin dynamics, and the kinetic and potential lattice energies. Calculations by this model also prove the possibility of the stratification. Lattice distortions play a stabilizing role, and the $J/t \geq 1$ condition is softened. Simulations are however performed for still smaller (4×4) clusters because of the complexity of the problem, which sharply decreases the accuracy of calculations [22]. The kinetics of transitions with formation and decay of stripe-structures cannot be described by such simulations.

The main goal of this work was to explain the results of our experiments [23] on biharmonic picosecond probing of the spectrum of states of thin superconducting $YBa_2Cu_3O_{7-\delta}$ films preliminarily excited by a short (20–40 ps) pumping pulse. These experiments show that, in the absence of pumping at $T < T_c$, the efficiency of self-diffraction of probe pulses with frequency detuning within the width of the energy gap decreases virtually to zero. This behavior persists for at least 1 ns after rapidly heating the initially superconducting sample to $T \approx T_c + (20–30)$ K. At the same time, it is known that under these conditions, superconductivity should disappear in times of the order of 1 ps [24, 25]. To explain this paradox, we construct a phenomenological model that describes self-organization of holes in the cuprate planes of the high-T_c superconductor as a result of their "magnetodipole" [26] interaction, that is, nonlocal interaction caused by spin-wave exchange processes. In all calculations, we only take into account the nonlocal part of the interaction potential obtained within the framework of spin-wave theory [27, 28]. In contrast to the local part of the interaction potential, described within the framework of the *t–J* model by compensation of the bond breaking in the antiferromagnet, the nonlocal part of the potential has a dipole character and is due to deformations of the antiferromagnetic environment. The corresponding interaction has the character of attraction in the spin-symmetric channel and repulsion—in the spin-antisymmetric channel. It was shown that, at $J/t \gg 1$, the interaction energy approximately equaled $8t^2/J$ [27], whereas calculations for $J/t \ll 1$ gave values of the order of *J* [28]. Precisely the last limiting case will be analyzed in what follows. We will also test the constructed model using the criterion of a correct description of the known stationary properties of stripe-structures and generalize this model to the conditions typical of nonlinear nonstationary spectroscopy experiments. We will show that the time necessary for the formation and decay of stripe-structures (and of the corresponding energy gap, pseudogap) can substantially exceed the time of superconductivity disappearance.

Note that a similar description of the formation of stripe-structures related to the specific angular dependence of the interaction potential was given by Khomskiĭ and Kugel [29], who, however, used another physical mechanism of self-organization (it was assumed that angular anisotropy was a consequence of deformation crystal lattice distortions by defects) and did not consider the kinetics of stripe-structures themselves.

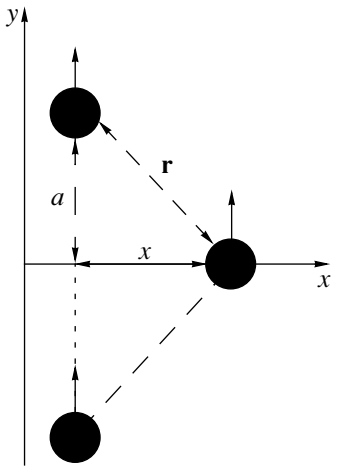


Fig. 1. Scheme of calculations of interaction energy $V_t(x)$ between a hole and its nearest neighbors. Arrows indicate the orientations of **d** vectors.

2. MODEL AND PRINCIPAL EQUATIONS

We assume that local antiferromagnetic ordering exists in copper-oxide high-T_c superconductors up to the highest possible δ values [6]. Supposing that Coulomb interaction is screened, we treat interactions of holes caused by deformation of their antiferromagnetic environments as magnetodipole interactions (see above) and describe them in the two-particle approximation by the potential energy function [27, 28].

$$V(\mathbf{r}, \alpha_1, \alpha_2) = B\cos(2\theta - \alpha_1 - \alpha_2)/r^2. \qquad (1)$$

Here, **r** is the radius vector connecting two holes, B is the constant related to the modulus of their effective dipole moments $|\mathbf{d}_{1,2}| = d$, and θ and $\alpha_{1,2}$ are the angles between the **r** and $\mathbf{d}_{1,2}$ directions and arbitrary axis y. Further, it is assumed that $\mathbf{d} \equiv 1$. At such a normalization, the B/a^2 constant in the $J/t \ll 1$ limiting case should be of the order of exchange interaction energy at a equal to the mean distance between holes [28]. From symmetry considerations, let y be directed along one of two possible crystallographic axes (stripe axes, see above). Suppose that the **d** vectors of all holes are oriented parallel to y and consider the displacement of one of the holes along axis x orthogonal to y (Fig. 1). Projecting the forces that act on this hole from its two nearest neighbors onto axis x, we obtain the potential energy of the hole in the form

$$V(x, a) = \frac{B}{a^2}\left[1 - \left(\frac{x}{a}\right)^2\right]\left[1 + \left(\frac{x}{a}\right)^2\right]^{-2}, \qquad (2)$$

where x and a are the projections of **r** onto axes x and y, respectively. Dependence (2) is shown in Fig. 2a. Next, we assume that the **d** vector can be oriented either along y or in the opposite direction and that $d(y) = d(-y)$; this is taken into account by two factors, $d(y) = \pm 1$. Let us rewrite the equation for total energy V_t (that is, the energy that takes into account forces acting on the hole under consideration from its more distant neighbors) in the form

$$V_t(x, a) = \sum_{j=1}^{\infty} d(0)d(ja)V(x, ja). \qquad (3)$$

Here, the argument of d numbers holes in the stripe. The orientations of d are partially disordered because of thermal fluctuations. If the number of stripe holes with "irregular" **d** orientations is small, the mean field approximation can be used. For this purpose, we introduce the mean local (along y) dipole moment $\langle d\rangle(y)$ in the stripe and use the assumption that $\langle d\rangle$ is a slow (on the a scale) function of y. Next, let $V(x, a)$ decrease as the distance increases faster that $\langle d\rangle$ varies. Assuming that $d(y) - \langle d\rangle(y) \ll \langle d\rangle(y)$, replacing the summation over j by the integration in y, and expanding $\langle d\rangle(y)$ into a series in the vicinity of arbitrary y, we obtain

$$V_t(x, y) \approx \frac{\pi^2}{6}\frac{B}{a_0^2}d(y) \times \left[\langle d\rangle(y) + a_e^2\frac{d^2\langle d\rangle(y)}{dy^2} + \ldots\right]\left[1 + \left(\frac{x}{a_0}\right)^2\right]^{-1}. \qquad (4)$$

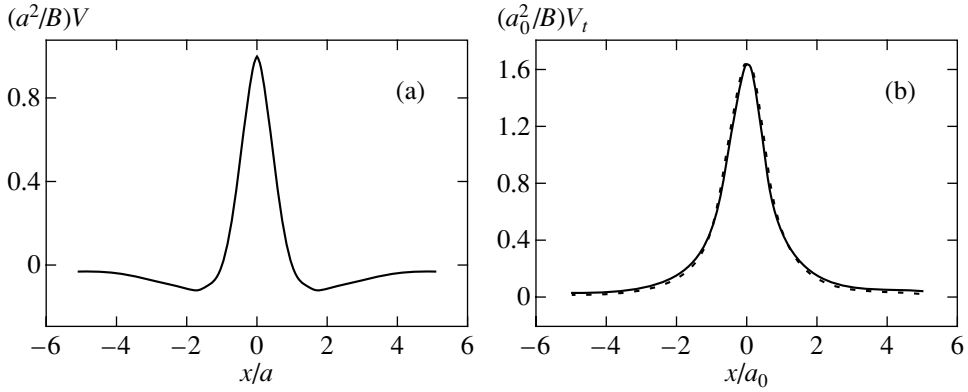


Fig. 2. (a) Potential energy $V(x, a)$ and (b) $V_t(x, a)$ as functions of hole displacements along the x axis: exact summation (solid line) and approximation (dashed line).

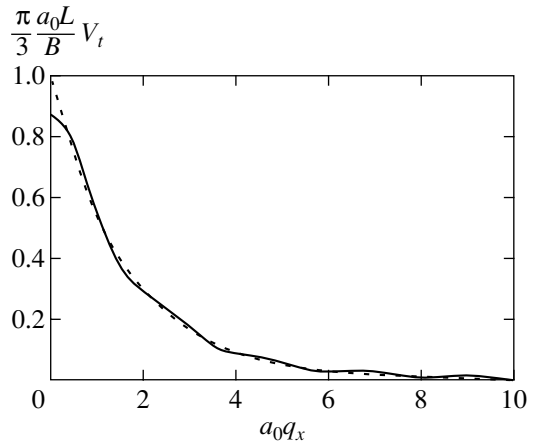


Fig. 3. Fourier transform of $V_t(q_x)$ according to exact (solid line) and approximate (dashed line) calculations at $L = 3a_0$.

Here, $a_0 = (6/\pi^2)a$ and a_e are constants of the order of the magnetic correlation length related to finiteness of the interaction radius. The $V_t(x, y)$ dependences calculated by (3) (solid line) and by approximation (4) (dashed line) are shown in Fig. 2b. It is easy to see that interaction within the stripe is well localized.

Next, we assume that the distribution of stripes over x is periodic. The L period of this distribution and the spatial frequency of its lowest harmonic, $q_{xm} = 2\pi/L$, will be treated as external problem parameters rigidly related to δ (and $\langle n\rangle$). Let us calculate the spatial distribution of potential energy $V_0(x, y)$ at an arbitrary distribution of stripes along the x axis. For this purpose, we will use approximation (4) to find the Fourier transform of the $V_t(x, y)$ function,

$$\begin{aligned} V_t(q_x, y) &= \frac{1}{L}\int_{-L/2}^{L/2} V_t(x, y)\exp(iq_x x)dx \\ &\approx \tilde{G}(y)\exp(-q_x a_0), \\ \tilde{G}(y) &= G_0 d(y)\left[\langle d\rangle(y) + a_e^2\frac{d^2\langle d\rangle(y)}{dy^2} + \ldots\right], \\ G_0 &= \frac{3}{\pi}\frac{B}{a_0 L}. \end{aligned} \tag{5}$$

Here, q_x is the spatial frequency. At small δ, the limits of the integration in (5) can be extended to infinity taking into account screening at distances of the order of several cell periods. The results of exact (solid line) and approximate (dashed line) calculations of $V_t(q_x, y)$ at $L = 3a_0$ are shown in Fig. 3a. At an arbitrary one-dimensional (along x) spatial distribution of stripes with identical $\langle d\rangle(y)$ distributions and $n(q_x)$ spectral densities, the spectral density of the $V_0(q_x, y)$ energy distribution can be written as

$$\begin{aligned} V_0(q_x, y) &= V(q_x, y)n(q_x) \\ &= \tilde{G}(y)\exp(-q_x a_0)n(q_x). \end{aligned} \tag{6}$$

After the inverse Fourier transform of (6) and the renormalization

$$\tilde{G}/E_0 \longrightarrow \tilde{G}, \quad x/a \longrightarrow x,$$
$$L/a \longrightarrow L, \quad a_{0,e}/a \longrightarrow a_{0,e},$$

where $E_0 = (2\pi\hbar)^2/2m_0a^2$ (m_0 is the effective mass), the stationary Shcrödinger equation can be written as

$$\left[\frac{d^2}{dx^2} + \tilde{G}(y)\exp\left(-ia_0\frac{d}{dx}\right)|\psi(x, y)|^2 + \varepsilon\right] \times \psi(x, y) = 0. \tag{7}$$

Here, $\psi(x, y)$ is the wave function of the hole and ε is the eigenvalue of the energy operator. The differentiation operator in the exponent in (7) takes into account nonlocal interactions and acts on $|\psi(x, y)|^2$. For two systems of parallel stripes (subscripts 1 and 2) with $\langle d_1\rangle(y) = -\langle d_2\rangle(y)$, (7) transforms into the system

$$\left\{\frac{d^2}{dx^2} + \tilde{G}(y)\exp\left(-ia_0\frac{d}{dx}\right) \times [|\psi_k(x, y)|^2 - |\psi_l(x, y)|^2] + \varepsilon\right\}\psi_k(x, y) = 0, \quad k, l = 1, 2; \quad k \neq 1. \tag{8}$$

Note that the ε and $\langle d_{1,2}\rangle(y)$ values present in the expression for $\tilde{G}(y)$ should be consistent with $\psi_{1,2}(x, y)$ for the solutions of (8) that are of interest to us. Averaging **d** in each stripe over two admissible directions taking into account thermal fluctuations in local (over y) thermodynamic equilibrium, we must require [30] that the equality

$$\langle d_{1,2}\rangle(y) = \pm\tanh\frac{\Delta(y)}{k_B T} \tag{9}$$

be fulfilled. Here, $2\Delta(y)$ is the "order" parameter (the difference of interaction energies for holes with **d** directed along and against the $\langle\mathbf{d}\rangle(y)$ direction in the stripe), k_B is the Boltzmann constant, and T is the temperature. Simultaneously, the equation

$$\Delta(y) = -\sum_{p=-\infty}^{\infty}\sum_{i,j=1,2}(-1)^{i-j}\tilde{G}(y) \times \exp(-|p|q_{xm}a_0)n_i(pq_{xm}, y)n_j(-pq_{xm}, y) \tag{10}$$

should be satisfied. In the calculations, the normalization of $\psi_{1,2}(x, y)$ corresponded to the arrangement of two holes with opposite **d** orientations along period L (in the "unit cell") of the stripe-structure [31],

$$\int_{-L/2}^{L/2}\int_{-L/2}^{L/2} dxdy|\psi_k(x, y)|^2 = 1, \quad k = 1, 2. \tag{11}$$

It follows that the probabilities of finding any hole in any stripe were equal.

Dynamic processes were described on the assumption that $\langle d\rangle$ explicitly depended not only on y but also on time t. The relaxation approximation [31]

$$\frac{\partial\langle d\rangle(y,t)}{\partial t} = -\frac{1}{\tau_d}[\langle d\rangle(y,t) - \langle d\rangle_0] \tag{12}$$

was used. Here, τ_d is the characteristic time of **d** flip and $\langle d\rangle_0$ is the equilibrium $\langle d\rangle$ value. Dynamic terms were not included into system (8) explicitly, and the evolution of $\psi_{1,2}(t)$ was described through the $\langle d\rangle(y, t)$ dependence in the adiabatic approximation (the time scale was set by the $\tau_F = L/v_F$ parameter, where v_F is the Fermi velocity) with the use of (12).

3. STATIONARY STRIPE-STRUCTURES

Taking (8) into account, let us write the Hamiltonian of the system in the form

$$\begin{aligned} H = &\sum_{q_x;\,(i,\,j)=1,2;\,(i\neq j)} q_x^2\psi_i(q_x,y)\psi_i(-q_x,y) \\ &- \tilde{G}(y)\exp(-q_x a_0)n_i(q_x,y)n_i(-q_x,y) \\ &+ \tilde{G}(y)\exp(-q_x a_0)n_i(q_x,y)n_j(-q_x,y). \end{aligned} \tag{13}$$

It follows from the Cauchy–Schwartz inequality [32] that H reaches a minimum at

$$\psi_i(q_x,y) = \begin{cases} -\psi_j(q_x,y), & q_x = \pm q_{xm}, \pm 3q_{xm}, \ldots \\ \psi_j(q_x,y), & q_x = 0, \pm 2q_{xm}, \ldots, \end{cases} \tag{14}$$

which is equivalent to the requirement

$$\psi_1(x,y) = \psi_2(x \pm L/2, y). \tag{15}$$

In the presence of stripe-structures, there should be at least one spatial harmonic with the amplitude $\psi_{1,2}(q_{xm}) \neq 0$. We can therefore take into account two lowest harmonics in $\psi_{1,2}$ to write

$$\psi_{1,2}(x,y) = \psi_0(y) \pm \psi(q_{xm},y)\cos(q_{xm}x). \tag{16}$$

The expressions for the energy of localization of holes, H and Δ, then take the form

$$\begin{aligned} H(y) = &\frac{2\psi^2(q_{xm},y)}{L}\left\{\pi^2 - 8L^2\psi_0^2(y)\right. \\ &\left.\times\left[\langle d\rangle(y) + a_e^2\frac{d^2\langle d\rangle(y)}{dy^2} + \ldots\right]G_0\exp(-q_{xm}a_0)\right\}, \end{aligned} \tag{17}$$

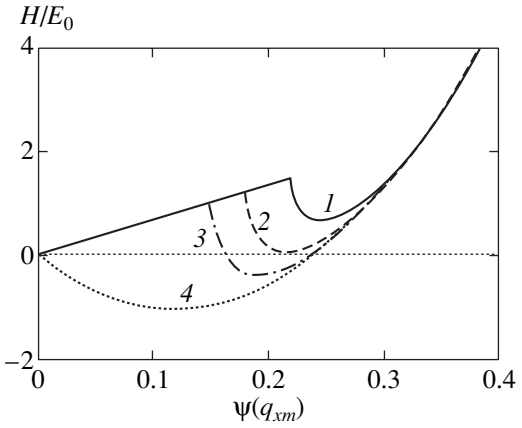


Fig. 4. Energy H of hole localization as a function of $\psi(q_{xm})$ at $L = 3$ and $G_0 = 4.5$ for $T = (1)\ 1.4T^*$, $(2)\ T^*$, $(3)\ 0.7T^*$, and $(4)\ 0$.

$$\begin{aligned} \Delta(y) = &\ 8L^2\psi_0^2(y)\psi^2(q_{xm},y) \\ &\times\left[\langle d\rangle(y) + a_e^2\frac{d^2\langle d\rangle(y)}{dy^2} + \ldots\right]G_0\exp(-q_{xm}a_0). \end{aligned} \tag{18}$$

Using (9), (11), (17), and (18), we can calculate the dependence of H (the width of the pseudogap in the spectrum of states) on $\psi(q_{xm})$ in the thermodynamically equilibrium state [$\psi_0(y) = \psi_0$ and $\psi(q_{xm}, y) = \psi(q_{xm})$], Fig. 4. Figure 4 shows that there is a region of parameter values within which H reaches a minimum for $\psi(q_{xm}) \neq 0$ at $T < T^* = \text{const}$. Precisely critical temperature T^* determines the transition point related to the formation of stripe-structures. The dependence of $\psi(q_{xm})$ on H is not single-valued, and excitation of metastable states is possible if T changes at a high rate (see below).

More rigorous calculations including six lowest harmonics of the Fourier transform of $\psi_{1,2}(x)$ were performed numerically. To determine $\psi_{1,2}(x)$, we constructed the auxiliary Hamiltonian

$$\begin{aligned} H' = &\sum_{p=-\infty}^{\infty}\sum_{i,j=1,2}[(pq_{xm})^2\psi_i(pq_{xm})\psi_i(-pq_{xm}) \\ &- \langle d\rangle_i(-1)^{i-j}G_0\exp(-|p|q_{xm}a_0)n_i(pq_{xm})n_j(-pq_{xm})] \\ &+ M\exp\{R[Ln_i(0)-1]^2\}, \end{aligned} \tag{19}$$

$$\langle d\rangle_i = \tanh\left[\frac{\langle d\rangle_i}{k_BT}\sum_{p=-\infty}^{\infty}\sum_{i,j=1,2}(-1)^{i-j}G_0 \times\exp(-|p|q_{xm}a_0)n_i(pq_{xm})n_j(-pq_{xm})\right].$$

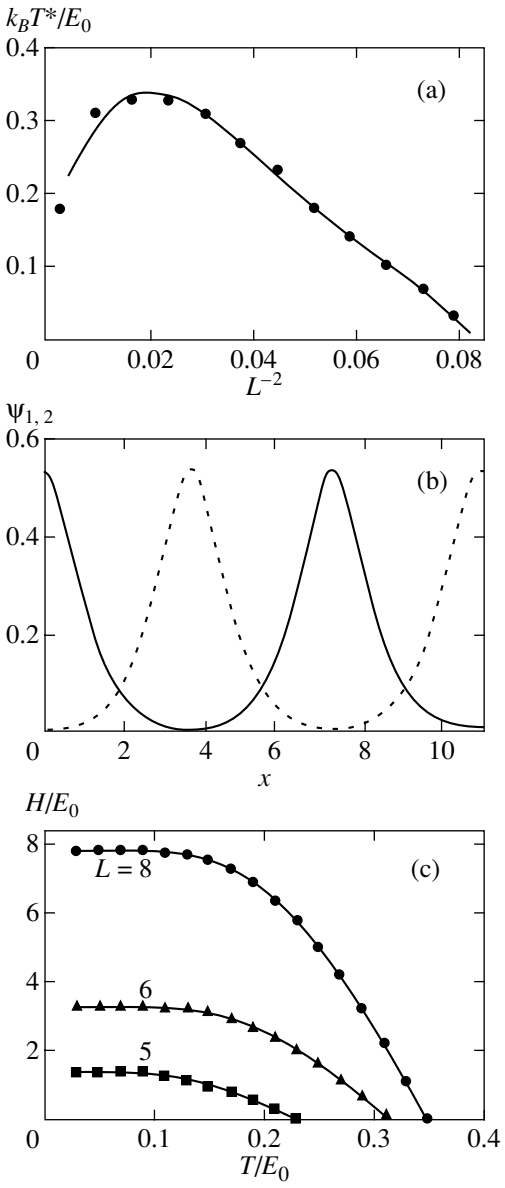


Fig. 5. (a) Calculated T^* versus L^{-2} dependence (solid circles) and its approximation by a polynomial of degree four (dashed line); (b) wave functions $\psi_{1,2}(x)$ (solid and dashed lines) at $L = 7.5$; and (c) temperature dependence of pseudogap width at $L = 5$, 6, and 8; $G_0 = 4.5$.

This Hamiltonian was minimized by the unrestricted simplex method [33], which ensured a correct normalization of $\psi_{1,2}(x)$ as $R \longrightarrow \infty$. The M and R parameters were selected to make the error in $\psi_{1,2}(x)$ smaller than 1%. The interaction energy was determined by substituting the obtained $\psi_{1,2}(x)$ functions into (13). The calculated $T^*(L^{-2})$ dependence is shown in Fig. 5a. The domain of the existence of stripe-structures is bounded by the limiting δ_m value for which $T^* = 0$. The $T^*(L^{-2})$ dependence contains a region of an almost linear growth of T^* as L^{-2} (that is, δ) decreases. The $\psi_{1,2}(x)$ wave functions are strongly anharmonic (Fig. 5b). The corresponding energy gap width monotonically decreases as T increases and vanishes at $T = T^*$ (Fig. 5c). This is indirect evidence that the $T = T^*$ point is a second-order phase transition point.

4. COEXISTENCE OF PHASES

Assuming that $\langle d\rangle(y) \neq \text{const}$, let us vary $\psi(q_x, y)$ [see (16)] taking into account the term proportional to $d^2\langle d\rangle(y)/dy^2$ and constantly tracing the position of the local extremum point. It is easy to show that H becomes minimum at

$$\psi(q_{xm}, y) = \frac{1}{2L}\left(1 - \frac{\gamma}{P}\right), \tag{20}$$

where

$$\gamma = \frac{1}{4g} = \frac{\pi^2}{4LG_0\exp(-q_{xm}a_0)},$$

$$P = \langle d\rangle(y) + a_e^2\frac{d^2\langle d\rangle(y)}{dy^2} + \ldots.$$

In addition,

$$\Delta(y, P) = \frac{G_0P}{L}\left[1 - \left(\frac{\gamma}{P}\right)\right]^2\exp(-q_{xm}a_0). \tag{21}$$

Substituting (21) into (9), linearizing the resulting equation with respect to P, and taking into account the self-consistent character of the problem, we obtain the closed equation

$$a_e^2\frac{d^2\langle d\rangle(y)}{dy^2} = E_b(T)\operatorname{arctanh}[\langle d\rangle(y)] + \gamma - \langle d\rangle(y), \tag{22}$$

$$E_b(T) = \frac{k_BTL}{2G_0\exp(-q_{xm}a_0)},$$

which can, by analogy with mechanics, be interpreted as the equation of motion in the potential

$$U(\langle d\rangle) = -E_b(T)\left[\langle d\rangle\operatorname{arctanh}(\langle d\rangle) + \frac{1}{2}\ln(1 - \langle d\rangle^2)\right] - \gamma\langle d\rangle + \frac{1}{2}\langle d\rangle^2. \tag{23}$$

This dependence is shown in Fig. 6a. At $T = T^*$, both maxima of the $U(\langle d\rangle)$ potential curve have equal values. In the problem under consideration, the system is conservative, and all processes that occur in it obey the law of the conservation of energy.

Consider the simplest example of inhomogeneous (with respect to y) solutions of the "phase switching" type. We assume that

$$\langle d\rangle(y)|_{y\to-\infty} = \langle d\rangle_1, \quad \langle d\rangle(y)|_{y\to+\infty} = \langle d\rangle_2 \neq \langle d\rangle_1$$

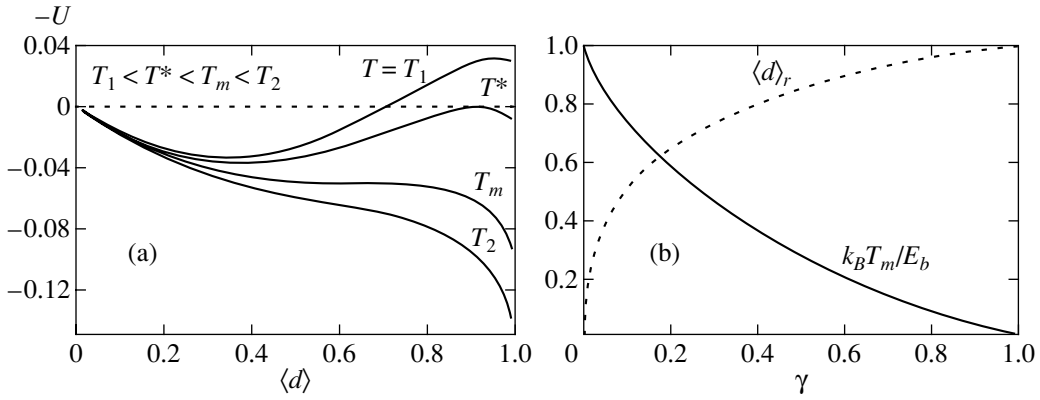


Fig. 6. (a) $U(\langle d\rangle)$ dependence calculated at several T values and (b) $T_m(\gamma)$ and $\langle d\rangle_r(\gamma)$ dependences calculated at $T = T_m$.

on two infinitely spaced (along y) cuprate plane sides. We are interested in the boundary between these two regions ("phases"). It follows from Fig. 6a that $U(\langle d\rangle(y))$ curves contain two local extrema at $T < T_m = \text{const} > T^*$. The first extremum ($U_1$) is localized at point $\langle d\rangle = 0$ at arbitrary T values. The position of the second extremum (U_2), $\langle d\rangle = \langle d\rangle_r$, changes with T, and $U_1 < U_2$ at $T^* < T < T_m$ and $U_1 > U_2$ at $T < T^*$. At $T = T^*$, the energies of the phases are equal, $U_1 = U_2$, and there exist solutions with the asymptotic

$$\left.\frac{d\langle d\rangle(y)}{dy}\right|_{y\to\pm\infty} = 0 \qquad (24)$$

transition from one "equilibrium" $\langle d\rangle_1 = 0$ value to the other, $\langle d\rangle_2 = \langle d\rangle_r$. The condition of the coexistence of homogeneous (here, along y) phases is known as the Maxwell rule [34].

The width of the transition region can be approximately estimated analytically, because the solution to (22) with condition (24) can be written in quadratures,

$$\zeta - \zeta_0 = \int_0^{\langle d\rangle_r} d\langle d\rangle \left\{ E_b(T^*)\left[\langle d\rangle \operatorname{arctanh}\langle d\rangle + \frac{1}{2}\ln(1-\langle d\rangle^2)\right] + \gamma\langle d\rangle - \frac{1}{2}\langle d\rangle^2 \right\}^{-1/2}, \qquad (25)$$

where $\zeta = y/a_e$ and $\zeta_0 = \text{const}$. Expanding the integrand into a series and truncating this series after quadratic terms, we obtain

$$\langle d\rangle(\zeta) = \frac{1}{2}\frac{\gamma}{1-E_b(T^*)} \times \{\sin[\sqrt{1-E_b(T^*)}(\zeta-\zeta_0)] + 1\}. \qquad (26)$$

at $E_b(T^*) \ll 1$. This gives the transition region width

$$\Delta y = \pi a_e \sqrt{\frac{1}{1-E_b(T^*)}}. \qquad (27)$$

Clearly, the obtained inhomogeneous solution is invariant with respect to an arbitrary $y \longrightarrow y + \delta y$ ($\delta y = \text{const}$) translation along the y axis.

5. METASTABLE STATES AND THE KINETICS OF PHASE TRANSITION

The second potential curve maximum (Fig. 6a) disappears at $T = T_m$. This temperature corresponds to simultaneous fulfillment of the conditions

$$\frac{\partial U(\langle d\rangle, T_m)}{\partial\langle d\rangle} = 0, \quad \frac{\partial^2 U(\langle d\rangle, T_m)}{\partial\langle d\rangle^2} = 0. \qquad (28)$$

The calculated $T_m(\gamma)$ dependence is shown in Fig. 6b. It is easy to see that $T_m > T^*$. Consider the situation when $T^* \neq T < T_m$ and the interphase boundary is mobile. As previously, it is assumed that $\langle d\rangle(y)|_{y\to-\infty} = 0$ and $\langle d\rangle(y)|_{y\to+\infty} = \langle d\rangle_r$ and the transition region is described by the profile of the phase switching wave $\langle d\rangle(y, t) = \langle d\rangle(\xi)$, which propagates along y at a constant velocity v. Here, $\xi = y - vt$ is the running coordinate. Preforming similar transformations and using (12), we obtain

$$a_e^2\frac{d^2\langle d\rangle}{d\xi^2} = E_b(T) \times \operatorname{arctanh}\left(\langle d\rangle - \tau_d v\frac{d\langle d\rangle}{d\xi}\right) + \gamma - \langle d\rangle. \qquad (29)$$

At $v = 0$, this equation transforms into (22). This means that, at $T = T^*$, the switching wave is immobile. Within the framework of the mechanical analogy, the term proportional to v describes nonlinear "friction." It follows that the phases cannot coexist at $T \neq T^*$. This conclusion conforms with the observation that the development of a switching wave in distributed systems

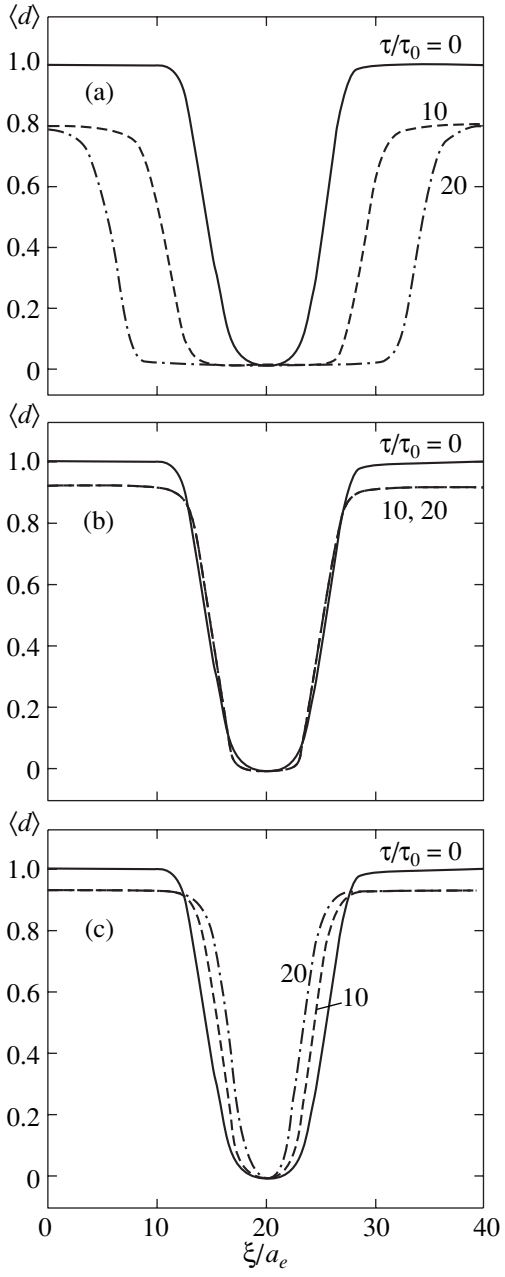


Fig. 7. Time evolution of initial $\langle d \rangle$ fluctuation at $T/T^* =$ (a) 1.22, (b) 1.00, and (c) 0.96; $g = 1.25$.

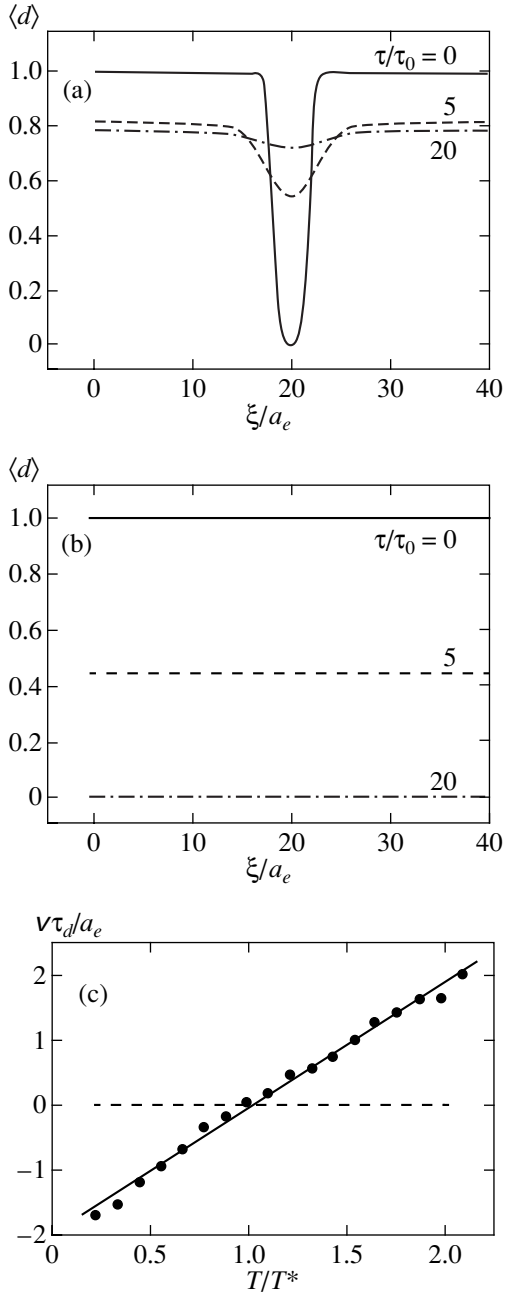


Fig. 8. (a) Dissipation of a narrow $\langle d \rangle$ fluctuation at $T/T^* =$ 1.22; (b) phase transition scenario at $T > T_m$ and $T/T^* =$ 1.57; and (c) dependence of v on T/T^*, $g = 1.25$.

requires a seed, that is, a fluctuation with characteristic dimensions of the order of the width of the transition region [35, 36]. Velocity v of the wave depends on the controlling parameter (here, T) value. At the critical point, velocity v equals zero and changes sign.

Equation (29) was solved numerically. The second derivative with respect to y was approximated by second-order finite differences. The sixth-order Runge–Kutta method with adaptive time step t was used. The evolution of the initial $\langle d \rangle = 1$ state with a perturbation of the hyper-Gaussian spatial profile was considered. The evolution of a broad (compared with Δy initial fluctuation at $T > T^*$ is shown in Fig. 7a. The system experiences a rapid transition to the metastable state $\langle d \rangle = \langle d \rangle_r$, and only then, the wave of switching to the stable phase $\langle d \rangle = 0$ is formed. The $\langle d \rangle_r$ value decreases as temperature increases. At $T = T^*$ (Fig. 7b) and $T < T^*$ (Fig. 7c) also, the transition to the metastable state $\langle d \rangle = \langle d \rangle_r$ occurs first, and the fluctuation then either freezes (a stationary interphase boundary is formed, $v = 0$) or dissipates ($v < 0$). A narrow (compared with Δy) fluctuation dissipates even at $T > T^*$ (Fig. 8a). At

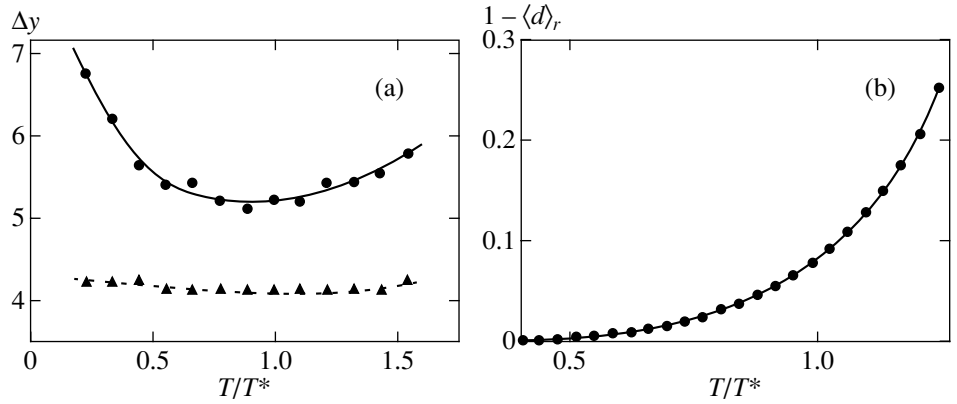


Fig. 9. (a) Temperature dependence of transition region width Δy at $g = 1.25$ (solid line) and 0.56 (dashed line) and (b) dependence of $1 - \langle d \rangle_r$ on T/T^* at $g = 1.25$.

$T > T_m$, the system always immediately transforms from the unstable $\langle d \rangle \neq 0$ to the stable $\langle d \rangle = 0$ state (Fig. 8b). Velocity v linearly depends on T/T^* and changes sign at $T = T^*$ (Fig. 8c).

It follows that, if two phases can coexist at $T = T^*$ (Fig. 7b), the seed of the stripe-structure can expand at $T < T^*$ ($v < 0$) because of the involvement of holes from regions with still unestablished order into the self-organization process (Fig. 7a). Assuming that vectors **d** in a stripe are ordered for approximately 10^2 holes and $\tau_d \sim 10^{-12}$–10^{-13} s, we obtain an estimate of the time of switching wave propagation, $\tau \sim 0.1$–1.0 ns. However, for such a wave to begin to propagate, the stripe seed should have dimensions exceeding Δy (see Figs. 7a, 8a). The calculated $\Delta y(T/T^*)$ dependences (over the level at 0.9 of the front height) are shown in Fig. 9a for $g = 1.25$ (solid line) and 0.56 (dashed line). If the minimum seed size (actually, the number of holes with ordered vectors **d**) is set to $\Delta y \sim 4$–6, the probabilistic estimate of time τ_f of the formation of the corresponding initial fluctuation takes the form

$$\tau_f = \frac{\tau_d}{(1 - \langle d \rangle_r)^{\Delta y - 1}}. \qquad (30)$$

It follows that $\tau_f \sim 1.0$ ns $\gg \tau_d$ at $T/T^* \approx 1.25$, when $1 - \langle d \rangle_r \approx 0.25$ (see Fig. 9b).

6. DISCUSSION AND CONCLUSION

To summarize, we constructed a simple phenomenological model of self-organization processes based on the assumption that stripe-structures in cuprate planes of oxide high-T_c superconductors are one-dimensional and periodic and are formed by magnetodipole (nonlocal spin-wave, see above) interaction forces. At a reasonable $G_0 \sim 100$ meV value, the model predicts the period of such structures to be several nanometers. The doping level at which these structures can be formed is bounded from above by the $\langle n \rangle_{\max} \approx 0.125$ value, in agreement with experimental data. The critical temperature of the existence of such structures is also close to the $T^* \sim 150$ K temperature determined in many experiments. What is more, the $T^*(\langle n \rangle)$ dependence predicted by our model has a well defined linear region at $\langle n \rangle < \langle n \rangle_{\max}$, which is also observed experimentally. Surprisingly, the calculated width of the corresponding energy gap and its temperature dependence closely agree with the ARPES data on pseudogaps in cuprates, which leads us to suggest that precisely this gap is the pseudogap. As far as the interpretation of the experimental data [23, 37, 38] on optical diagnostics of the kinetics of phase transitions in high-T_c superconductors is concerned, this suggestion brings to the fore the fact that, in all these experiments, samples with the optimal doping level δ_{opt} providing the highest T_c value were studied. In follows from the data obtained in [12] that, in the experiments described in [23, 37, 38], the $T_c \approx T^*$ approximate equality was satisfied, and the states with superconducting and pseudogaps were indistinguishable. When initially superconducting samples were rapidly heated to $T > T_c$, superconductivity had to disappear virtually instantaneously in all these experiments [25]. The samples were, however, heated to moderate temperatures satisfying the inequalities $T^* < T < T_m \approx (1.4$–$1.5)T^*$, and metastable stripe-structures had to persist in the samples for a fairly long time (1 ns or longer, see above). During the whole this time period, probing by the experimental scheme of [23] had to detect the energy gap, and probing by the scheme of [37, 38] had to well "remember" the initial (superconducting or nonsuperconducting) state of the object of study.

ACKNOWLEDGMENTS

In conclusion, the authors wish to express their sincere thanks to V.I. Emel'yanov for fruitful discussions that helped us to complete this investigation. This study was supported by the Russian Foundation for Basic

Research (project no. 00-15-96726) and the State Scientific–Technical Programs "Fundamental Metrology" and "Physics of Quantum and Wave Phenomena" for financial support.

REFERENCES

1. J. G. Bendnorz and K. A. Müller, Z. Phys. B **64**, 189 (1986).
2. A. Fujimori, E. Takayama-Muromachi, and Y. Uchida, Solid State Commun. **63**, 857 (1987).
3. J. Fink, N. Nucker, E. Pellegrin, *et al.*, J. Electron Spectrosc. Relat. Phenom. **66**, 395 (1994).
4. Z. Schlesinger, L. D. Rotter, R. T. Collins, *et al.*, Physica C (Amsterdam) **185–189** (1), 57 (1991).
5. R. J. Birgeneau and G. Shirane, in *Physical Properties of High Temperature Superconductors I*, Ed. by D. M. Ginsberg (World Scientific, Singapore, 1989), p. 151.
6. J. Rossat-Mignod, L. P. Regnault, C. Vettier, *et al.*, Physica C (Amsterdam) **185–189** (1), 86 (1991); J. M. Tranquada, P. M. Gehring, G. Shirane, *et al.*, Phys. Rev. B **46**, 5561 (1992).
7. G. Shirane, Physica C (Amsterdam) **185–189** (1), 80 (1991).
8. M. A. Kastner, R. J. Birgeneau, T. R. Thurston, *et al.*, Phys. Rev. B **38**, 6636 (1988).
9. *Problems in High-Temperature Superconductivity*, Ed. by V. L. Ginzburg and D. A. Kirzhnits (Nauka, Moscow, 1977).
10. J. F. Annett, N. Goldenfeld, and A. J. Leggett, J. Low Temp. Phys. **105**, 473 (1996).
11. A. Moreo, A. Nazarenko, S. Haas, *et al.*, J. Phys. Chem. Solids **56**, 1645 (1995).
12. M. R. Norman, H. Ding, M. Randeria, *et al.*, Nature **392**, 157 (1998).
13. V. V. Kabanov, J. Demsar, B. Podobnik, *et al.*, Phys. Rev. B **59**, 1497 (1999).
14. J. M. Tranquada, B. J. Sternlieb, J. D. Axe, *et al.*, Nature **375**, 561 (1995).
15. B. Buchner, M. Breuer, A. Freimuth, *et al.*, Phys. Rev. Lett. **73**, 1841 (1994).
16. J. L. Cohn and J. Karpinski, Phys. Rev. B **58**, 14 617 (1988).
17. G. Baskaran, Z. Zou, and P. W. Anderson, Solid State Commun. **63**, 973 (1987).
18. E. Dagotto, Rev. Mod. Phys. **66**, 763 (1994).
19. W. O. Putikka, M. U. Luchini, and T. M. Rice, Phys. Rev. Lett. **68**, 538 (1992).
20. M. S. Hybertsen, E. B. Stechel, M. Schluter, *et al.*, Phys. Rev. B **41**, 11 068 (1990).
21. K. Yonemitsu, A. R. Bishop, and J. Lorenzana, Phys. Rev. B **47**, 12 059 (1993).
22. *Monte Carlo Methods in Statistical Physics*, Ed. by K. Binder (Springer-Verlag, Berlin, 1986).
23. A. N. Zherikhin, V. A. Lobastov, V. M. Petnikova, and V. V. Shuvalov, Phys. Lett. A **179**, 145 (1993); Physica C (Amsterdam) **221**, 311 (1994).
24. *Nonequilibrium Superconductivity*, Ed. by D. N. Langenberg and A. I. Larkin (North-Holland, Amsterdam, 1986).
25. N. Bluzer, Phys. Rev. B **44**, 10 222 (1991).
26. B. P. Stojkovic, Z. G. Yu, A. L. Chernyshev, *et al.*, cond-mat/9911380.
27. D. M. Frenkel and W. Hanke, Phys. Rev. B **42**, 6711 (1990).
28. B. I. Shraiman and E. D. Siggia, Phys. Rev. B **40**, 9162 (1989).
29. D. I. Khomskiĭ and K. I. Kugel, cond-mat/0103317.
30. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 5: *Statistical Physics* (Pergamon, Oxford, 1980; Nauka, Moscow, 1995).
31. A. Animalu, *Quantum Theory of Crystalline Solids* (Prentice Hall, Englewood Cliffs, 1977; Mir, Moscow, 1981).
32. G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers* (McGraw-Hill, New York, 1961; Nauka, Moscow, 1970).
33. J. A. Nelder and R. Mead, Comput. J. **7**, 308 (1965).
34. E. M. Lifshitz and L. P. Pitaevskiĭ, *Physical Kinetics* (Nauka, Moscow, 1979; Pergamon, Oxford, 1981).
35. N. N. Rozanov, *Optical Bistability and Hysteresis in Distributed Nonlinear Systems* (Nauka, Moscow, 1997).
36. V. A. Vasil'ev, Yu. M. Romanovskiĭ, and V. G. Yakhno, *Autowave Processes* (Nauka, Moscow, 1987).
37. M. E. Gershenzon, V. V. Golovlev, I. B. Kedich, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **52**, 1189 (1990) [JETP Lett. **52**, 602 (1990)]; S. V. Chekalin, V. M. Farztdinov, V. V. Golovlyov, *et al.*, Phys. Rev. Lett. **67**, 3860 (1991).
38. A. S. Kazeroonian, T. K. Cheng, S. D. Brorson, *et al.*, Solid State Commun. **78**, 95 (1991).

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 93, No. 5, 2001, pp. 1101–1104.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 5, 2001, pp. 1268–1272.

SOLIDS
Electronic Properties

Photoinduced Drift of Electrons in Thin Magnetic Films

N. N. Krupa

Institute of Magnetism, National Academy of Sciences of Ukraine, Kiev, 02142 Ukraine
e-mail: krupa@imag.kiev.ua

Received April 27, 2001

Abstract—The results are presented of experimental studies of variations in the polarization of light reflected from multilayer thin metal films containing nonmagnetic Bi or Ti films and a TbFe magnetic film magnetized in the direction perpendicular to the film plane. An additional optical rotation of light reflected from bismuth and titanium films was observed upon their irradiation by intense nanosecond pulses from a semiconductor laser. This optical rotation is attributed to the photoinduced drift of electrons with polarized spins from the magnetic TbFe film. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Photoinduced electron drift, or electron entrainment by powerful laser radiation, was first observed in semiconductors [1, 2], and since then this effect was investigated mainly in such materials.

Electron entrainment results in a strong increase in their concentration in the region of emergence of a laser beam from a sample. This causes a great change in the refractive index ($|\Delta n| = 0.01$–0.1) and provides quite a broad scan of nanosecond and picosecond laser pulses on the total internal reflection face of semiconductor crystals [3].

The photoinduced electron drift should also occur in metals; however, because of the high absorption coefficient of metals and the high concentration of conduction electrons in them, its observation in metals is greatly complicated.

In this paper, variations in the reflection of light from thin multilayer metal films are experimentally studied upon their irradiation by intense nanosecond pulses. The irradiation results in the optical rotation of light reflected from nonmagnetic metal films due to photoinduced drift of electrons with polarized spins.

2. EXPERIMENTAL

Multilayer Bi/SiC/TbFe/SiC, Ti/SiC/TbFe/SiC, and SiC/TbFe/SiC films were studied. The films were deposited by the electron-beam evaporation method or by magnetron sputtering on polished substrates made of optical glass of a thickness of 1.2 mm. The TbFe films had a high degree of perpendicular magnetic anisotropy, which was provided by the appropriate deposition technology [4] and was controlled with a Kerr hysteresigraph.

Experimental samples contained layers of the following thickness: 5- or 10-nm-thick Bi or Ti films, 3- or 5-nm-thick SiC barrier-layer films, 50- to 60-nm-thick SiC protective coatings, and 40-nm-thick TbFe films.

Optical rotation was studied upon reflection of light from two sides of multilayer films at the point of irradiation by intense pulses from a semiconductor laser. Fig. 1 shows the optical scheme of the experimental setup.

Radiation from a semiconductor laser *1* passed through a polarization Nicol prism *2* and a special interference mirror *3* and was focused with a microobjective *4* on a film on a substrate *5*.

Two types of microobjectives were used: a standard objective with an aperture of 0.65 for focusing a laser beam directly on the film and a special microobjective with an aperture of 0.52 for focusing laser radiation on the film through the substrate. These two microobjectives were used in a pair. The standard objective was used for focusing radiation from a semiconductor laser, while the special objective *6* focused radiation from a helium–neon laser *12* on the same region of the multilayer film but from the opposite side.

The polarization of reflected radiation of both lasers was measured using polarization Senarmont prisms *8*, high-frequency photodetectors *9*, differential amplifiers, and two oscilloscopes.

Optical filters *7* cut off radiation transmitted by the film. The accuracy of focusing of the both beams opposite to each other was controlled by signals from two- to four-area photodiodes *10*.

The measurements were performed without a magnetic field or in a magnetic field directed perpendicular to the film plane. The direction and magnitude of the magnetic field were varied from 0 to 32 000 A/m using Helmholtz coils. The film was initially magnetized in a strong magnetic field.

The maximum output cw power of the semiconductor laser at 0.81 μm was 100 mW. The power density produced by 20- to 100-ns pulses from the semiconduc-

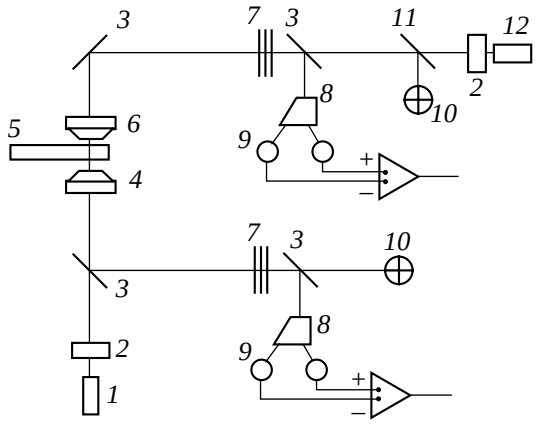


Fig. 1. Scheme for studying photoinduced electron drift: (*1*) semiconductor laser; (*2*) Nicole prism; (*3*) interference mirror; (*4*) standard microobjective; (*5*) substrate with a film; (*6*) special microobjective; (*7*) optical filters; (*8*) Senarmont prism; (*9*) recording photodiodes; (*10*) four-area photodiodes; (*11*) semitransparent mirror; (*12*) helium–neon laser.

tor laser on the film plane was varied from 1 kW/cm^2 to 10 MW/cm^2, while the power density produced by the helium–neon laser was about 500 W/cm^2.

The lasers operated in a single-frequency mode and the polarization plane of the light incident on the film lay in the horizontal plane.

Special interference mirrors *3* transmitted more than 50% of light polarized in the plane of incidence and reflected about 90% of light with the orthogonal polarization, resulting in the increase in the sensitivity and accuracy of measurements of the Kerr angle upon reflection of light from films.

3. RESULTS AND DISCUSSION

Our experiments showed that the reflection of light by multilayer Bi/SiC/TbFe/SiC and Ti/SiC/TbFe/SiC changes after irradiation of these films by intense pulses from a semiconductor laser. The change in the light reflection depends on the direction of the laser beam incident on the film and on the observation plane (entrance or exit with respect to the laser beam).

When a pulse from the semiconductor laser is incident from the side of the protective SiC cover, the polarized signal of laser radiation reflected from the magnetic TbFe film weakly decreases with increasing incident power, which is manifested in the change in the laser-pulse shape (Fig. 2, curves *3* and *5*).

At the same time, the polarization reflection signal for the helium–neon laser first increases during the pulse from the semiconductor laser (Fig. 2, curve *2*) and then slowly decreases after the termination of the semiconductor-laser pulse. As the power of semiconductor-laser pulses increases, the duration of polarization signals strongly decreases due to a shortening of their trailing edge (Fig. 2, curve *6*).

These variations in reflection are manifested more distinctly with increasing the duration of the semiconductor-laser pulse (Fig. 3). The duration of polarization signals for both lasers becomes shorter than that of intense pulses from the semiconductor laser.

The magnetic field, whose direction coincides with the direction of magnetization of the TbFe film does not

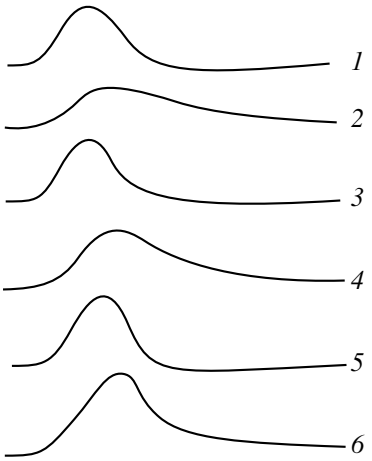


Fig. 2. Variations in the radiation pulses from (curves *1*, *3*, *5*) semiconductor and (curves *2*, *4*, *6*) helium–neon lasers reflected from the Ti/SiC/TbFe/SiC multilayer film for different intensities of 20-ns pulses from the semiconductor laser in the absence of a magnetic field. The semiconductor-laser pulse is incident from the side of the SiC film: curves *1*, *2* correspond to I = 1 MW/cm^2; *3*, *4*, to I = 3 MW/cm^2; and *5*, *6*, to I = 10 MW/cm^2.

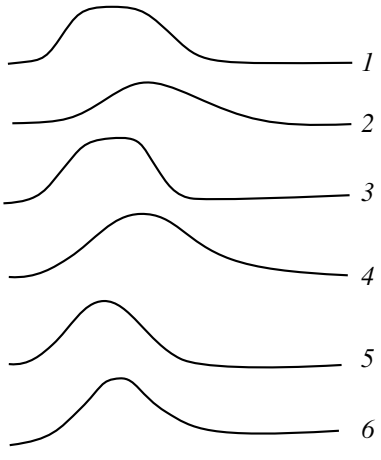


Fig. 3. Variations in the radiation pulses from (curves *1*, *3*, *5*) semiconductor and (curves *2*, *4*, *6*) helium–neon lasers reflected from the Ti/SiC/TbFe/SiC multilayer film for different intensities of 80-ns pulses from the semiconductor laser in the absence of a magnetic field. The semiconductor-laser pulse is incident from the side of the SiC film: curves *1*, *2* correspond to I = 1 MW/cm^2; *3*, *4*, to I = 3 MW/cm^2; and *5*, *6*, to I = 10 MW/cm^2.

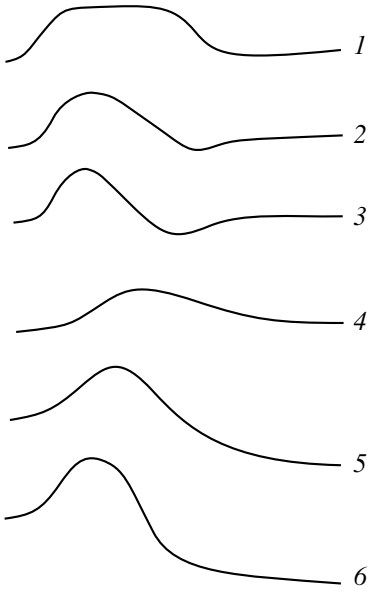


Fig. 4. Variations in the radiation pulses from (curves *1*, 2, *3*) semiconductor and (curves *4*, *5*, *6*) helium–neon lasers reflected from the Ti/SiC/TbFe/SiC multilayer film for different intensities of 120-ns pulses from the semiconductor laser. The semiconductor-laser pulse is incident from the side of the SiC film: curves *1*, *4* correspond to $I = 1$ MW/cm^2; *2*, *5*, to $I = 3$ MW/cm^2; and *3*, *6*, to $I = 10$ MW/cm^2. The magnetic field $H = 25\,000$ A/m is directed oppositely to magnetization of the TbFe film.

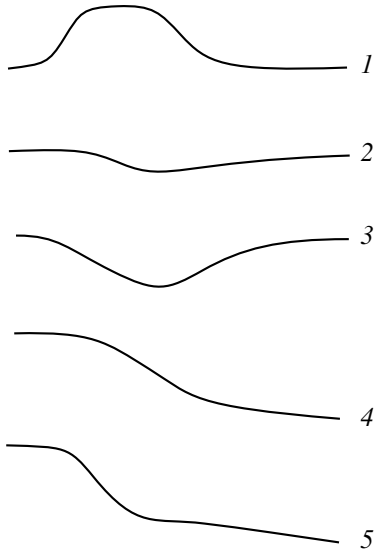


Fig. 5. Variations in the radiation pulses from (curves *2*–*5*) the helium–neon laser reflected from the SiC/TbFe/SiC multilayer film for different intensities of pulses (curve *1*) from the semiconductor laser in the absence of a magnetic field (curves *2*, *3*) and in the magnetic field $H = 25\,000$ A/m (curves *4*, *5*) directed oppositely to the magnetization of the TbFe film: curve *2* corresponds to $I = 1$ MW/cm^2; *3*, to $I = 3$ MW/cm^2; and *5*, to $I = 10$ MW/cm^2.

almost affect these processes. In the opposite magnetic field, the shape of reflection signals becomes much more complicated (Fig. 4).

Upon reflection of intense pulses of the semiconductor laser from the TbFe film, the reflected signal not only decreases to zero during the action of the laser pulse but becomes negative (Fig. 4, curve *3*). These variations are even more distinctly manifested upon reflection of helium–neon laser radiation from Bi or Ti films (Fig. 4, curves *5* and *6*).

The polarization reflection signal of radiation from the helium–neon laser in SiC/TbFe/SiC films almost does not increase upon irradiation of these films by intense pulses from the semiconductor laser. In the absence of a magnetic field, this signal decreases almost to zero and then returns to the initial level (Fig. 5). In the magnetic field whose direction is opposite to the direction of magnetization of the TbFe film, the signal decreases to zero; then it becomes negative, and its amplitude slowly increases to the level close to that of the initial positive signal.

The variations in reflection signals observed for multilayer Bi/SiC/TbFe/SiC and Ti/SiC/TbFe/SiC films irradiated by the semiconductor laser from the side of bismuth or titanium were analogous to those observed for SiC/TbFe/SiC films.

We believe that the variations observed upon reflection of laser radiation from the multilayer films are caused by the photoinduced drift of electrons with high spin polarization. Such electrons are transferred by light pulses of the semiconductor laser from the magnetized TbFe film to bismuth or titanium films and locally magnetize them, which results in the optical rotation of radiation of the helium–neon laser reflected from these magnetized local regions. For this reason, we observe the enhancement of polarization reflection signals for this radiation.

As the power and duration of semiconductor-laser pulses increase, the TbFe film is heated, resulting in the decrease in the angle φ_k of Kerr optical rotation and in the degree of polarization of spins of the electrons transferred by the laser beam to bismuth or titanium films. This reduces polarization signals for both lasers.

In the presence of the magnetic field directed oppositely to the direction of magnetization of the TbFe film, the heating of the film drastically reduces its coercion force and results in the film magnetic reversal. In this case, the polarization signal for the semiconductor laser vanishes and then becomes negative.

The optical rotation angle of the reflected radiation of the helium–neon laser increases at the onset of an intense pulse from the semiconductor laser due to the photoinduced drift of spin-polarized electrons. Then, this angle weakly decreases and changes its sign due to the heating of the TbFe film. In this case, the reflection signal vanishes and then becomes negative.

The magnitude of the photoinduced electron drift can be estimated, as in [2], from the expression

$$j_c = -I\alpha\beta\frac{2e^3 n_0}{m_e c^2 \omega}, \quad (1)$$

where j_c is the current density upon the photoinduced electron drift; e and m_e are the electron charge and effective mass, respectively; c is the velocity of light; I, ω, and α are the intensity, frequency, and absorption coefficient of laser radiation, respectively; and β is a coefficient whose value depends on the height and width of a barrier upon the transition of electrons from TbFe films to Bi or Ti films.

The intensity of the polarization signal observed upon reflection of radiation of the helium–neon laser, neglecting absorption dichroism, is proportional to the double angle $2\varphi_1$ of the additional Kerr effect, which arises due to the appearance of a high concentration of electrons with a given polarization of spins in bismuth or titanium films.

Taking into account Eq. (1) and the proportionality of the Kerr angle to the nondiagonal terms of the conductivity tensor [5], we can estimate φ_1 from the expression

$$\varphi_1 \approx \alpha\beta I\frac{e^5 \tau N_\sigma}{m_e^2 c^2 \omega(1 + \omega^2\tau^2)}, \quad (2)$$

where τ is the relaxation time and is N_σ is the concentration of electrons with polarized spins.

Thus, the results presented in this paper showed that nonmagnetic films in multilayer magnetic films could be magnetized due to the photoinduced drift of electrons with polarized spins.

REFERENCES

1. A. M. Danishevskiĭ, A. A. Kastal'skiĭ, S. M. Ryvkin, and I. D. Yaroshetskiĭ, Zh. Éksp. Teor. Fiz. **58**, 544 (1970) [Sov. Phys. JETP **31**, 292 (1970)].
2. N. A. Bryzhnykh, A. A. Grinberg, and É. É. Imamov, Fiz. Tekh. Poluprovodn. (Leningrad) **5**, 1735 (1971) [Sov. Phys. Semicond. **5**, 1516 (1971)].
3. N. N. Krupa and A. N. Pogorelyĭ, Zh. Tekh. Fiz. **68** (4), 121 (1998) [Tech. Phys. **43**, 457 (1998)].
4. N. N. Krupa, Metallofiz. Noveĭshie Tekhnol. **22**, 35 (2000).
5. R. Pittini and P. Wachter, J. Magn. Magn. Mater. **186**, 306 (1998).

Translated by M. Sapozhnikov

Journal of Experimental and Theoretical Physics, Vol. 93, No. 5, 2001, pp. 1105–1112.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 5, 2001, pp. 1273–1281.

SOLIDS
Electronic Properties

Electrodynamic Features of Anisotropic Hard Superconductors

I. F. Voloshin[a], A. V. Kalinov[a], L. M. Fisher[a,*], A. V. Aksenov[b], and V. Ya. Yampol'skiĭ[b]

[a]*State Research Center "All-Russia Electrical Engineering Institute," Moscow, 111250 Russia*
[b]*Institute of Radiophysics and Electronics, National Academy of Sciences of Ukraine, Kharkov, 310085 Ukraine*
**e-mail: lfisher@online.ru; fisher@vei.ru*

Received May 14, 2001

Abstract—The low-frequency electromagnetic response of superconducting plates characterized by strong anisotropy of current-carrying capacity in the plane of the sample is studied experimentally and theoretically. Measurements are made on polycrystalline textured plates of the Y-123 system with the **c** axis lying in the plane of the sample and on a single crystal with a single preferred direction of twinning boundaries. It is shown that the shape of the curves describing the dependence of the relative losses q on the ac field amplitude h_0 is quite sensitive to the orientation of vector $\mathbf{h}_0$ in the sample plane. As in the case of isotropic samples, the $q(h_0)$ dependence is characterized by a single size peak if vector $\mathbf{h}_0$ is oriented along one of the principal symmetry directions of the anisotropic critical current density. If $\mathbf{h}_0$ deviates considerably from principal directions, two size peaks are observed on the $q(h_0)$ curve. A detailed analysis of the evolution of the $q(h_0)$ curves upon a rotation of vector $\mathbf{h}_0$ in the sample plane is carried out. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

A distinguishing feature of high-temperature superconductors is the layered structure of their crystal lattice. Owing to this structure, HTSC systems are characterized by a well-defined anisotropy of almost all of their physical properties. It is well known that the main parameters of superconductors such as coherence length ξ, London penetration depth λ_L, and the lower and upper critical fields H_{c1} and H_{c2} along the **c** axis and on the **ab** plane differ substantially. The parameters of the vortex lattice, its phase state, and elastic moduli are quite sensitive to the orientation of the external magnetic field relative to the crystallographic axes of the sample. All these circumstances, as well as the presence of ordered planar and linear defects, lead to strong anisotropy of the current-carrying capacity of high-temperature superconductors.

In view of the importance of a correct interpretation of the properties of HTSC materials, many groups of scientists paid serious attention to the study of anisotropy of electrical and magnetic characteristics of superconductors almost immediately after their discovery; such investigations are being carried out even at present (see, for example, reviews [1, 2] and articles [3–8]). Nevertheless, a generally accepted theoretical model providing a correct description of the electrodynamic properties of anisotropic superconductors has not been developed so far. The dependence of the magnitude and direction of current density on the orientation of the exciting electric field remains unclear even under the simplest conditions, when the critical state model is applicable.

Recently, new phenomena associated with the peculiar penetration of the magnetic field into anisotropic superconductors have been discovered. For example, anomalous behavior of the size effect in electromagnetic absorption associated with magnetic field penetration to the middle of a superconducting plate was observed in [9]. For samples of the Y-123 system with the **c** axis lying in the sample plane, the size effect for certain directions of the exciting ac magnetic field $\mathbf{h}_0\cos(\omega t)$ oriented in the plane of the plate is manifested in the form of two peaks on the dependence of relative losses q on the amplitude h_0. The term relative losses is applied to the average power dissipated in the sample, normalized to $h_0^2\omega d/8\pi$, where d is the plate thickness. It should be recalled that the value of q in the isotropic case coincides with the real component of the dynamic magnetic susceptibility χ', which has a single peak for amplitude h_0 close to the penetration field H_p. The observed effect was interpreted under the assumption that the critical current density J_c is a second-rank tensor whose principal values J_{cy} and J_{cz} correspond to the critical current densities along the **c** axis and in the direction perpendicular to it. Two peaks of the function $q(h_0)$ are observed when vector $\mathbf{h}_0$ does not coincide with the directions of the principal axes of y and z of the critical current density tensor. These peaks correspond to successive penetration of the two ac field components (h_y and h_z), each of which is independently screened by the critical currents J_{cz} and J_{cy}.

Such a peculiarity of magnetic field penetration into an isotropic superconductor is manifested in the emergence of additional extrema on static magnetization

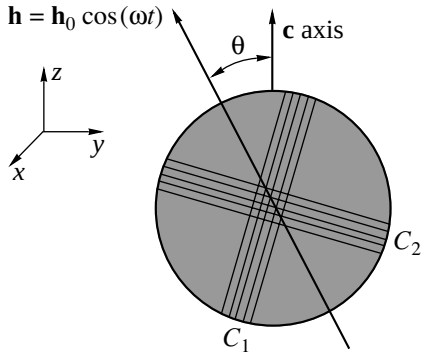


Fig. 1. Schematic diagram of the experiment.

curves. In the isotropic case, the presence of extrema on the curve describing the dependence of the sample magnetization M on the magnetic field H is associated with a decrease in current J_c and with the magnetic flux penetration to the bulk of the superconductor upon an increase in the external field H. In anisotropic superconductors, the successive penetration of the two components of field **H** to the middle of the plate leads to the emergence of new singularities on the $M(H)$ curves, i.e., to a new type of the peak effect experimentally observed in [10]. While the traditional peak effect is associated with the nonmonotonic dependence of the critical current density on the magnetic induction, the new effect is due entirely to the anisotropy of J_c.

In order to describe the effects discovered in [9, 10] consistently, a more detailed investigation of the behavior of relative losses $q(\mathbf{h}_0)$ for different orientations of vector $\mathbf{h}_0$ relative to the principal anisotropy axes of the sample is required and a theoretical model correctly taking into account the peculiar nature of magnetic field penetration into anisotropic superconductors must be constructed.

In the present work, we investigate the evolution of the behavior of relative losses $q(\mathbf{h}_0)$ upon a change in the orientation of vector **h** relative to the principal anisotropy axes of the sample. The values of critical angles for which the $q(h_0)$ curves change their topology (namely, the curves with a single peak are transformed into peculiar curves with two size peaks) are established. The results of observation are interpreted in the framework of the proposed model of the critical state generalized to the anisotropic case.

2. EXPERIMENT

2.1. Samples and Measurements

This work mainly aims at the study of the effect of anisotropy in the critical current density of superconductors on their electrodynamic properties. For this reason, the experiments were made on samples with clearly manifested anisotropic properties. As the basic model sample, we used an yttrium superconducting disk with the **c** axis lying in its plane. A sample of diameter 1.3 mm and thickness 0.23 mm was cut from an ingot of fused grain-oriented ceramic of the Y-123 system synthesized at a high temperature using a seed mounted at the top. In order to select a perfect monodomain region of the ingot, the latter was observed through an optical microscope in polarized light and the frozen magnetic flux was mapped with the help of a Hall probe. The model sample was cut from the chosen monodomain with a diamond cutter so that the **c** axis was in the plane of the sample. Before measurements, the sample was mechanically polished. The superconducting transition temperature T_c of the sample was 89 K and the transition width was 0.5 K. Some measurements were made on a single crystal having a size of $1.1 \times 0.6 \times 0.04$ mm, belonging to the same system, and prepared by Obolenskiĭ and Bondarenko at the Kharkov State University. The twinning boundaries in this crystal were predominantly directed along the [110] crystallographi**c** axis. For this crystal, the value of T_c was 91 K and the transition width was 0.3 K.

We studied experimentally the low-frequency electromagnetic response of a superconductor placed in an external static magnetic field **H** created by an electromagnet. The experiments were aimed at an analysis of electromagnetic absorption for various orientations of the exciting ac magnetic field $\mathbf{h}(t) = \mathbf{h}_0\cos(\omega t)$, parallel to the sample surface, relative to its crystallographic axes. Measurements were usually made at the frequency $\omega = 2\pi f = 2\pi \times 343\ \text{s}^{-1}$. Field $\mathbf{h}(t)$ could be rotated continuously with the help of two exciting coils with mutually orthogonal axes. Synphase currents with amplitudes $I_0\sin\psi$ and $I_0\cos\psi$ were passed through the coils. The ac field $\mathbf{h}(t)$ is obviously tilted through an angle ψ relative to the axis C_2 of the coil. The value of angle ψ was monitored using a computerized measuring system.

The response of the superconductor to an electromagnetic excitation was recorded with the help of another pair of mutually orthogonal detector coils tightly wound on the sample. This system is shown schematically in Fig. 1 (coils C_1 and C_2). After synchronously detecting the voltages from each of the detecting coils, the amplitudes of signals synphase with field $h(t)$ were singled out. Thus, our measuring system enabled us to measure the magnitude and direction of the component $\mathscr{E}$ (which was synphase with $h(t)$) of the electric field vector $\mathbf{E}(t)$ induced on the surface of the superconductor:

$$\mathscr{E} = \frac{\omega}{2\pi}\int_0^{2\pi/\omega} \mathbf{E}(t)\cos(\omega t)dt. \qquad (1)$$

Vector $\mathscr{E}$ carries important physical information. Its component ($\mathscr{E}_\perp = \mathbf{n}\cdot[\mathscr{E}'\times\mathbf{h}_0]/h_0$) in the direction perpendicular to the external exciting field $\mathbf{h}_0$ is in fact the average value of the Poynting vector on the sample sur-

face, i.e., the power dissipated in the superconductor. Here, **n** is the unit vector of the normal to the sample surface. It is convenient for the subsequent analysis to introduce the dimensionless function

$$q(h_0) = \frac{4c\mathscr{E}_\perp}{h_0\omega d} \quad (2)$$

describing the relative losses in the sample (dissipated power normalized to $h_0^2\omega d/8\pi$). In the case of an isotropic superconductor, this function coincides with the imaginary component of the dynamic magnetic susceptibility $\chi''(h_0)$.

The component of vector $\mathscr{E}$ along the direction parallel to the external field $\mathbf{h}_0$ ($\mathscr{E}_\parallel = \mathscr{E}\cdot\mathbf{h}_0/h_0$) vanishes in the isotropic case. Consequently, the nonzero value of the dimensionless function

$$a(h_0) = \frac{4c\mathscr{E}_\parallel}{h_0\omega d} \quad (3)$$

indicates anisotropy of the screening properties of the superconductor under investigation.

In our experiments, we studied the behavior of the relative losses q and the anisotropy signal a as functions of the amplitude and direction of the ac magnetic field $\mathbf{h}(t)$ as well as the magnitude and direction of the static magnetic field **H**. In each experiment, we set the direction of field **H** parallel to the plane of the sample by tilting the electromagnet. The parallelism of the field corresponded to the minimum of the signal $q(\mathbf{H})$ being measured. All measurements were made on the sample at temperature $T = 77$ K in the zero field cooling mode.

2.2. Experimental Results

We began our analysis of the effect of anisotropy of the current-carrying capacity of our samples on their electromagnetic response by studying the anisotropy signal $a(\mathbf{h}_0)$ for various orientations of the static magnetic field **H**. The results of measurements of the dependence of a on the angle θ of rotation of vector $\mathbf{h}_0$ for a superconducting disk are presented in Fig. 2. It can be seen that the anisotropy signal oscillates upon a change in angle θ. Function a vanishes only at four characteristic points over the entire period $0 < \theta < 2\pi$. The positions of these points are independent of the orientation of the static magnetic field. It was found that they correspond to the directions of vector $\mathbf{h}_0$ along the **c** axis and perpendicularly to it. In Fig. 2 and in the subsequent analysis, we will measure angle θ of rotation of field $\mathbf{h}_0$ from the direction of the **c** axis. The results of measurements of the angular dependence $a(\mathbf{h}_0)$ show that in spite of a strong anisotropy, there exist two preferred directions of the ac magnetic field (which will be referred to as the principal directions) for which the sample behaves as an isotropic system. For the sake of definiteness, we will refer to the direction along **c** axis as the z axis and the other principal direction lying in the **ab** plane as the y axis.

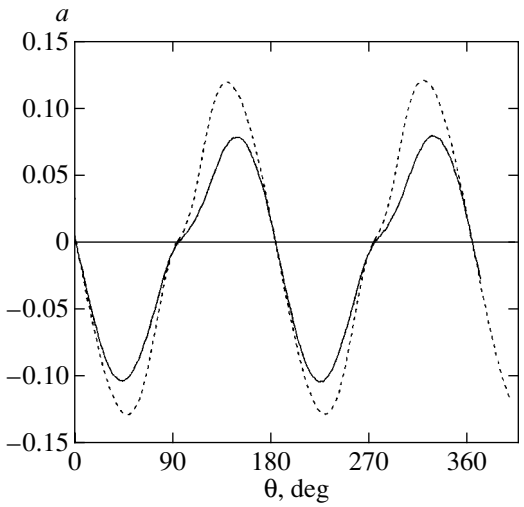


Fig. 2. Dependence of the dimensionless anisotropy signal a on the angle θ of rotation of the exciting field $\mathbf{h}_0$ in the plane of the sample for various orientations of a static magnetic field of 5 kOe at temperature 77 K for a superconducting disk. The solid curve corresponds to the **H** || **c** orientation and the dashed curve, to the **H** || **ab** orientation.

Before analyzing the effect of anisotropy in the current-carrying capacity of the sample on the relative losses $q(\mathbf{h}_0)$, we made contactless measurements of the dependence of the critical current density along the principal directions on the magnitude and direction of the static magnetic field **H**. For this purpose, we determined the dependence of quantity q on field **H** for its orientation along and at right angles to the **c** axis for $h_0 \ll H$. The magnetic-field dependence of the critical current density was reconstructed from the results of measurements of the function $q(H)$ using the method described in detail in [11, 12]. The obtained dependences of the critical current densities J_{cz} and J_{cy} along the **c** axis and in the **ab** plane are presented in Fig. 3. The most important features of the current-carrying capacity of the superconducting disk are worth noting. First, it can be seen that the critical current density in the **ab** plane is quite large and decreases relatively slowly upon an increase in the magnetic field. As expected, strong anisotropy of the critical current is observed: the current density in the **ab** plane exceeds the value of current along the **c** axis by more than an order of magnitude. As the static magnetic field H increases, the anisotropy becomes weaker. Finally, both current components (J_{cy} and J_{cz}) are quite sensitive to the direction of vector **H**. This circumstance is illustrated in Fig. 4, showing the dependence of components J_{cy} and J_{cz} on the angle of rotation of field **H** in the plane of the sample. Angle θ is measured from the **c** axis.

The results presented in Figs. 3 and 4 visually indicate the complex nature of anisotropy of the current-

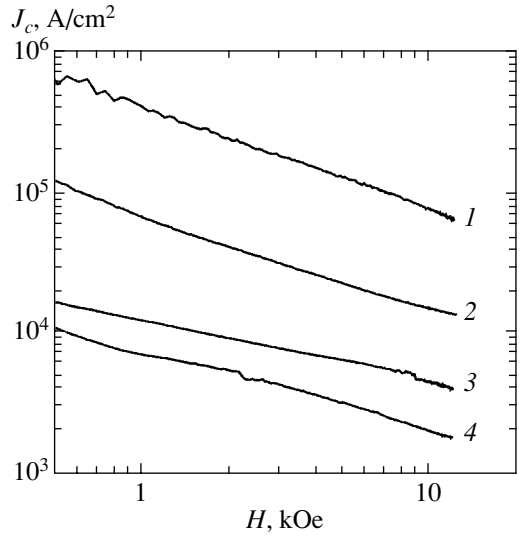


Fig. 3. Dependence of the critical current density J_c in the **ab** plane and along the **c** axis on the magnetic field oriented in the **ab** plane and along the **c** axis at 77 K for a superconducting disk. Curve *1* corresponds to the conditions $\mathbf{J}_c \parallel \mathbf{ab}$, $\mathbf{H} \parallel \mathbf{ab}$; (*2*) to $\mathbf{J}_c \parallel \mathbf{ab}$, $\mathbf{H} \parallel \mathbf{c}$; (*3*) to $\mathbf{J}_c \parallel \mathbf{c}$, $\mathbf{H} \parallel \mathbf{ab}$, and (*4*) to $\mathbf{J}_c \parallel \mathbf{c}$, $\mathbf{H} \parallel \mathbf{c}$.

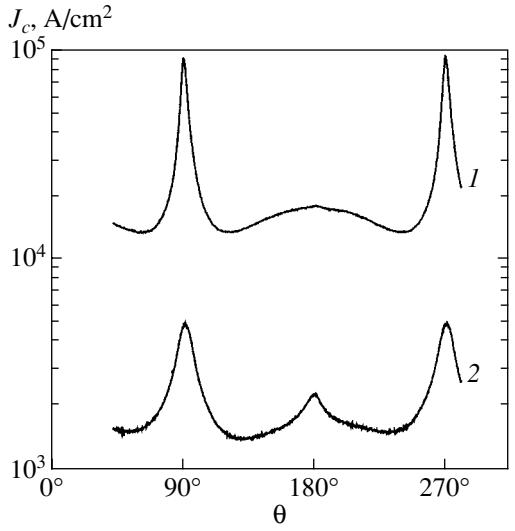


Fig. 4. Angular dependences of the critical current density J_c in the **ab** plane (curve *1*) and along the **c** axis (curve *2*) upon a variation of the orientation of the magnetic field **H** in the plane of the superconducting disk; T = 77 K, H = 10 kOe. The angle is measured from the **c** axis.

carrying capacity of our sample. Hence, it is expedient to analyze the effect of this anisotropy on the peculiarities of the magnetic field penetration into the superconductor. For this purpose, we carried out a series of measurements of the dependence of the relative losses q on the amplitude h_0 of the ac field oriented at an arbitrary angle to the **c** axis for various magnitudes and directions of the static magnetic field. First of all, we consider the simplest case when field $\mathbf{h}(t)$ is oriented along one of the principal directions. The corresponding results are presented in Fig. 5. Curve *1* in this figure corresponds to the direction of $\mathbf{h}(t)$ along the **c** axis, while curve *2* was obtained for $\mathbf{h}(t)$ parallel to the **ab** plane. Each curve has a peak associated with the size effect, which can be described as follows. It is well known that the penetration depth of an ac field in hard superconductors in the critical state increases with the amplitude h_0, and the field reaches the middle of the plate for the value of h_0 equal to the penetration field

$$H_p = 2\pi J_c d/c. \qquad (4)$$

The amplitude $h_0 = h_m$ for which the relative losses attain the maximum value is connected with the penetration field through the relation $h_m = (4/3)H_p$. A comparison of Figs. 5 and 3 shows that the peaks in Fig. 5 are obviously connected with the size effect. Indeed, the ac field oriented along the **c** axis is screened by strong currents of density J_{cy}, while the field $\mathbf{h}(t) \parallel \mathbf{ab}$ is screened by weak currents of density J_{cz}. Accordingly, the peak on curve *1* is observed for higher values of h_0 than the peak on curve *2*. Similarly, the peak on the dashed curve obtained in a weak static magnetic field (for large values of the critical current density) than on curve *1* is displaced towards larger values of amplitude h_0. Direct calculations show that the positions of the peaks correlate to a high degree of accuracy with formula (4) and with the corresponding values of the critical density of screening currents. Finally, the height of the peaks is approximately equal to 0.24,

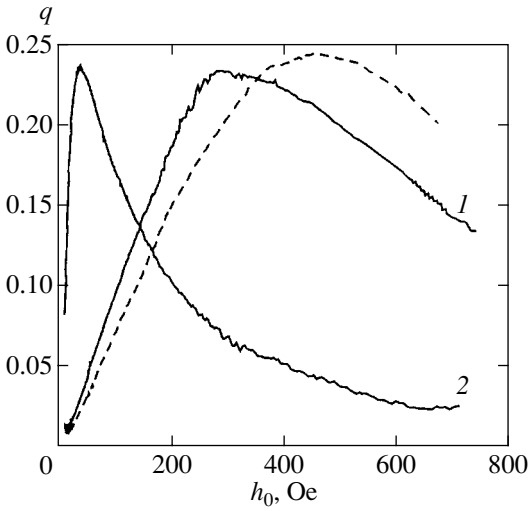


Fig. 5. Dependence of the relative losses q in a superconducting disk on the amplitude of the ac magnetic field $\mathbf{h}_0$ oriented along one of the principal directions for the orientation of **H** along the **c** axis. Curve *1* corresponds to $\mathbf{h}_0 \parallel \mathbf{c}$, and curve *2* to $\mathbf{h}_0 \parallel \mathbf{ab}$. Both curves were obtained for **H** = 10 kOe. The dashed curve corresponds to H = 5 kOe, $\mathbf{h}_0 \parallel \mathbf{c}$, and T = 77 K.

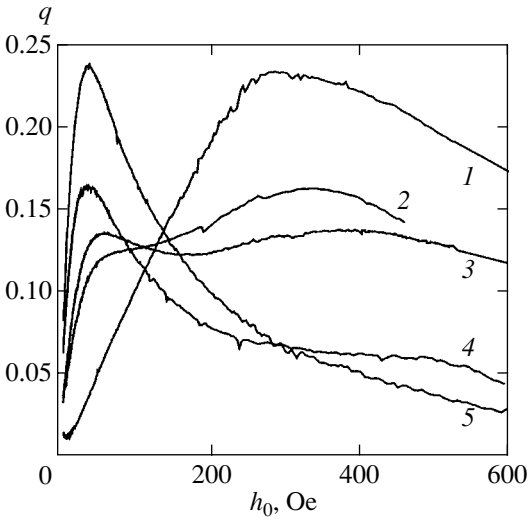


Fig. 6. Dependence of relative losses q in a superconducting disk on the amplitude of the ac magnetic field h_0 for various directions of the ac magnetic field in a constant magnetic field H = 10 kOe oriented along the **c** axis, T = 77 K. Curves *1*–*5* correspond to angles θ equal to 0°, 38°, 45°, 60°, and 90°.

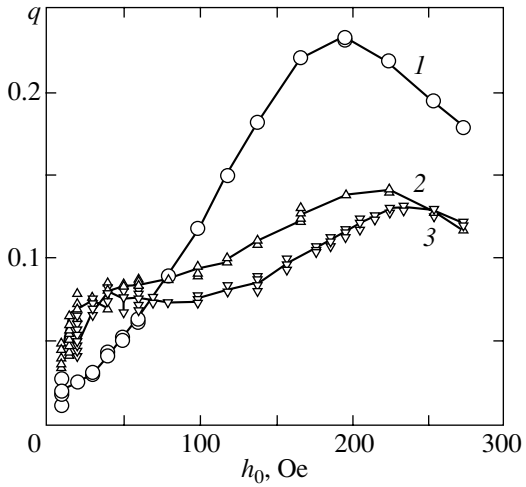


Fig. 7. Dependence of the relative losses q on the amplitude of the varying magnetic field h_0 for various directions of the ac magnetic field for a single crystal with a single preferred direction of twinning boundaries at T = 77 K: h_0 is directed perpendicularly to the twinning boundary, H = 4 kOe (curve *1*), $\mathbf{h}_0$ forms the angle of 45° with the boundary, H = 4 kOe (curve *2*), and $\mathbf{h}_0$ also forms the angle of 45° with the twinning boundary, H = 6 kOe (curve *3*).

which corresponds to the height of the size-effect peak observed for isotropic superconductors.

The behavior of the $q(h_0)$ curves changes significantly upon a deviation of vector $\mathbf{h}(t)$ from the principal directions. The transformation of the $q(h_0)$ curves upon an increase in the angle θ between $\mathbf{h}_0$ and the **c** axis is demonstrated in Fig. 6. For certain values of θ, the relative losses $q(h_0)$ display two peaks instead of one (e.g., curve *3* in Fig. 6). These peaks are observed virtually for the same values of h_0 as the peaks on curves *1* and *5*, corresponding to the orientation of $\mathbf{h}_0$ along the principal directions. Measurements made for other magnitudes and directions of the static magnetic field proved that each peak is associated with the size effect. The penetration field H_p (4) corresponding to the left peak is such as if the field were screened by a single (small) critical current density component J_{cz}. The other peak corresponds to the screening with the strong current (J_{cy}). The situation is interesting in that it is as if the ac field penetrates to the middle of the sample twice upon an increase in the amplitude. This apparent contradiction is removed if we assume that the peaks correspond to successive penetration of two different components (y and z) of the ac field. The difference in the penetration fields of the induction components B_y and B_z may be due to the fact that they are screened independently: each component is screened by only one ("own") component of the critical current density (J_{cz} and J_{cy}, respectively).

Let us consider in greater detail the transformation of the $q(h_0)$ curves upon a deviation of $\mathbf{h}_0$ from the **c** axis. First, an increase in angle θ slightly displaces the peak towards larger values of amplitude h_0, and the peak height becomes smaller. For θ ≈ 38°, a point of inflection with a horizontal tangent (curve *2* in Fig. 6) appears on the left of the peak on the $q(h_0)$ curve. For θ > 38°, the left peak appears so that the relative losses are characterized by two peaks in the angular interval 38° < θ < 60°. The right peak continues its displacement towards larger amplitudes h_0 and its height decreases. For θ = 60°, this peak vanishes (curve *4*). Conversely, the left peak emerging for θ = 38° becomes higher and is displaced towards smaller amplitudes. Finally, for θ = 90° (curve *5*), the left peak acquires the extreme left position.

Such a dynamics of the variation of the shape of the $q(h_0)$ curves was observed for all the samples under investigation with the **c** axis in their plane. However, the specific values of the critical angles for which additional peaks emerge or vanish were different for different samples.

The experiments proved that the size effect manifested in the form of two peaks on the dependence of the relative losses on the ac field amplitude is observed for superconductors with different origin of anisotropy in the critical current density in the plane of the sample. By way of illustration, Fig. 7 shows the results of measurements of $q(h_0)$ for a single crystal with a single preferred direction of twinning boundaries which are sources of anisotropy.

3. DISCUSSION

3.1. Model of the Critical State of Anisotropic Superconductors

The relative losses in superconductors in the critical state are usually calculated on the basis of Bean's model [13]. According to this model, the magnetic induction distribution $\mathbf{B}(\mathbf{r}, t)$ is determined by a single isotropic parameter, viz., the critical current density $J_c(\mathbf{B})$. The equation for the critical state can be written in the form

$$\operatorname{curl}\mathbf{B} = \frac{4\pi}{c} J_c \frac{\mathbf{E}}{E}, \tag{5}$$

where $\mathbf{E}$ is the electric field strength. Bean's equation correctly describes the static and low-frequency electrodynamic properties of hard isotropic superconductors in the case when the external field is considerably higher than the lower critical field.

In order to describe the experimental results described by us here, we will use a simple model [9] generalizing Bean's model of the critical state to the anisotropic case. Let us consider a plane-parallel infinitely large superconducting plate in an external field $\mathbf{H}$ directed along its surface. We assume that all the fields and currents are functions of only one spatial coordinate x perpendicular to the plane of the plate. The origin $x = 0$ is located on the sample surface. We will formulate the anisotropic model of the critical state for the simplest anisotropic case characterized by two mutually orthogonal principal directions y and z in the plane of the plate. These directions coincide with the crystallographic axes of the sample or with the characteristic directions of the defect structure. If the applied field $\mathbf{h}(t)$ is oriented along one of such directions (y or z), there exists a single component of screening currents (J_{cz} or J_{cy}) perpendicular to $\mathbf{h}(t)$. In this case, only the magnitude $\mathbf{B}(x, t)$ of magnetic induction changes in the sample. Its distribution, like in the isotropic case, is determined by Bean's equation (5) with the corresponding current density. In the chosen geometry, the critical state is described by one of the following equations:

$$\begin{aligned} -\frac{\partial B_z}{\partial x} &= \frac{4\pi}{c} J_{cy}(B_y, B_z)\operatorname{sgn} E_y, \\ \frac{\partial B_y}{\partial x} &= \frac{4\pi}{c} J_{cz}(B_y, B_z)\operatorname{sgn} E_z \end{aligned} \tag{6}$$

depending on the principal direction along which the vector $\mathbf{h}(t)$ is oriented. On the basis of precisely these equations, we calculated the relation between the imaginary component χ'' of the dynamic magnetic susceptibility and the critical current density used for constructing the curves in Figs. 2 and 3.

If the ac magnetic field is tilted relative to the principal directions, both components of the critical current density (J_{cy} and J_{cz}) participate in its screening. In this case, the equations of the critical state can be written in the form

$$\begin{aligned} -\frac{\partial B_z}{\partial x} &= \frac{4\pi}{c} J_{cy}(B_y, B_z)\cos\phi(x), \\ \frac{\partial B_y}{\partial x} &= \frac{4\pi}{c} J_{cz}(B_y, B_z)\sin\phi(x), \end{aligned} \tag{7}$$

where $\phi(x)$ is the angle between the vector $\mathbf{E}(x)$ and the y axis.

As the conventional model of the critical state, the system of equations (7) takes into account an important property of hard superconductors. The density of the current flowing in a certain direction decreases if an orthogonal component of the current is excited at the same point of the sample. The factors $\cos\phi$ and $\sin\phi$ in Eqs. (7) take this circumstance into account. The above-mentioned property of hard superconductors leads to an interesting effect: the collapse of the transport current [14] and of the static magnetic moment [15] under the action of the ac magnetic field.

Equations (7) should be supplemented with equations for the electric field. We can write Faraday's law in this geometry in the form

$$-\frac{\partial E_y}{\partial x} = \frac{1}{c}\frac{\partial B_z}{\partial t}, \quad \frac{\partial E_z}{\partial x} = \frac{1}{c}\frac{\partial B_y}{\partial t}. \tag{8}$$

Maxwell's equations (7) and (8) must be solved together with the boundary conditions

$$\mathbf{B}(0, t) = \mathbf{B}(d, t) = \mathbf{H} + \mathbf{h}(t). \tag{9}$$

The proposed model describes an interesting scenario of the penetration of an electromagnetic field into an anisotropic superconductor. It follows from the boundary conditions (9) that the varying component of the magnetic induction at the sample surface coincides in the magnitude and direction with the field $\mathbf{h}(t)$. In accordance with Eqs. (7), each magnetic induction component is screened by critical currents of various density. Thus, as the ac field penetrates to the bulk of the sample, the spatial orientation of vector $\mathbf{B}(x, t)$ changes and, hence, vector $\mathbf{E}(x, t)$ rotates. This means that angle ϕ becomes a function of coordinate x. Such a rotation of vector $\mathbf{E}(x, t)$ proceeds until one of the components (y or z) of the varying magnetic induction component vanishes. For large values of x, the penetrating varying magnetic flux is presented only by the second (nonattenuating) component (z or y). Consequently, different magnetic induction components penetrate into an anisotropic superconductor to different depths. Ultimately, this determines the experimentally observed peculiarities in the behavior of relative losses as a function of the ac field amplitude.

3.2. Calculation of Relative Losses $q(h_0)$

The calculation of the relative losses $q(h_0)$ in the framework of the anisotropic model of the critical state

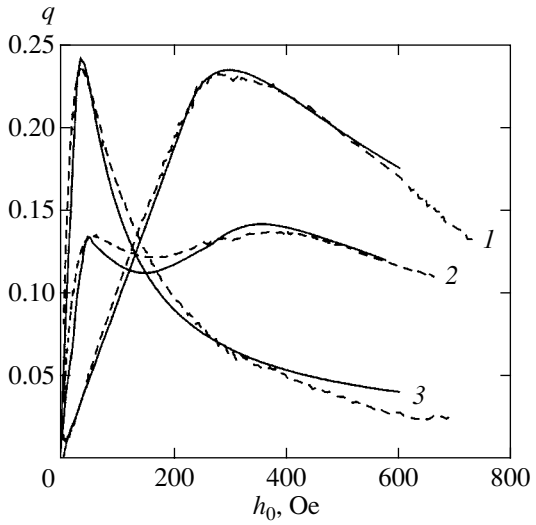


Fig. 8. Calculated dependences of relative losses q on the amplitude of the ac magnetic field h_0 for different directions of $\mathbf{h}(t)$: field $\mathbf{h}(t)$ is oriented along the **c** axis $\theta = 0°$ (*1*), 45° (*2*), and 90° (*3*). The dashed curves correspond to the results of measurement of the losses in the superconducting disk.

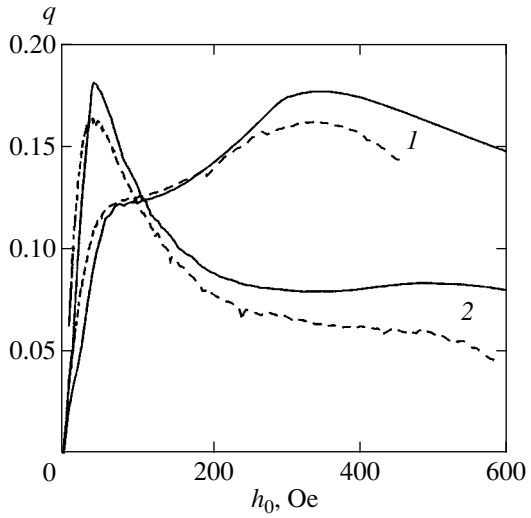


Fig. 9. Calculated dependences of relative losses q on the amplitude of the ac magnetic field h_0 for different orientations of $\mathbf{h}(t)$ in the vicinity of threshold angles: field $\mathbf{h}(t)$ is directed at angle $\theta = 38°$ to the **c** axis (*1*) and 60° (*2*). The dashed curves correspond to experiment.

is a complicated nonlinear problem. In accordance with the experimental conditions, this problem should be simplified assuming that the dependence of both principal components of the critical current density on the ac magnetic field can be disregarded for $h_0 \ll H$. Besides, we can neglect the coordinate dependence of the constant component of the magnetic induction since the experimental value of the static magnetic field was much larger than the penetration field H_p. Thus, the current densities $J_{cy}(B_y, B_z)$ and $J_{cz}(B_y, B_z)$ in Eqs. (7) should be replaced by their values in the static field **H**.

The proposed model was used in [9] for interpreting of the observed size effect in an anisotropic superconductor. In this publication, an additional simplification was used for obtaining the dependence of the relative losses on the amplitude of the ac field in analytic form. Instead of the angle ϕ between the electric field and the y axis, which is a function of coordinate x, the constant angle θ formed by vector $\mathbf{h}_0$ with the z axis was substituted into Eqs. (7). In spite of the fact that such a substitution is justified only for isotropic samples, the results obtained in [9] were in qualitative agreement with experiment. In the present work, we calculated the relative losses $q(h_0)$ for the first time in the framework of the anisotropic model of the critical state without using some uncontrollable approximations.

The results of calculation of functions $q(h_0)$ for angles $\theta = 0°$, 45°, 90° are presented by solid curves in Fig. 8. While calculating such parameters as the sample thickness, the values of critical current densities along the principal directions of anisotropy in the field of 10 kOe were assumed to be in accord with the results of measurements for a superconducting disk (see Fig. 3). For the sake of comparison, the dashed curves in the same figure represent the experimental curves *1*, *3*, and *5* from Fig. 6, which were obtained for the same directions of the ac field. As expected, the $q(h_0)$ curves for the orientation of $\mathbf{h}_0$ along the principal directions of anisotropy have only one peak corresponding to the penetration of the ac field to the middle of the sample. For $\theta = 45°$, the theoretical curve displays two size peaks each of which is associated with the penetration of one of the magnetic flux components. The result obtained earlier in [9] was qualitatively the same. However, the position of the theoretical peak in [9] was displaced towards smaller amplitudes h_0 upon the deviation of the magnetic field $\mathbf{h}_0$ from the **c** axis, which contradicted the experimental results. In the present work, the position and height of the peaks vary with increasing angle θ in the same way as on the experimental curve owing to the correctly estimated spatial variation of the orientation of vector $\mathbf{E}(x)$. The calculated values of the threshold angles θ for which a second peak appears or disappears on the $q(h_0)$ curves are also in conformity with experiment. The quantitative behavior of the $q(h_0)$ curves for the threshold angles is also in satisfactory agreement with experiment (see Fig. 9).

Thus, the proposed anisotropic model of the critical state makes it possible to describe an aggregate of electromagnetic properties of hard superconductors not only qualitatively, but also quantitatively. This model is found to be convenient not only for interpreting the results of low-frequency experiments, but also for explaining the observed features of the static magnetization curves, which are associated with anisotropy of the superconductor [10]. This model operates with simple phenomenological parameters having clear physical

meaning, namely, the critical current density components along the principal directions of sample anisotropy. As in the conventional Bean's model, we disregard here the nature of the current-carrying capacity of the superconductor and the sources of its anisotropy. It appears that the model can be applied for superconductors with anisotropy of the crystallographic origin as well as for samples with a clearly manifested anisotropy of the defect structure. Indeed, the dependence of the relative losses $q(h_0)$ measured on a single crystal with a single preferred direction of twinning boundaries (see Fig. 7) is qualitatively the same as in a grain-oriented sample with the **c** axis lying in the plane of the sample. In both cases, we have a pair of two mutually orthogonal anisotropy directions in the plane of the sample, and the proposed model (7) takes into account precisely this type of anisotropy. In principle, the proposed anisotropic model of the critical state can be easily generalized to superconductors with another symmetry of the critical current density.

ACKNOWLEDGMENTS

The authors are grateful to V.E. Shapovalov for his help in computer calculations.

This work was supported financially by the Russian Program on Superconductivity (project nos. 100 292 and 100 199), INTAS-RFBR (project no. IR-97-1394), and the Russian Foundation for Basic Research (project nos. 00-02-17145 and 01-02-6211).

REFERENCES

1. G. Blatter, M. V. Feigel'man, V. B. Geshkenbein, *et al.*, Rev. Mod. Phys. **66**, 1125 (1994).
2. L. F. Cohen and H. J. Jensen, Rep. Prog. Phys. **60**, 1581 (1997).
3. L. W. Conner and A. P. Malozemoff, Phys. Rev. B **44**, 403 (1991).
4. A. A. Zhukov, G. K. Perkins, Yu. V. Bugoslavsky, and A. D. Caplin, Phys. Rev. B **56**, 2809 (1997).
5. I. M. Babich and G. P. Mikitik, Phys. Rev. B **58**, 14 207 (1998).
6. I. F. Voloshin, A. V. Kalinov, K. I. Kugel', *et al.*, Zh. Éksp. Teor. Fiz. **111**, 2158 (1997) [JETP **84**, 1177 (1997)].
7. B. Rosenstein and A. Knigavko, Phys. Rev. Lett. **83**, 844 (1999).
8. L. M. Fisher, A. V. Kalinov, J. Mirković, *et al.*, Appl. Supercond. **2**, 639 (1994).
9. L. M. Fisher, A. V. Kalinov, S. E. Savel'ev, *et al.*, Physica C (Amsterdam) **350**, 152 (2001).
10. I. F. Voloshin, A. V. Kalinov, L. M. Fisher, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **73**, 324 (2001) [JETP Lett. **73**, 285 (2001)].
11. I. F. Voloshin, L. M. Fisher, N. M. Makarov, and V. A. Yampol'skiĭ, Pis'ma Zh. Éksp. Teor. Fiz. **51**, 225 (1990) [JETP Lett. **51**, 255 (1990)].
12. L. M. Fisher, V. S. Gorbachev, N. M. Makarov, *et al.*, Phys. Rev. B **46**, 10 986 (1992).
13. C. P. Bean, Phys. Rev. Lett. **8**, 250 (1962).
14. I. V. Baltaga, L. M. Fisher, N. M. Makarov, *et al.*, Fiz. Nizk. Temp. **21**, 411 (1995) [Low Temp. Phys. **21**, 320 (1995)].
15. L. M. Fisher, A. V. Kalinov, I. F. Voloshin, *et al.*, Solid State Commun. **97**, 833 (1996).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 93, No. 5, 2001, pp. 1113–1118.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 54, 2001, pp. 1282–1287.

MISCELLANEOUS

Double Potts Chain and Exact Results for Some Two-Dimensional Spin Models¶

M. A. Yurishchev[a, b]

[a]*The Abdus Salam International Centre for Theoretical Physics, Trieste, 34100 Italy*
[b]*Vasilsursk Laboratory, Radiophysical Research Institute,*
Vasilsursk, Nizhni Novgorod oblast, 606263 Russia
e-mail: yur@itp.ac.ru

Received April 12, 2001

Abstract—An exact analytical solution for the q-state Potts model on a $2 \times \infty$ ladder with arbitrary two-, three-, and four-site interactions in a unit cell in presented is a closed form. This solution is used to show that the finite-size internal energy equation [6] yields an accurate value of the critical temperature for the triangular Potts lattice with three-site interactions in alternate triangular faces. It is argued that the above equation is exact at least for self-dual models on isotopic strips. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The methods that allow one to extract information about a multidimensional system from solutions of its counterparts of lower-dimension play an important role in statistical physics. One of the most well-known examples of this kind is the finite-size scaling approach [1, 2].

There are cases that evoke particular interest when the critical properties of a system experiencing a phase transition can be exactly determined from the data pertaining to its subsystems. For instance, for the Ising strips, the intersection point of the partition function zero locus in a complex temperature plane with the real positive axis yields the exact value of the critical temperature for the two-dimensional Ising model [3]. Exact critical temperatures for the $S = 1/2$ Ising models on square, triangular, honeycomb, and centered square (Union Jack) anisotropic lattices are obtained by using strip clusters when an effective field is applied to one side of the strip only [4]. Another exotic way of estimating the critical point of the square-lattice Ising model was proposed in [5]. The authors of this paper showed that in the quasidiagonal form of a transfer matrix of a finite-width strip, all coefficients of the characteristic equation for the sub-block containing the largest eigenvalue have an extremum located precisely at the exact value of the phase transition temperature of the infinite lattice.

In the present paper, we concentrate our attention on the method to calculate the critical temperature proposed by Wosiek [6] (see also [7–12]). The author of [6] introduced a maximum criterion for the ratio of moments of the transfer matrix and obtained the following equation for determining the critical point position in a d-dimensional system:

$$u_1(K_c) = u_2(K_c). \quad (1)$$

Here, u_1 and u_2 are the respective internal energies of $(d-1)$-dimensional and two coupled $(d-1)$-dimensional subsystems and K_c is the critical coupling (the normalized inverse critical temperature) of the d-dimensional system.

It is remarkable that, at $d = 2$, Eq. (1) (see [6]) yields the exact value of K_c for the isotropic square and triangular Ising lattices, as well as for the three-site Potts model on the square lattice with isotropic interactions. Subsequently, several other models were added to the list, which now includes another isotropic Baxter model (two square Ising lattices coupled by four-particle interactions), the Baxter–Wu model (triangular lattice with three-site interactions of Ising spins) [10], and the q-state Potts model on an isotropic square lattice with an arbitrary value of q [12]. The physical nature of Eq. (1) can be elucidated when it either yields an exact solution or admits an approximate estimate, or does not give any solution at all for a given model.

For a two-dimensional system, Eq. (1) connects the internal energies of infinitely long linear and double chains. Therefore, in order to test Eq. (1) rigorously, it is necessary to have analytical solutions for such subsystems.

In Section 2, we give an exact analytical solution for the two-chain Potts strip with a large number of independent parameters. As a special case, it contains a solution for the linear Potts chain.

Our solution for the double Potts chain enables us to cover all the previously known cases where Eq. (1) exactly reproduces the critical temperatures for the two-dimensional Ising, Baxter–Wu, and Potts models. In addition, we discover (Section 3) a new model for which Eq. (1) yields the exact result. This is the q-state Potts model on the triangular lattice with purely three-site interactions in one-half of the triangular faces [13].

¶This article was submitted by the authors in English.

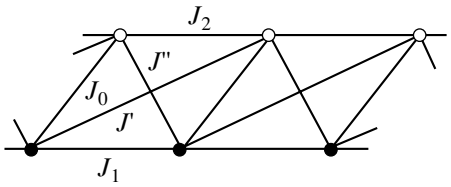


Fig. 1. Geometry of two-site couplings in the double q-state Potts chain with the $\mathbf{S}_q$ symmetry.

In Section 4, we discuss the results. In particular, we show that duality is a sufficient condition for the validity of Eq. (1) for isotropic spin lattices. In Section 5, we summarize the results obtained in the work.

2. SOLUTION OF THE DOUBLE q-STATE POTTS CHAIN WITH THE $\mathbf{S}_q$ SYMMETRY

We consider a two-chain (ladder) lattice with spin variables σ_l^i attached to its sites (i = 1, 2 is the chain index and i = 1, 2, 3, ... labels the sites in the longitudinal direction of the ladder); the spin variables take the values 1, 2, ..., q.

We write the Hamiltonian of the system as

$$\mathcal{H} = -\sum_l H(\sigma_l^1, \sigma_l^2; \sigma_{l+1}^1, \sigma_{l+1}^2). \quad (2)$$

The locality of interactions in this Hamiltonian allows us to introduce the transfer matrix V with the elements

$$\langle \sigma_1, \sigma_2 | V | \sigma_1', \sigma_2' \rangle = \exp \frac{H(\sigma_1, \sigma_2; \sigma_1', \sigma_2')}{k_B T} \quad (3)$$

(where T is the temperature and k_B is the Boltzmann constant) and reduce the problem of calculating the free energy density f of an infinitely long strip to finding the largest eigenvalue λ_1 of the matrix V:

$$f = \frac{1}{2} \ln \lambda_1. \quad (4)$$

Transfer matrix (3) has the size $q^2 \times q^2$. It is real and all its elements are positive, but the matrix is not symmetric in general ($V_{ij} \neq V_{ji}$).

To solve the eigenvalue problem for the transfer matrix, we use the group-theoretical approach (see, e.g., [14], where this approach was applied to a quasidiagonalization of the Ising model transfer matrix on parallelepipeds $L \times L \times \infty$). In order to obtain a solution for the two-leg spin ladder (in which we are particularly interested) in the most general form, we proceed in the reverse order. Namely, we first select a symmetry group in the space $|\sigma_1, \sigma_2\rangle$, which enables us to quasidiagonalize the transfer matrix up to sub-block secular equations that can be solved analytically; only then do we expand the Hamiltonian density H into a series in the invariants of the symmetry group.

We take a model that is invariant, e.g., under transformations of the symmetric group $\mathbf{S}_q$ of the degree q. For the Potts model, this means that we are dealing with a system in the zero external field. Fortunately, the field is not required to test Eq. (1).

It is known (see, e.g., [15]) that the largest eigenvalue of the transfer matrix is located in the sub-block of the identity irreducible representation. In accordance with group theory, the basis vectors ψ_i of the identity irreducible representation can be obtained by successively acting with the permutation operators of the $\mathbf{S}_q$ group on the orths $|1, 1\rangle$, $|1, 2\rangle$, … $|q, q\rangle$. Acting by elements of the symmetric group first on the orth $|1, 1\rangle$ and then on $|1, 2\rangle$, we find that the two linear combinations obtained involve all the orths. The normalized basis vectors are given by

$$\psi_1 = \frac{1}{\sqrt{q}} \sum_{i=1}^{q} |i, i\rangle, \quad \psi_2 = \frac{1}{\sqrt{q(q-1)}} \sum_{i,j=1}^{q}{}' |i, j\rangle \quad (5)$$

(the prime at the second sum indicates that the terms with $i = j$ are omitted). Hence, the sub-block of the identity irreducible representation has the size 2×2, and therefore, its eigenvalues (one of which is λ_1) can be easily obtained by solving an algebraic equation of only the second degree. We note that if we take the group $\mathbf{S}_q \times \mathbf{C}_s$ (where $\mathbf{C}_s$ is the group of mirror reflections in the plane placed between the chains of the two-leg ladder), the sub-block corresponding to the identity irreducible representation again has the size 2×2, and therefore, this symmetry (which only reduces the number of independent parameters in the Hamiltonian) does not justify itself in the given case.

We now represent Hamiltonian (2) as a sum of terms that are invariant under transformations of the (4) group $\mathbf{S}_q$:

$$\begin{aligned}\mathcal{H} = -\sum_l [& J_1 \delta_{\sigma_l^1 \sigma_{l+1}^1} + J_2 \delta_{\sigma_l^2 \sigma_{l+1}^2} + J_0 \delta_{\sigma_l^1 \sigma_l^2} \\ & + J' \delta_{\sigma_l^1 \sigma_{l+1}^2} + J'' \delta_{\sigma_l^2 \sigma_{l+1}^1} + J_3 \delta_{\sigma_l^1 \sigma_l^2 \sigma_{l+1}^1} + J_3' \delta_{\sigma_l^1 \sigma_{l+1}^1 \sigma_{l+1}^2} \\ & + \tilde{J}_3 \delta_{\sigma_l^1 \sigma_l^2 \sigma_{l+1}^2} + \tilde{J}_3' \delta_{\sigma_l^2 \sigma_{l+1}^1 \sigma_{l+1}^2} + J_4 \delta_{\sigma_l^1 \sigma_l^2 \sigma_{l+1}^1 \sigma_{l+1}^2}].\end{aligned} \quad (6)$$

The Kronecker symbols entering here are defined as

$$\delta_{\sigma_1 \ldots \sigma_k} = \begin{cases} 1, & \text{if } \sigma_1 = \ldots = \sigma_k \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

The structure of the two-site couplings in Hamiltonian (6) is shown in Fig. 1. Matrix elements of the original

transfer matrix are written as

$$\langle \sigma_1, \sigma_2 | V | \sigma_1', \sigma_2' \rangle = \exp\Big[K_1 \delta_{\sigma_1\sigma_1'} + K_2 \delta_{\sigma_2\sigma_2'} + \frac{1}{2} K_0 (\delta_{\sigma_1\sigma_2} + \delta_{\sigma_1'\sigma_2'}) + K' \delta_{\sigma_1\sigma_2'} + K'' \delta_{\sigma_2\sigma_1'} + K_3 \delta_{\sigma_1\sigma_2} \delta_{\sigma_1\sigma_1'} + K_3' \delta_{\sigma_1\sigma_1'} \delta_{\sigma_1\sigma_2'} + \tilde{K} \delta_{\sigma_1\sigma_2} \delta_{\sigma_1\sigma_2'} + \tilde{K}' \delta_{\sigma_2\sigma_2'} \delta_{\sigma_2\sigma_1'} + K_4 \delta_{\sigma_1\sigma_1'} \delta_{\sigma_1\sigma_2} \delta_{\sigma_1\sigma_2'} \Big], \quad (8)$$

where

$$K_0 = J_0/k_BT, \quad K_1 = J_1/k_BT, \quad K_2 = J_2/k_BT,$$
$$K' = J'/k_BT, \quad K'' = J''/k_BT, \quad K_3 = J_3/k_BT,$$
$$K_3' = J_3'/k_BT, \quad \tilde{K}_3 = \tilde{J}_3/k_BT, \quad \tilde{K}_3' = \tilde{J}_3'/k_BT,$$
$$K_4 = J_4/k_BT.$$

Using Eqs. (5) and (8), we calculate the matrix elements

$$Q_{ij} = \psi_i^+ V \psi_j$$

of the sub-block corresponding to the identity irreducible representation:

$$\begin{aligned} Q_{11} &= [q - 1 + \exp(K_1 + K_2 + K' + K'' + K_3 + K_3' + \tilde{K}_3 + \tilde{K}_3' + K_4)] \exp K_0, \\ Q_{12} &= (q-1)^{1/2} [q - 2 + \exp(K_1 + K'' + K_3) + \exp(K_2 + K' + \tilde{K}_3)] \exp(K_0/2), \\ Q_{21} &= (q-1)^{1/2} [q - 2 + \exp(K_1 + K' + K_3') + \exp(K_2 + K'' + \tilde{K}_3')] \exp(K_0/2), \\ Q_{22} &= (q-2)(q - 3 + e^{K_1} + e^{K_2} + e^{K'} + e^{K''}) + \exp(K_1 + K_2) + \exp(K' + K''). \end{aligned} \quad (9)$$

As a result, we find that the largest eigenvalue of the transfer matrix of the double q-state Potts chain with Hamiltonian (6) is given by

$$\lambda_1^{(2)} = \frac{1}{2}(Q_{11} + Q_{22}) + \Big[\frac{1}{4}(Q_{11} - Q_{22})^2 + (q-1) A \exp K_0 \Big]^{1/2}, \quad (10)$$

where

$$A = [q - 2 + \exp(K_1 + K'' + K_3) + \exp(K_2 + K' + \tilde{K}_3)] \times [q - 2 + \exp(K_1 + K' + K_3') + \exp(K_2 + K'' + \tilde{K}_3')]. \quad (11)$$

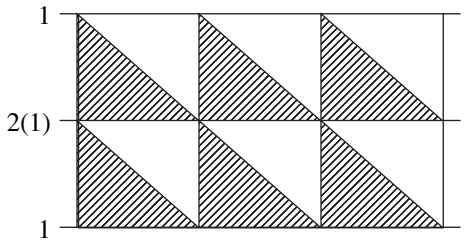


Fig. 2. Fragment of the Potts lattice with three-site interactions in alternate triangular faces (shaded).

The versions of the double Potts chains solved previously [3, 12, 16–18] correspond to a particular choice of the interaction constants. Setting

$$J_0 = J'(= J)$$

with all the other interaction constants vanishing, we arrive at the solution for the linear Potts chain [19],

$$\lambda_1^{(1)}(K) = e^K + q - 1. \quad (12)$$

3. THE TRIANGULAR POTTS LATTICE WITH THREE-SITE INTERACTIONS ON ALTERNATE TRIANGLE FACES

A large number of independent parameters in the model solved in the previous section enables us to test Eq. (1) for a wide class of two-dimensional spin systems.

In addition to the cases listed in the Introduction, in which Eq. (1) is satisfied exactly, we consider the Potts model on a triangular lattice with three-site interactions in each up-triangle (Fig. 2). The position of the critical point in this model was found with both three- and two-site interactions [13]. However, it is known [12] that for the triangle lattice with pair couplings, Eq. (1) yields the exact result only for the Ising case ($q = 2$). We therefore discuss the model with purely three-site interactions. In this case,

$$K_c = \ln(1 + q). \quad (13)$$

We now show that this value satisfies Eq. (1) by subsystems in the shape of strips with the periodic boundary condition in the transverse direction.

The internal energy of the one-dimensional subsystem is

$$u_1(K) \equiv \frac{\partial f_1}{\partial K} = [(q-1)e^{-K} + 1]^{-1}. \quad (14)$$

Substituting $K_3 = \tilde{K}_3$ with all the other interaction constants vanishing, we obtain from (4), (10), and (11) the

free energy density of the double Potts chain:

$$f_2(K) = \frac{1}{2}\ln\left[\frac{1}{2}(e^{2K}+q^2-1) + \left[\frac{1}{4}(e^{2K}-(q-1)^2)^2 + q(q-1)(2e^K+q-2)\right]^{1/2}\right]. \quad (15)$$

The internal energy is given by

$$u_2(K) = \frac{\partial f_2}{\partial K}.$$

Differentiating Eq. (15) with respect to K, we find the expression for $u_2(K)$.

The analysis shows that the dependences $u_1(K)$ and $u_2(K)$ have a crossing point that lies exactly at

$$K = K_c = \ln(1+q)$$

both for integer and noninteger q. The internal energy of the system at the critical point is given by

$$u_\infty(K_c) = u_1(K_c) = u_2(K_c) = \frac{1}{2}(1+q^{-1}). \quad (16)$$

Thus, using solutions for only the linear and double Potts chains, Eq. (1) has enabled us to extract the exact value of K_c for the bulk two-dimensional Potts model on a triangular lattice with alternating faces that interact by three-site forces.

4. DISCUSSION

In [8], Eq. (1) was extended to

$$u_L(K_c) = u_{L'}(K_c), \quad L, L' = 1, 2, 3, \ldots, \quad (17)$$

where u_L is the internal energy per site of L coupled ($d-1$)-dimensional subsystems. In the two-dimensional case, L denotes the width of the strip.

The validity of condition (17) for arbitrary L and L' means the absence of a "singular" (i.e., L-dependent) part of the internal energy density at the critical point,

$$u_L(K_c) = \text{const on } L. \quad (18)$$

In other words, the amplitudes of all finite-size corrections to the critical internal energy of the system $u_\infty(K_c)$ are equal to zero.

For the square isotropic Ising lattice, the derivative of the inverse correlation length $\kappa_L(K)$ with respect to the temperature-like variable K has a similar property [20, 21],

$$\left.\frac{\partial \kappa_L}{\partial K}\right|_{K=K_c} = \left.\frac{\partial \kappa_{L'}}{\partial K}\right|_{K=K_c}, \quad (19)$$

i.e., $\partial\kappa_L/\partial K|_c$ does not depend on L. This property has enabled us to exactly determine the value of the thermal critical exponent $y_t(=1)$ for this model using only the finite-size data [20, 21].

Equations (1) and (17) are valid for the ferromagnetic isotropic square Potts lattices. These models are self-dual and their critical coupling (in the anisotropic case) is determined from the condition

$$(\exp K_x - 1)(\exp K_y - 1) = q. \quad (20)$$

For the antiferromagnetic square-lattice Potts model, the criticality condition is [22]

$$(\exp K_x + 1)(\exp K_y + 1) = 4 - q, \quad (21)$$

where $K_x < 0$ and $K_y < 0$. We performed a verification and found that in the antiferromagnetic case, the curves $u_1(K)$ and $u_2(K)$ do not have any self-crossing point, and therefore, Eq. (1) does not lead to the exact value that follows from Eq. (21), nor to any approximate estimate for the critical point.

It is not difficult to show that if the model is self-dual and the dual point therefore coincides with the original one, Eqs. (1) and (17) are valid.

Indeed, we consider for instance the Ising model on the isotropic square lattice $L \times N$ with toroidal boundary conditions. The partition function of this system has a fundamental property: it is invariant (up to a multiplicative factor exponentially depending on L) under the duality transformation (see [23]),

$$Z_{L,N}(K^*) = (\sinh 2K)^{-LM} Z_{L,M}(K), \quad (22)$$

where K and K^* are related by

$$\tanh K^* = \exp(-2K). \quad (23)$$

(We here used another normalization of the exchange constant in the Ising model, namely $J_{\text{Potts}} = 2J_{\text{Ising}}$.) In the limit of an infinitely long strip ($N \longrightarrow \infty$), Eq. (22) transforms to the duality condition for the largest eigenvalue:

$$\lambda_1^{(L)}(K^*) = (\sinh 2K)^{-L}\lambda_1^{(L)}(K). \quad (24)$$

This implies that the values of the normalized internal energy in dually conjugated points (K and K^*) are related by

$$u_L(K^*)\frac{\partial K^*}{\partial K} = u_L(K) - 2u_0(K), \quad (25)$$

where the additive term $u_0(=\coth 2K)$ does not depend on L. Another important feature related to the isotropy of the lattice is that the dually conjugated points K and K^* merge into one point at criticality,

$$K^* = K = K_c. \quad (26)$$

Using Eq. (23), we find that at the critical point

$$K_c = \frac{1}{2}\ln(1+\sqrt{2}),$$

the derivative

$$\left.\frac{\partial K^*}{\partial K}\right|_c = -1.$$

Consequently,

$$u_L(K_c) = u_0(K_c) = \sqrt{2}. \quad (27)$$

Thus, the critical internal energy per site $u_L(K_c)$ of an Ising cylinder with isotropic square cells satisfies condition (18) for all $L = 1, 2, \ldots$. This, in turn, leads to the validity of Eqs. (1) and (17).

Similarly, Eqs. (1) and (17) can be derived for other isotropic spin model partition functions that satisfy a functional equation like

$$Z_L(K^*) = [g(K)]^L Z_L(K). \quad (28)$$

In the cases where the model is self-dual but the critical manifold is a line or a surface (as, e.g., for anisotropic lattices), Eqs. (1) and (17) no longer hold. This is not difficult to prove if we again consider the two-dimensional Ising model. For the anisotropic square lattice, the duality condition becomes

$$\lambda_1^{(L)}(K_x^*, K_y^*) = [\sinh(2K_x)\sinh(2K_y)]^{-L/2}\lambda_1^{(L)}(K_x, K_y) \quad (29)$$

with

$$\tanh K_x^* = \exp(-2K_y) \text{ and } \tanh K_y^* = \exp(-2K_x). \quad (30)$$

It then follows that on the critical line

$$\sinh(2K_x)\sinh(2K_y) = 1, \quad (31)$$

condition (29) relates the values of the free energy at distinct (dually conjugated) points (K_x, K_y) and (K_y, K_x),

$$f_L(K_x, K_y) = f_L(K_y, K_x) + \frac{1}{2}\ln[\sinh(2K_x)\sinh(2K_y)]. \quad (32)$$

This violates Eqs. (1) and (17), which identify the internal energies at the same point.

The critical internal energy density of the strip $L \times \infty$ cut out from an anisotropic lattice depends on the size L. This is easy to verify using the results of Section 2 if one calculates the values $u_1(K_c)$ and $u_2(K_c)$ for the anisotropic Ising and Potts lattices.

On the other hand, we can establish the same property if we take the Onsager solution [24] for the two-dimensional Ising model. The dominant eigenvalue of the transfer matrix of the cylinder $L \times \infty$ with spatially anisotropic interactions is given by

$$\lambda_1^{(L)}(K; \alpha) = [2\sinh(2K)]^{L/2} \times \exp\left[\frac{\gamma_1 + \gamma_3 + \ldots + \gamma_{2L-1}}{2}\right], \quad (33)$$

where $\alpha = J_y/J_x$ is the lattice anisotropy parameter and γ_r are positive solutions of the equations

$$\cosh\gamma_r = \cosh(2\alpha K)\coth(2K) - \frac{\sinh(2\alpha K)}{\sinh(2K)}\cos\left(\frac{\pi r}{L}\right). \quad (34)$$

From this, we obtain the internal energy per site

$$u_L(K; \alpha) = \coth(2K) + \frac{1}{2L}\left(\frac{\partial\gamma_1}{\partial K} + \frac{\partial\gamma_3}{\partial K} + \ldots + \frac{\partial\gamma_{2L-1}}{\partial K}\right). \quad (35)$$

The functions $\gamma_r(K)$ have a smooth extremum (minimum) that in the isotropic case ($\alpha = 1$) lies exactly at $K = K_c$, and therefore,

$$\left.\frac{\partial\gamma_r}{\partial K}\right|_{K = K_c} = 0 \quad (r \neq 0). \quad (36)$$

As a result, the second term in Eq. (35) disappears and the critical internal energy ceases to depend on L. When $\alpha \neq 1$, Eq. (36) is not valid and $u_L(K; \alpha)$ depends on the strip width in a complicated way. This explains the failure of the exact calculations of K_c from Eq. (1) in the anisotropic Ising lattice [10].

In closing this section, we note that in spite of Eqs. (1) and (17), Eq. (19) cannot be deduced from the dual invariance of the system.

5. CONCLUSIONS

Using the group-theoretical approach, we obtained the exact analytical solution for the double Potts chain with Hamiltonian (6). The solution allows for examining Eq. (1) for a large number of models with Ising ($q = 2$) and arbitrary Potts spins (including noninteger q). The validity of Eq. (1) for the triangular Potts lattice with pure by three-site interactions in alternate triangular faces was established.

We have also shown that Eqs. (1) and (17) are a consequence of the duality symmetry of models for which the critical point coincides with its dual image.

As far as the author knows, the inverse theorem has not been proved. Duality plus isotropy or, more precisely, self-conjugation of the critical point are not necessary conditions for Eq. (1). In general, therefore, there can exist systems that are not invariant under the duality transformation or a combination of the dual and star–triangle transformations, but for which all amplitudes of finite-size corrections to the critical internal energy (or to some other quantity) are equal to zero.

ACKNOWLEDGMENTS

The author thanks A.A. Belavin, A.A. Nersesyan, and A.M. Sterlin for fruitful discussions and comments. I am also grateful to the Abdus Salam International

Centre for Theoretical Physics (Trieste, Italy) for kind hospitality where this work was finished. The research presented in this paper is supported in part by the Russian Foundation for Basic Research (project no. 99-02-16472) and CRDF (grant no. RP1-2254).

REFERENCES

1. M. N. Barber, in *Phase Transitions and Critical Phenomena*, Ed. by C. Domb and J. L. Lebowitz (Academic, London, 1983), Vol. 8, p. 145.
2. V. Privman, in *Finite Size Scaling and Numerical Simulation of Statistical Systems*, Ed. by V. Privman (World Sci., Singapore, 1990), p. 1.
3. D. W. Wood, R. W. Turnbull, and J. K. Ball, J. Phys. A **20**, 3495 (1987).
4. A. Lipowski and M. Suzuki, J. Phys. Soc. Jpn. **61**, 4356 (1992).
5. A. I. Bugrij and V. N. Shadura, E-print archives BITP-84-66R (Kiev, 1984).
6. J. Wosiek, Phys. Rev. B **49**, 15 023 (1994).
7. Z. Burda and J. Wosiek, Nucl. Phys. B (Proc. Suppl.) **34**, 677 (1994).
8. A. Pelizzola, Phys. Rev. B **51**, 12 005 (1995).
9. P. Bialas, Z. Burda, P. Sawicki, and J. Wosiek, Phys. Rev. B **51**, 12 007 (1995).
10. I. Souza, C. Lobo, and J. Seixas, Phys. Rev. B **55**, 6356 (1997).
11. P. Sawicki, Nucl. Phys. B (Proc. Suppl.) **63A–C**, 628 (1998).
12. P. Sawicki, hep-lat/9801006.
13. R. J. Baxter, H. N. V. Temperly, and S. A. Ashley, Proc. R. Soc. London, Ser. A **358**, 535 (1977).
14. M. A. Yurishchev, Phys. Rev. B **50**, 13 533 (1994).
15. H. W. J. Blöte and M. P. Nightingale, Physica A (Amsterdam) **112**, 405 (1982).
16. R. Shrock, Physica A (Amsterdam) **283**, 388 (2000).
17. S.-C. Chang and R. Shrock, Physica A (Amsterdam) **286**, 189 (2000).
18. S.-C. Chang and R. Shrock, cond-mat/0007505.
19. R. B. Potts, Proc. Cambridge Philos. Soc. **48**, 106 (1952); C. Domb, J. Phys. A **7**, 1335 (1974).
20. M. A. Yurishchev, Nucl. Phys. B (Proc. Suppl.) **83–84**, 727 (2000).
21. M. A. Yurishchev, Zh. Éksp. Teor. Fiz. **118**, 380 (2000) [JETP **91**, 332 (2000)].
22. R. J. Baxter, Proc. R. Soc. London, Ser. A **383**, 43 (1982).
23. R. J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic, London, 1982).
24. L. Onsager, Phys. Rev. **65**, 117 (1944).

Journal of Experimental and Theoretical Physics, Vol. 93, No. 5, 2001, pp. 1119–1136.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 5, 2001, pp. 1288–1308.

MISCELLANEOUS

Thermodynamics and Topology of Disordered Systems: Statistics of the Random Knot Diagrams on Finite Lattices

O. A. Vasilyev[a, *] and S. K. Nechaev[a, b, **]

[a]*Landau Institute of Theoretical Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia*

[b]*UMR 8626, CNRS-Université Paris XI, LPTMS, Bat. 100, Université Paris Sud, 91405 Orsay Cedex, France*

**e-mail: vasilyev@itp.ac.ru*

***e-mail: nechaev@itp.ac.ru*

Received May 7, 2001

Abstract—The statistical properties of random lattice knots, the topology of which is determined by the algebraic topological Jones–Kauffman invariants, was studied by analytical and numerical methods. The Kauffman polynomial invariant of a random knot diagram was represented by a partition function of the Potts model with a random configuration of ferro- and antiferromagnetic bonds, which allowed the probability distribution of the random dense knots on a flat square lattice over topological classes to be studied. A topological class is characterized by the highest power of the Kauffman polynomial invariant and interpreted as the free energy of a q-component Potts spin system for $q \longrightarrow \infty$. It is shown that the highest power of the Kauffman invariant correlates with the minimum energy of the corresponding Potts spin system. The probability of the lattice knot distribution over topological classes was studied by the method of transfer matrices, depending on the type of local junctions and the size of the flat knot diagram. The results obtained are compared to the probability distribution of the minimum energy of a Potts system with random ferro- and antiferromagnetic bonds. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

New interesting problems are formulated, as a rule, in the boundary regions between traditional fields. This is clearly illustrated by the statistical physics of macromolecules, which arose due to the interpenetration of solid state physics, statistical physics, and biophysics. Another example of a new, currently forming field is offered by the statistical topology born due to merging of statistical physics, the theory of integrable systems, algebraic topology, and group theory. The scope of the statistical topology includes, on the one hand, mathematical problems involved in the construction of topological invariants of the knots and entanglements based on some solvable models and, on the other hand, the physical problems related to determination of the entropy of random knots and entanglements. In what follows, we will predominantly pay attention to problems of the latter kind that can be rather conventionally separated into a subfield of "probabilistic-topological" problems [1]. Let us dwell on this class of problems in more detail.

Consider a lattice embedded in a three-dimensional space for which Ω is the ensemble of all possible closed self-nonintersecting trajectories of N steps with a fixed point. Let ω denote some particular realization of such a trajectory. The task is to calculate the probability $P(\omega)$ for a knot on this trajectory $\omega \in \Omega$ to belong to a certain topological type. This can be formally expressed as

$$P\{\mathrm{Inv}\} = \frac{1}{\mathcal{N}(\Omega)} \sum_{\{\omega \in \Omega\}} \delta\{\mathrm{Inv}(\omega) - \mathrm{Inv}\}, \qquad (1)$$

where $\mathrm{Inv}(\omega)$ is a functional representation of the invariant for a knot on trajectory ω, Inv is a particular topological invariant characterizing the given topological type of the knot or entanglement, $\mathcal{N}(\Omega)$ is the total number of trajectories, and δ is the delta function. The probability under consideration should obey the usual normalization condition

$$\sum_{\{\mathrm{Inv}\}} P\{\mathrm{Inv}\} = 1.$$

The entropy S of a knot of the given topological type is defined as

$$S\{\mathrm{Inv}\} = \ln P\{\mathrm{Inv}\}. \qquad (2)$$

Based on the above definitions, it is easy to note that the probabilistic-topological problems are similar to those encountered in the physics of disordered systems and sometimes, as will be demonstrated below, of the thermodynamics of spin glasses. Indeed, the topological state plays the role of a "quenched disorder" and the $P\{\mathrm{Inv}\}$ functional is averaged over the trajectory fluctuations at a fixed "quenched topological state," which is similar to the calculation of a partition function for a

spin system with "quenched disorder" in the coupling constant. In the context of this analogy, a question arises as to whether the concepts and methods developed over many years of investigation into the class of disordered statistical systems can be transferred to the class of probabilistic-topological problems, in particular, whether the concept of self-averaging is applicable to the knot entropy S.

The main difference between the systems with topological disorder and the standard spin systems with disorder in the coupling constant is a strongly nonlocal character in the former case: a topological state is determined only for the entire closed chain, is a "global" property of this chain, and is difficult to determine for any arbitrary subsystem. Therefore, we may speak of the topology of a part of some closed chain only in a very rough approximation. Nevertheless, below we consider a lattice model featuring a unique relationship between the topological disorder and the disorder in the local coupling constant for a certain disordered spin system on the lattice.

Every time we deal with topological problems, there arises the task of classification of the topological states. A traditional topological invariant, known as the Gauss invariant, is inapplicable because this Abelian (commutative) characteristic takes into account only a cumulative effect of the entanglements, not reflecting the fact that the topological state depends significantly on the sequence in which a given entanglement was formed. For example, when some trial trajectory entanglements with two (or more) obstacles, there may appear configurations linked with several obstacles simultaneously, while not being linked to any one of these obstacles separately. In this context, it is clear that the most evident questions concerning calculation of the probability of a knot formation as a result of the random closing of the ends of a given trajectory cannot be solved in terms of the Gauss invariant because this characteristic is incomplete.

A very useful method of the classification of knots was offered by a polynomial invariant introduced by Alexander in 1928. A breakthrough in this field took place in 1975–1976, when it was suggested to use the Alexander invariants for classification of the topological state of a closed random trajectory computer-simulated by the Monte Carlo method. The results of these investigations showed that the Alexander polynomials, being much stronger invariants compared to the Gauss integral, offer a convenient approach to the numerical investigation of the thermodynamic characteristics of random walks with topological constraints. The statistical–topological approach developed in [2] has proved to be very fruitful: the main results gained by now were obtained using this method with subsequent modifications.

An alternative polynomial invariant for knots and entanglements was suggested by Jones in 1984 [3, 4]. This invariant was defined based on the investigation of the topological properties of braids [5, 6]. Jones succeeded in finding a profound connection between the braid group relationships and the Yang–Baxter equations representing a necessary condition for commutativity of the transfer matrix [7]. It should be noted that neither the Alexander, Jones, and HOMFLY invariants [8], nor their various generalizations, are complete. Nevertheless, these invariants are successfully employed in solving statistical problems. A clear geometric meaning of the Alexander and Jones invariants was provided by the results of Kauffman, who demonstrated that the Jones invariants are related to a partition function of the Potts spin model [9] and that the Alexander invariants can be represented by a partition function of the free fermion model [10].

We employed the analogy between the Jones–Kauffman polynomial invariant and the partition function of the Potts model with ferro (f-) and antiferromagnetic (a-) entanglements [11, 12] to study the statistical properties of knots. In particular, the method of transfer matrices developed in [13–15] was used to determine the probability $P\{f_K\}$ of finding a randomly generated knot K in a particular topological state characterized by the invariant f_K.

The idea of estimating $P\{f_K\}$ is briefly as follows: (i) the Jones–Kauffman invariant is represented as a partition function of the Potts model with disorder in the coupling constant; (ii) the thermodynamic characteristic of the ensemble of knots are calculated using the method of transfer matrices for the Potts spin system.

The paper is arranged as follows. Section 2 describes the model, defines the Jones–Kauffman polynomial invariant, and shows a connection of this invariant to the Potts model. Section 3 presents the numerical methods employed and introduces the necessary supplementary constructions. The results of numerical modeling are discussed and the conclusions are formulated in Section 4.

2. KNOTS ON LATTICES: MODEL AND DEFINITIONS

Consider an ensemble of randomly generated dense knots[1] on a lattice in a three-dimensional space. A knot is called "dense" if the string forming this knot is tightly fit to the lattice, not fluctuating in the space. In this case, the knots of various topological types possess no configuration entropy and the probability of formation of a knot belonging to a given topological type is determined only by a local topology of the system.

Not taking into account the fluctuational degrees of freedom of the lattice sites, this model is obviously very simplified. Nevertheless, this approach adequately reflects a physical situation, for example, in the statistical physics of condensed (globular) state of polymer

[1] Below we will use the term "knots" implying both knots and entanglements. Where necessary, the difference is specified.

macromolecules. Virtually all presently existing models taking into account topological restrictions with respect to the admissible spatial configurations of polymer chains assume a small density of the polymer, that is, describe a situation far from the compact state of globules. A hypothesis of the essentially new globular phase of a ring knotless macromolecule (crumpled globule) existing at a large polymer density (i.e., in the state of dense knots) was formulated in [16]. Despite an indirect experimental evidence for validity of the crumpled globule hypothesis [17], direct observation of such objects in real experiments or computer simulations encounters considerable technical difficulties. In this context, investigation of the distribution of randomly generated dense knots of various topological types may provide for a better understanding of the structure of the phase space of knotted polymers in the globular state.

Analysis of a string configuration in the three-dimensional space is not very convenient for determining the topological type of a knot. The standard method consists in projecting the knot onto a plane in the general position (with no more than two knot segments intersecting at each point of the plane) and determining which segment line passes above (overcrossing) or below (undercrossing) in accordance with the knot topology. Such a projection, featuring over- and undercrossings, is called the diagram of a knot. In what follows we deal only with statistics of the knot diagrams. Evidently, this description implies an additional simplification of the model, but we believe that (in the phase of dense knots under consideration) the additional restrictions are not very significant since, as noted above, the knot entropy contains no contribution due to the string fluctuations.

Thus, let us consider a square lattice of size $N = L \times L$ on a plane, which is rotated for convenience by the $\pi/4$ angle. The lattice is filled by a (dense) trajectory featuring intersections at each lattice site with clear indications of the threads passing above and below (see an example of the 3×3 lattice in Fig. 1). As can be readily verified, a dense trajectory on a lattice with an odd L is unique, that is, represents a knot with the topology unambiguously determined by the pattern of over- and undercrossings and the boundary conditions. In this case, the probability of realization of the knot belonging to a given topological type is determined by the distribution of over- and undercrossings in all lattice sites. The problem considered in this study is to describe the distribution of dense knots over topological classes for various kinds of the over- and undercrossing distributions.

2.1. Reidemeister Moves and Definition of the Kauffman Invariant

In solving any topological problem, the main point consists in comparing the knots. A knot diagram on the plane obey the following Reidemeister theorem [18]: *Two knots in the tree-dimensional space can be continuously transformed into each other if and only if the diagram of one knot can be transformed into that of the other knot by a sequence of local transformations (moves) of types I, II, and III* (see Fig. 2).

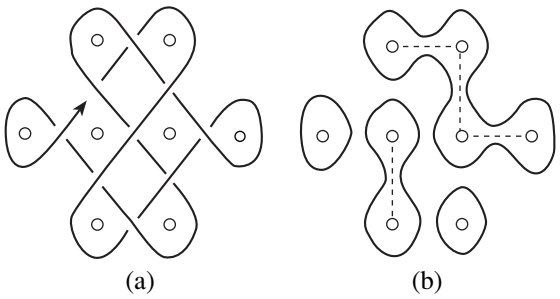


Fig. 1. (a) A knot diagram on the $N = 3 \times 3$ lattice and (b) the diagram splitting. Open circles indicate the spin positions in the Potts model; dashed lines show the graphs on the Potts lattice.

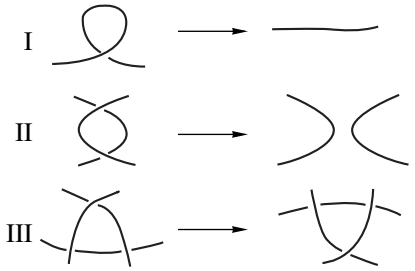


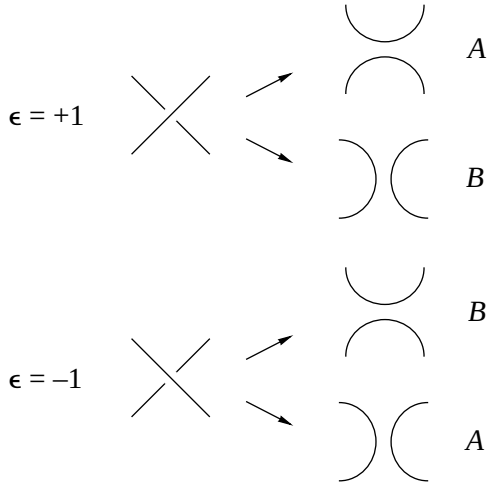
Fig. 2. The Reidemeister moves I, II, and III.

As can be seen in Fig. 2, the Reidemeister move I leads to the formation of a singularity on the plane during continuous fastening of the loop; this move is forbidden for smooth trajectories on the plane. Two knots are called regularly isotopic if their plane diagrams can be transformed into each other by means of the Reidemeister moves II and III. When the mutual transformation of the knot diagrams requires using the Reidemeister moves of all three types, the knots are called ambiently isotopic.

Consider a two-dimensional knot diagram as a graph in which all intersection points (vertices) are characterized by the order of crossing (over- and undercrossing). Then each intersection point belongs to one of the two possible crossing modes. Let a kth point of the graph be characterized by the variable ϵ_k acquiring the values ± 1, depending on the mode of crossing.

Let us define the algebraic Kauffman invariant as a sum over all possible variants of splitting the diagram at the vertices. According to this, each cut is ascribed a certain statistical weight by the following rule: a vertex with $\epsilon = +1$ is given the weight A or B corresponding to the horizontal or vertical splittings, respectively; for a vertex with $\epsilon = -1$, the weight B corresponds to the hor-

izontal cut and the weight A, to the vertical cut. This definition can be illustrated by the following scheme:



(3)

Thus, there are 2^N various microstates for the diagram of a knot possessing N vertices determined by specifying a horizontal or vertical splitting at each of the N vertices. Each state ω of the knot diagram represents a set of nonintersecting and self-nonintersecting cycles. The manifold of all microstates is denoted by $\{\omega\}$ (below, the braces $\{\ldots\}$ will denote summing over these states).

Let $S(\omega)$ be the number of cycles for the microstate ω. Consider a partition function

$$f_{KR} = \sum_{\{\omega\}} d^{S(\omega)-1} A^{l(\omega)} B^{N-l(\omega)}, \tag{4}$$

where the sum is taken over all 2^N possible splittings of the diagram; $l(\omega)$ and $N - l(\omega)$ are the numbers of vertices with the weights A and B, respectively, for a given set of splittings of the microstate ω. Kauffman derived the following statement [9]: a polynomial f_{KR} of the variables A, B, and d representing a partition function (4) is a topological invariant of the regularly isotopic knots, if and only if the parameters A, B, and d obey the relationships:

$$AB = 1, \quad ABd + A^2 + B^2 = 0.$$

The proof of this statement in [9] is based on verification of the invariance of the partition function f_{KR} with respect to the Reidemeister moves II and III. (The invariant for all tree Reidemeister moves is defined below.) The latter relationships pose the following limitations on the parameters A, B, and d in Eq. (4):

$$\begin{aligned} B &= A^{-1}, \\ d &= -A^2 - A^{-2}, \end{aligned} \tag{5}$$

which imply that invariant (4) is a polynomial of the single variable A.

2.2. A Partition Function of the Potts Model as a Bichromatic Polynomial

Consider an arbitrary flat graph containing N vertices. Let each ith vertex be characterized by a spin variable σ_i ($1 \le \sigma_i \le q$) and each edge of the graph, connecting the ith and jth spins ($1 \le \{i, j\} \le N$), by the coupling constant $J_{i,j}$. The energy of the Potts model is defied as [7]

$$E = -\sum_{\{i,j\}} J_{i,j}\delta(\sigma_i, \sigma_j),$$

where the sum over $\{i, j\}$ is taken only for the adjacent spins connected by edges of the graph. Then, the partition function can be expressed as

$$Z = \sum_{\{\sigma\}} \exp\left(\sum_{\{i,j\}} \frac{J_{i,j}}{T}\delta(\sigma_i, \sigma_j)\right),$$

where $\{\sigma\}$ denotes summation over all possible spin states, and the sum over $\{i, j\}$ is taken as indicated above. The last expression can be written in the following form:

$$\begin{aligned} Z &= \sum_{\{\sigma\}} \prod_{\{i,j\}} (1 + v_{i,j}\delta(\sigma_i, \sigma_j)), \\ v_{i,j} &= \exp\left(\frac{J_{i,j}}{T}\right) - 1. \end{aligned} \tag{6}$$

A pair of adjacent spins σ_i and σ_j introduces into the product term a contribution equal to $\exp(J_{i,j}/T)$ for $\sigma_i = \sigma_j$ and a unity contribution for $\sigma_i \ne \sigma_j$. Let us perform a procedure according to the following rules to a given spin configuration and the corresponding graph: (1) an edge is removed from the graph if a contribution to the above product from the spins connected by this edge is unity; (2) an edge is retained in the graph if the contribution from the spins σ_i and σ_j connected by this edge is $\exp(J_{i,j}/T)$. This procedure ensures a one-to-one mapping between the spin configuration corresponding to a product term in the sum (6) and the related set of graph components.

Consider a graph G containing M edges and C coupled components (an isolated vertex is considered as a separate component). Upon summing over all possible spin configurations and the corresponding subdivisions of the graph G, we may present the sum (6) in the following form:

$$Z = \sum_{\{G\}} q^C \prod_{\{i,j\}}^{M} v_{i,j}, \tag{7}$$

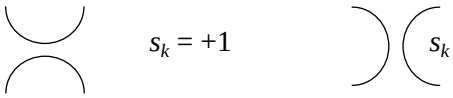
where $\{G\}$ denotes summing over all graphs and $\prod_{\{i,j\}}^{M}$ is the product over all edges of the graph G. Note that expression (7) can be considered as an analytic continuation of the Potts spin system to noninteger and even complex q values.

For $v_{i,j} \equiv v$, expression (7) coincides with a well-known representation of the partition function of the Potts model in the form of a bichromatic polynomial (see, e.g., [7, 19]). The same expression is involved in the correspondence between the Potts model and the model of correlated percolation over edges and bonds suggested by Fortuin and Kastelleyn [20], which serves as a basis for the Monte Carlo cluster algorithms developed by Swedsen and Wang [21] and Wolff [22] for the Potts model.

2.3. Kauffman Invariant Represented as a Partition Function for the Potts Model

The Kauffman invariant of a given knot can be represented as a partition function for the Potts model on a graph corresponding to an arbitrary plane diagram of this knot, but we restrict the consideration below to an analysis of the knots on lattices.

Let us rewrite the Kauffman invariant in the form of a partition function for the Potts model determined in the preceding section, with M denoting the lattice knot diagram (Fig. 1a). The auxiliary variables $s_k = \pm 1$ describe the mode of the knot splitting in each kth lattice site, irrespective of the values of variables $\epsilon_k = \pm 1$ in the same vertices:



$s_k = +1$ $\qquad$ $s_k = -1.$

Let $\omega = \{s_1, s_2, \ldots, s_N\}$ be the set of all variables characterizing the mode of splitting in the lattice containing N intersections. The Kauffman invariant (4) can be written in the following form:

$$f_{KR}(\epsilon_1, \ldots, \epsilon_N) = \sum_{\{\omega\}} (-A^2 - A^{-2})^{S(\omega)-1} \times \exp\left(\ln A \sum_{k=1}^{N} \epsilon_k s_k\right). \tag{8}$$

Here, $\{\omega\}$ indicates summing over all values of the variables s_k (i.e., over all modes of splitting thy lattice diagram M) and the variables ϵ_k characterize a particular disorder "quenched" in the system. Now we will demonstrate that configurations obtained as a result of splitting the diagram are in a one-to-one correspondence to configurations of the Potts model on a dual lattice.

Consider the Potts model lattice Λ corresponding to the lattice M (see Fig. 1b, where open circles indicate positions of the Potts spins). The $b_{i,j}$ edge of lattice Λ corresponds to the kth site of the lattice diagram M. Therefore, disorder in the vertices of diagram M set by the variables ϵ_k coincides with the disorder in edges $b_{i,j}$ of the lattice Λ, that is, with the disorder in coupling constants. Let us define the disorder of $b_{i,j}$ for the lattice Λ via the coupling constants at the corresponding kth site of the lattice diagram M:

$$b_{i,j} = \begin{cases} -\epsilon_k, & \text{for vertical } (i,j) \text{ edges}, \\ \epsilon_k, & \text{for horizontal } (i,j) \text{ edges}. \end{cases} \tag{9}$$

It should be recalled that the definition of the Kauffman invariant (4) is based on splitting the lattice diagram M into polygons representing a system of closed densely packed nonintersecting contours (Fig. 1b). For a given configuration of splitting the lattice diagram M and the corresponding dual lattice Λ, we take the following agreement: all edges of the lattice Λ not crossed by polygons of the lattice M are labeled. In Fig. 1b, the labeled edges are indicated by dashed lines. All the remaining edges of Λ are unlabeled. In these terms, the partition function $\sum s_k \epsilon_k$ in Eq. (8) can be written as follows:

$$\begin{aligned} \sum_k s_k \epsilon_k &= \sum_{\text{lab}} s_k \epsilon_k + \sum_{\text{unlab}} s_k \epsilon_k \\ &= \sum_{\text{lab}}^{\text{hor}} s_k \epsilon_k + \sum_{\text{lab}}^{\text{ver}} s_k \epsilon_k + \sum_{\text{unlab}}^{\text{hor}} s_k \epsilon_k + \sum_{\text{unlab}}^{\text{ver}} s_k \epsilon_k \\ &= \sum_{\text{lab}}^{\text{hor}} b_{i,j} + \sum_{\text{lab}}^{\text{ver}} b_{i,j} - \sum_{\text{unlab}}^{\text{hor}} b_{i,j} - \sum_{\text{unlab}}^{\text{ver}} b_{i,j} \\ &= -\sum_{\text{all}} b_{i,j} + 2\sum_{\text{lab}} b_{i,j}, \end{aligned} \tag{10}$$

where

$$\sum_{\text{lab}} b_{i,j} + \sum_{\text{unlab}} b_{i,j} = \sum_{\text{all}} b_{i,j}.$$

Let m_ω be the number of labeled edges and C_ω is the number of connected components in the labeled graph ω with N_p vertices (each vertex corresponding to a Potts spin). The Euler relationship for this graph is

$$S(\omega) = 2C_\omega + m_\omega - N_p.$$

Now we can readily transform Eq. (8) to

$$\begin{aligned} f_{KR}(A, \{b_{i,j}\}) &= (-A^2 - A^{-2})^{-(N_p+1)} \\ &\times \prod_{\text{all}}^{N} (A^{-b_{i,j}}) \sum_{\{G\}} (-A^2 - A^{-2})^{2C_\omega} \\ &\times \prod_{\text{lab}}^{m_\omega} (A^{2b_{k,l}}(-A^2 - A^{-2})), \end{aligned} \tag{11}$$

by using relationships (10) and the fact that N is odd. Comparing expressions (11) and (6), we obtain the equality

$$\sum_{\{G\}}(A^2+A^{-2})^{2C_\omega}\prod_{\text{lab}}^{m_\omega}(A^{2b_{i,j}}(-A^2-A^{-2})) \equiv \sum_{\sigma}\prod_{\{i,j\}}(1+v_{i,j}\delta(\sigma_i,\sigma_j)), \tag{12}$$

in which the right-hand part coincides with a partition function of the Potts model represented in the form of a bichromatic polynomial. From this we obtain

$$v_{i,j} = A^{2b_{i,j}}(-A^2-A^{-2}) = -1-A^{4b_{i,j}},$$
$$q = (A^2+A^{-2})^2.$$

Since the disorder constants may acquire only discrete values $b_{i,j}=\pm 1$, we may write the following expression for the coupling constants $J_{i,j}$:

$$J_{i,j} = T\ln(1-(A^2+A^{-2})A^{2b_{i,j}}) = T\ln(-A^{4b_{i,j}}). \tag{13}$$

Thus, we arrive at the following statement. The topological Kauffman invariant $f_{KR}(A)$ of the regularly isotopic knots on the lattice M can be represented in the form of a partition function for a two-dimensional Potts model on the corresponding lattice Λ:

$$f_{KR}(A,\{b_{i,j}\}) = K(A,\{b_{i,j}\})Z(q(A),\{J_{i,j}(b_{i,j},A)\}), \tag{14}$$

where

$$K(A,\{b_{i,j}\}) = (A^2+A^{-2})^{-(N_p+1)} \times \exp\left(-\ln A\sum_{\{i,j\}}b_{i,j}\right) \tag{15}$$

is a trivial factor independent of the Potts spins. Here, the partition function of the Potts model is

$$Z(q(A),\{J_{i,j}(b_{i,j},A)\}) = \sum_{\{\sigma\}}\exp\left(\sum_{\{i,j\}}\frac{J_{i,j}(b_{i,j},A)}{T}\delta(\sigma_i,\sigma_j)\right), \tag{16}$$

with the coupling constant $J_{i,j}$ and the number of states q given by the formulas

$$\frac{J_{i,j}}{T} = \ln(-A^{4b_{i,j}}), \quad q = (A^2+A^{-2})^2, \tag{17}$$

and the variable $b_{i,j}$ expressing disorder on edges of the lattice Λ corresponding to the lattice M. A relationship between $b_{i,j}$ and ϵ_k is defined in Eq. (9).

A specific feature of the partition function (16) is the existence of a relationship between the temperature T and the number of spin states q. For this reason, T and q cannot be considered as independent variables. Once a positive q value is fixed, the variable A can formally acquire the complex values according to Eqs. (17), where a logarithm can be taken of a complex argument. The appearance of complex quantities in the partition function can be interpreted in two ways. On the one hand, this implies expansion of the domain of the partition function to the complex plane. On the other hand, the parameters T and $J_{i,j}$ do not enter explicitly into an expression for the Kauffman invariant and, hence, their complex values do not require any special consideration. Below we will be interested mostly in the probability distribution of the maximum power in the polynomial invariant and, therefore, can digress from particular values of the variables A, T, and $J_{i,j}$.

Thus, we have determined an invariant f_{KR} for the regularly isotopic knots, the diagrams of which are invariant with respect to Reidemeister moves II and III. In order to obtain an invariant f_{KI} for the ambiently isotopic knots with oriented diagram, the corresponding partition function has to be invariant with respect to the Reidemeister moves of all types. Let us characterize each oriented intersection by a variable $c_k=\pm 1$ according to the following rule:

(a)  $c_k=-1$, (b)  $c_k=-1$.

In addition, we define the knot twisting $Tw(\omega)$ as a sum of the c_k values taken over all intersections:

$$Tw(\omega) = \sum_k c_k.$$

The invariant $f_{KI}(\omega)$ of the ambiently isotopic knots can be expressed as follows [9]:

$$f_{KI}(\omega) = f_{KR}(\omega)(-A)^{3Tw(\omega)}. \tag{18}$$

It should be noted that our boundary conditions imply that $c_k = b_k$. As is known, the Kauffman polynomial invariant $f_{KI}(\omega; A)$ of isotopic knots as a function of variable A is equivalent to the Jones polynomial invariant $f_J(\omega; x)$ of the variable $x = A^4$. Now we can use formulas (14) and (15) to express the Jones invariant through a partition function of the Potts model. Let a partition function of the Potts model have the form

$$Z(t;q) = \sum_E H(E,q)t^{-E},$$

where $t = e^\beta$, E are the energy levels over which the sum is taken, and $H(E, q)$ is the degree of degeneracy of an energy level E for a given q value. Taking into account that

$$x = A^4 = -t$$

and using formulas (14), (15), and (18), we obtain the following expression for the Jones polynomial invariant:

$$f_J = \sum_E H(E; q = 2 + x + x^{-1})(-x)^{-E} \times (1 + x)^{-N_p - 1}(-\sqrt{x})^{N_p + 1 + \sum_{\{i,j\}} b_{i,j}}. \quad (19)$$

In what follows, by the maximum power of a polynomial invariant we imply the maximum power of variable x in the Jones polynomial invariant f_J. It will be borne in mind that the power of the Kauffman polynomial invariant f_{KI} of the ambiently isotopic knots is obtained from the corresponding power of the Jones polynomial invariant through simply multiplying by a factor of four.

As indicated above, our task is to calculate the probability $P\{f_J\}$ of finding a knot on a lattice in the topological state with a preset Kauffman invariant $f_J(x, \{\epsilon_k\})$ among all the 2^N possible disorder realizations $\{\epsilon_k\}$, $k = 1, \ldots, N$. This probability can formally be written as

$$P_N\{f_J\} = \frac{1}{2^N}\sum_{\{\epsilon_k\}} \delta(f_J(x, \{\epsilon_1, \ldots, \epsilon_N\}) - f_J).$$

Thus, the topological disorder determined by a random independent selection of intersections of the $\epsilon = +1$ and $\epsilon = -1$ types represents a random quenched external field.

When we use the Jones invariant, each topological class (homotopic type) is characterized by a polynomial. In this case, precise identification of the homotopic type for a knot on a lattice containing N intersections will require N variables. Since the number of various homotopic types increases as 2^N, it is very difficult to study the probability of each separate homotopic type characterized by N variables. For this reason, we will introduce a simplified characteristic of a knot on a lattice—the maximum power n of the polynomial invariant $f_J(x)$:

$$n = \lim_{|x| \to \infty} \frac{\ln f_J(x)}{\ln x}. \quad (20)$$

For a trivial knot $n = 0$ and, as the knot complexity grows, the maximum power increases (not exceeding N). Thus, the whole ensemble of knots on a lattice can be divided into subclasses characterized by the power n ($0 \le n \le N$) of the Jones polynomial invariant. In these terms, we will study the probability that a randomly selected knot belongs to one of these subclasses (and is characterized by the maximum power n of the Jones invariant).

Random knots can be generated by two methods: (1) place a fixed number of intersections with $\epsilon = -1$ on the knot diagram with N vertices (accordingly, the other vertices belong to intersections of the $\epsilon = +1$ type); (2) place intersections of the $\epsilon = -1$ and $\epsilon = +1$ types in each vertex with the probabilities p and $1 - p$, respectively.

As can be readily seen, a trivial knot having all intersections of the $\epsilon = +1$ type corresponds to a partition function with the ferro (f-) and antiferromagmetic (a-) entanglements distributed in accordance with rule (9). An impurity (corresponding to $\epsilon_k = -1$ of the knot diagram) will be considered as a change in the sign of $b_{i,j}$ relative to the values characterizing trivial knots. Note that we must differentiate between the notion of impurity (a change in the sign of $b_{i,j}$) from the a-entanglement with $b_{i,j} = -1$. For a trivial knot with all $\epsilon_k = +1$, the lattice contains no impurities while containing the a-entanglements.

3. AUXILIARY CONSTRUCTIONS AND NUMERICAL METHODS

3.1. The Form of a Lattice for the Potts Model and the Positions of Ferro- and Antiferromagnetic Entanglements

Now we will describe the geometry of a lattice for the Potts model corresponding to a knot diagram of size $N = L \times L$. This study is restricted to square lattices, although all considerations remain valid for rectangular lattices as well. Figure 3a shows an example of the trivial knot on a lattice with $N = 5 \times 5$. Positions of the Potts spins corresponding to this lattice diagram are indicated by circles. The Potts spin lattice corresponding to this knot is depicted in Fig. 3b, where the f-entanglements (that would be horizontal in Fig. 1) are indicated by solid lines and the a-entanglements (that would be vertical in Fig. 1) are indicated by dashed lines. In Fig. 3, the same lattice is transformed into a rectangular one of the $L_h \times L_v$ type, where $L_h = L + 1$ and $L_v = (L + 1)/2$ (in this particular case, $L_v = 3$, $L_h = 6$). The partition function of the Potts model will be studied below for a rectangular lattice of this type.

The fact that the height of the lattice for the Potts model is half that of the lattice knot diagram (see Figs. 3a–3c) is very convenient in the case of using the method of transfer matrices, where the computational time expenditure exponentially depends on the lattice height. Let $N = L \times L$ be the total number of entanglements in the lattice, with

$$N_+ = \sum_{\{i,j\}} \delta(b_{i,j}, 1)$$

representing the number of f-entanglements and

$$N_- = \sum_{\{i,j\}} \delta(b_{i,j}, -1)$$

the number of a-entanglements. An impurity corresponding to $\epsilon_k = -1$ in the initial lattice is described by a change in sign of the corresponding constant $b_{i,j}$ for the Potts model lattice.

It should be emphasized once again that by impurity we imply a change in the type of intersection from $\epsilon_k = +1$ to $\epsilon_k = -1$, which corresponds to a change in sign of

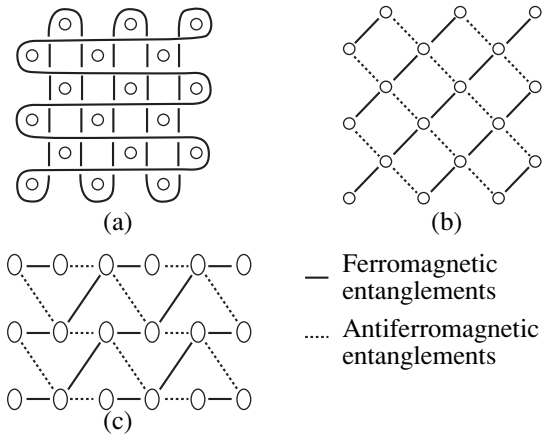


Fig. 3. (a) A knot on the $N = 5 \times 5$ lattice, (b) the Potts spin configuration corresponding to this lattice, and (c) the same configuration reduced to a rectangular lattice $L_h \times L_v = 3 \times 6$.

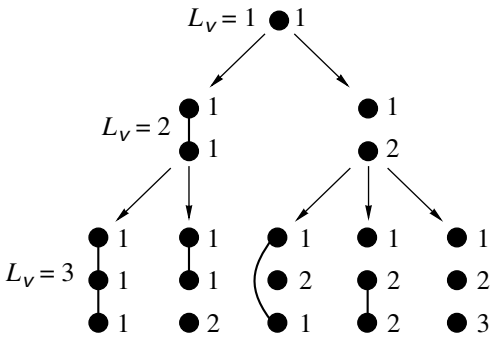


Fig. 4. A schematic diagram illustrating the step-by-step generation of subdivisions into clusters for a column of spins corresponding to the knot state vectors.

the coupling constant $b_{i,j}$ for this intersection, rather than to the a-entanglement (such entanglements are present in the Potts model lattice of the trivial knot). A trivial knot is characterized by the absence of impurities on the Potts model lattice

$$N_+ = \frac{N+1}{2}, \quad N_- = \frac{N-1}{2}.$$

Note that the Kauffman invariants for a knot and its mirror image are different. This is due to the following property of the Jones–Kauffman invariant: the invariant of the mirror image of a knot is obtained by substituting $t \longrightarrow t^{-1}$ (and $A \longrightarrow A^{-1}$), after which the distribution of powers in the polynomial is asymmetric with respect to the substitution $p \longrightarrow 1 - p$.

3.2. The Method of Transfer Matrix

It should be recalled that we are interested in determining the probability distribution of the maximum power n (averaged over various types of intersections in the knot diagram) of the Jones–Kauffman invariant. According to definition (20), the coefficient at the term of the maximum power n in the Jones polynomial of a randomly generated knot is insignificant. Let us fix a certain value n_0, generate an ensemble of random knot diagrams, and determine the fraction of knots the polynomials of which contain the maximum power $n = n_0$.

The traditional approach to numerical analysis of a q-component Potts spin system assumes that each possible state of the column L_v spins corresponds to an eigenvector of the q^{L_v}-dimensional transfer matrix. With computers of the P-II-300 type and a reasonable computational time, the maximum possible size of the transfer matrix is 100–200. This corresponds to a spin strip width $L_v = 8$ for $q = 2$ and $L_v = 4$ for $q = 3$. Obviously, this method cannot be used for investigation of the Potts model with large q values.

A method of transfer matrices applicable to the case of arbitrary q was developed by Blote and Nightingale [13]. The basic idea of this method is that each eigenvector corresponds to a subdivision of the column into clusters of "coupled" spins. By definition, the spins in each cluster are parallel. Therefore, the number of "colors" q can be considered simply as a parameter taking any values (including non-integer and even complex). A detailed description of this method can be also found in [14, 15], where this approach was used to study the Potts model with impurities on the entanglements and to determine the roots of a partition function for the Potts model in the complex plane.

Below we will briefly describe the principles of determining a basis set in the space of coupled spins and constructing transfer matrices in this basis. This description follows the results reported in [15].

A basis set corresponds to a set of various subdivisions of the spin column into clusters of coupled spins. We must take into account that the subdivision should admit realization in the form of a flat graph on a half-plane. Now we will describe a recursive procedure for determining the basis set. Let a basis set be available for a column with the height L_v. Upon adding one more spin from below, we obtain a column with the height $\tilde{L}_v$. Then we take, for example, a subdivision $\nu_1(L_v)$ corresponding to the first basis vector L_v and begin to generate subdivisions corresponding to the basis vectors for the column $\tilde{L}_v$ by attaching the added spin to those existing in $\nu_1(L_v)$. In each step, we check for the possibility of attaching this spin within the framework of the flat graph on a half-plane. After an attempt to attach the added spin to all existing clusters, we generate a subdivision in which the added cluster is separate. Then we take the next subdivision (for example, $\nu_2(L_v)$) for the L_v column and repeat the generation procedure.

The procedure begins with a single spin as depicted in Fig. 4. This spin is assigned the index of unity as belonging to the first cluster. Then another spin is added from below to the same column. Accordingly, the strip

of two spins may occur in one of the two states: $\nu_1(L_\nu = 2)$, in which the spins are coupled and belong to cluster 1 or $\nu_2(L_\nu = 2)$, in which the spins are not coupled and belong to clusters 1 and 2 (Fig. 4). A basis set for the strip of three spins is obtained by adding another spin from below and coupling this spin to all clusters. The clusters are enumerated by integers top to bottom, while vectors in the basis set are enumerated in the order of generation. As can be readily seen, the first basis vector corresponds to a state in which all spins are coupled to each other and belong to the same cluster, while the last basis vector corresponds to the state where all spins are uncoupled and the number of clusters equals to the number of spins in the column. The total number of vectors in the basis set for a strip of m spins is determined as the Catalan number

$$C_m = \frac{1}{m+1}\binom{2m}{m}$$

with the corresponding index number (see [15]).

Let us denote by $\nu_i(\backslash k)$ a subdivision obtained upon separating the kth spin from the subdivision ν_i to form a separate cluster; $\nu_i(\{k, l\})$ will denote a subdivision in which the clusters containing kth and lth spins are combined to form a common cluster. Following a method described in [15], let us form the matrix

$$D_{i,j}(k) = \delta(\nu_j(\backslash k), \nu_i) + q\delta(\nu_j(\backslash k), \nu_j).$$

As can be seen, an element of the matrix $D_{i,j}(k)$ is nonzero if the kth spin is not coupled to any other spin in the subdivision ν_i. When the same spin is not coupled to any other in the jth subdivision as well, the matrix element is q; otherwise it is taken equal to unity.

We will also introduce the matrix $C(k, l)$ defined as

$$C_{i,j}(k, l) = \delta(\nu_j(\{k, l\}), \nu_i),$$

the elements of which are equal to unity if and only if the subdivision ν_i is obtained from the subdivision ν_j by combining clusters containing the kth and lth spins.

Now let us consider a lattice structure with the first two rows such as depicted in Fig. 5d (for an even index $j + 1$ in the second column) and Fig. 5e (for an odd index $j + 1$ in the second column). The spin columns are enumerated by index j ($1 \le j \le L_h$), while the spin position in a column is determined by index i ($1 \le i \le L_\nu$).

According to these constructions, the transfer matrices of the type $T^{\text{even}}(j + 1 = 2k)$, which corresponds to adding an even column $j + 1 = 2k$ after an odd one $j = 2k - 1$ (we add the columns from the right-hand side), contain entanglements between the spins $\sigma_{i, j = 2k-1}$ and $\sigma_{i+1, j = 2k}$. The transfer matrices of the type $T^{\text{odd}}(2j + 1)$, which corresponds to adding an odd column $j + 1 = 2k + 1$ after an even one $j = 2k$, contain entanglements between the spins $\sigma_{i, j = 2k}$ and $\sigma_{i-1, j+1 = 2k+1}$.

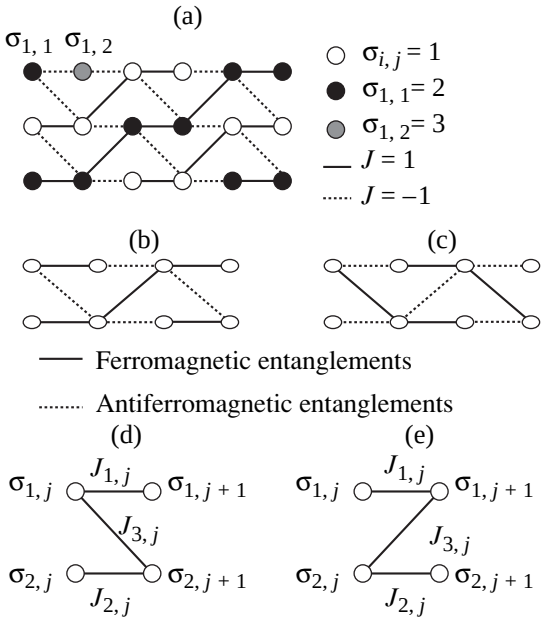


Fig. 5. (a) An example of the configuration of entanglements in which a minimum energy of the Potts model is not reached for $q = 2$; (b, c) the arrangement of f- and a-entanglements on the Potts lattice with $N = 3 \times 3$ and the impurity concentrations $p = 0$ and 1, respectively; (c, d) the arrangement of entanglements between columns with even and odd numbers, respectively.

Let the contribution to a statistical weight corresponding to the f-entanglement $b(\sigma_{i,j}, \sigma_{i',j'}) = 1$ be expressed as

$$t = \exp(\beta J),$$

where $\beta = 1/T$. Then the a-impurity $b(\sigma_{i,j}, \sigma_{i',j'}) = -1$ corresponds to the weight

$$t^{-1} = \exp(-\beta J),$$

and the total contribution of a given configuration to the statistical weight of the system is $t^{b(\sigma_{i,j}, \sigma_{i',j'})}$. Let us define a matrix for the horizontal entanglement $\sigma_{i,j}, \sigma_{i,j+1}$ as

$$P(i, j) = I(t^{b(\sigma_{i,j}, \sigma_{i,j+1})} - 1) + D(i, j)$$

and the matrices for "sloped" entanglements as

$$R^{\text{even}}(i, j) = I + C(i, j)(t^{b(\sigma_{i,j}, \sigma_{i+1,j+1})} - 1),$$

$$R^{\text{odd}}(i, j) = I + C(i, j)(t^{b(\sigma_{i,j}, \sigma_{i-1,j+1})} - 1),$$

where I is the unit matrix.

Then the transfer matrices can be expressed as follows:

$$T^{\text{even}}(j) = P(L_\nu, j)R^{\text{even}}(L_\nu - 1, j)$$
$$\times P(L_\nu - 1, j)\ldots R^{\text{even}}(1, j)P(1, j),$$

$$T^{\text{odd}}(j)$$
$$= P(1, j)R^{\text{odd}}(2, j)P(2, j)\ldots R^{\text{odd}}(L_\nu, j)P(L_\nu, j),$$

Table 1. Mean values of the correlation coefficient $\mathrm{corr}(n_{\max}, |E_{\min}|)$ and the corresponding statistical error for several lattices with the linear size L

L	$\mathrm{corr}(n, \lvert E_{\min}\rvert)$	$\Delta\mathrm{corr}(n, \lvert E_{\min}\rvert)$
3	0.4871	0.0021
5	0.6435	0.0022
7	0.7205	0.0007
9	0.7692	0.0013
11	0.7767	0.0129

and the partition function is

$$Z(t) = u^T \times \prod_{k=1}^{L_h/2-1} T^{\mathrm{even}}(L_h)(T^{\mathrm{odd}}(2k+1)T^{\mathrm{even}}(2k))\nu_{\max},$$

where $\nu_{\max}$ is a basis vector with the maximum number corresponding in our representation to the state in which all spins belong to different clusters and

$$u^T = \{q^{N_{CL}(\nu_1)}, q^{N_{CL}(\nu_2)}, \ldots, q^{N_{CL}(\nu_{\max})}\},$$

so that

$$u^T \nu_i = q^{N_{CL}(\nu_i)},$$

where $N_{CL}(\nu_i)$ is the number of various clusters in the subdivision ν_i. For the first basis vector, $N_{CL}(\nu_1) = 1$ (a single cluster) and the last vector corresponds to $N_{CL}(\nu_{\max}) = L_\nu$ different clusters.

The general algorithm of the calculation is as follows.

1. Generate basis set vectors for a spin strip of the required width.

2. Use this basis set to generate the matrices $R(i, j)$ and $P(i, j)$ in cases of f- and a-entanglements (σ_i, σ_j).

3. Generate a distribution of impurities on the lattice using a random number generator.

4. Generate the transfer matrices for the Potts spin model and the polynomial invariant.

5. Calculate a partition function for the Potts model, the polynomial invariant, the minimum energy, and the maximum power of the polynomial.

6. Repeat points 3–5 and perform averaging over various realizations of the distribution of impurities on the lattice.

The error was determined upon averaging over ten various series of calculations. For each realization of the impurity distribution, the program simultaneously determines both the polynomial invariant and the partition function of the Potts model for an arbitrary (but fixed) q value. This method allowed us to study correlations between the maximum power of the polynomial and the minimum energy within the framework of the Potts model on the corresponding lattice.

4. RESULTS OF CALCULATIONS

4.1. A Correlations between the Maximum Power of the Jones Polynomial of the Lattice Knot and the Minimum Energy of the Corresponding Potts Model

Let us consider dependence of the maximum power of the Jones polynomial on the type of the partition function for the Potts model. Formula (19) explicitly relating the Jones–Kauffman invariant to the Potts model shows that, if the variable x were not entering into an expression for the degeneracy of the energy level $H(E; q = 2 + x + x^{-1})$, the maximum power of the polynomial invariant would always correspond to a term of the partition function with the minimum energy.

Taking into account dependence of the degree of degeneracy on the variable x, we can see that contributions to the coefficient at this power in some cases mutually cancel each other and this coefficient turns zero. A simple example is offered by a system free of impurities, in which case the minimum energy is

$$E_{\min} = -(N+1)/2,$$

and the Jones polynomial is identically unity (with the maximum power being zero).

Nevertheless, there is a strong correlation between the maximum power n of the Jones polynomial and the minimum energy $E_{\min}$ of the corresponding Potts model. Since the minimum energy $E_{\min}$ of the Potts model is always sign-definite and cannot be positive, we will use below a positive quantity representing the absolute value of the minimum energy $|E_{\min}| = -E_{\min}$ Figure 6a shows a joint probability distribution $P(n/N, |E_{\min}|/N)$ of the normalized maximum power n/N and the normalized minimum energy $|E_{\min}|/N$ obtained for a lattice with $N = 49$ and an impurity concentration of $p = 0.5$ by averaging over $N_L = 10^5$ impurity configurations. Here and below we use only the normalized quantities determined on the [0, 1] segment for the energy and [–1, 1] for the maximum power of the polynomial (the latter value can be negative). This allows us to plot the curves for various lattice dimensions on the same figure. In Fig. 6a, the probability distribution is described by the level curves, the spacing between which corresponds to a probability difference of 0.001. As can be seen, there is a strong correlation between the maximum power of the polynomial invariant and the minimum energy of the corresponding Potts model. This relationship can be quantitatively characterized by the correlation coefficient.

It should be recalled that the coefficient of correlation between random quantities x_1 and x_2 with mathematical expectations $\langle x_1\rangle$ and $\langle x_2\rangle$ and dispersions

$$\Delta x_1 = \langle x_1^2\rangle - \langle x_1\rangle^2, \quad \Delta x_2 = \langle x_2^2\rangle - \langle x_2\rangle^2$$

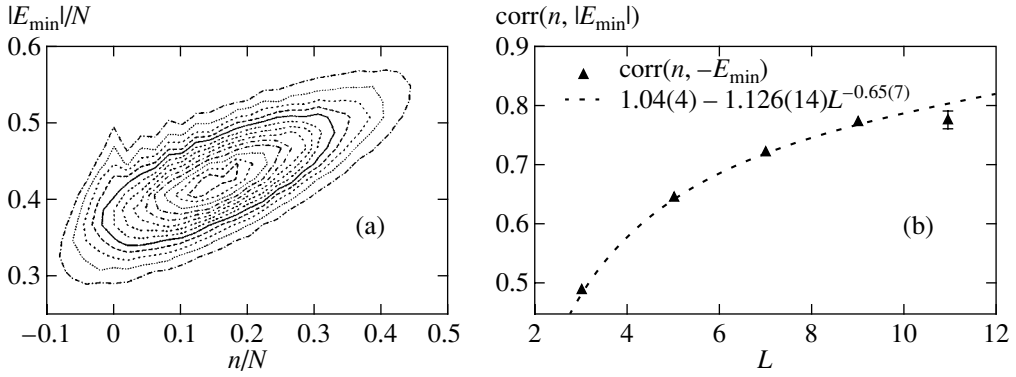


Fig. 6. (a) A joint probability distribution $P(n/N, |E_{min}|/N)$ of the normalized maximum polynomial power n/N and the normalized minimum Potts energy $|E_{min}|/N$ described by the level curves with a step of 0.0015 for a lattice with $N = 49$ and an impurity concentration of $p = 0.5$. (b) The coefficient of correlation between the maximum power of the polynomial invariant and the minimum energy as a function of the lattice size L; dashed curve shows the approximation of these data by a power function of L.

(where $\langle \ldots \rangle$ denotes averaging), is determined by the formula

$$\mathrm{corr}(x_1, x_2) = \frac{\langle x_1 x_2 \rangle - \langle x_1 \rangle \langle x_2 \rangle}{\sqrt{\Delta x_1 \Delta x_2}}.$$

The correlation coefficient equal to ±1 corresponds to a linear relationship between x_1 and x_2.

The values of the coefficient of correlation between the maximum power n_{max} of the polynomial invariant and the minimum energy $|E_{min}|$ for the impurity concentration $p = 0.5$ on lattices with various the linear dimensions $L = 3$, 5, 7, 9, and 11 are presented in Table 1. These values were obtained by averaging over $N_L = 10^5$ impurity realizations for $L = 3$–7, $N_L = 2 \times 10^4$ for $L = 9$, and $N_L = 10^3$ for $L = 11$. As is seen, the correlation increases with the lattice size.

Figure 6b shows the approximation of these data by a power function

$$\mathrm{corr}\left(\frac{n}{N}, \frac{|E_{min}|}{N}\right) = 1.04(4) - 1.126(14)L^{-0.65(7)}.$$

Naturally, the correlation coefficient cannot be greater than unity. The last expression is the result of ignoring the higher terms of expansion in powers of $1/L$. The approximation shows that there is a relationship between the maximum power of the polynomial invariant and the minimum energy of the Potts model, the degree of correlation increasing with the lattice size. In what follows, the results for the maximum power of the polynomial invariant will be accompanied by data for the minimum energy.

The presence of a correlation between the maximum power n of the Kauffman polynomial invariant and the minimum energy E_{min} of the corresponding Potts model for each particular impurity realization on the lattice is of interest from the theoretical standpoint and can be used in the numerical experiments. At present, the limiting lattice size L used in the method of transfer matrix does not exceed $L = 11$, which is determined by the large volume of necessary computations. At the same time, the minimum energy for the Potts model can be calculated within the framework of the standard Monte Carlo algorithm, which poses much smaller requirements on the computational facilities and, hence, can be readily applied to lattices of significantly greater size.

4.2. The Probability Distribution of the Maximum Power of the Polynomial Invariant and the Minimum Energy of the Corresponding Potts Model

Here we present the results of determining the probability distribution $P(n)$ of the maximum polynomial power n. Figure 7a shows the $P(n)$ curves for $L = 7$ and the impurity concentrations $p = 0.1$ (depicted by crosses), 0.2 (squares), and 0.5 (circles). The data were obtained by averaging over $N_L = 10^5$ impurity realizations. The statistical errors are smaller than the size of symbols.

As can be seen, the $P(n)$ curve for small impurity concentrations is nonmonotonic. As the p value increases, the probability distribution function becomes monotonic and approaches the Gauss function in shape. Figure 7b shows the corresponding probability distribution of the normalized minimum energy E_{min}/N obtained for the same lattice size ($L = 7$) and impurity concentrations (denoted by the same symbols). This function appears as more monotonic and approaches the Gauss function already at small impurity concentrations ($p \approx 0.2$).

The shapes of the probability distribution observed for a fixed impurity concentration ($p = 0.5$) and various lattice dimensions are shown in Fig. 8a (for the maximum polynomial power) and Fig. 8b (for the minimum energy). Data for the lattice size $L = 3$ (squares), 5 (squares), 7 (circles), and 9 (triangles) were averaged over $N_L = 10^5$ ($L = 3$, 5, and 7) and data for $L = 9$, over

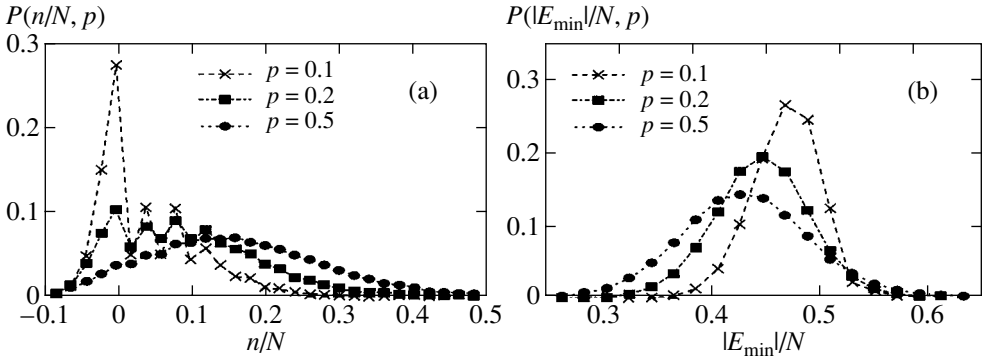


Fig. 7. The distributions of (a) probability $P(n/N, p)$ of the normalized maximum polynomial power n and (b) probability $P(|E_{\min}|/N, p)$ of the normalized minimum energy modulus $E_{\min}$ for a 7×7 lattice and various impurity concentrations $p = 0.1$, 0.2, and 0.5.

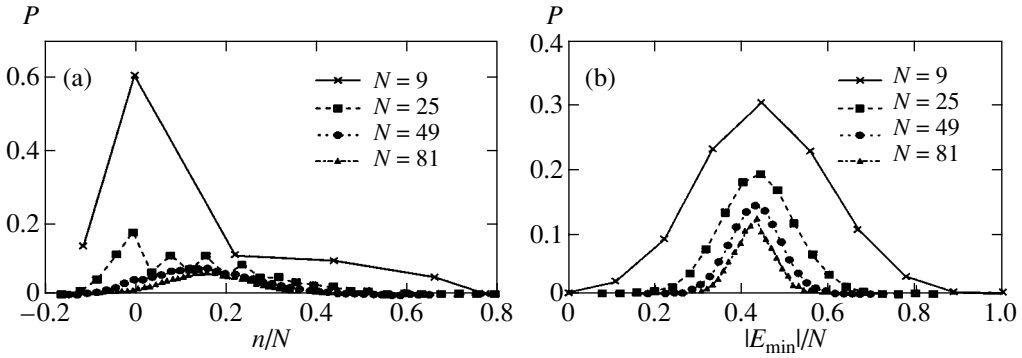


Fig. 8. The probability of (a) the normalized maximum polynomial power n/N and (b) the normalized minimum energy modulus $|E_{\min}|/N$ for the square lattices with $N = 9$, 25, 49, and 81 and an impurity concentration of $p = 0.5$.

$N_L = 5 \times 10^3$ impurity realizations. The probability was normalized to unity:

$$\sum_n P(n) = 1.$$

For this reason, an increase in the lattice size is accompanied by a growing number of the values that can be adopted by the normalized maximum power of $P(n/N)$, while the value of the probability distribution function decreases approximately as $1/N$. As is seen, the probability distribution for lattices of smaller size ($L = 3$, 5) is nonmonotonic; as the L value increases, the distribution becomes a smooth Gaussian-like function.

Thus, we may conclude that the probability distribution $P(n)$ of the maximum power of the polynomial invariant for small-size lattices is nonmonotonic as a result of the boundary effects even for a considerable impurity concentration. As the lattice size increases, the nonmonotonic character disappears and the probability distribution becomes a smooth function. Apparently, we may ascertain that, however small the impurity concentration p, there is a lattice size N ($N \gg 1/p$) such that the corresponding probability distribution is smooth and Gaussian-like.

Thus, we may suggest that the probability distribution $P(n)$ on lattices of large size is determined by the Gauss function, the main parameters of which are the mathematical expectation (mean value) of the maximum polynomial power and the dispersion. Plots of the mean value of the maximum polynomial power $\langle n/N \rangle$ and the corresponding dispersion W^2_{knot} as functions of the impurity concentration p are presented in Figs. 9a and 9b, respectively, for the lattices size of $L = 3$ (crosses), 5 (squares), and 7 (circles). The data for each point were obtained by averaging over $N_L = 10^5$ impurity realizations for $L = 3$ and 5 and $N_L = 5 \times 10^4$ for $L = 7$.

As expected, both the mean value and the dispersion of the maximum polynomial power turn zero for $p = 0$ and 1 (trivial knot) and reach maximum at $p \approx 0.5$. Note that these functions are not symmetric with respect to the transformation $p \longrightarrow 1 - p$, since the Jones polynomial invariant of a mirror knot (with all overcrossings changed for undercrossings and vice versa) is obtained by changing variables $x \longrightarrow x^{-1}$, whereby the maximum polynomial power of the mirror knot corresponds to the minimum power of the original polynomial,

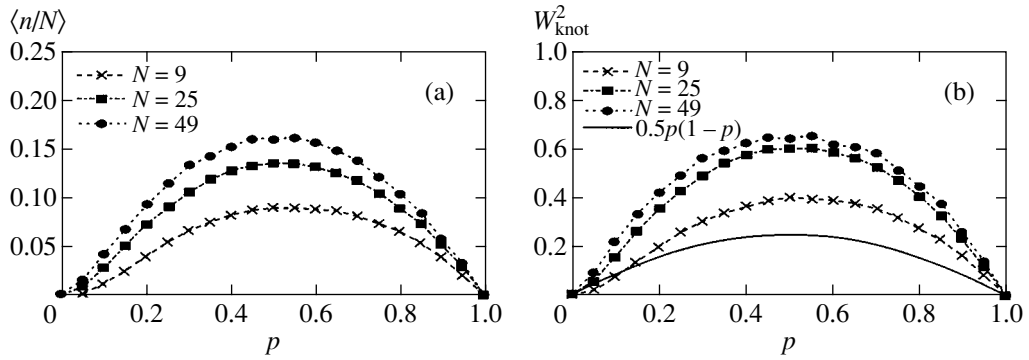


Fig. 9. Plots of (a) the mean normalized maximum polynomial power $\langle n/N\rangle$ and (b) the dispersion $W^2_{\text{knot}}(p)$ of the n/N value distribution versus the impurity concentration p in the range from 0 to 10.

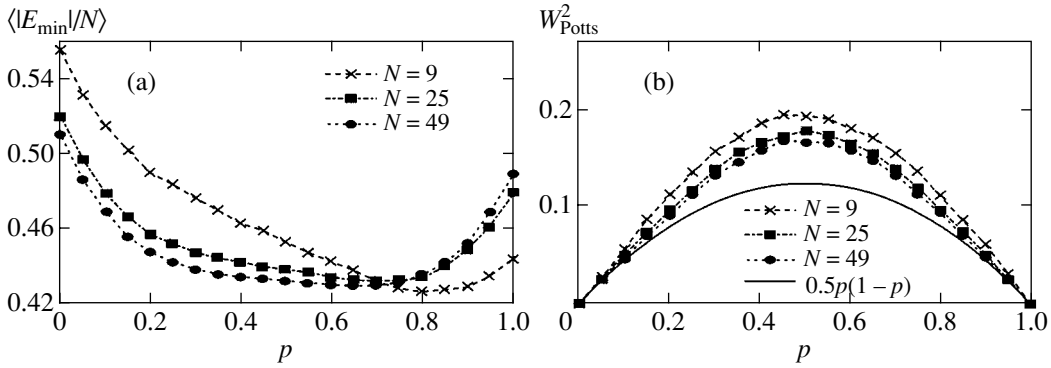


Fig. 10. Plots of (a) the mean normalized minimum energy modulus $\langle|E_{\min}|/N\rangle$ and (b) the dispersion $W^2_{\text{Potts}}(p)$ of the $|E_{\min}|/N$ value distribution versus the impurity concentration p in the range from 0 to 10.

taken with the minus sign. However, this asymmetry disappears with increasing lattice size N as a result of increase in the amount of impurities and in the number of possible impurity realizations employed in the averaging. For the comparison, Fig. 9b shows the function $(1/2)p(1-p)$ representing the dispersion of the distribution function in the hypothetical case when the maximum polynomial power n is a linear function of the number of impurities M. Thus, the difference between W^2_{knot} and $(1/2)p(1-p)$ characterizes dispersion of the distribution of the maximum polynomial power at a fixed number of impurities M (it should be recalled that the impurity occupies each lattice site with a probability p and the total number of impurities M fluctuates).

Figures 10a and 10b (with the parameters and notations analogous to those in Fig. 9) shows data for the absolute value of the mean minimum energy $\langle|E_{\min}|/N\rangle$ and the corresponding dispersion W^2_{Potts}. The asymmetry of the mean minimum energy plot is related to the fact that the number of f-entanglements is greater than that of the a-entanglements by one at $p = 0$ and is smaller by one, at $p = 1$. The probability distribution of the minimum energy is treated in more detail in Subsection 2 of the Appendix.

We have studied dependence of the mean normalized maximum polynomial power on the lattice size L for $p = 0.5$ (Fig. 11a). The results were averaged over $N_L = 10^5$ impurity realizations for $L = 3$, 5, and 7 and over $N_L = 1.5 \times 10^4$ realizations for $L = 9$. As can be seen, the $\langle n/N\rangle$ value tends to a certain limit with increasing $L = \sqrt{N}$. We approximated the results by a power function to obtain

$$\left\langle \frac{n}{N} \right\rangle \approx 0.334(8) - 0.41(2)L^{-0.48(5)}. \qquad (21)$$

Analogous data for the normalized minimum energy in the Potts model are depicted in Fig. 11b. The mean absolute value of the minimum energy decreases with increasing lattice size and can be approximated by a power function of the type

$$\left\langle \frac{|E_{\min}|}{N} \right\rangle \approx 0.4185(7) + 0.119(3)L^{-1.11(4)}.$$

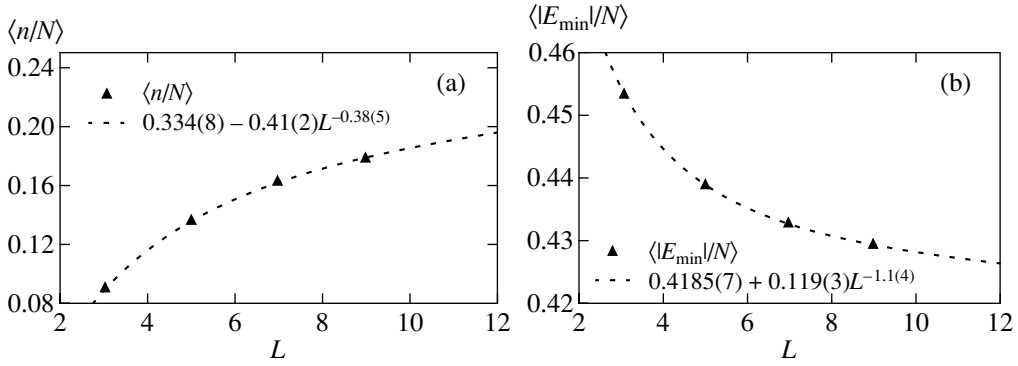


Fig. 11. Plots of (a) the mean normalized maximum power $\langle n/N\rangle$ of the polynomial invariant and (b) the normalized mean minimum energy modulus $\langle|E_{\min}|/N\rangle$ for the square lattices with $L = \sqrt{N} = 3$, 5, 7, and 9 and an impurity concentration of $p = 0.5$; dashed curve in (b) shows the results of approximation by the power function $0.334(8) - 0.41(2)L^{-0.38(5)}$.

Some features of the minimum energy distribution in the Potts model with random ferro- and antiferromagnetic entanglements are treated in the Appendix for $q \geq 4$, in which case analytical expressions can be obtained for small ($p \approx 0$) and large ($p \approx 1$) impurity concentrations.

5. CONCLUSIONS

The results of this investigation can be formulated as follows.

1. An analysis of the relationship between the partition function of the Potts model and the Jones–Kauffman polynomial invariant, in combination with the results of numerical calculations for the Potts model, allowed us to study the probability distribution $P(n/N; p, N)$ of the maximum polynomial power n for the probability p of under- and overcrossings in the sites of a square lattice with $N = L \times L$ for $L = 3$–11.

2. For the lattices of small size with small impurity concentrations p, the probability distribution $P(n/N; p, N)$ of the maximum polynomial invariant power n is not smooth, showing an alternation of more or less probable states. This behavior indicates that the knots of certain topological types are difficult to realize for $\frac{1}{N} \leqslant p \ll 1$ on a square lattice as a result of geometric limitations.

3. As the impurity concentration (probability) p increases, the probability distribution $P(n/N; p, N)$ of the maximum polynomial invariant power n becomes smooth, with the shape approaching that of the Gauss function. A typical value of the "knot complexity" η for $p = 0.5$ and sufficiently large lattices ($N \gg 1$) can be obtained by extrapolating the expression

$$\eta = \lim_{N\to\infty}\left\langle \frac{n}{N}\right\rangle \approx 0.334.$$

4. There is a correlation between the maximum power of the polynomial invariant of a knot and the minimum energy of the corresponding Potts lattice model for $q \geq 4$.

5. An analytical expression was obtained for the probability distribution function $P(E_{\min}/N; p, N)$ at relatively small ($p \sim 1/N$) impurity concentrations.

We have studied the knots with a square lattice diagram. However, all calculations can be generalized to the case of rectangular lattices. In connection with this, it would be of interest to check whether the distribution of knots over the topological types only depends on the number of intersections on the lattice diagram or whether it depends on the diagram shape as well. We may expect that the probability distribution function for a strongly elongated rectangular diagram would differ from the distribution for a square diagram with the same number of intersections.

The method developed in this study is applicable, in principle, to the investigation of knots with arbitrary diagrams (including the case of knots with diagrams not tightly fit to a rectangular lattice), provided that a configuration of the Potts system corresponding to this diagram is known.

We believe that the proposed combination of analytical and numerical methods for the investigation of topological problems using the models of statistical physics offers both a promising means of solving such topological problems and a new approach to the standard methods of investigation of disordered systems. This can be illustrated by the following fact. One of the main concepts in statistical physics is the principle of additivity of the free energy of a system, that is, proportionality of the free energy to the system volume N. By interpreting the free energy as a topological characteristic of the "complexity" of a knot, we may conclude that the complexity of a typical knot increases linearly with the system volume N. This property is, in turn, well known in the topology (outside the context of statistical physics), being a fundamental manifestation of the non-Abelian (non-commutative) character of the phase space of knots.

ACKNOWLEDGMENTS

This study is a logical development of the ideas formulated by one of the authors in collaboration with A.Yu. Grosberg in 1992–1993 [12]. The authors are grateful to A.Yu. Grosberg for his interest in this study and fruitful remarks.

The study was partly supported by the Russian Foundation for Basic Research, project no. 00-15-99302. One of the authors (O.A.V.) is grateful to the L.D. Landau Grant Committee (Forschungzentrum, KFA Jülich, Germany) for support.

APPENDIX

Minimum Energy Distribution in the Potts Model with Random Ferro- and Antiferromagnetic Entanglements

1. The minimum energy of a spin system is independent of q for $q \geq 4$. In this study of the topological invariants of random lattice knots, we are interested in determining the behavior of the system for $x \longrightarrow \infty$ (see Eq. (20)). With allowance for the relationship

$$q = 2 + x + x^{-1},$$

this implies $q \longrightarrow \infty$ for the corresponding Potts model. Some features of the dependence of the free energy of the Potts spin system on the number of states q can be established based on simple considerations.

In cases when the lattice contains no impurities and the a-entanglements are arranged as depicted in Figs. 3b and 3c, the minimum energy is independent of q for $q \geq 2$. Under these conditions, two spin values $\sigma = \{1, 2\}$ are sufficient to "create" a configuration corresponding to the minimum energy. In the presence of impurities, the minimum energy depends on q for $q \geq 2$. This is illustrated in Fig. 5a for a 5×5 lattice with a single impurity on the entanglement between $\sigma_{1,1}$ and $\sigma_{1,2}$ spins. For the spin configuration energy to reach a minimum value of $E_{\min} = -12$ for the given distribution of entanglements, it is necessary that all spin variables in the lattice sites connected by f-entanglements $J = 1$ (in Fig. 5a, these sites are indicated by open and filled circles) would acquire the same values, while the spin variables in the sites $\sigma_{1,1}$ and $\sigma_{1,2}$ connected by a-entanglements $J = -1$ (in Fig. 5a, these sites are indicated by filled and hatched circles) would acquire different values (e.g., $\sigma_{1,1} = 2$, $\sigma_{1,2} = 3$). However, it is impossible to assign the values of spin variables in a system with $q = 2$ so that the spins in clusters $\sigma_{1,1}$, $\sigma_{1,2}$ (as well as in the adjacent cluster) indicated by unlike symbols (black versus open or hatched) were different. For this reason, a minimum energy of the spin state for $q = 2$ is $E_{\min} = -11$ (instead of –12).

As was demonstrated, a minimum energy of the spin configuration in the Potts model corresponding to a given distribution of a-entanglements may depend on the number of spin states q. This fact can be represented as follows: to reach the state with minimum energy, it is necessary that the spin variables in the lattice sites connected by f-entanglements would acquire for the most part the same values (so as to form clusters), while the spin variables in the sites connected by a-entanglements would be possibly different. Thus, a given distribution of the a-entanglements corresponds to a subdivision of the lattice into independent clusters of spins. Spins belonging to the same cluster are connected predominantly by the f-entanglements, while spins of different clusters are connected by a-entanglements.

The energy reaches minimum if the adjacent clusters in a given subdivision possess different values of the spin variable. We may bring each value of the spin variable into correspondence with a certain color. Then a minimum energy corresponds to the lattice subdivision into clusters painted so that all spins in one cluster are of the same color, whereas adjacent clusters have different colors. Figure 5a shows an example of the configuration which cannot be painted in this way using two colors, while three colors allow reaching the goal.

In mathematics, there is a theorem concerning the task of "painting maps," according to which any configuration on a surface possessing a spherical topology can be painted using four (or more) colors. In other words, any subdivision of a lattice into clusters can be painted using four (or more) colors so that the adjacent clusters would possess different colors. If each color corresponds to a certain value of the spin variable $q = \{1, 2, \ldots\}$, we may assign the q values (for $q \geq 4$) so that spins in the adjacent clusters would possess different values (colors). There is no impurity configuration (and the corresponding lattice subdivision into clusters) such that four values of the spin variable would be insufficient to reach the state of minimum energy.

The above considerations allow us to formulate the following statement: *For an arbitrary configuration of a-impurities on a Potts model lattice, a minimum energy of the spin system is independent of the number of spin states q for $q \geq 4$.*

As is known, the Potts model exhibits a first-order phase transition at $q \geq 4$. Note also that the parameter

$$x = \frac{1}{2}(q - 2 \pm \sqrt{q(q-4)})$$

in the Kauffman invariant becomes real just for $q \geq 4$.

2. Distribution of the minimum energy at small ($p \approx 0$) and large ($p \approx 1$) impurity concentrations. Consider the Potts model on an $L_v \times L_h$ lattice with a total number of lattice sites N and $q \geq 4$. Let M be a fixed number of impurities of the ϵ_k types. In the absence of impurities ($M = 0$), the number of f-entanglements is

$$N_+ = \frac{N+1}{2},$$

and the number of a-entanglements is

$$N_- = \frac{N-1}{2}.$$

Table 2. The probability distribution $P(E_{\min}; M, N)$ for a fixed number of impurities $M = 0, 1, 2, N-2, N-1$, and N on a lattice with N sites ($C = \sqrt{N}$)

I	II	III	IV
M	Entanglement type	$E_{\min}$	$P(E_{\min}; M)$
0		$-\frac{N+1}{2}$	1
1	+	$-\frac{N+1}{2}+1$	$\frac{N+1}{2N}$
1	–	$-\frac{N+1}{2}$	$\frac{N-1}{2N}$
2	++	$-\frac{N+1}{2}+2$	$\frac{(N+1)(N-1)}{4N(N-1)}$
2	+–	$-\frac{N+1}{2}+1$	$\frac{(N+1)(N-1)-16(C-1)}{2N(N-1)}$
2	+–&––	$-\frac{N+1}{2}$	$\frac{(N-1)(N-3)+16(C-1)}{4N(N-1)}$
2	––	$-\frac{N+1}{2}-1$	$\frac{4(C-2)}{N(N-1)}$
2	––	$-\frac{N+1}{2}-2$	$\frac{4}{N(N-1)}$
$N-2$	++	$-\frac{N-1}{2}+2$	$\frac{(N-1)(N-3)}{4N(N-1)}$
$N-2$	+–	$-\frac{N-1}{2}+1$	$\frac{(N-1)(N-3)-16(C-1)}{2N(N-1)}$
$N-2$	+–&––	$-\frac{N-1}{2}$	$\frac{(N+1)(N-1)+16(C-1)-4}{4N(N-1)}$
$N-2$	+–&––	$-\frac{N-1}{2}-1$	$\frac{2(N-1)+2(C-1)}{N(N-1)}$
$N-2$	––	$-\frac{N-1}{2}-2$	$\frac{2}{N(N-1)}$
$N-1$	+	$-\frac{N-1}{2}+1$	$\frac{N-1}{2N}$
$N-1$	–	$-\frac{N-1}{2}$	$\frac{N-3}{2N}$
$N-1$	–	$-\frac{N-1}{2}-1$	$\frac{2}{N}$
N		$-\frac{N-1}{2}$	1

The arrangement of a-entanglements is illustrated in Fig. 5a. The minimum energy in the absence of impurities is

$$E_{\min} = -N_+.$$

We may calculate the probability distribution $P(E_{\min}; M, N)$ of the minimum energy for $M = 1$ and 2 by trials of the possible variants. In the case of a single impurity ($M = 1$), there is a probability

$$N_+/N = (N+1)/2N$$

for this impurity to fall on the f-entanglement, after which the minimum energy increases by unity:

$$P(E = -N_+ + 1) = (N+1)/2N.$$

There is also the probability

$$N_-/N = (N-1)/2N$$

for the impurity to fall on the a-entanglement, after which the minimum energy remains unchanged:

$$P(E = -(N+1)/2) = (N-1)/N.$$

Similar considerations can be conducted for $M = 2$. Let $C = \sqrt{N}$ denote the number of spin clusters with the same q on a lattice without impurities. The results of this analysis are summarized in Table 2, where each line indicates (I) the number of impurities M, (II) the coupling constant of a entanglement on which the impurity falls (with the signs "+" and "–" corresponding to $b_{i,j} = 1$ and –1, respectively), (III) the minimum energy $E_{\min}$, and (IV) the probability $P(E_{\min}; M)$ of obtaining a given minimum energy value.

With p denoting the probability that a entanglement contains the impurity ($J' = -1$), the probability of finding M impurities is

$$P(M; p) = \frac{N!}{M!(N-M)!}P^M(1-p)^{N-M}.$$

Thus, a mean value of the minimum energy E in the Potts model is

$$\langle E_{\min}\rangle = \sum_{M=0}^{N}\frac{N!}{M!(N-M)!} \times p^M(1-p)^{N-M}P(E; M, N). \quad (22)$$

Calculating the first three terms for this sum using the data in Table 2, we obtain an approximate formula for the mean minimum energy

$$\langle E_{\min}\rangle \approx (1-p)^{N-2}((1-p)^2P(E; M=0) + Np(1-p)P(E; M=1) + p^2P(E; M=2)) + o(p^2). \quad (23)$$

This formula is applicable if the probability for three impurities to appear on the lattice is small, that is, if

$$\frac{1}{6}N(N-1)(N-2)(1-p)^{N-3}p^3 \ll 1.$$

Upon calculating the probability of the impurity concentration for $M = 1$ and 2 and using expression (22), we obtain a formula for $\langle E_{\min}\rangle$ at small p. For example, in the case of $M = 1$ this yields

$$\langle E_{\min}(M=1)\rangle = \left(-\frac{N+1}{2}\right)\left(\frac{N-1}{2N}\right) + \left(-\frac{N+1}{2}+1\right)\left(\frac{N+1}{2N}\right) = -\frac{N+1}{2}+\frac{N+1}{2N}.$$

By the same token, for $M = 2$ we obtain

$$\langle E_{\min}(M=2)\rangle = -\frac{N+1}{2}+\frac{N+1}{N}-\frac{12C-8}{N(N-1)}.$$

Using these expressions and taking into account that $C = \sqrt{N}$, we obtain an expression for the mean normalized minimum energy $\langle e_{\min}\rangle = \langle E_{\min}/N\rangle$ as a function of the impurity concentration p (in the case of $p \ll 1$)

$$\langle e_{\min}(p)\rangle = \frac{1}{N}\Big[(1-p)^NP(E; M=0) + Np(1-p)^{N-1}P(E; M=1) + \frac{1}{2}N(N-1)p^2(1-p)^{N-2}P(n; M=2) + o(p^2)\Big] \quad (24)$$

$$= -\frac{1}{2}\left(1-\frac{1}{N}\right)+p\frac{1}{2}\left(1-\frac{3}{N}\right) - p^2\frac{6(\sqrt{N}-4)}{N}+o(p^2).$$

Analogous calculations for the concentrations $1 - p \ll 1$ yield

$$\langle E_{\min}(M=N-1)\rangle = -\frac{N-1}{2}+\frac{N-5}{2N},$$

$$\langle E_{\min}(M=N-2)\rangle = -\frac{N-1}{2}+\frac{N-5}{N}-\frac{10C-14}{N(N-1)},$$

and

$$\langle e_{\min}(p)\rangle = \frac{1}{N}\Big[p^NP(E; M=N) + N(1-p)p^{N-1}P(E; M=N-1) + \frac{1}{2}N(N-1) \times (1-p)^2p^{N-2}P(n; M=N-2) + o(p^2)\Big] \quad (25)$$

$$= -\frac{1}{2}\left(1-\frac{1}{N}\right)+(1-p)\frac{1}{2}\left(1-\frac{5}{N}\right) - (1-p)^2\frac{5\sqrt{N}-7}{N}+o((1-p)^2).$$

The results of calculations using the formulas presented in Table 2 and the results of numerical analysis for $q = 4$, $L_v = 3$, and $L_h = 6$ (corresponding to a knot with a 5×5 square lattice diagram) are plotted in Fig. 12 in coordinates of the minimum energy modulus versus probability. Figure 12a shows the results for a fixed number of impurities $M = 1$ (squares) and 2 (circles), while Fig. 12b presents data for the fixed impurity concentrations $p = 0.005$ (crosses), 0.01 (squares), and 0.15 (circles). The results of calculations using the formulas from Table 2 and relationship (23) are depicted by lines and the values obtained by the Monte Carlo method are represented by symbols. The data were averaged over $N_L = 10^5$ realizations; the statistical error for most of the points is smaller than the size of the symbols. As can be seen, the numerical data fit remarkably to the analytical curves. In Fig. 12b, the analytical calculations are illustrated only in the range $|E_{\min}| = 11$–15, where the results could be obtained by expansion into series with terms on the order of p^2. Thus, we have numerically verified the results of analytical calculations performed in this section.

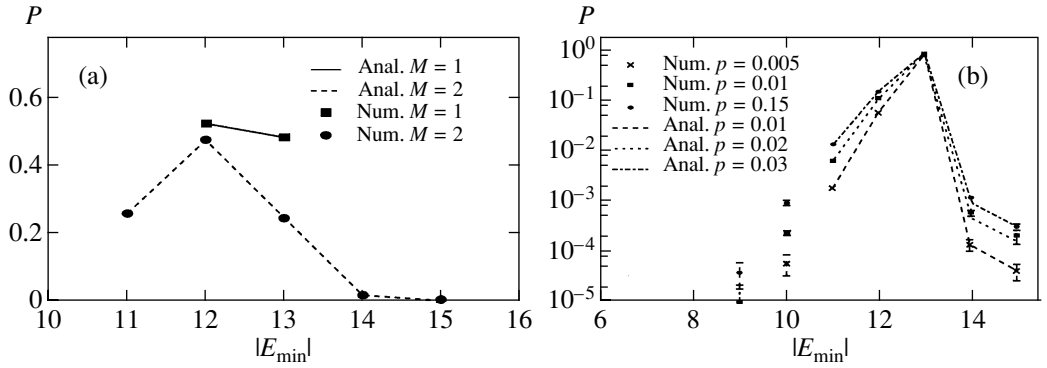


Fig. 12. A comparison of the results obtained by analytical (anal.) and numerical (num.) methods for the minimum energy probability distributions on a lattice with $N = 5 \times 5$: (a) $P(|E_{\min}|; M)$ with $M = 1$ and 2; (b) $P(|E_{\min}|; p)$ with $p = 0.005$, 0.01, and 0.15.

Let us make some remarks concerning the form of the normalized minimum energy

$$\langle e_{\min}(p; N)\rangle = \left\langle \frac{|E_{\min}(p; N)|}{N} \right\rangle$$

for $N \longrightarrow \infty$. As can be seen in Fig. 9b, the plots of $e_{\min}(p; N)$ for the lattices with $N = 25$ and 49 appear as asymmetric troughs. As $N \longrightarrow \infty$, the values of this function at $p = 0$ and 1,

$$\langle e_{\min}(p = 0; N)\rangle = 0.5\left(1 + \frac{1}{N}\right),$$

$$\langle e_{\min}(p = 1; N)\rangle = 0.5\left(1 - \frac{1}{N}\right),$$

will be equal and the profiles will be symmetric with respect to the transformation $p \longrightarrow 1 - p$. The trough bottom will occur on a level of

$$\langle e(p = 0.5; N \longrightarrow \infty)\rangle = 0.415(7)$$

(see Fig. 11b). As can be seen from formulas (24) and (25), the second derivative of this function with respect to p at the points $p = 0$ and 1 becomes zero for $N \longrightarrow \infty$. Probably, all derivatives of the higher orders will become zero as well. In this case, the derivative $\frac{d}{dp}e(p; N = \infty)$ exhibits a break at the points corresponding to the trough "corners."

REFERENCES

1. S. K. Nechaev, Usp. Fiz. Nauk **168**, 369 (1998) [Phys. Usp. **41**, 313 (1998)].
2. M. D. Frank-Kamenetskii and A. V. Vologodskii, Usp. Fiz. Nauk **134**, 641 (1981) [Sov. Phys. Usp. **24**, 679 (1981)]; A. V. Vologodskii *et al.*, Zh. Éksp. Teor. Fiz. **66**, 2153 (1974) [Sov. Phys. JETP **39**, 1059 (1974)]; A. V. Vologodskii, A. V. Lukashin, and M. D. Frank-Kamenetskii, Zh. Éksp. Teor. Fiz. **67**, 1875 (1974) [Sov. Phys. JETP **40**, 932 (1975)]; M. D. Frank-Kamenetskii, A. V. Lukashin, and A. V. Vologodskii, Nature (London) **258**, 398 (1975).
3. V. F. R. Jones, Ann. Math. **126**, 335 (1987).
4. V. F. R. Jones, Bull. Am. Math. Soc. **12**, 103 (1985).
5. J. Birman, *Knots, Links, and Mapping Class Groups* (Princeton Univ. Press, Princeton, 1976), Ann. Math. Studies **82**.
6. A. B. Sosinskii and V. V. Prasolov, *Nodes, Lockings, Braids, and Three-Dimensional Manifolds* (Nauka, Moscow, 1997).
7. R. J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic, New York, 1982; Mir, Moscow, 1985).
8. W. B. R. Likorish, Bull. London Math. Soc. **20**, 558 (1988); M. Wadati, T. K. Deguchi, and Y. Akutso, Phys. Rep. **180**, 247 (1989).
9. L. H. Kauffman, Topology **26**, 395 (1987).
10. L. H. Kauffman and H. Saleur, Commun. Math. Phys. **141**, 293 (1991).
11. F. Y. Wu, J. Knot Theory Ramif. **1**, 47 (1992).
12. A. Yu. Grosberg and S. Nechaev, J. Phys. A **25**, 4659 (1992); A. Yu. Grosberg and S. Nechaev, Europhys. Lett. **20**, 613 (1992).
13. H. W. Blote and M. P. Nightingale, Physica A (Amsterdam) **112**, 405 (1982).
14. J. L. Jacobsen and J. Cardy, cond-mat/9711279.
15. J. Salas and A. D. Sokal, cond-mat/0004330.
16. A. Yu. Grosberg, S. K. Nechaev, and E. I. Shakhnovich, J. Phys. (Paris) **49**, 2095 (1988).
17. B. Chu, Q. Ying, and A. Grosberg, Macromolecules **28**, 180 (1995); A. Grosberg, Y. Rabin, S. Havlin, and A. Neer, Europhys. Lett. **23**, 373 (1993); J. Ma, J. E. Straub, and E. I. Shakhnovich, J. Chem. Phys. **103**, 2615 (1995).
18. K. Reidemeister, *Knotentheorie* (Springer-Verlag, Berlin, 1932).
19. F. Y. Wu, Rev. Mod. Phys. **54**, 235 (1982).
20. C. M. Fortuin and P. M. Kastelleyn, Physica (Amsterdam) **57**, 536 (1972).
21. R. H. Swedsen and J. S. Wang, Phys. Rev. Lett. **58**, 86 (1987).
22. U. Wolff, Phys. Rev. Lett. **62**, 361 (1989).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 93, No. 5, 2001, pp. 903–919.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 5, 2001, pp. 1043–1061.

GRAVITATION, ASTROPHYSICS

Two-Component Cosmological Models with a Variable Equation of State of Matter and with Thermal Equilibrium of Components

V. M. Zhuravlev

Institute of Theoretical Physics, Ul'yanovsk State University, Ul'yanovsk, 432700 Russia
e-mail: zhuravl@sv.uven.ru

Received May 23, 2001

Abstract—A class of models describing the evolution of the homogeneous and isotropic spatially flat Universe filled with a scalar field and matter and changing the equation of state during its evolution from a vacuum-like form to an ideal liquid is proposed under the assumption that both components of matter are in thermal equilibrium. The main characteristics of such models are analyzed and their asymptotic behavior in the vicinity of a cosmological singularity and at the postinflation stage is investigated. It is shown that the thermal equilibrium condition and the requirement of asymptotic decrease of the field with time unambiguously lead to secondary inflation at the final stage of evolution, which is accompanied by accelerated expansion of the Universe and an increase in the temperature of matter. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Modern cosmological models describe the scenario of evolution of the Universe from its origination to the present time as a chain of epochs in which a certain form of matter dominates, which replaced one another in a series of phase transitions. The presence of the inflation epoch is one of the central elements of modern cosmological scenarios. The type of matter filling the Universe at the early stages of its evolution in the course of inflation is known quite precisely. This type of matter is a uniform scalar field similar to quantum vacuum, whose effective equation of state has a form close to the so-called quasi-vacuum equation of state: $p = -\varepsilon$, where p is the effective pressure and ε is the energy density of matter. It is this matter with such properties that makes it possible an exponentially rapid expansion of the Universe. This fact, which is known since 1917 [1], was used in the theory of inflation as a method for solving the entire set of problems in the standard cosmological model (see [2–4]). It is assumed that at the modern epoch, the Universe is filled with matter in the form of a substance with an equation of state close to that for a dust-like substance, i.e., with $p \sim 0$ and with an isotropic Planck's electromagnetic radiation at temperature $T \sim 2.75$ K. At the same time, recent results of observations [5] indicate that the matter component with a quasi-vacuum type equation of state with $p < 0$ is preserved in the Universe at the present epoch, which is responsible for the observed accelerated expansion of the Universe at the modern stage of its evolution. Besides, the results of observations point to the existence of a fourth component of matter (the so-called dark matter) at the present epoch. This explains why the values of mass of matter in the structure of individual galaxies and their clusters, obtained with the help of dynamic measurements and luminosity measurements, differ by a factor of several units.

Apart from the inflation epoch, modern observation data indicate the existence of the radiation dominance era in the evolution process, when the Universe was filled by an isotropic electromagnetic radiation with the equation of state $p = \varepsilon/3$, and the matter dominance era which corresponds to the present state of the Universe.

The existence of the inflation epoch and the radiation dominance epoch has been established quite reliably. An authentic description of physical mechanisms governing the change of epochs continues to be the main problem in the construction of the general scenario of evolution of the Universe from the preinflation epoch to the present time. The most important point in this general problem is the description of the mechanism of emergence of the Universe from the inflation stage [2]. Until now, this problem was solved under the assumption of the dominance of one form of matter (scalar field) at the inflation stage (see, for example, [7–10]). In the framework of such a "one-component" approach, a general principle of modification of inflation models to models with the transition from inflation to the Friedman stage of expansion was obtained in [9, 10]. However, certain difficulties encountered in an analysis of one-component models with the transition from inflation (see [10]) as well as the known data concerning the multicomponent nature of matter at the modern epoch mentioned above necessitate the analysis of multicomponent cosmological models at the inflation stage.

In this connection, we devote the present work to an analysis of cosmological models with matter having several components: a scalar field and a matter changing its equation of state from the "quasi-vacuum" type ($p \sim -\varepsilon$) to the equation of state $p = \gamma\varepsilon$ with $\gamma > 0$ in the

course of evolution of the Universe. The ideology of the present work concerning the study of the dynamics of the scale factor, the form of the self-action potential, and the effective equation of state of matter is mainly associated with the method of "fine tuning". Various versions of this method were used earlier by many authors (see, for example, [7, 11–13] and the literature cited therein). In the present work (Sections 2–4), this method is generalized to the case of the two-component model, which makes it possible to study in detail the mutual influence of different forms of matter by using some auxiliary considerations.

It was mentioned above that the transition from one epoch to another can be interpreted in broad sense as a chain of phase transition in the structure of the matter. The nature of phase transitions occurring in the Universe at the turn of epochs can be grasped in detail by studying the evolution of perturbations of the density of matter and the parameters of the gravitational field, which constitutes a separate and complicated problem. If, however, we confine our analysis only to the assertion about the existence of phase transition and to studying the variation of the most general properties of matter during these periods of time without going into details of the phase-transition processes, we must choose the velocity of sound in the matter components as a convenient parameter. The calculation of the velocity of sound does not require the application of perturbation theory; however, it is precisely this quantity that determines to a considerable extent the dynamic properties of perturbations of the density of matter at all stages of the evolution of the Universe. The idea of such an approach is based on the results obtained in [14], where the equations of the perturbation dynamics were derived with definite indication of the role of the velocity of sound in their dynamics. In the present work (Sections 5, 8), this approach is applied to the problems of describing the properties of two-component models (Section 7) for separating different epochs.

According to modern results of observations, The completion of the inflation stage was followed by a secondary heating of the matter filling the Universe at this period of time. Such heating is usually attributed (see [4, 6]) to the transition of the inflanton field dynamics to the mode of attenuating field oscillations in the vicinity of the local minimum of the self-action potential. New types of inflation models were proposed in [9, 10, 13] in the framework of the standard model with a self-acting scalar field with the natural transition to the Friedman expansion mode. The transition to the Friedman mode in such models is accompanied by a monotonic increase in the scalar field according to a logarithmic law (with a monotonic decrease in its energy density) and does not presume a transition to the oscillating mode. Consequently, the secondary heating in such models must be of some other origin than in the standard models with oscillating mode. In the present work, we propose another interpretation of the mechanism of matter heating during the inflation as well as at the postinflation stage. This mechanism is determined by the form of the effective equation of state of matter, which is natural for the inflation stage, and the corresponding laws of thermodynamic equilibrium between the forms of matter. We consider here a family of one-component (Section 6) models of this type and analogous two-component models with a variable equation of state of the matter (Sections 7–10). Special attention in this work is paid to the derivation of basic equations describing the behavior of all main parameters of the model and to their asymptotic analysis in the periods close to the creation of the Universe and its emergence from the inflation mode.

2. "SLOW SLIDE" REPRESENTATION

In the case of spatially flat homogeneous and isotropic Universe, the standard cosmological model including the inflation stage is described by the system of equations (see, for example, [4])

$$H^2 = \frac{\kappa}{3}\left(\left[\frac{1}{2}\dot{\phi}^2 + V(\phi)\right] + \varepsilon\right),$$
$$\ddot{\phi} + 3H\dot{\phi} = -\frac{d}{d\phi}V(\phi), \tag{1}$$

where ϕ is a scalar field, $V(\phi)$ is the corresponding self-action potential, $H(t) = \dot{R}/R$ is the Hubble parameter, $R(t)$ is the scale factor, κ is the gravitational constant, and ε is the energy density of matter. These two equations should be supplemented with the equation for the pressure p of an ideal liquid, which has the form

$$p = -\frac{1}{\kappa}\left(2\frac{\ddot{R}}{R} + \frac{\dot{R}^2}{R^2}\right) - \frac{1}{2}\dot{\phi}^2 + V(\phi). \tag{2}$$

These three equations completely describe the evolution of the homogeneous spatially flat Universe filled with a scalar field and an ideal liquid with a given equation of state $p = p(\varepsilon)$. Before constructing and studying various types of cosmological models, it is expedient to transform Eqs. (1) to a form more convenient for analysis and to introduce the concept of the effective equation of state of matter.

It was shown in [9, 10] that in the absence of conventional matter ($\varepsilon = 0$), exact equations (1) can be presented in the form of truncated equations known as the equations of "slow slide" in the cosmological inflation theory. According to [9, 10], such a transition is carried out through the introduction of the effective self-action potential in accordance with the rule

$$W(\phi) = V(\phi) + \frac{1}{2}U^2(\phi), \tag{3}$$

where

$$\dot{\phi} = U(\phi), \tag{4}$$

and $U(\phi)$ is a certain function of ϕ. This relation always holds since $\phi = \phi(t)$. Function $W(\phi)$ is the total energy density of the field. Using Eq. (3) we can reduce the initial equations (1) to the form

$$H^2 = \frac{\kappa}{3}W(\phi), \quad 3H\dot{\phi} = -\frac{d}{d\phi}W(\phi). \tag{5}$$

These equations coincide with the truncated equations in the "slow slide" approximation except for the replacement of $V(\phi)$ by $W(\phi)$. The scaling factor as a function of t can be calculated as follows:

$$R(t) = R_0 \exp\left[\int_{t_0}^{t} \sqrt{\frac{\kappa}{3}W(t)}\,dt\right]. \tag{6}$$

Such a representation of equations will be referred to as the "slow slide" representation.

3. THE EFFECTIVE EQUATION OF STATE OF MATTER

Equations (5) lead to the relation

$$\sqrt{3\kappa}UW^{1/2} = -W', \tag{7}$$

which describes, together with Eq. (3), the relation between functions $W(\phi)$ and $U(\phi)$. It can be seen that only one of these functions is arbitrary. This relation makes it possible to derive the general relation between the total energy density W of the scalar field and its effective pressure:

$$\mathcal{P} = -V(\phi) + \frac{1}{2}U^2(\phi).$$

Since $U^2 = W + \mathcal{P}$, we obtain from Eq. (7)

$$\mathcal{P} = -W - \frac{1}{\sqrt{3\kappa}}\frac{\dot{W}}{\sqrt{W}}. \tag{8}$$

Here, $\dot{W} = dW/dt$. Equation (8) can be regarded as the effective equation of state of matter, which is in accord with the dynamics of the gravitational field and matter (scalar field in the present case). This equation indicates the close relation between the rate of decrease in the field energy density and the effective equation of state. For example, for $\dot{W} = 0$, we obtain a quasi-vacuum state of matter $\mathcal{P} = -W$, and the condition

$$\dot{W} = -(\gamma + 1)\sqrt{3\kappa}W^{3/2}$$

corresponds to all equations of state of matter of the form $\mathcal{P} = \gamma W$. Relation (7) also indicates that the rate of decrease in the energy density unambiguously determines the form of the self-action potential $V(\phi)$ of the scalar field as well. Thus, apart from the equation of state $p = p(\varepsilon)$ for an ideal liquid, we can speak of the effective equation of state of the scalar field in each specific model.

4. GENERAL PROPERTIES OF MODELS WITH COMBINED MATTER: SCALAR FIELD + IDEAL LIQUID

Let us consider a model including a scalar field and a matter. In this case, the evolution equations in the slow slide representation have the form

$$H^2 = \frac{\kappa}{3}(W(\phi) + \varepsilon), \tag{9}$$

$$3H\dot{\phi} = -\frac{d}{d\phi}W(\phi), \tag{10}$$

and Eq. (7) can be written as

$$\sqrt{3\kappa}U\sqrt{W(\phi) + \varepsilon} = -W'. \tag{11}$$

Using these relations, we obtain the following expression for the evolution of the scaling factor:

$$R(t) = R_0 \exp\left[\int_{t_0}^{t} \sqrt{\frac{\kappa}{3}(W(t) + \varepsilon)}\,dt\right]. \tag{12}$$

Thus, the evolution of the scaling factor entirely depends on the total energy of matter, $E = W(t) + \varepsilon$. As in the previous section, $\mathcal{P}$ stands here for the effective pressure of the scalar field. Using expression (12), we can write Eq. (2) in the form

$$p = -\varepsilon - \frac{1}{\sqrt{3\kappa}}\frac{\dot{\varepsilon}}{\sqrt{E}}. \tag{13}$$

This relation can be treated as the generalized equation of state of matter. Taking into account Eq. (11), we can transform Eq. (8) to a similar form:

$$\mathcal{P} = -W - \frac{1}{\sqrt{3\kappa}}\frac{\dot{W}}{\sqrt{E}}. \tag{14}$$

Let us introduce the total effective pressure

$$P = p + \mathcal{P}.$$

Combining Eqs. (13) and (14), we obtain the relation

$$P = -E - \frac{1}{\sqrt{3\kappa}}\frac{\dot{E}}{\sqrt{E}}, \tag{15}$$

connecting the total effective pressure with the total energy of matter. Eliminating the total energy E from Eqs. (13) and (14), we arrive at one more useful relation

$$\frac{d\varepsilon}{\varepsilon + p} = \frac{dW}{W + \mathcal{P}} = \frac{dE}{E + P}, \tag{16}$$

which establishes a relation between the increments of the energy density of matter, field, and total energy.

The derived relations can easily be generalized to the case of several scalar field components. For this purpose, we must take for the energy density W of the scalar field and the effective pressure P of the scalar field in formulas (9)–(15) the total values of energy

density and effective pressure for all the scalar field components:

$$W = \sum_{k=1}^{N} W_k(\phi_k) = \sum_{k=1}^{N}\left(\frac{1}{2}\dot{\phi}_k^2 + V_k(\phi_k)\right),$$

$$\mathcal{P} = \sum_{k=1}^{N} \mathcal{P}_k(\phi_k) = \sum_{k=1}^{N}\left(\frac{1}{2}\dot{\phi}_k^2 - V_k(\phi_k)\right).$$

In this case, $E = \varepsilon + W$ as before, and the effective equations of state for each component have the same form (14):

$$\mathcal{P}_k = -W_k - \frac{1}{\sqrt{3\kappa}}\frac{\dot{W}_k}{\sqrt{E}}, \quad k = 1, \ldots, N.$$

Relation (16) assumes the form

$$\frac{d\varepsilon}{\varepsilon + p} = \frac{dW_1}{W_1 + \mathcal{P}_1} = \ldots = \frac{dW_N}{W_N + \mathcal{P}_N} = \frac{dE}{E + P}.$$

It should be noted that for a complete analysis of cosmological model, it is sufficient to consider the set of equations (14), (13), and (12) or an equivalent set of equations (15), (16), and (12). Indeed, these relations describe the variation of the scaling factor as well as the form of the self-action potential and the change in the energy density of matter. Proceeding from the form of the effective equation of state of matter, we can obtain in explicit form the change in the effective temperature of matter in the course of evolution under the assumption of the quasi-stationary nature of thermodynamic processes accompanying this evolution. The only important element which cannot be analyzed on the basis of the obtained relations is the increase in the matter density perturbations in such a model. In order to analyze this quantity, we must solve additional equations describing the evolution of the scalar field perturbations, the metric, and the density of matter [4, 15]. However, a number of the most important features of evolution of perturbations can be established from an analysis of the equation of state, namely, from an analysis of the behavior of the velocity of sound in such models. In other words, without solving the equations describing the evolution of density perturbations, we can determine a number of their important properties. Such a possibility stems from the results obtained in [14]. In the subsequent analysis, we will bear in mind this possibility without analyzing the evolution of perturbations directly. We will start from simple equations of inflation.

5. INFLATION COSMOLOGIES WITH A SCALAR FIELD

Proceeding from the above analysis of equations of cosmological dynamics, we consider several special models of inflation with matter in the form of a scalar field without an ideal liquid. According to the results obtained in [9, 10], an important equation in the inflation cosmology is the equation of state of matter, corresponding to an exponentially rapid decrease in the energy density of matter:

$$\dot{W} = -\beta W \tag{17}$$

for $\beta = \text{const}$. In this case, in accordance with Eq. (8), the effective equation of state has the form

$$\mathcal{P} = \frac{\beta}{\sqrt{3\kappa}} W^{1/2} - W. \tag{18}$$

This equation follows from two different models of inflation, having peculiar properties. The first model was constructed in [13] on the basis of the variational determination of the slow slide mode. The second model of inflation, characterized by an effective equation of state of matter of the type (18), corresponds to the Higgs self-action potential of the form

$$V(\phi) = \frac{\lambda}{4}(\phi - \phi_\infty)^4 - \frac{1}{2}M_h^2(\phi - \phi_\infty)^2. \tag{19}$$

Here, $\lambda = M_h\sqrt{3\kappa}$ and M_h is the Planck mass which in the given case is equal to the mass of a Higgs boson. A model of this type was considered in [9, 10] in connection with the construction of models with the emergence from inflation to the Friedman stage.

A specific feature of the model with the Higgs potential is that at a certain instant of evolution, which is of the inflation type, the potential energy of self-action of the field becomes negative. It can easily be verified that the energy dominance principle is violated for negative values of self-action potential. Starting from this instant, the model behaves "abnormally." First, in this model, the Universe tends in the asymptotic limit to a state of unstable equilibrium corresponding to a local peak of $V(\phi)$ (this was noted in [10]), and second (which is even more surprising), the velocity of sound in this model attains after a certain time the value of the velocity of light and subsequently even exceeds this velocity (this effect was not mentioned in [10]). It was indicated in the Introduction that the velocity of sound is one of the simplest indicators of the occurrence of phase transitions in the course of evolution. We can easily calculate the velocity of sound for model (18):

$$c_s^2 \equiv \frac{dp}{d\rho} = c^2\left(-1 + \frac{\beta}{2\sqrt{3\kappa}} W^{-1/2}\right). \tag{20}$$

Here, c is the velocity of light, $W = c^2\rho$, and ρ is the mass density. The behavior of the square of the velocity of sound c_s^2 corresponding to Eq. (20) is presented in Fig. 1. The monotonic increase in the velocity of sound upon a decrease in the energy density of matter is the most remarkable effect accompanying the expansion of the Universe in the inflation scenario with a given type of matter. There exist two types of the threshold energy

density. It can be seen from Eq. (20) that as $W \longrightarrow \infty$, the velocity of sound becomes an imaginary quantity. For this reason, density perturbations either attenuate or build up rapidly. As the density decreases to a certain value W_0, the value of c_s becomes zero. After the attainment of this value of energy density, the velocity of sound assumes real values, the dynamics of perturbations becomes oscillatory, and the inflation stage terminates. The second threshold value corresponds to the attainment of the critical density W_c for which the velocity of sound becomes equal to the velocity of light ($c_s = c$). The critical density corresponding to the velocity of sound (20) is given by

$$W_c = \frac{1}{4}W_0 = \frac{\beta^2}{12\kappa}.$$

A subsequent decrease in the energy density in the model under investigation leads to a further increase in the velocity of sound and to the transition of the field dynamics to the supercritical region: $c_s > c$.

Thus, the violation of the energy dominance principle in this model leads to the emergence of processes with a rate exceeding the velocity of light. It should also be noted that this effect is not associated with the field peculiarities of the model under investigation, but is entirely determined by the violation of the energy dominance principle. It should be recalled in this connection that the model of a self-acting scalar field is equivalent to the model of an ideal liquid except the notation. Note that a transition of the model to the supercritical region is observed not immediately after the violation of energy dominance, but after a certain finite time interval. It follows hence that the model becomes inapplicable for describing actual processes only after the attainment of the boundary of the supercritical region for the velocity of sound. If we now recollect that the evolution modes associated with violation of the energy dominance principle correspond to time intervals during which phase transformations of matter take place, the natural assumption arises that the models under investigation should be interpreted as prototypes of models with phase transitions. Henceforth, we will refer to this type of models as the models with a phase transition.

The main idea can be formulated as follows. In the region of supercritical densities, the conditions for free propagation of acoustic perturbations are not observed since the velocity of acoustic perturbations cannot exceed the velocity of light. Consequently, wave perturbations of the density of matter (field in the present case) must rapidly attenuate, transferring energy to some other form of matter which must be intensely formed as the threshold density is approached. It follows hence that the attainment of the critical density must lead to a considerable change in the form of the equation of state of matter, i.e., to a phase transition. Since the velocity of sound in the vicinity of the critical value of energy density is equal to the velocity of light,

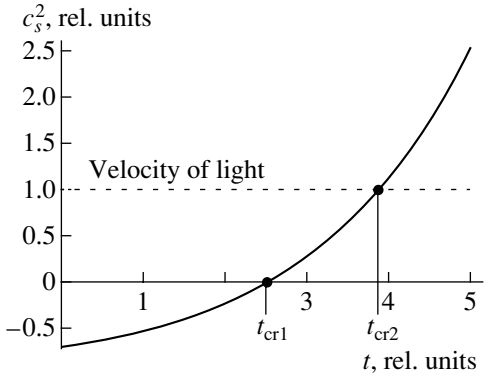


Fig. 1. Time dependence of the square of the velocity of sound for the model with a Higgs self-action potential ($\alpha = 1$).

acoustic vibrations (density perturbations) must apparently transfer the energy to the electromagnetic waves most effectively during this period of time. Thus, near the critical density, a phase transition to the matter mainly in the form of isotropic radiation with the equation of state $p = \varepsilon/3$. However, the model becomes inapplicable for describing processes at the Friedman stage and requires modification.

One of the versions of such a modification was proposed in [9, 10]. It is based on the "correction" of the asymptotic behavior of the field and self-action potential as functions of time for $t \longrightarrow \infty$ and, hence, the form of the self-action potential as a function of the field. This is achieved by supplementing the function $\dot{\phi}$ with a term decreasing according to the law t^{-1} under the assumption that the remaining elements of the functional representation of $\dot{\phi}$ decrease more rapidly. In this case, $\phi \sim \ln t$. If the asymptotic behavior of parameters is taken into account correctly, the period of energy dominance violation vanishes in the history of evolution in such a model, its basic properties being preserved at the inflation stage. In this form, the model becomes applicable for describing the global evolution from the inflation stage to the modern epoch.

The only unsatisfactory element in such a model is that the entire matter is presented by a single scalar field. This scalar field possesses different properties at the inflation stage and at the stage of the Friedman expansion. It was shown in [13] that the transition to the Friedman stage in models with a single scalar field is necessarily associated with a logarithmic increase in this field: $\phi \sim \ln t$. However, in view of the presence of elementary particles with the Higgs mechanism of the particle mass formation in the standard model, the field value is associated with the particle mass: $m \sim \phi^2$. This indicates an increase in the particle mass at the Friedman stage according to the law $\sim(\ln t)^2$. Since this effect has not been observed experimentally, we have to assume that the transition to the Friedman stage must be ensured not by the scalar field, but by another compo-

nent of matter, which has the form of a substance rather than a field at the instant of completion of inflation (see [10]). It would be more acceptable to assume that matter and scalar field have different sources and to consider two components of matter from the very outset in the model. In the present work, we propose precisely this version of model modification, which takes into account explicitly at least two forms of matter on the entire interval of evolution of the Universe from the inflation stage to the modern epoch.

6. THE GENERAL CLASS OF MODELS OF INFLATION WITH A PHASE TRANSITION

Before analyzing two-component models, let us consider in greater detail all types of one-component models in which the above-mentioned phase transition may take place according to an analogous scenario. Such models can be described by the equation of energy density variation in the form

$$\frac{dW}{dt} = -kW^{\alpha}, \tag{21}$$

where k and α are constants. The law describing the decrease in the energy density in this case has the form

$$W(t) = [kt(\alpha-1)]^{-1/(\alpha-1)} = w_0 t^{-m}. \tag{22}$$

Here, the following notation has been introduced: $m = (\alpha-1)^{-1}$, $w_0 = (k(\alpha-1))^{-m}$. For the zeroth instant of time, we chose the instant corresponding to singularity: $W = \infty$. It should be noted that it is expedient to consider this relation for $\alpha \geq 1$. Otherwise, energy does not decrease during expansion, but increases since $m < 0$; i.e., the total energy of the Universe increases, which is possible only if the energy dominance principle is violated.

In connection with contemporary data on accelerated expansion in the modern epoch [5], it is also expedient to consider models with the law of the energy density decrease not to its zero value, but to a certain small constant value $\Lambda > 0$, which can be interpreted as a cosmological constant [16]. In this case, Eq. (21) should be replaced by the equation

$$\frac{dW}{dt} = -k(W-\Lambda)^{\alpha}, \tag{23}$$

whose solution has the form

$$W(t) = [kt(\alpha-1)]^{-1/(\alpha-1)} + \Lambda = w_0 t^{-m} + \Lambda. \tag{24}$$

In the subsequent analysis, we will consider the effect of the introduction of the cosmological constant in the model, but for the time being, we assume that $\Lambda = 0$.

The effective equations of state of the matter, which correspond to Eq. (22), have the form

$$\mathcal{P} = -W + \frac{k}{\sqrt{3\kappa}} W^{\alpha-1/2}. \tag{25}$$

An analysis of this relation shows that as $t \longrightarrow \infty$, the effective pressure $\mathcal{P}$ decreases together with W only for the values of the parameter $\alpha \leq 3/2$. Indeed, for $\alpha < 3/2$. The first term in Eq. (25) is smaller than the second term for quite large values of $t > 0$; in this limit, we have

$$\mathcal{P} \sim W^{\alpha-1/2}\frac{k}{\sqrt{3\kappa}} \longrightarrow 0.$$

For $\alpha = 3/2$ (which corresponds to Friedman's models), $\mathcal{P} = \gamma W \longrightarrow 0$, but the sign of pressure is determined by constants k and α. If, however, $\alpha > 3/2$, the effective pressure increases indefinitely for large values of $t > 0$, while $W \longrightarrow 0$.

Consequently, the "Friedman" value of the parameter $\alpha = 3/2$ divides all models into two classes. Models of the first type correspond $1 \leq \alpha < 3/2$, while models of the second class correspond to $\alpha > 3/2$. The exponent in Eq. (22) for first-type models satisfies the inequality $m = (\alpha-1)^{-1} > 2$. This means that the energy density in such models decreases for $t \longrightarrow \infty$ at a higher rate than in Friedman's models with the equation of state $\mathcal{P} = \Gamma W$ corresponding to $\alpha = 3/2$ and $W \sim t^{-2}$ [13]. The Friedman law corresponds to the situation when the "number of particles of matter" in each element of space does not change and the energy decreases only due to the extension of the unit volume of the space. For $\alpha > 3/2$, we have $m < 2$ and the energy decreases more slowly than in the Friedman modes. As a result, models with $\alpha > 3/2$ describe an incessant inflation, which rules out their transition to the Friedman mode of expansion (see [10]). For $1 \leq \alpha < 3/2$, the rate of energy decrease increases upon a decrease in α in accordance with Eq. (22) and attains its maximum value for $\alpha = 1$, which corresponds to an exponential decrease in the energy density of the field. Thus, in the region $1 < \alpha < 3/2$, the "decay" of scalar field "quanta" takes place and the evolution of the scaling factor slows down. This conclusion is also illustrated by the fact that the scalar field decreases monotonically for models with $1 < \alpha < 3/2$, which indicates that the particle masses approach a constant value. It should be recalled (see above and [10]) that for $\alpha = 3/2$, the field increases according to the law $\ln t$, which leads to an increase in the mass of particles according to the law $(\ln t)^2$. For this reason, it is expedient to consider models precisely with $1 \leq \alpha < 3/2$.

The velocity of sound for these equations of state depends on energy density as follows:

$$c_s^2 \equiv \frac{dp}{d\rho} = c^2\left(-1 + \frac{\kappa(2\alpha-1)}{2\sqrt{3\kappa}} W^{\alpha-3/2}\right).$$

The range $1 < \alpha < 3/2$ is also distinguished by the fact that the velocity of sound on this interval increases upon a decrease in the energy density and attains at a certain instant the critical value equal to the velocity of light. This is possible even in a wider range (for all $\alpha \geq 1$, $\alpha \neq 3/2$). Accordingly, the critical density is given by

$$W_c = \left[\frac{k(2\alpha-1)}{4\sqrt{3\kappa}}\right]^{2/(3-2\alpha)},$$

and the instant at which the critical density is attained is given by

$$t_c = \frac{1}{k(\alpha-1)}\left[\frac{4\sqrt{3\kappa}}{k(2\alpha-1)}\right]^{2(1-\alpha)/(2\alpha-3)}.$$

For $\alpha = 3/2$, the critical density is never attained except for matter with an extremely stringent equation of state of the substance $p = \varepsilon$. Thus, all the models permitting the attainment of the critical density satisfy the conditions $1 \le \alpha$, $\alpha \ne 3/2$. It was noted above that the limiting value of the parameter $\alpha = 3/2$ corresponds to Friedman's models. The other limiting case $\alpha = 1$ corresponds to inflation models (19) in which the energy density decreases exponentially (17). It should also be noted that in all these models with $1 < \alpha < 3/2$, as well as in the model with $\alpha = 1$, there exists a threshold energy density W_0 above which the velocity of sound is imaginary and below which it is real-valued. This threshold value corresponds to $c_s = 0$. The value of W_0 determined by the formula

$$W_0 = \left[\frac{k(2\alpha-1)}{2\sqrt{3\kappa}}\right]^{2/(3-2\alpha)} = 2^{2/(3-2\alpha)}W_c.$$

Another distinguishing feature of the class of models under investigation is the behavior of temperature. Knowing the effective equation of state of matter, we can establish the temperature dependence of the energy density under the assumption that matter (aggregate of field quanta) is in thermal equilibrium and is an analogue of a relativistic gas, but with the equation of state (25). In this case, the internal energy $U = W\mathcal{V}$, where $W = W(T)$ and $\mathcal{V}$ is the volume of the region occupied by matter. It follows from the second law of thermodynamics of reversible processes that U and $\mathcal{P}$ are connected through the equation

$$\frac{\partial U}{\partial \mathcal{V}} = -\mathcal{P} + T\frac{\partial \mathcal{P}}{\partial T}. \quad (26)$$

This leads to the following expression for $W(T)$:

$$W = -\mathcal{P} + T\frac{\partial \mathcal{P}(W)}{\partial T} = -\mathcal{P} + T\frac{c_s^2}{c^2}\frac{\partial W}{\partial T}. \quad (27)$$

For the equation of state (25), integration of Eq. (27) leads to the relation

$$W\exp\left[-\frac{4\sqrt{3\kappa}W^{3/2-\alpha}}{k(3-2\alpha)(2\alpha-1)}\right] = \left(\frac{T}{T_0}\right)^{2/(2\alpha-1)}. \quad (28)$$

It follows hence that in models with $1 < \alpha < 3/2$, the temperature in the initially singular state is equal to zero: $T \longrightarrow 0$ for $W \longrightarrow \infty$. At the instant $t = 0$ of its creation, the Universe is cold. The later statement appears natural since initially the Universe is filled with a field and fluctuations of these fields are of quantum origin in this period. After the birth of the Universe, the temperature increases upon a monotonic decrease in the energy density, attains its maximum value, and then again tends to zero ($T \longrightarrow 0$ as $W \longrightarrow 0$). The temperature peak is attained at the instant of attainment of the threshold density corresponding to the condition $c_s = 0$. It was mentioned above that this instant corresponds to the exit of the universe from the inflation process and actually coincides with the secondary heating presumed in standard scenarios.

In this respect, the given scenarios differ from those in the standard model. In the latter scenarios, the Universe is initially hot and its temperature decreases at the inflation stage, after which the secondary heating takes place due to energy transfer from field oscillations to the substance. In the models considered by us here, the initially cold Universe is heated during the inflation stage and attains the maximum temperature at the end of inflation. For this reason, the secondary heating in these models actually coincides with the primary heating and does not require a transition to the oscillatory mode of the field.

The temperature increases in the course of inflation due to the fact that wave perturbations cannot be generated until the threshold density W_0 is attained, and the energy density decreases more slowly than in the case when the number of field quanta is conserved. For this reason, the entire field decay energy is spent on increasing the internal energy of the aggregate of field quanta as a gaseous condensate. Then the temperature drops in the segment of evolution from the values of density W_0 to W_c. After this, the model becomes unable to describe the evolution of the Universe correctly.

Let us now analyze the evolution of the scaling factor and the form of the self-action potential corresponding to the models under investigation. Substituting Eq. (22) into Eq. (6), we find the law of evolution of the scaling factor for models with the equations of state (25) for $1 < \alpha < 3/2$:

$$R(t) = R_0\exp\left[\int_{t_0}^{t}\sqrt{\frac{\kappa}{3}W(t)}dt\right] = R_\infty\exp\left(-\frac{\sigma}{\nu t^\nu}\right). \quad (29)$$

Here,

$$\nu = \frac{3-2\alpha}{2(\alpha-1)} > 0$$

for $1 < \alpha < 3/2$ and $\nu < 0$ for $\alpha > 3/2$,

$$\sigma = \sqrt{\frac{\kappa}{3}}[k(\alpha-1)]^{-1/(\alpha-1)}, \quad R_0 = R(t_0),$$

$$R_\infty = R_0\exp\left(\frac{\sigma}{\nu t_0^\nu}\right).$$

It follows hence that in models with $1 < \alpha < 3/2$, the Universe expands, passing asymptotically to a steady

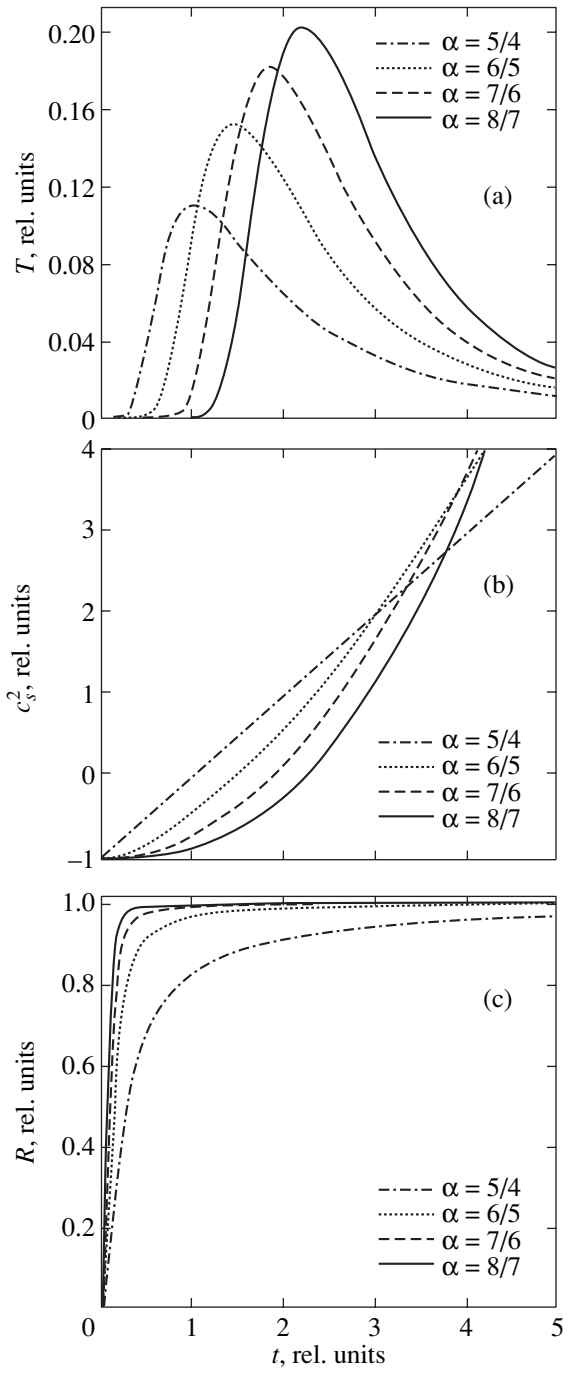


Fig. 2. Time dependence of (a) temperature, (b) square of the velocity of sound, and (c) scaling factor for various values of the parameter α ($\kappa = 1$, $k = 3$).

state. For $\alpha > 3/2$, the expansion occurs in the accelerating mode. The cases $\alpha = 1$ and $\alpha = 3/2$ are special. For $\alpha = 1$, the evolution of the scaling factor has the form [9, 10]

$$R(t) = R_\infty \exp(-qe^{-2\mu t}). \quad (30)$$

For $\alpha = 3/2$, the evolution of the scaling factor obeys the power law $R(t) = R_1 t^s$.

Let us consider the general characteristics of the scaling factor evolution laws (29) for $1 < \alpha < 3/2$. For $t \longrightarrow 0$, we have $R(t) \longrightarrow 0$. This means that at instant $t = 0$, the Universe is in a singular state: $\varepsilon \longrightarrow \infty$ for $t \longrightarrow 0$. In this case, all the time derivatives of the scaling factor are equal to zero for $t \longrightarrow +0$. The Universe starts expanding at zero rate. Then the universe expands almost jumpwise to a certain finite size R_∞ to which it tends asymptotically as $t \longrightarrow \infty$. The limiting case $\alpha = 1$ corresponds to an analogous evolution except that the Universe in an extremely compressed state $R = 0$ for $t \longrightarrow -\infty$. As in the case of $\alpha = 1$, the nearly "jumpwise" change in the scaling factor in the vicinity of a certain instant of time for all the models under investigation can be naturally attributed to the inflation process. The model with $\alpha = 1$ was considered in [9, 10], where it was shown that expansion in this model is of the inflation type. The rate of expansion in models with $1 < \alpha < 3/2$ is lower that in the model with $\alpha = 1$, but the inflation nature of expansion is preserved.

The behavior of the basic parameters of the models with a scalar field under investigation is presented in Figs. 2a–2c for some values of parameter α.

Summarizing the results of our analysis, we can propose that the epochs in the Universe evolution should be singled out according to the velocity of sound. Inflation is the epoch during which $c_s^2 < 0$. The instant of its completion corresponds to the attainment of the critical density for which $c_s^2 = 0$. The "wave epoch" is the period of time during which $0 < c_s^2 < c^2$. It must correspond to the era of radiation dominance (at least, its final stage). During this period, secondary heating and transition to the model epoch of dominance of substance and dark matter must take place in accordance with the standard scenario. The definitions of these epochs will be refined below in the framework of the two-component model, but the definition of inflation remains unchanged in this case.

Let us now consider the expressions for the self-action field potential ensuring the given expansion modes in the framework of their field interpretation. It should be noted that for $\alpha = 1$, the self-action potential has the form (19), which is standard for describing the Higgs mechanism of spontaneous symmetry breaking in the quantum field theory (see [9, 10]). Integrating the equation

$$U^2 = -\frac{\dot{W}}{\sqrt{3\kappa W}}, \quad (31)$$

which follows from Eq. (7), with respect to $W(t) = W(\phi(t))$, we obtain an expressions for $U(t)$ and $\phi(t)$.

Eliminating time with the help of the explicit dependence $\phi = \phi(t)$,

$$\phi(t) = \phi_0 \pm \frac{4\sqrt{\alpha-1}}{(3-2\alpha)(3\kappa)^{1/4}}[k(\alpha-1)]^{-1/4(\alpha-1)}t^{-l},$$

$$l = \frac{3-2\alpha}{4(\alpha-1)},$$

in accordance with Eq. (3), we obtain

$$V(\phi) = A(\alpha)(\phi-\phi_0)^M - B(\alpha)(\phi-\phi_0)^N, \qquad (32)$$

where

$$A(\alpha) = [k(\alpha-1)]^{-1/(\alpha-1)} \times\left\{\frac{2\alpha-3}{4\sqrt{\alpha-1}}(3\kappa)^{1/4}[k(\alpha-1)]^{1/4(\alpha-1)}\right\}^M,$$

$$B(\alpha) = \frac{1}{(\alpha-1)\sqrt{3\kappa}}[k(\alpha-1)]^{-1/2(\alpha-1)} \times\left\{\frac{2\alpha-3}{4\sqrt{\alpha-1}}(3\kappa)^{1/4}[k(\alpha-1)]^{1/4(\alpha-1)}\right\}^n,$$

$$M = \frac{4}{3-2\alpha}, \quad N = 2\frac{2\alpha-1}{3-2\alpha},$$

$$n = -\frac{4(\alpha-1)(2\alpha-1)}{(3-2\alpha)^2}.$$

For instance, for $\alpha = 1$, we obtain $M = 4$, $N = 2$, which corresponds to the standard Higgs potential (19). For $\alpha = 5/4$, we have $M = 8$, $N = 6$, while for $\alpha = 4/3$, we have $M = 12$, $N = 10$. For the Friedman value $\alpha = 3/2$, the potential has a different form; viz., $V(\phi) = c\exp(-2\phi)$.

Thus, the self-action potential for all the models under investigation has a typical form of the Higgs potential, but with a slightly different functional form of potential wells ensuring the Higgs mechanism of spontaneous symmetry breaking. It should be noted, however, that for $1 \le \alpha < 3/2$, there exists a period of time in the course of evolution, when $V(\phi) < 0$ and the energy dominance principle is violated. It is for this reason that the field tends to the value ϕ_0 corresponding to the maximum of the potential $V(\phi)$, i.e., to an unstable state, while the velocity of sound becomes larger than the velocity of light. These defect of the models with $1 \le \alpha < 3/2$ can be eliminated by introducing a special modification of the model, similar to the model with $\alpha = 1$ considered in [10]. Such a modification gives a monotonic decrease in potential for $t \longrightarrow \infty$ leading to the transition to the Friedman mode of the evolution.

7. TWO-COMPONENT MODELS OF INFLATION

In order to close the system of equations for the model with combined matter, we must first establish the equation of state of the substance (ideal liquid). This equation is usually chosen in the form $p = \gamma\varepsilon$. The quantity γ determines the type of the substance. For $\gamma = 1/3$, this is usually isotropic radiation, for $\gamma = 0$, this is dust, and so on. In the most general case, $\gamma = \gamma(t)$ is a function of time, which describes the change in the state of the substance filling the Universe through a sequence of phase transitions. Second, we must indicate the conditions specifying the field dynamics. Such conditions can be (a) the explicit representation of the self-action potential $V(\phi)$ as a function of field ϕ and, perhaps, of t; (b) the effective equation of state of the field: $\mathcal{P} = \mathcal{P}(W)$; and (c) the law of the time evolution of the field, e.g., $\dot{W} = -kW^\alpha$.

The approach based on condition (a) is usually applied in quantum models and is determined by physical quantum properties of the field. Its main drawback is that upon the time variation of the self-action potential $V = V(\phi, t)$, which corresponds to phase transitions into the states of early Universe, this time evolution in homogeneous models does not differ from the complicated dependence

$$V = V(\phi, t) = \tilde{V}(\phi).$$

For this reason, such a method of defining the field dynamics does not specify rigorously all physical mechanisms in the model.

Approach (b) is an attempt to interpret field as a material substance of the type of a gas of relativistic particles, probably, with an exotic equation of state. The above analysis shows that the effective equation of state can be chosen in the form (25). Such a choice corresponds to quite definite mechanisms of the evolution of field perturbations (e.g., the evolution of the velocity of sound for the field component of matter is fixed). Another disadvantage of this method is the ambiguity in the choice of the equation of state and the complexity of the description of kinetics. The last and the simplest approach (c) rigidly fixes the time evolution of the field energy density and, hence, unambiguously determines the evolution of all the remaining components of the model. This can be interpreted as the condition that the field evolution is determined only by its intrinsic properties (decay of the quanta of this field). For this reason, we will henceforth use the third method of determination of the field dynamics.

8. THE FIXED LAW OF ENERGY DENSITY EVOLUTION FOR A SCALAR FIELD

The system of equations of the model in the case when the evolution rate of the field energy density

is given has the form

$$\dot{W} = -kW^{\alpha}, \quad p = \gamma\varepsilon,$$
$$\mathcal{P} = -W + \frac{k}{\sqrt{3\kappa}}\frac{W^{\alpha}}{\sqrt{\varepsilon + W}}, \tag{33}$$
$$\frac{\dot{\varepsilon}}{\varepsilon + p} = -\sqrt{3\kappa}\sqrt{W + \varepsilon},$$

$$R(t) = R_0 \exp\left[\int_{t_0}^{t} \sqrt{\frac{\kappa}{3}(W(t) + \varepsilon)}\,dt\right]. \tag{34}$$

Here, we assume that $\gamma = \text{const}$. Using the first equation of this system, we derive the law of evolution of the field energy density, which coincides with Eq. (22). Then we derive the equation of the energy density for the substance:

$$\frac{\dot{\varepsilon}}{(\gamma + 1)\varepsilon\sqrt{\varepsilon + W(t)}} = -\sqrt{3\kappa}. \tag{35}$$

In order to determine the form of solution ε for models of field evolution with $W \sim t^{-m}$, we consider first the case when $W = W_0\text{const}$, which corresponds to the quasi-vacuum equation of state for field component of matter: $\mathcal{P} = -W = \text{const}$. In this specific case, Eq. (35) can be integrated exactly and its solution can be written in the form

$$\varepsilon(t) = \frac{W_0}{\sinh^2(\Omega t)}, \tag{36}$$

where

$$\Omega = (\gamma + 1)\frac{\sqrt{3\kappa W_0}}{2}.$$

This solution shows that for a constant level of energy density of the field component of matter, the energy density of an ideal liquid increases as t^{-2} for $t \longrightarrow 0$ and decreases in proportion to $e^{-2\Omega t}$ for $t \longrightarrow \infty$. The Universe in this case does not pass to the Friedman mode, but follows the de Sitter scenario after a certain time.

The general property of the solutions of Eq. (35) for $W \sim t^{-m}$ is the monotonic decrease of $\varepsilon(t)$ upon an increase in t. Indeed, since $\varepsilon \geq 0$ and $W \geq 0$, we have $\dot{\varepsilon} < 0$. Assuming that ε decreases more slowly than W in the limit $t \longrightarrow \infty$ and hence disregarding the value of W as compared to ε in the denominator of Eq. (35), we arrive at the asymptotic equation

$$\frac{d\varepsilon}{\varepsilon^{3/2}} = -(\gamma + 1)\sqrt{3\kappa}\,dt. \tag{37}$$

It follows hence that as $t \longrightarrow \infty$, we have

$$\varepsilon(t) \sim \frac{q}{t^2}, \tag{38}$$

where

$$q = \frac{4}{3(\gamma + 1)^2\kappa}. \tag{39}$$

This result does not contradict the initial assumption that $W < \varepsilon$ for $t \longrightarrow \infty$ since $W(t) \sim t^{-m}$ and $m = (\alpha - 1)^{-1} > 2$ for $1 \leq \alpha < 3/2$. Thus, in accordance with the last equation in system (33), the Universe asymptotically passes to the Friedman mode.

Let us now consider the behavior of ε in the vicinity of instant $t = 0$, i.e., near the $W(t)$ singularity. It is natural to assume that at the initial instant, the entire energy is in the field form of matter and conventional matter appears as a result of the decay of the field. For this reason, the value of ε near $t = 0$ must be bounded or, at least, the condition $W \gg \varepsilon$ must be satisfied. However, an analysis of Eq. (37) near $t = 0$ indicates a quite complex behavior of the model variables near $t = 0$. In analogy with Eq. (36), we will seek the solution for ε in the form

$$\varepsilon(t) = \frac{W(t)}{\sinh^2\Theta(t)}. \tag{40}$$

Substituting this relation into Eq. (37), we arrive at the following equation for the function $\Theta(t)$:

$$\dot{\Theta} = \frac{\dot{W}}{2W}\tanh\Theta + \operatorname{sgn}\Theta\frac{\gamma + 1}{2}\sqrt{3\kappa W}. \tag{41}$$

Note the fact that the second term in this equation contains a discontinuous function (the sign of function Θ). This determines the main properties of the asymptotic forms of Θ. Substituting $W(t)$ from Eq. (22) into this equation, we obtain

$$\dot{\Theta} = -\frac{m}{2t}\tanh\Theta + \operatorname{sgn}\Theta\frac{\gamma + 1}{2}\sqrt{3\kappa w_0}\,t^{-m/2}. \tag{42}$$

Let us assume that in the range of small values of $t \sim 0$, the function Θ is positive and quite large, and $\tanh\Theta \sim 1$. In the range of such values of t, we have

$$\Theta \longrightarrow \Theta_0 - \frac{m}{2}\ln t - \frac{\gamma + 1}{m - 2}\sqrt{3\kappa w}\,t^{1 - m/2}. \tag{43}$$

For $1 < \alpha < 3/2$, we have

$$2 - m = 2 - \frac{1}{\alpha - 1} = \frac{2\alpha - 3}{\alpha - 1} < 0.$$

Consequently, the principal term of the asymptotic form (43) is negative and increases indefinitely in magnitude as $t \longrightarrow 0$:

$$\Theta \longrightarrow \frac{\gamma + 1}{2 - m}\sqrt{3\kappa w}\,t^{1 - m/2} \longrightarrow -\infty.$$

It follows hence that for $t \longrightarrow 0$, there exists an instant at which Θ vanishes. Let us now assume the converse: let the function Θ be smaller than zero in a certain range of $t \sim 0$. Then the principal term of the asymptotic form

of Θ is positive and increases indefinitely. Consequently, for $t \longrightarrow 0$, Θ becomes equal to zero again at a certain instant. It can also be seen that Eq. (42) has the exact solution $\Theta(t) = 0$. All that has been said above can be formulated as follows. Equation (42) for all $t > 0$ has only integral curves originating from points on the abscissa axis at a certain fixed instant τ, which is determined by the initial condition, i.e., from points $(\tau, \Theta(\tau) = 0)$, and then are continued in the region of positive (see Fig. 3) or negative values of $\Theta(t)$.

In accordance with Eq. (40), this means that before the instant τ, the energy density of matter is infinitely large and the Universe is in a singular state. At instant τ, when an integral curve emerges from point $(\tau, \Theta(\tau) = 0)$, the Universe leaves the singular state, passes through the inflation stage, and then arrives at the Friedman mode. Thus, in the scenarios under investigation with $1 < \alpha < 3/2$, the formation of the Universe from a singular state occurs after a finite time interval following the "decay" of the singular state with an infinitely large energy density of the field at instant $t = 0$. Actually, this conclusion has a more general meaning and holds not only for scenarios of $W = W(t)$ with $1 < \alpha < 3/2$. The behavior of ε analyzed above is connected with the sign of the quantity $\gamma + 1$ on the right-hand side of Eq. (37). The above analysis is valid for all scenarios with $\gamma + 1 > 0$ and with W decreasing as $t \longrightarrow \infty$.

This result seem to be rather unexpected. An analysis of the behavior of the integral curves of Eq. (42) for $\gamma + 1 > 0$ indicates that Eq. (35) has no real solutions satisfying the initial condition $\varepsilon(t_0) = 0$ for finite values of t_0. This means that the second component of matter (if it is included in the model) is necessarily in the singular state on a finite time interval in the past. Consequently, if we want to construct a model in which conventional substance is absent at the instant of the decay of the singular state and is generated with time as a result of the decay of field "quanta," we must specify more exactly what is meant under the condition of the absence of substance at the initial stage of the formation of the Universe. The fulfillment of this condition corresponds to two essentially different physical situations. One of them corresponds to the requirement $\varepsilon = 0$ for a given equation of state at instant $t = 0$ and consequently cannot be realized (see above). The other corresponds to the requirement that at instant $t = 0$, the equation of state of the substance resembles the field equation, i.e., corresponds to the quasi-vacuum case: $p \sim -\varepsilon$. In other words, at the initial stages, the second component of matter can be treated as a second scalar field which is transformed into a substance in the course of the evolution of the Universe. The second situation can be realized in the framework of the model under investigation as follows. The equation of state of the second component can vary with time, e.g., according to the law $p = \gamma(t)\varepsilon$ so that the quantity $\gamma(t) \sim -1$ at the initial stages of the evolution, passing to a value $\gamma > 0$ at the postinflation stage.

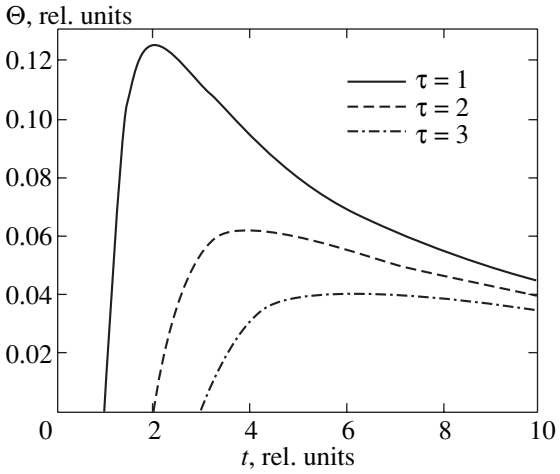


Fig. 3. Typical integral curves for equation of function $\Theta(t)$ in the model with $\alpha = 5/4$.

Let us make a few remarks concerning the interpretation of the dependence $\gamma = \gamma(t)$. It is clear from quantum-mechanical considerations that each value of the quantity γ corresponds to a definite type of substance or radiation. For example, typical values are $\gamma = 1/3$ for radiation with zero-mass particles or $p = 0$ for a dustlike matter. Consequently, the time variation of γ could indicate a smooth transition of substance or radiation from one form to another, i.e., a smooth phase transition. The latter is ruled out in view of the quantum nature of radiation and substance. However, this dependence can also be interpreted in a different way. If we assume that the second component of matter is comprised of several forms of matter in thermal equilibrium, the equation of state of this mixture can be treated as the weighted mean equation of state. Indeed, let us suppose that each component of the mixture has an equation of state in the form

$$p_i = \gamma_i \varepsilon_i, \quad i = 1, 2, \ldots,$$

where $\gamma_i = \text{const}$. In this case, the total pressure of all components is the sum of partial pressures and the total energy density is the sum of the partial densities of the components:

$$p = \sum_i p_i = \sum_i \gamma_i \varepsilon_i, \quad \varepsilon = \sum_i \varepsilon_i.$$

As a result, we can introduce the following equation of state for the mixture:

$$p = \frac{\sum_i \gamma_i \varepsilon_i}{\sum_i \varepsilon_i} = \gamma(t)\varepsilon,$$

where

$$\gamma(t) = \frac{\sum_i \gamma_i \varepsilon_i(t)}{\sum_i \varepsilon_i(t)}.$$

The dependence of parameter γ on t *i*s determined by the variation of the relative concentration of individual components in the mixture, which also reflects, in fact, phase transitions occurring in the substance, but in the form of a smooth variation of the percentage of the components. In our subsequent analysis, we will stick to just this concept of time variation of γ.

Thus, the scenario under consideration presumes that at the inflation stage, the Universe has two components of matter of the field form. One of these components behaves as field during the entire evolution of the Universe, while the other component has a variable equation of state (i.e., the variable composition of the mixture of the components constituting it). At the inflation stage, the energy density of the second component decreases more slowly than that of the first (purely field) component. In the second component, the energy is stored, which is transformed into a substance with $\gamma \geq 0$ as a result of the phase transition after the completion of the inflation stage.

9. EVOLUTION OF PARAMETER $\gamma(t)$ IN TWO-COMPONENT MODELS IN THERMAL EQUILIBRIUM

We assume that for $\gamma = \gamma(t)$, the form of the evolution of the Universe for $t = 0$ corresponds to values of energy density $\varepsilon \ll W$ and $W = \infty$. The subsequent evolution must follow a scenario in which the energy density ε must decrease more slowly than W during a certain finite time interval Δt: $\dot{\varepsilon} > 0$, $t \in \Delta t$. As a result, by the end of the interval, a sufficient amount of matter of the Universe will be transformed into a substance upon a change in the equation of state. At the present stage, the substance has an equation of state close to that for a dustlike matter: $\gamma \sim 0$. Consequently, the following asymptotic conditions must be satisfied for γ as a function of time: $\gamma \longrightarrow -1$ for $t \longrightarrow 0$, where $\delta(t) \leq 0$, and $\gamma(t) \longrightarrow 0$ as $t \longrightarrow \infty$. Besides, it should be noted that there existed epochs in the course of evolution, when the Universe was filled with radiation with the equation of state $\gamma = 1/3$. Consequently, the dependence $\gamma = \gamma(t)$ must attain a positive value of γ close to 1/3 in the time interval $(0, \infty)$.

The main difficulty encountered while solving the problem of evolution of $\gamma(t)$ lies in the formulation of the physical conditions governing this evolution. The variation of $\gamma(t)$ as a quantity characterizing the properties of the structure of one of the components of matter can be directly connected only with the thermodynamic parameters of the evolution: the energy densities of both components, their pressure, and temperature. The most important question here is whether or not the two components of matter are in thermodynamic equilibrium with each other. It is natural to assume that equilibrium can be absent only on certain short intervals of time during which rapid phase transitions in one or both components of matter take place. Consequently, the thermal equilibrium condition, which is equivalent here to the condition of the equality of the temperatures of both components, can be regarded as most suitable for describing the process of a simultaneous evolution of the two components. It turns out that this condition is sufficient for describing the evolution of $\gamma(t)$ also.

Let us use Eq. (26) for describing the time dependence of the common temperature of the matter components under the assumption of their thermal equilibrium. The equation of state for the first component has the form (33), while for the second component, we have $p = \gamma(t)\varepsilon$. Writing Eq. (26) for both components separately, we obtain the following system of equations for calculating the dependences $T(t)$ and $\gamma(t)$:

$$\frac{d}{dS}\left(\frac{W^\alpha}{\sqrt{W+\varepsilon}}\right) - \frac{W^\alpha}{\sqrt{W+\varepsilon}} = \frac{\sqrt{3\kappa}}{k}\frac{dW}{dS}, \tag{44}$$

$$\frac{d(\gamma\varepsilon)}{dS} = (\gamma+1)\varepsilon, \tag{45}$$

where $S = \ln T$. Eliminating the time derivative $\dot{S}$ from these equations and using Eqs. (33) and (35), we obtain the following equation for $\delta(t) = \gamma(t) + 1$:

$$\dot{\delta} = -\delta A(t) + \delta^2 B(t), \tag{46}$$

where

$$A(t) = \frac{kW^{\alpha-1}}{2(W+\varepsilon)}[(2\alpha-1)W + 2\alpha\varepsilon],$$

$$B(t) = \sqrt{3\kappa}\frac{2W+3\varepsilon}{2\sqrt{W+\varepsilon}}.$$

For a given evolution $\varepsilon = \varepsilon(t)$, this equation can be integrated in quadratures. Its general solution has the form

$$\delta(t) = -\frac{\Delta_0 F_1(t)}{\Delta_0 F_2(t) - 1}, \tag{47}$$

where

$$F_1(t) = \exp\left[-\int_{t_0}^{t} A(t)dt\right],$$

$$F_2(t) = \int_{t_0}^{t} F_1(t)B(t)dt,$$

and $\Delta_0 = \delta(t_0)$ is the value of function $\delta(t)$ for $t = t_0$. The joint solution of Eqs. (46) and (35) cannot be obtained in an explicit form. However, we are interested in the

asymptotic behavior of the solutions of these equations for $t \longrightarrow 0$ and $t \longrightarrow \infty$ as well as in the existence of local extrema of function δ for $t > 0$. Such an analysis can be carried out on the basis of the preset properties of the evolution $\varepsilon(t)$ in these limits.

Let us first carry out an asymptotic analysis of the solutions of Eq. (46) for $t \longrightarrow \infty$ on the basis of the following two assumptions. The first corresponds to the condition $\varepsilon \longrightarrow \varepsilon_\infty$ for $t \longrightarrow \infty$, and the second, to $\varepsilon \longrightarrow qt^{-2}$ in the same limit. The former case corresponds to the return of the second component to a quasi-vacuum state and the latter to the return to an ideal liquid.

In both cases, we have $W \ll \varepsilon$ for $t \longrightarrow \infty$. Let us suppose that $\varepsilon \longrightarrow \varepsilon_\infty$ in this case. This gives

$$A(t) \longrightarrow kw_0^{\alpha-1}t^{-1}\alpha = \alpha(\alpha-1)^{-1}t^{-1},$$

$$B(t) \longrightarrow \sqrt{3\kappa\varepsilon_\infty},$$

and Eq. (46) assumes the form

$$\dot{\delta} = -\alpha_0 t^{-1}\delta + d_0\delta^2,$$

where $a_0 = \alpha(\alpha-1)^{-1} > 0$ and $b_0 = \sqrt{3\kappa\varepsilon_\infty} > 0$. Integrating the last equation with the help of the substitution

$$\delta = \mu\frac{d\ln\chi}{dt}, \tag{48}$$

we obtain

$$\delta \longrightarrow -\frac{\alpha-1}{b_0}\frac{1}{C_1 t^{a_0} - t} \longrightarrow 0 \quad \text{for} \quad t \longrightarrow \infty$$

(C_1 is the integration constant). It follows hence that $\gamma \longrightarrow -1$ in this limit, as expected. Consequently, this solution is allowed for the given model with thermal equilibrium.

The importance of the obtained solution is determined by an analysis of a more general situation. Let us consider general solutions of the form $\varepsilon \sim \varepsilon_\infty t^\beta$ for $t \longrightarrow \infty$ and for $\varepsilon \gg W$. We will seek the asymptotic form for δ in the power form also: $\delta \sim \delta_0 t^\nu$. Substituting these functions into the equation for ε and δ, we find that $\beta = -(\alpha-1)^{-1} < 0$ and $\gamma = -\beta/2 - 1 > 0$. It follows hence that the thermal equilibrium condition in the general case corresponds to the same asymptotic behavior of the two components of matter: $W \sim t^{-1/(\alpha-1)}$ and $\varepsilon \sim t^{-1/(\alpha-1)}$. The deviation from this behavior is possible only in the special case for $\varepsilon \longrightarrow \varepsilon_\infty > 0$. Besides, it follows from the obtained asymptotic solution that δ increases indefinitely as $\varepsilon \longrightarrow 0$. This inevitably leads with time to a violation of the energy dominance principle, which in unacceptable from the physical point of view. Consequently, the only correct solution of the equations for ε and δ are those satisfying the condition $\varepsilon \longrightarrow \varepsilon_\infty > 0$ for $t \longrightarrow \infty$. It was proved above that in this case, $\delta \longrightarrow 0$ as $t \longrightarrow \infty$. The only exception from this general rule may be associated with a search for the solution for the case when $\varepsilon \longrightarrow qt^{-2}$.

In the limit $t \longrightarrow \infty$, for $\varepsilon \longrightarrow qt^{-2}$ we have

$$A(t) \sim \frac{a_0}{t},$$

$$B(t) \sim \frac{b_1}{t},$$

where $a_0 = \alpha(\alpha-1)^{-1}$ and $b_1 = 3\sqrt{3\kappa q}$. Equation (35) in this limit leads to the requirement

$$\delta \longrightarrow \delta_\infty = \frac{2}{\sqrt{3\kappa q}} = \text{const} \tag{49}$$

while Eq. (46) assumes the form

$$t\dot{\delta} = -a_0\delta + b_1\delta^2.$$

Integrating this equation, we obtain

$$\delta \longrightarrow \frac{\alpha-1}{\alpha b_1}\frac{1}{1 - C_1 t^{a_0}} \quad \text{for} \quad t \longrightarrow \infty.$$

Here, C_1 is also the integration constant. Since $a_0 > 0$, $\delta \longrightarrow 0$ if $C_1 \neq 0$ Thus, this solution matches with the initial assumption $\delta \longrightarrow \delta_\infty > 0$ only when $C_1 = 0$. In this case, we have

$$\delta \longrightarrow \delta_\infty = \frac{\alpha-1}{\alpha b_1} > 0.$$

Equating the obtained asymptotic value of δ_∞ to the value from Eq. (49), we find that this is possible only if

$$\alpha = 3/2.$$

Thus, as expected, there exists only one model of the evolution of the field component of type (33), corresponding to $\alpha = 3/2$, for which the Universe can pass to the Friedman scenario after the completion of inflation under the condition of thermal equilibrium between the matter components. In the case with $\alpha = 3/2$ corresponding to an ideal liquid, inflation is absent (see above). For this reason, this version should be discarded.

Before formulating the final conclusions, let us consider the asymptotic form of model parameters for $t \longrightarrow 0$.

An analysis of the behavior of δ for $t \sim 0$ indicates that

$$A(t) \sim A_0 t^{-1}, \quad B(t) \sim B_0 t^{-m/2},$$

where

$$A_0 = \frac{2\alpha-1}{2(\alpha-1)} > 0, \quad B_0 = \sqrt{3\kappa w_0} > 0.$$

Equation (46) for δ in this limit has the form

$$t^{m/2}\dot{\delta}(t) = A_0 t^{(m-2)/2}\delta + B_0\delta^2.$$

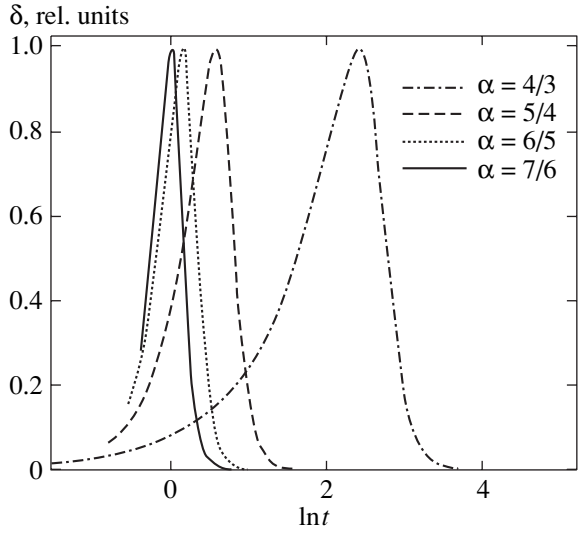


Fig. 4. Variation of parameter δ normalized to the maximum value for models with various values of parameter α.

The solution of this equation is the function

$$\delta(t) = D_0 \frac{t^{m/2-1}}{1 + C_1 t^{(A_0+1)(m/2-1)}},$$

where

$$D_0 = \frac{2}{\sqrt{3\kappa w_0}(3-2\alpha)} > 0. \quad (50)$$

It follows from this expression that for $t \longrightarrow 0$, the value of $\delta(t)$ tends to zero, remaining positive. This means that this component of matter as well as the first one, is initially in a quasi-vacuum state. However, the parameter $\delta(t) = \gamma(t) + 1$ subsequently increases and the substance acquires the properties of an ideal liquid.

It should be observed that, in accordance with Eq. (46), there exists at least one maximum of δ on the evolution interval $0 < t < \infty$. If we denote by t_* the instant of attainment of this maximum of δ ($\dot{\delta}(t_*) = 0$, $\delta(t_*) = \delta_{max}$), we obtain from Eq. (46)

$$\delta_{max} = \frac{A(t_*)}{B(t_*)} > 0.$$

This inequality is a consequence of the fact that we have $A(t) > 0$ and $B(t) > 0$ on the entire interval $t > 0$. As was noted above, it should be expected that $\delta_{max} \sim 4/3$.

The above asymptotic analysis confirms the results of the numerical analysis of the system of equations, which is reduced for convenience to the following form:

$$\xi' = -\xi\sqrt{1+\xi}\Delta - \frac{2}{3-2\alpha\tau}\xi, \quad (51)$$

$$\Delta' = \frac{1}{(3-2\alpha)\tau}\frac{(2\alpha-1)+2\alpha\xi}{1+\xi}\Delta - \frac{2+3\xi}{2\sqrt{1+\xi}}\Delta^2. \quad (52)$$

Here, $\xi = \varepsilon/W$, $\Delta = \delta\Delta_0$, and

$$\Delta_0 = \sqrt{3\kappa w_0}\tau = \frac{2(\alpha-1)t^{-(3-2\alpha)/2(\alpha-1)}}{3-2\alpha}.$$

In this form, the condition $\tau \longrightarrow 0$ corresponds to $t \longrightarrow \infty$ and, conversely, $\tau \longrightarrow \infty$ corresponds to $t \longrightarrow 0$. Such an approach makes it possible to formulate approximate boundary conditions for the model parameters for $t \longrightarrow \infty$, replacing the asymptotic conditions by approximate ones. This, however, complicates the analysis for large values of t. It is impossible to extend the solution beyond the minimum step in τ. At the same time, this facilitates the analysis for $t \longrightarrow 0$.

The results of numerical solution of these equations for δ for some values of parameter α are presented in Fig. 4. These solutions were obtained for the initial conditions $\delta(\tau_*) = \delta_*$ and $\ln(\varepsilon(\tau_*)/W(t_*)) = 0$ where $\tau_* = 0.01$ is a small quantity, which corresponds to a long time interval t_*, and $\delta_* = 0.00001$, which is also a quite small value for numerical calculations.

Let us now consider the behavior of ε for $t \longrightarrow 0$, assuming, as before, that $W \gg \varepsilon$ in this limit. Taking into account relation (50), we can write Eq. (35) in the form

$$\dot{\varepsilon} = -\Delta_0 t^{-1}\varepsilon = -\frac{2}{3-2\alpha}t^{-1}\varepsilon.$$

It follows hence that as $t \longrightarrow 0$,

$$\varepsilon \sim t^{-D},$$

where

$$D = \frac{2}{3-2\alpha} > 0.$$

Condition $W \gg \varepsilon$ holds if

$$\frac{2}{3-2\alpha} < \frac{1}{\alpha-1}.$$

This condition is equivalent to

$$\alpha < 5/4.$$

Figure 5 shows the results of numerical analysis of the model equations in the function $\ln(\varepsilon/W)$ for the same values of parameter α and initial conditions as in Fig. 4. It can be seen from the figure that the rates of variation of ε and W in the vicinity of $t \sim 0$ differ insignificantly, but these values themselves differ by many orders of magnitude (as expected, $\varepsilon \ll W$). As $t \longrightarrow \infty$, the function $\ln(\varepsilon/W)$ increases linearly, which means that the quantity ε starts dominating in this limit. In this solution, δ tends to a small value for large t (see Fig. 4) and, hence, $\varepsilon \longrightarrow \varepsilon_\infty$ for large values of t, which follows from the asymptotic analysis. Since $W \longrightarrow 0$ in this case, the value of ε_∞ corresponds to the cosmological constant [16].

In accordance with Eqs. (9) and (12), the evolution of the Hubble parameter and of the scaling factor is determined exclusively by the behavior of the total energy. Consequently, there is no need to analyze their behavior separately. The behavior of the scaling factor these models on the whole can be characterized as follows. First, during the inflation epoch, the scaling factor increases exponentially rapidly. Then, in the epoch with $\delta > 1$, the increase in the scaling factor slows down and the evolution resembles the Friedman mode, but at the final stage, for $\delta \longrightarrow 0$, the exponential increase is resumed.

Concluding this section, let us briefly consider the properties of models with a cosmological constant, which correspond to Eqs. (23), (24). It is important to estimate the role played by the cosmological constant in the light of the above analysis. The most significant in these models is the behavior of the components of matter for $t \longrightarrow \infty$. In the case $\Lambda = 0$, the condition $\varepsilon \gg W$ holds starting from a certain (probably, large) instant of time τ, while for $\Lambda \neq 0$, the converse condition $\varepsilon \ll W \sim \Lambda$ holds in this limit since $\varepsilon \sim t^{-2}$ and $\Lambda = \text{const}$. In this case, as $t \longrightarrow \infty$, the second component of matter is transformed into a field. Vacuum "dissolves" the substance. If, however, the value of Λ is small, quite a long time interval may exist in reality (as long as $\varepsilon \gg W$), on which the above scenario is realized. For this reason, models with small but nonzero values of Λ are equivalent on the whole to models with $\Lambda = 0$ up to very long lifetimes of the Universe.

10. THE EVOLUTION OF TEMPERATURE AND VELOCITY OF SOUND

Let us now consider the behavior of temperature in the model under investigation. For this purpose, we reduce Eq. (44) to the form

$$\frac{d}{dt}S(t) = \sqrt{3\kappa}\sqrt{W+\varepsilon} - k\alpha W^{\alpha-1} + \frac{kW^{\alpha}}{2(W+\varepsilon)} + \frac{\delta(t)\sqrt{3\kappa}}{2\sqrt{W+\varepsilon}}\varepsilon. \quad (53)$$

As expected, for $t \sim 0$, when $W \gg \varepsilon$, the behavior of temperature does not differ strongly from its behavior in purely field models (see Eq. (28)). For this reason, we consider the behavior of temperature for $t \longrightarrow \infty$, when $W \ll \varepsilon$. In this case, Eq. (53) can be transformed to

$$\frac{d}{dt}S(t) \sim \sqrt{3\kappa\varepsilon} - \frac{\alpha}{\alpha-1}t^{-1} + \frac{kW^{\alpha}}{2\varepsilon} + \frac{\delta(t)}{2}\sqrt{3\kappa\varepsilon}. \quad (54)$$

Substituting the asymptotic forms $\delta \longrightarrow 0$ and $\varepsilon \longrightarrow \varepsilon_{\infty}$ obtained for $t \longrightarrow \infty$, we arrive at the asymptotic form for $S = \ln T$ for $t \longrightarrow \infty$:

$$\frac{d}{dt}S(t) \longrightarrow \sqrt{3\kappa\varepsilon_{\infty}} > 0. \quad (55)$$

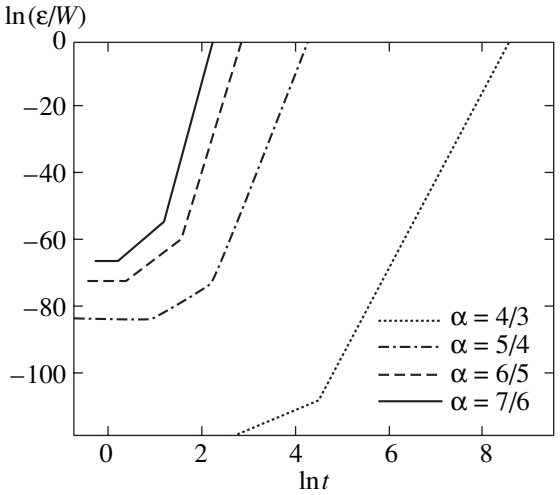


Fig. 5. Variation of parameter $\ln(\varepsilon/W)$ for models with various values of parameter α.

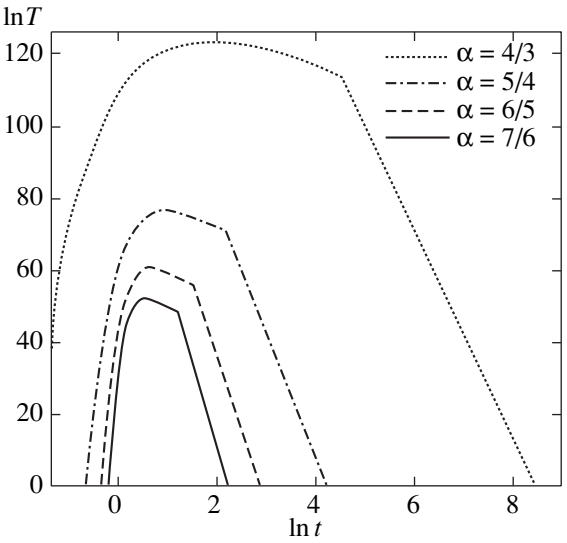


Fig. 6. Temperature variation for models with various values of parameter α.

Thus, the temperature in this model increases with time according to an exponential law but with a very small growth rate equal to $\sqrt{3\kappa\varepsilon_{\infty}}$. Consequently, in the models under investigation, secondary heating of substance due to the transition of the Universe to the stage of accelerated expansion takes place. The mechanism of this secondary heating is the same as at the primary inflation stage. This secondary heating does not coincide in time with the secondary heating in standard models and takes place a considerable time after the emergence of the Universe at the secondary inflation stage.

Figure 6 shows the results of numerical analysis of the equation for $\ln T$ for the same values of parameter α

and the same boundary values as in the previous figures. We solved the equation

$$S' = \Delta_0\sqrt{1+\xi} - \frac{\alpha}{(\alpha-1)\tau} + \frac{1}{(3-2\alpha)\tau}\frac{(2\alpha-1)+2\alpha\xi}{1+\xi} + \frac{\Delta\xi}{\sqrt{1+\xi}}, \quad (56)$$

in which the notation is the same as in Eqs. (51) and (52). No secondary asymptotic increase in temperature is observed on the curves in view of the specific definition of the initial conditions for the equations in variables τ. The asymptotic growth is beyond the time interval on which the solution was obtained. A more detailed numerical solution of the problem is required for its observation, which was impossible at the present stage of investigations.

In connection with our investigations, it would be interesting to analyze the evolution of the velocity of sound in the two-component model. By virtue of the equation of state $p = \gamma(t)\varepsilon$, the velocity of sound in the substance obeys the law $c_s = \sqrt{\gamma(t)}\,c = \sqrt{\delta(t)-1}\,c$. For the field component of matter, we obtain from Eqs. (14) and (21)

$$\mathcal{P} = -W + \frac{k}{\sqrt{3\kappa}}\frac{W^\alpha}{\sqrt{E}}.$$

This leads to the following expression for the velocity of sound:

$$c_s^2 = \frac{d\mathcal{P}}{dW} = c^2\left\{-1 + \frac{k}{\sqrt{3\kappa}}\left[\frac{\alpha W^{\alpha-1}}{\sqrt{E}} - \frac{1}{2}\frac{W^\alpha}{\sqrt{E^3}}\left(1+\frac{d\varepsilon}{dW}\right)\right]\right\}$$

$$= c^2\left[-1 + \frac{k}{\sqrt{3\kappa}}\frac{W^{\alpha-1}}{\sqrt{E}}\left(\alpha - \frac{W}{2E}\right) - \frac{\gamma+1}{2}\frac{\varepsilon}{E}\right].$$

In the limit $t \longrightarrow \infty$, for $\varepsilon \longrightarrow \varepsilon_\infty$ and $\delta \longrightarrow 0$, we obtain $c_s^2 \longrightarrow -1$, as expected. The asymptotic increase in temperature established above can be attributed in this case to the transition of the Universe to the de Sitter scenario, which ultimately takes place.

11. CONCLUSIONS

The above analysis shows that the aggregate of the available data on the evolution of the Universe at the inflation stage and after it, as well as physically reasonable considerations, necessitate an analysis of cosmological models with at least two components of matter. One of these components is purely of the field type, while the other must have a varying equation of state of substance. Under the quite natural assumption that the components of matter are in thermal equilibrium, the set of equations of the model leads to the virtually unambiguous conclusion that, as energy of the purely field component decreases monotonically, the second component, whose initial equation of state is close to the quasi-vacuum equation, is transformed with time to a substance with the equation of state $p = \gamma(t)\varepsilon$ for $\gamma > 0$. After a finite time interval, the second component inevitably returns to the quasi-vacuum state and the Universe passes to the accelerated expansion mode. In such scenarios, the temperature increases at the initial inflation stage, attains its peak value when the velocity of sound becomes zero, then decreases to a certain finite value, after which it again increases exponentially, when the second component with a variable equation of state approaches the quasi-vacuum state again.

The asymptotic transition to the de Sitter stage can be interpreted as the transformation of substance into a vacuum-like state ("thawing," or dissolution of the substance in vacuum). It is well known that the substance does not disintegrate spontaneously (protons and electrons are stable under normal conditions). Consequently, "thawing" of the substance must occur under specific conditions corresponding to very high energies (e.g., in the interior of neutron stars or white dwarfs). If an alternative to the model considered above does not exist, the condensation of substance into neutron stars and white dwarfs from the cosmological point of view is the gradual conversion of matter in their central regions into a vacuum-like field. The accelerated expansion of the Universe discovered recently [5] and mentioned in the Introduction may serve as an indirect proof of this model. In the model considered by us here, this means that at the modern epoch, the value of parameter γ is close to zero (the epoch of substance dominance), but is already negative so that the Universe gradually passes to the de Sitter stage. Moreover, since the formation of a "new vacuum" occurs in a dense matter, this vacuum must engulf primarily such objects as galaxies. We can also assume that "dark matter" in the light of the given model is a quasi-vacuum field "recovered" in galaxies in the course of prolonged evolution.

It has been proved by us here that an alternative to the model with a transition to the de Sitter stage may be scenarios with a violation of the energy dominance principle, for which the value of $\gamma(t)$ increases indefinitely as $t \longrightarrow \infty$. Besides, there exists a preferred scenario in which the scalar field has an effective equation of state equivalent to the equation of state $\mathcal{P} = \gamma_0 W$ for an ideal liquid, for which $W \sim t^{-2}$, which corresponds to $\alpha = 3/2$. Such a scenario is possible for the self-action potential $V(\phi) \sim e^{-2\phi}$. In this case, the second component also tends asymptotically to the similar equation of state $p = \gamma_\infty\varepsilon$; however, in this case, the above-mentioned problem with an asymptotic increase in the particle mass arises (see also [10]).

Another alternative can be models in which individual components of matter are not in thermal equilibrium. Probably, this assumption corresponds to the initial stages of the evolution (especially, to epochs of phase transitions), but as $t \longrightarrow \infty$, the equilibrium must

set in. In fact, the absence of equilibrium for $t \longrightarrow \infty$ indicates that the components of matter do not interact directly or indirectly, which is too exotic a requirement. On the other hand, the violation of thermal equilibrium at some epochs in the past causes only a temporary deviation from the scenario considered above and does not affect the asymptotic behavior of the Universe. It can be assumed, however, that thermal equilibrium does not set in simultaneously in the entire Universe. In this case, the asymptotic transition to the secondary inflation occurs only in its individual parts, in which the equilibrium stabilizes more quickly. Such a scenario resembles some types of global inflation scenarios [4] and is probably the most admissible approach to the interpretation of the results obtained in the present work.

The author has not been able to think of any other reasonable alternative scenarios which would describe a different asymptotic behavior. Nevertheless, in order to confirm the validity of the above conclusions, it is necessary to calculate the perturbation spectrum of the density of matter (field and substance) in the framework of the given model in order to be able to draw a final conclusion about the effectiveness of the model for explaining the experimental facts.

ACKNOWLEDGMENTS

The author is grateful to S.V. Chervon and V.K. Shchigolev for fruitful discussions of the subject of this paper.

This work was partly supported financially by the Russian Foundation for Basic Research (project no. 01-02-17935).

REFERENCES

1. W. de Sitter, Proc. R. Acad. Sci. Amsterdam **18**, 1217 (1917); **20**, 229 (1917).
2. A. H. Guth, Phys. Rev. D **23**, 347 (1981).
3. A. D. Linde, Phys. Lett. B **108**, 389 (1982).
4. A. D. Linde, *Physics of Elementary Particles and Inflation Cosmology* (Nauka, Moscow, 1990).
5. S. Perlmutter *et al.*, astro-ph/9812133; astro-ph/9812473; A. G. Riess *et al.*, astro-ph/9804065.
6. A. A. Starobinsky, Phys. Lett. B **91**, 99 (1980).
7. B. Spokoiny, Phys. Lett. B **315**, 40 (1993).
8. R. Maartens, D. R. Taylor, and N. Roussos, Phys. Rev. D **52**, 3358 (1995).
9. S. V. Chervon and V. M. Zhuravlev, gr-qc/9907057 (1999).
10. V. M. Zhuravlev and S. V. Chervon, Zh. Éksp. Teor. Fiz. **118**, 259 (2000) [JETP **91**, 227 (2000)].
11. S. V. Chervon, in *Nonlinear Fields in Theory of Gravitation and Cosmology* (Srednevolzhsk. Nauchn. Tsentr, Ul'yanovsk, 1997), p. 191.
12. S. V. Chervon, V. M. Zhuravlev, and V. K. Shchigolev, Phys. Lett. B **398**, 269 (1997).
13. V. M. Zhuravlev, S. V. Chervon, and V. K. Shchigolev, Zh. Éksp. Teor. Fiz. **114**, 179 (1998) [JETP **87**, 223 (1998)].
14. J. Garriga and V. F. Mukhanov, Phys. Lett. B **458**, 219 (1999); gr-qc/9904176.
15. A. A. Starobinsky, Pis'ma Zh. Éksp. Teor. Fiz. **68**, 757 (1998) [JETP Lett. **68**, 757 (1998)].
16. V. Sahni and A. A. Starobinsky, astro-ph/9904398; Int. J. Mod. Phys. D **9**, 373 (2000).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 93, No. 5, 2001, pp. 920–925.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 5, 2001, pp. 1062–1068.

GRAVITATION, ASTROPHYSICS

Axial Magnetostatics of a Ring Current in a Kerr Field

A. A. Shatskiy

Astrospace Center, Lebedev Institute of Physics, Russian Academy of Sciences,
Profsoyuznaya ul. 84/32, Moscow, 119991 Russia
e-mail: shatskiy@lukash.asc.rssi.ru

Received March 28, 2001

Abstract—The electromagnetic fields generated by a ring current around a Kerr black hole have been found. The acceleration of a charged particle by a force electric field along the rotation axis is investigated in the constructed model, as applied to the astrophysics of quasars. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Studying the interaction of electromagnetic fields with the gravitational field of a rotating black hole is of great importance in understanding the astrophysics of quasars. Quasars manifest themselves as compact, intense sources of electromagnetic radiation, which occasionally have huge narrowly directed jets and which are most likely active galactic nuclei. The Blandford–Znajek process [1] is one of the models that accounts for observational manifestations of quasars. A magnetohydrodynamic (MHD) model of plasma accreting onto a rotating black hole underlies this process. Through the Bardeen–Petterson process [2], accretion can proceed only from the equatorial plane; therefore, it makes sense to model accretion as a superposition of equatorial ring currents. Such modeling is proper if the pair production by an induced electric field gives rise to currents that are much weaker than the source ring current. The Hawking effect of particle production on the horizon is negligible in this model, because it gives a negligible correction when the Compton particle wavelength is much smaller than the radius of space curvature. For electrons, this corresponds to a black hole with a mass larger than about $10^{-16}M_{\odot}$.

In all the cited studies, calculations were performed by using 3 + 1 formalism. In contrast to these studies, we use the general covariant formalism of general relativity and do not use the approximation of MHD magnetic field line freezing-in in plasma, which leads to the condition for the scalar product of the electric and magnetic fields being equal to zero (force-free field).

We introduce same notation for electromagnetic and gravitational quantities as in [3].[1]

[1] Below, we use the system of units $c = 1$, the speed of light, and $G = 1$, the gravitational constant, for convenience.

2. SPECIFYING BOUNDARY CONDITIONS

Let us specify the boundary conditions that must be imposed on the electromagnetic-field tensor components required to determine the latter.

First, the classical boundary conditions must be satisfied: all field components must become zero at infinity.

Second, the boundary conditions on the horizon must also be added to the classical ones. They are required so that, in the frame of reference associated with a freely falling observer (FFO), no anomalies arise in its motion as it flies up to the horizon due to the electromagnetic-field acceleration (anomalies result in the violation of the condition for the FFO electric charge being a test one.[2] Therefore, to find the conditions imposed on the field components on the horizon, we must write out the 4-vector of FFO acceleration and establish which of its components have the anomalies on the horizon related to the electromagnetic field.

Below, we use the coordinates that are at rest with respect to an infinitely remote, static observer.

The FFO 4-acceleration components are[3]

$$W^i = w^i + \frac{e}{m}F^{ij}u_j. \qquad (1)$$

Here, e and m are the FFO test charge and mass, respectively; $w^i = -\Gamma^i_{km}u^k u^m$, where Γ^i_{km} are the Christoffel symbols; F^{ij} are the contravariant tensor components of the electromagnetic field; and u^k are the FFO 4-velocity components. Let us calculate the 4-acceleration for a radially falling FFO in the Schwarzschild metric. The components of the Christoffel symbols and the FFO 4-velocity for this metric are given in [3]. Denoting the

[2] A more detailed discussion of this requirement can be found in [4, 5].

[3] Unless stated otherwise, the Latin indices run the series 0, 1, 2, 3, while the Greek indices run the series 1, 2, 3.

Schwarzschild radius by $r_g = 2M$, where M is the black-hole mass, we have

$$u^0 = -g_{rr}, \quad u^r = -\sqrt{\frac{r_g}{r}}, \quad w^0 = -\frac{1}{r}\left(\frac{r_g}{r}\right)^{3/2},$$
$$w^1 = \frac{r_g}{2r^2}(2 + g_{rr}).$$

The remaining u^i and w^i are zero. When approaching the horizon ($r \longrightarrow r_g$), the metric tensor component $g_{rr} \longrightarrow -\infty$. Therefore, let us write out the asymptotics of the squares of the observed 4-acceleration components in the principal (in g_{rr}) order expressed in terms of the contravariant field components:

$$\begin{aligned} W_0 W^0 &\longrightarrow -g_{rr}\left(\frac{e}{m}\right)^2 \frac{r_g}{r}(F^{r0})^2, \\ W_r W^r &\longrightarrow g_{rr}\left(\frac{r_g}{2r^2}g_{rr} + \frac{e}{m}F^{r0}\right)^2, \\ W_\theta W^\theta &\longrightarrow g_{rr}^2\left(\frac{e}{m}\right)^2 \frac{r_g}{r} g_{\theta\theta}(F^{r\theta})^2, \\ W_\varphi W^\varphi &\longrightarrow g_{rr}^2\left(\frac{e}{m}\right)^2 \frac{r_g}{r} g_{\varphi\varphi}(F^{r\varphi})^2. \end{aligned} \tag{2}$$

We thus see that the singularity in the radial acceleration component does not result in the violation of the condition for the charge being a test one for the following reasons: (1) the gravitation near the horizon in the radial direction acts a factor of g_{rr} more strongly than the electromagnetic field; (2) the zero W component has a weak singularity; and (3) the tangential acceleration components have a strong singularity, which can significantly change the FFO trajectory near the horizon and can violate the condition for the FFO charge being a test one: $e \ll m$. The contravariant components of the tangential magnetic field must be set equal to zero on the horizon, lest this happen. Similarly, the asymptotics of the squares of the 4-acceleration components dependent on the covariant field components is

$$\begin{aligned} W_0 W^0 &\longrightarrow -g_{rr}\left(\frac{e}{m}\right)^2 \frac{r_g}{r}(F_{r0})^2, \\ W_r W^r &\longrightarrow g_{rr}\left(\frac{r_g}{2r^2}g_{rr} + \frac{e}{m}F_{r0}\right)^2, \\ W_\theta W^\theta &\longrightarrow g_{rr}^2\left(\frac{e}{m}\right)^2 g^{\theta\theta}(F_{\theta 0})^2, \\ W_\varphi W^\varphi &\longrightarrow g_{rr}^2\left(\frac{e}{m}\right)^2 g^{\varphi\varphi}(F_{\varphi 0})^2. \end{aligned} \tag{3}$$

We thus see that the covariant components of the tangential electric field with the strongest singularity must become equal to zero on the horizon. However, the Kerr field rather than the Schwarzschild field is of physical interest. Let us write out the Kerr metric and its determinant in a nonrotating (relative to remote stars) frame:

$$\begin{aligned} ds^2 &= \left(1 - \frac{r_g r}{\rho^2}\right)dt^2 - \frac{\rho^2}{\Delta}dr^2 - \rho^2 d\theta^2 \\ &- \left(r^2 + a^2 + \frac{r_g r a^2}{\rho^2}\sin^2\theta\right)\sin^2\theta d\varphi^2 + \frac{2r_g r a}{\rho^2}\sin^2\theta d\varphi dt, \\ -g &= \rho^4 \sin^2\theta \end{aligned} \tag{4}$$

(here, $\rho^2 = r^2 + a^2\cos^2\theta$, $\Delta = r^2 + a^2 - r_g r$, and a is the Kerr parameter). In this case, in the Kerr field, both the space itself and the FFO are drawn into rotation when approaching a black hole. The conditions for the tangential electromagnetic field components in the Schwarzschild field then change to the conditions for the same components in the frame of reference observers in the Kerr field.[4] Expressions (2) and (3) give a summary of the boundary conditions for the electromagnetic field on the horizon in the Kerr field:

$$\begin{aligned} F''_{\theta 0} &\longrightarrow 0, \quad F''_{\varphi 0} \longrightarrow 0, \\ F''^{r\theta} &\longrightarrow 0, \quad F''^{r\varphi} \longrightarrow 0. \end{aligned} \tag{5}$$

Let us write out formulas that can be of use in the subsequent analysis. Designating

$$\kappa = -g_{00}/g_{\varphi\varphi}, \quad \kappa + \Omega^2 = \Delta\sin^2\theta/g_{\varphi\varphi}^2,$$

we then have

$$\begin{aligned} g^{\theta\theta} &= 1/g_{\theta\theta}, \quad g^{rr} = 1/g_{rr}, \quad g^{00} = -\frac{g_{\varphi\varphi}}{\Delta\sin^2\theta}, \\ g^{\varphi\varphi} &= -\frac{g_{00}}{\Delta\sin^2\theta}, \quad g^{0\varphi} = \frac{g_{0\varphi}}{\Delta\sin^2\theta}. \end{aligned} \tag{6}$$

3. PASSAGE TO THE REFERENCE FRAME

Since the current is axisymmetric and stationary, the system has two Killing vectors: along the translations in time and along the translations in angle φ. Therefore, all fields in frame (4) are stationary. In general, this stationarity is not conserved when passing to a frame of reference rotating with an arbitrary angular velocity. To show this, let us write out the coordinate transformations to a frame of reference rotating with a spatially nonuniform angular velocity $\omega(r, \theta)$ required in the subsequent analysis:

$$dx^i = dx'^k[\delta_k^i + \omega\delta_\varphi^i\delta_k^0 + t\delta_\varphi^i\delta_k^\beta\partial_\beta\omega]. \tag{7}$$

[4] In what follows, by the reference frame we mean a frame in which $g''_{0\varphi} = 0$ (see [4, 5]). This is the frame in which the reference observers rotate with angular velocity $\Omega = -g_{0\varphi}/g_{\varphi\varphi}$ relative to remote stars and have a zero angular momentum.

The corresponding transformations of the contravariant field components (see [3]) are

$$F'^{\alpha 0} = F^{\alpha 0}, \quad F'^{0\varphi} = F^{0\varphi} + tF^{\alpha 0}\partial_\alpha \omega,$$
$$F'^{\alpha\varphi} = F^{\alpha\varphi} - \omega F^{\alpha 0}, \quad F'^{r\theta} = F^{r\theta}. \tag{8}$$

For the covariant components, we derive[5]

$$F'_{\alpha\varphi} = F_{\alpha\varphi}, \quad F'_{\alpha 0} = F_{\alpha 0} + \omega F_{\alpha\varphi},$$
$$F'_{0\varphi} = F_{0\varphi}, \quad F'_{r\theta} = F_{r\theta} + te^{\alpha\beta\varphi}F_{\alpha\varphi}\partial_\beta \omega. \tag{9}$$

In expressions (8) and (9), α runs the values of r and θ. We see from these expressions that only the frames of reference rigidly rotating relative to another stationary frame (relative to remote stars) are stationary. The stationary frame of reference rotating with the horizon angular velocity Ω_H coincides with the FFO frame on the horizon. Therefore, the boundary conditions in this horizon frame, which below is marked by a tilde,[6] are the same as (5):

$$\tilde{F}_{\theta 0} \longrightarrow 0, \ \tilde{F}_{\varphi 0} \longrightarrow 0, \ \tilde{F}^{r\theta} \longrightarrow 0, \ \tilde{F}^{r\varphi} \longrightarrow 0. \tag{10}$$

4. MAXWELL EQUATIONS

Suppose that the current has a density with a delta-shaped distribution function in the meridional plane:

$$j^i(r, \theta) = J^\varphi \delta^i_\varphi [\delta(r - r_0)\delta(\theta - \theta_0)]/\sqrt{-g}. \tag{11}$$

Since the other currents are assumed to be negligible and since the frame is stationary, the toroidal electromagnetic-field components are zero. Therefore, the axisymmetric Maxwell equations for the covariant field components outside the horizon are

$$e^{\alpha\beta\varphi}\partial_\beta F_{\alpha\varphi} = 0,$$
$$e^{\alpha\beta\varphi}\partial_\beta F_{\alpha 0} = 0. \tag{12}$$

We thus see that the covariant tensor components of the electromagnetic field can be represented as

$$F_{\alpha\varphi} = \partial_\alpha A_\varphi, \quad F_{\alpha 0} = \partial_\alpha A_0. \tag{13}$$

Here, A_i are the covariant components of the 4-vector electromagnetic-field components. Let us now write the axial Maxwell equations for the contravariant field components without electric charges:

$$\frac{1}{\sqrt{-g}}\partial_\alpha(\sqrt{-g}F^{\alpha 0}) = 0,$$
$$\frac{1}{\sqrt{-g}}\partial_\alpha(\sqrt{-g}F^{\alpha\varphi}) = 4\pi j^\varphi. \tag{14}$$

The right-hand part of the second equation in (14) is given by expression (11).

5. DETERMINING THE MAGNETIC-FIELD COMPONENTS

Let us determine the magnetic field in the Schwarzschild metric [$a = 0$ in (4)]. Because of axial symmetry, the vector potential of the magnetic field has the toroidal component alone. According to (13), we have for its covariant part

$$F^{\gamma\varphi} = g^{\gamma\beta}g^{\varphi\varphi}\partial_\beta A_\varphi. \tag{15}$$

Denoting the tensor components in Euclidean space by the subscript 0, we then have, according to (6),

$$F^{\gamma\varphi} = g_0^{\gamma\beta}\left(\delta^\gamma_\beta - \frac{r_g}{r}\delta^r_\beta\delta^\gamma_r\right)g_0^{\varphi\varphi}\partial_\beta A_\varphi$$
$$= F_0^{\beta\varphi}\left(\delta^\gamma_\beta - \frac{r_g}{r}\delta^r_\beta\delta^\gamma_r\right). \tag{16}$$

We see from (15) and (16) that if the function A_φ is smooth and has no singularities, then the boundary conditions (5) are satisfied with the required asymptotics (2). The second equation in (14) can then be rewritten as

$$\frac{1}{\sqrt{-g}}\partial_\alpha(\sqrt{-g}F_0^{\alpha\varphi}) = 4\pi j^\varphi_{\text{tot}}, \tag{17}$$

where we designated

$$j^\varphi_{\text{tot}} = j^\varphi + \frac{r_g}{4\pi r^2}\partial_r(rF_0^{r\varphi}).$$

Next, let us introduce the physical vector components by the definition (see [3])

$$\hat{B}^\alpha = B^\beta\sqrt{|g_{\alpha\beta}|}. \tag{18}$$

In these components, Eq. (17) has the form of the Poisson equation in Euclidean space for the vector potential $\hat{A}^\gamma$ with source $\hat{j}^\gamma_{\text{tot}}$. The solution of this equation is known to be (see [3])

$$\hat{A}^\alpha(r, \theta) = \delta^\alpha_\varphi\hat{e}^\varphi\int_{r_g}^{\infty}\int_0^\pi\int_{-\pi}^{\pi} dr'd\theta'd\varphi'\frac{\hat{j}'^\gamma_{\text{tot}}\hat{e}_\gamma}{|\mathbf{r} - \mathbf{r}'|}r'^2\sin\theta'. \tag{19}$$

Here, $\hat{e}^\gamma$ is a unit vector[7] in the direction of angle φ,

$$|\mathbf{r} - \mathbf{r}'|^2 = r^2 + r'^2 - 2rr'(\cos\theta\cos\theta' + \sin\theta\sin\theta'\cos\varphi'),$$

θ is the inclination of vector $\mathbf{r}$ to the z axis, θ' is the inclination of vector $\mathbf{r}'$ to the z axis, and φ' is the angle

[5] $e^{\alpha\beta\gamma} = e_{\alpha\beta\gamma}$ is the Levi–Civita symbol.

[6] A tilde denotes the electromagnetic field components in a stationary frame of reference that passes to the horizon reference frame, i.e., to the frame rotating with the FFO angular velocity on the horizon, $\Omega_H = a/r_H r_g$.

[7] $\hat{j}^\gamma_{\text{tot}}\hat{e}_\gamma = j'^\varphi_{\text{tot}}r'\sin\theta'\cos\varphi'$.

between the projections of vectors **r** and **r'** onto the plane perpendicular to the *z* axis. The integration is performed in Euclidean space outside the sphere of radius r_g. Solution (19) is obtained by iterations; we assume that $r_g = 0$ in the initial iteration.

Let us determine the magnetic vector

$$H_\lambda = -\frac{\sqrt{-g}}{2} e_{\lambda\beta\gamma} F^{\beta\gamma} \tag{20a}$$

or

$$F^{\alpha\beta} = -\frac{1}{\sqrt{-g}} H_\lambda e^{\lambda\alpha\beta}. \tag{20b}$$

According to (18) and (20), the components A_φ, H_r, and H_θ are given by

$$\begin{aligned} A_\varphi &= -r\sin\theta\hat{A}^\varphi, \quad H_r = \frac{\partial_\theta(\sin\theta\hat{A}^\varphi)}{r\sin\theta}, \\ \frac{H_\theta}{r} &= -\frac{1 - r_g/r}{r}\partial_r(r\hat{A}^\varphi). \end{aligned} \tag{21}$$

For the initial iteration, according to (21), the following expressions can be derived for the magnetic-field components from (19) and because of the presence of delta functions in (11):

$$\begin{aligned} H_r^0(r,\theta) &= \frac{J^\varphi}{r_0} \\ &\times \int_{-\pi}^{\pi} d\varphi' \frac{\cos\theta\cos^2\varphi'(1 + x^2 - x\sin\theta\cos\varphi')}{x\sin\theta(1 + x^2 - 2x\sin\theta\cos\varphi')^{3/2}}, \\ \frac{H_\theta^0(r,\theta)}{r} &= \frac{J^\varphi}{r_0}\int_{-\pi}^{\pi} d\varphi' \frac{x\sin\theta\cos^2\varphi' - \cos\varphi'}{x(1 + x^2 - 2x\sin\theta\cos\varphi')^{3/2}}. \end{aligned} \tag{22}$$

Here, we designated $x = r/r_0$. In particular, we obtain

$$\begin{aligned} \lim_{\theta\to 0}\left(\frac{H_\theta^0}{r}\right) &= 0, \quad \lim_{x\to\infty}\left(\frac{H_\theta^0}{r}\right) = \frac{\pi J^\varphi \sin\theta}{r_0 x^3}, \\ \lim_{\theta\to 0} H_r^0 &= \frac{2\pi J^\varphi\cos\theta}{r_0(1 + x^2)^{3/2}}. \end{aligned} \tag{23}$$

6. DETERMINING THE ELECTRIC POTENTIAL

To calculate the electric potential A_0, we use the formula

$$F^{\alpha 0} = \{F_{\beta 0} g^{00} + F_{\beta\varphi} g^{0\varphi}\} g^{\alpha\beta}. \tag{24}$$

However, the components $F^{\alpha 0}$ must be expressed in terms of $\tilde{F}_{\beta 0}$ and $\tilde{F}^{\beta\varphi}$, because we know the boundary conditions only for them in stationary frames of reference. For the poloidal electromagnetic-field components, using formulas (6), (8), and (9), and the expression $g^{\alpha i} g_{\gamma i} = g^{\alpha\beta} g_{\gamma\beta} = \delta^\alpha_\gamma$,[8] we obtain

$$F^{\alpha 0} = L_f[g^{00} g^{\alpha\beta}\tilde{A}_{0,\beta} + \tilde{F}^{\alpha\varphi}(\Omega_H - \Omega)/(\kappa + \Omega^2)], \tag{25}$$

where $L_f = [1 - (\Omega - \Omega_H)^2/(\kappa + \Omega^2)]^{-1}$ has the meaning of FFO Lorentz factor relative to the stationary frame of the horizon.

Hence, in the second approximation in $a^* = a/M = L/M^2$, the dimensionless black-hole angular momentum ($0 < a^* < 1$), the first equation in (14) can be reduced to

$$\Delta\tilde{A}_0 = -4\pi(f(H_\alpha) + U(\tilde{A}_0)). \tag{26}$$

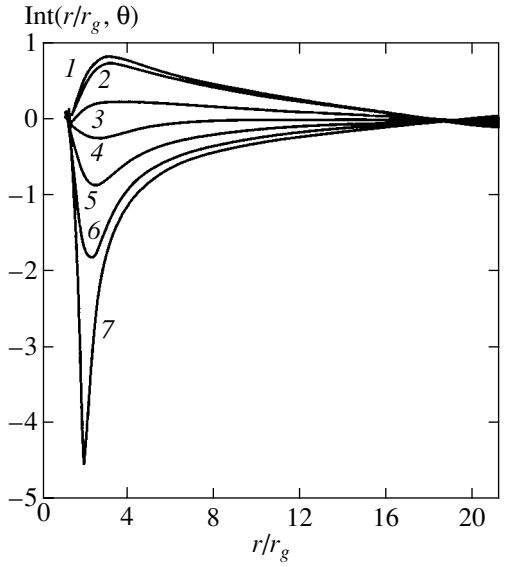
Here, Δ is the Laplace operator in Euclidean space and H_α are the Schwarzschild magnetic-field components derived in the preceding section:

$$\begin{aligned} U(\tilde{A}_0) &= -\frac{r_g}{4\pi r^3}\partial_r(r^2\partial_r\tilde{A}_0), \\ f(H_\alpha) &= \frac{a}{4\pi r_g^2}\left(1 - \frac{r_g}{r}\right) \\ &\times\left[\left(1 + \frac{r_g}{r} + \frac{r_g^2}{r^2}\right)(2H_r\cos\theta - 4\pi j^\varphi r^2\sin^2\theta)\right. \\ &\left. -\left(2 + \frac{r_g}{r}\right)\frac{H_\theta}{r}\sin\theta\right]. \end{aligned}$$

The expression in round brackets on the right-hand side of Eq. (26) is an analog of the electric charge density in the Poisson equation. However, because of the second term, the $\tilde{A}_0$-dependent function *U*, it can be solved by the iteration method assuming that $U = 0$ in the initial iteration. According to the boundary conditions (10), solving Eq. (26) is equivalent to calculating the potential $\tilde{A}_0$ produced by the density of electric charge $\rho = f + U$ around a conductive sphere of radius r_g in Euclidean space. This external problem for the Poisson equation can be solved by the image method [6]. The solution that satisfies the boundary conditions (10) is

$$\begin{aligned} \tilde{A}_0(r,\theta) &= \int_{r_g}^{\infty}\int_0^\pi\int_{-\pi}^{\pi} dr' d\theta' d\varphi' \rho(r',\theta') \\ &\times\left\{\frac{1}{|\mathbf{r} - \mathbf{r}'|} - \frac{r_g/r'}{\left|\mathbf{r} - \mathbf{r}'\frac{r_g^2}{r'^2}\right|}\right\} r'^2\sin\theta'. \end{aligned} \tag{27}$$

[8] For α running the values of *r* and θ, we have $g^{0\alpha} = 0$ and $g_{0\alpha} = 0$.



The dependence Int(r/r_g, θ) for $r_0/r_g = 2$ and θ = 0 (curve *1*), π/12 (*2*), π/6 (*3*), π/4 (*4*), π/3 (*5*), 5π/12 (*6*), and π/2 (*7*).

The integration is performed in Euclidean space outside the sphere of radius r_g [see (19)]. According to (7), we obtain the potential A_0 from (27)

$$A_0 = \tilde{A}_0 - \Omega_H \tilde{A}_\varphi, \quad A_\varphi = \tilde{A}_\varphi. \tag{28}$$

We can factor aJ^φ/r_0 outside integral (27). Denoting the magnetic-field strength at the center of the system in the absence of a black hole by $H_0 = 2\pi J^\varphi/r_0$ [see (23)], we derive for $\tilde{A}_0$ and A_0

$$\begin{aligned} \tilde{A}_0 &= \widetilde{\mathrm{Int}}\left(\frac{r}{r_g}, \theta\right)\frac{a^* M H_0}{2\pi}, \\ A_0 &= \mathrm{Int}\left(\frac{r}{r_g}, \theta\right)\frac{a^* M H_0}{2\pi}. \end{aligned} \tag{29}$$

Here, $\widetilde{\mathrm{Int}}(x)$ and $\mathrm{Int}(x)$ are dimensionless functions, which can be numerically calculated using expressions (27) and (28). The corresponding results are shown in the figure. Some important characteristics of the solution can also be established analytically.

(1) Since the system has a mirror symmetry relative to the inversion in the equatorial plane,[9] the electric-field direction on the axis depends on the coincidence of the directions of black-hole angular momentum and ring-current moment.

(2) We see from (27) that $\tilde{A}_0 \longrightarrow 0$ when $r \longrightarrow r_H$ (the expression in curly braces becomes zero). This important result has a simple explanation.

[9] This can be seen from (26) and from the expression for f ($\sin\theta$ and H_θ are even, while $\cos\theta$ and H_r are odd relative to the mirror inversion).

Since the total charge under the horizon is zero, the electric field must pierce the horizon in different directions. Therefore, there must be an equipotential surface that separates these directions, pierces the horizon, and goes to infinity. Since the potential $\tilde{A}_0$ changes neither on this surface, nor on the horizon (in view of the boundary conditions), the horizon potential is equal to the potential at infinity (zero). This conclusion has an important implication: the existence of a local extremum for the potential $\tilde{A}_0$. Similar reasoning also applies to A_0. It might seem that a nonzero electric charge density must arise near the extremum. In reality, however, this is not the case: the charge density is determined by the contravariant electric-field components ($F^{\alpha 0}$ rather than $F_{\alpha 0}$), while the derivation of (27) is based on the first equation in (14), a zero charge density in the entire space. We emphasize that there is no extremum in the well-known solution of Wald [7] for a black hole placed in a uniform magnetic field aligned along the symmetry axis of the black hole. This is because the field does not vanish at infinity in this solution, and this nonphysical boundary condition wipes out the extremum.

7. THE KINETIC ENERGY OF A CHARGED PARTICLE EMERGING ALONG THE AXIS

The conserved mass–energy of a charged particle with a zero angular momentum component along the z axis is given by [4, 5]

$$\begin{aligned} E &= m_0\gamma_L\sqrt{\frac{\rho^2\Delta}{(r^2+a^2)^2 - a^2\Delta\sin^2\theta}} \\ -qA_0 &= m_0\gamma_L^\infty, \end{aligned} \tag{30}$$

where γ_L is the Lorentz factor of the particle, and q is its charge.

Consider the acceleration of an emerging charged particle along the z axis by an electric field. Since $qA_0 = -|q\tilde{A}_0|$ in this case, the kinetic energy of this particle at infinity with γ_L, g_{00}, and A_0 specified at any point of the z axis is

$$E_k = E - m_0 = m_0(\sqrt{g_{00}}\gamma_L - 1) + |qA_0|. \tag{31}$$

Since the electromagnetic energy of a charged particle for quasars is much larger than the corresponding rest energy, it is convenient to represent the result for A_0 as the energy (in electronvolts) that an elementary charge emerging along the axis acquires. For magnetic fields $H_0 \approx 10^4$ G and mass $M \approx 10^8 M_\odot$, we obtain from (29) and (31)

$$E_k\ [\mathrm{eV}] \approx 10^{19} a^* \mathrm{Int}\left(\frac{r}{r_g}, 0\right).$$

8. CONCLUSION

We have calculated the electric field generated by a rotating black hole that interacts with an external magnetic field. As can be seen from the above discussion, this field is quadrupole in nature, being actually an analog of the pulsar electric field (unipolar inductor model [8]).

The situation on the z axis was not chosen by chance. Because of strong magnetic fields, the Larmor radii of a charged particle must be of the order of the gravitational radius of the system, and the particle acceleration mechanism will be effective only in directions close to the z axis, where the Lorentz force does not act.

The reader may ask a legitimate question: Will a strong electric field produce electron–positron plasma near the horizon and will it destroy the force field $\mathbf{E} \cdot \mathbf{H} \neq 0$? At large gradients in A_0, this can actually happen, and the problem should then be solved in the force-free approximation (see review papers [9–11] on this subject). However, as numerical estimates show, the electric-field strength in the model does not exceed 10^7 V cm^{-1}, while a strength of the order of 10^9 V cm^{-1} is required for the particle production.[10] In addition, in the force-free approximation, particles cannot be accelerated along the z axis, because there is no electric field in this direction. Besides, in any direction in a force-free field, the electric and magnetic fields equally act on the particle. Therefore, even if its trajectory goes to infinity, the particle loses most of its energy in the process. Recall that jets are observed in quasars precisely in the directions along and opposite to the axis. For this reason, solving the problem under consideration in the force-field approximation is of considerable importance.

[10]The production of particles by inverse-Compton-type effects is not considered, because the particle number density around the black hole is assumed to be low.

ACKNOWLEDGMENTS

I wish to thank N.S. Kardashev, V.S. Beskin, V.N. Lukash, B.V. Komberg, M.I. Zel'nikov, Yu.Yu. Kovalev, and all the remaining staff of the theoretical departments at the Astrospace Center and the Lebedev Institute of Physics, as well as participants of workshops for taking an active part in preparing the article and for important remarks. This study was supported by the Russian Foundation for Basic Research (project nos. 00-15-96698 and 01-02-17829).

REFERENCES

1. R. D. Blandford and R. L. Znajek, Mon. Not. R. Astron. Soc. **179**, 433 (1977).
2. J. M. Bardeen and J. A. Petterson, Astrophys. J. Lett. **195**, L65 (1975).
3. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Pergamon, Oxford, 1975; Nauka, Moscow, 1988).
4. *Black Holes: the Membrane Paradigm*, Ed. by K. S. Thorne, R. H. Price, and D. A. Macdonald (Yale Univ. Press, New Haven, 1986; Mir, Moscow, 1988).
5. I. D. Novikov and V. P. Frolov, *Physics of Black Holes* (Nauka, Moscow, 1986).
6. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media* (Nauka, Moscow, 1982; Pergamon, New York, 1984).
7. R. M. Wald, Phys. Rev. D **10**, 1680 (1974).
8. P. Goldreich and W. Julian, Astrophys. J. **157**, 869 (1969).
9. V. S. Beskin, Usp. Fiz. Nauk **167**, 689 (1997) [Phys. Usp. **40**, 659 (1997)].
10. V. S. Beskin, Ya. N. Istomin, and V. I. Par'ev, Astron. Zh. **69**, 1258 (1992) [Sov. Astron. **36**, 642 (1992)].
11. J. S. Heyl, Phys. Rev. D **63**, 064028 (2001); gr-qc/0012007; A. Tomimatsu and H. Koyama, Phys. Rev. D **61**, 124010 (2000); gr-qc/0002020.

Translated by V. Astakhov

Journal of Experimental and Theoretical Physics, Vol. 93, No. 5, 2001, pp. 926–940.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 5, 2001, pp. 1069–1085.

GRAVITATION, ASTROPHYSICS

Discrete Quantum Gravity and Anomalies

S. N. Vergeles

Landau Institute for Theoretical Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia
e-mail: vergeles@itp.ac.ru

Received May 10, 2001

Abstract—A new version of quantum gravity on discrete spaces (simplicial complexes) is proposed. A theory of gravitation interacting with Dirac field is considered. This theory is shown to be free of reparametrization anomaly. The problem of axial gauge anomaly and the associated problem of the doubling of fermion states on a lattice are discussed. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Since the traditional methods of quantization of gravity in four dimensions prove to be inconsistent because of ultraviolet divergences, a natural idea has arisen that qualitatively different physics takes place at supersmall distances (of the order of the Planck length and smaller).[1]

At present, a predominant opinion among theoretical physicists is that the superstring theory is a fundamental physical theory. In a ten-dimensional space, the superstring theory is self-consistent and free of divergences. The superstring theory involves gravitational interaction. Therefore, we have the following qualitative situation within the string theory: the quantum theory of gravity represents a part (the long-wavelength limit) of the superstring theory. Thus, the problems of ultraviolet divergences and the definition of the quantum theory of gravity are solved.

However, one encounters an extremely difficult problem within the string ideology, the problem of the compactification of six dimensions and the construction of a long-wavelength physics in four dimensions. Therefore, recently, actual progress in solving many problems of the quantum theory of gravity and quantum cosmology in this line has not been made.

The aforesaid justifies the existence of certain other ideas underlying the fundamental quantum field theory and, first of all, the theory of gravity. In our opinion, the most interesting idea is the idea of discrete space–time, which is the main subject of the present paper.

The idea of the discreteness of space–time (as applied to the theory of gravity) was first formulated in the pioneer work by Regge [2] long before the appearance of string theory. According to Regge, the role of smooth Riemannian spaces is played by piecewise flat spaces, namely, simplicial complexes (necessary information from the theory of simplicial complexes is given at the beginning of the next section). Each one-dimensional simplex (edge) is assigned its length, so that, if the set of three edges forms a boundary of a two-dimensional simplex (a triangle), then the lengths of these edges satisfy the triangle inequalities. Thus, the geometry of a complex is completely defined. The quantity representing an analogue of the Riemann tensor on a smooth Riemannian space proves to be nonzero only on a set of $(D-2)$-dimensional simplices (D is the dimension of the simplicial complex); i.e., the curvature tensor becomes a distribution.

Consider a two-dimensional piecewise flat surface in greater detail. Let α^r be the number of r-dimensional simplices of a complex $\mathfrak{K}$. In the two-dimensional case, $r = 0$, 1, and 2, and the lengths of all one-dimensional simplices are specified. Let us show that such a surface can be embedded into a three-dimensional Euclidean space $\mathbb{R}^3$. First, assume that our surface is embedded into $\mathbb{R}^N$. Then, the total number of available degrees of freedom is equal to $N\alpha^0$ (locations of vertices) minus the number of constraints α^1 (the lengths of edges). This number must be no less than the number of parameters of the group of translations and rotations of the space $\mathbb{R}^N$. Thus, we have the following estimate:

$$N\alpha^0 - \alpha^1 \geq N + \frac{1}{2}N(N-1).$$

When the complexes are too large, we can neglect the right-hand side, so that

$$N \geq \alpha^1/\alpha^0.$$

Note that

$$\chi = \alpha^0 - \alpha^1 + \alpha^2$$

is the Euler characteristic of the complex, which is invariant for a given topology [3, 4]. Therefore, for very large complexes, we can set

$$\alpha^0 - \alpha^1 + \alpha^2 = 0.$$

[1] In this sense, the quantum theory of gravity presented in [1] should be considered as a phenomenological theory useful for studying concrete problems (for example, the theory of a black hole inside a horizon) rather than as a fundamental theory

Moreover, the following relation always holds:

$$2\alpha^1 = 3\alpha^2.$$

Combining the relations obtained, we derive the estimate $N \geq 3$. Thus, we can state that any two-dimensional piecewise flat surface is embedded into a three-dimensional Euclidean space.

Consider in $\mathbb{R}^3$ a neighborhood of a certain vertex p of the complex $\mathfrak{K}$. Suppose that the vertex p is connected by edges with other n vertices of the complex with numbers $1, 2, \ldots, n$. Moreover, suppose that the triangles

$$\Delta_1 = pn1, \quad \Delta_2 = p12, \ldots, \quad \Delta_n = p(n-1)n$$

belong to the complex (a triangle with vertices p, i, and j is denoted by pij). Let us place the vertex p into the origin of coordinates of the space $\mathbb{R}^3$. Then, the plane containing a triangle Δ_i is defined by the equation

$$n_x(i)x + n_y(i)y + n_z(i)z = 0,$$

where $n_x(i)$, $n_y(i)$, and $n_z(i)$ are the components of a unit vector normal to Δ_i. Suppose that $n_z(i) \neq 0$ for all i. Then, the coordinates x and y may serve as local coordinates for all triangles Δ_i, and the vertex p has the coordinates $x = y = 0$. By definition, a metric in this piecewise flat surface is induced by the Euclidean metric of space $\mathbb{R}^3$. Therefore, the Riemann tensor R_{xyxy} vanishes in the interiors of the triangles. Obviously, the Riemann tensor also vanishes at the boundaries between the triangles Δ_i and Δ_{i+1}, except possibly at their vertices. Indeed, a part of the complex $\mathfrak{K}$ near the vertex p can be considered as a limit of a family of smooth cones with an induced metric and, hence, with a zero Riemann tensor everywhere except for the vertex.

Let θ_i be the interior angle of the triangle $p(i-1)i$ at the vertex p. Define the defect ε_p of the angle at the vertex p as

$$\sum_{i=1}^{n} \theta_i = 2\pi - \varepsilon_p.$$

According to the aforesaid, the Riemann tensor on a piecewise flat surface vanishes everywhere, except possibly at its vertices. Therefore, the Riemann and Ricci tensors, as well as a scalar curvature, are distributions with supports at the vertices of a complex. It is clear that the definition of a scalar curvature by the equality

$$\sqrt{|g|}R = 2\sum_i \varepsilon_i \delta(x - x_i)\delta(y - y_i) \qquad (*)$$

(where x and y are local coordinates on a piecewise flat surface, i is the number of a vertex of the surface, ε_i is the defect of the angle at this vertex, $g_{\mu\nu}$ is the metric tensor, and R is a scalar curvature) is reasonable. Indeed, first, relation (*) is invariant with respect to the change of local coordinates. Second, when $|\varepsilon_p| \ll 1$, a parallel translation of a vector belonging to a tangent space of a piecewise flat space around the vertex p results in the rotation of this vector through the angle ε_p, which is defined by the following formula in the orthonormal basis:

$$\varepsilon_p = \int_\sigma R_{1212}\,dx\,dy.$$

In this case, $g_{\mu\nu} = \delta_{\mu\nu}$ and $R = 2R_{1212}$. Here, σ is a domain bounded by the contour, containing the vertex p, along which the vector is translated. The equality obtained can be rewritten in the form independent of the local coordinates:

$$\varepsilon_p = \frac{1}{2}\int d^2x\sqrt{|g|}R.$$

This result confirms the equality (*).

In the D-dimensional case, the situation is analogous to that described above: the Riemann tensor is nonzero only on $(D-2)$-simplices. If we denote a $(D-2)$-simplex by s and its $(D-2)$-volume by $\bar{s}$, then the analogue of the equality (*) can be expressed as

$$\int d^Dx\sqrt{|g|}R = 2\sum_s \bar{s}\varepsilon_s, \qquad (1)$$

where ε_s is the defect of the angle around the simplex s.

The more detailed account of the Regge calculus is given in [5]. An approach to discrete geometry similar to the Regge calculus can also be found in [6, 7].

Despite obvious elegance of the Regge calculus, action (1) proves to be very inconvenient when passing to quantum theory. Indeed, the independent variables that determine the right-hand side of (1) are the lengths of one-dimensional simplices subject to a large number of constraints, namely, to triangle inequalities. Moreover, the introduction of Dirac fields to the theory creates a new difficulty consisting in the absence of orthonormal bases in the explicit form. Possibly, this is the reason why the variant of discrete gravity based on the so-called *B–F*-theory is currently being developed more intensively.

To formulate the ideas developed on the basis of the *B–F*-theory, we express the action for a four-dimensional gravity theory with a Λ-term and a massless Dirac field in the form of an integral of a 4-form:

$$S = \int \varepsilon_{abcd}\left\{-\frac{1}{4l_P^2}\right.$$

$$\times\left(R^{ab}\wedge\omega^c\wedge\omega^d + \frac{1}{3}\Lambda\omega^a\wedge\omega^b\wedge\omega^c\wedge\omega^d\right)$$

$$\left. + \frac{1}{6}\Theta^a\wedge\omega^b\wedge\omega^c\wedge\omega^d\right\}, \qquad (2)$$

$$d\omega^{ab} + \omega_c^a \wedge \omega^{cb} = \frac{1}{2}R^{ab},$$

$$\Theta^a = \frac{i}{2}(\bar{\psi}\gamma^a \mathcal{D}_\mu \psi - \mathcal{D}_\mu \bar{\psi}\gamma^a \psi)dx^\mu.$$

Here,

$$\omega^{ab} = \omega_\mu^{ab} dx^\mu$$

is a connection 1-form in a certain orthonormal basis $\omega^a = e_\mu^a dx^\mu$, $g_{\mu\nu} = e_{a\mu}e_\nu^a$ is a metric tensor, l_P is the Planck length, ψ is a Dirac field, γ^a are the Dirac matrices, and $\mathcal{D}_\mu\psi$ is the covariant derivative of the Dirac field ψ.

Now, let us consider the theory of gravity without Λ-term and without matter in a three-dimensional space. The action in this theory is given by the simple formula

$$S = -\frac{1}{l_P}\int B^a \wedge F_a. \quad (3)$$

Here, the following standard notation is introduced:

$$B^a = e_\mu^a dx^\mu, \quad F_a = \frac{1}{2}\varepsilon_{abc}R^{bc}.$$

The equations of motion derived from this action are given by

$$\nabla_\mu e_\nu^a - \nabla_\nu e_\mu^a = 0, \quad (4)$$

$$F_a = 0. \quad (5)$$

Thus, the curvature in the *B*–*F*-theory (3) is equal to zero, and this theory is topological. The *B*–*F*-theory can easily be formulated in the space of arbitrary dimension. For example, in four dimensions, the action of this theory is similar to the action (3):

$$S = -\frac{1}{l_P^2}\int B^{ab} \wedge F_{ab}, \quad (6)$$

where

$$B^{ab} = -B^{ba}, \quad F_{ab} = \frac{1}{2}\varepsilon_{abcd}R^{cd}$$

are 2-forms. Obviously, in the case of theory (6), just as in the case of the *B*–*F*-theory, the curvature tensor is equal to zero in any dimension of the space; i.e., the *B*–*F*-theory is always topological. However, while the *B*–*F*-theory really describes gravity in a three-dimensional space, this is not so in higher dimensional spaces. For example, in a four-dimensional space, theory (6) will describe gravity if we set

$$B^{ab} = \frac{1}{2}\omega^a \wedge \omega^b. \quad (7)$$

Of course, the curvature in this case is nonzero.

The circumstances described above underlie another approach to the construction of a discrete quantum theory of gravity. In this approach, at the first stage, one considers a discrete analogue of the *B*–*F*-theory in any space of dimension *D*. Since the field *B* (see (3) and (6)) is a $(D-2)$-form, each $(D-2)$-simplex in the discrete variant of the theory is assigned an independent element of the field *B*. The integration over the field *B* in the continuum theory yields the following expression for the transition amplitude:

$$Z \sim \int D\omega_\mu^{ab} \prod_x \delta(F(x)). \quad (8)$$

In the discrete variant, one directly analyzes the generalization of expression (8) for the transition amplitude. To simplify this generalization, one constructs a dual lattice of the original simplicial complex. The vertices v_i of the dual lattice are situated in the middle of *D*-dimensional simplices; i.e., they represent certain interior points of the *D*-dimensional simplices. Denote by e_s the line segments connecting adjacent vertices of the dual lattice, where *s* enumerates these segments. Thus, each dual edge e_s intersects once a certain $(D-1)$-dimensional simplex that is common for two *D*-simplices whose centers are connected by a dual edge e_s. Let us fix a $(D-2)$-simplex and denote by *f* its dual two-dimensional polygon that is bounded by dual edges $e_s f$, intersects a given $(D-2)$-simplex at one point, and does not intersect any other $(D-2)$-simplex. Thus, there is one-to-one correspondence between the $(D-2)$-simplices of the original simplicial complex and dual two-dimensional polygons. Note that, in the notation $e_s f$, the index *f* corresponds to the membership of a dual edge in the dual polygon *f*, where $s = 1, 2, \ldots, N$ is the number of this edge (the number *N* can be arbitrarily large). Let us assign an orientation to each dual polygon. This means that the direction of traversing the polygon boundary is specified. Thus, the orientation of a dual polygon obviously determines the orientation of each dual edge belonging to this polygon. Since one edge is shared by many polygons, its orientation may be changed depending on its membership in one or another polygon.

Consider a polygon *f*. Let us enumerate the dual edges $e_1 f, e_2 f, \ldots, e_N f$ as they appear when traversing the boundary of polygon *f* in the positive (counterclockwise) direction, which simultaneously defines the orientation of edges. Let us assign to each dual edge $e_s f$ an element $g_{e_s f}$ of the holonomy group *G* of connection $\omega_{b\mu}^a$. For example, in the four-dimensional case, the group *G* is either the Lorentz group SO(3, 1) or the orthogonal group SO(4) (in the case of Euclidean quantization). The change of the orientation of a dual edge

results in the change of element $g_{e_s f}$ to its inverse. In particular, if $e_s f$ and $e_{s'} f'$ represent the same dual edge belonging to two adjacent polygons f and f' and the orientations of $e_s f$ and $e_{s'} f'$ are opposite, then

$$g_{e_s f} = g^{-1}_{e_{s'} f'}.$$

The following gauge group acts in the space of elements $g_{e_s f}$:

$$g_{e_s f} \longrightarrow h_{v_+} g_{e_s f} h^{-1}_{v_-}. \tag{9}$$

Here, h_{v_i} is an arbitrary element of the holonomy group G that is assigned independently to each vertex v_i, h_{v_+} and h_{v_-} are the corresponding elements at the beginning and the end of the oriented edge $e_s f$.

By analogy with (8), in a discrete analogue of the *B*–*F*-theory, for any dual polygon f, we obtain the following conditions after the integration over the degrees of freedom contained in the field B:

$$g_{e_1 f} g_{e_2 f} \ldots g_{e_N f} = 1. \tag{10}$$

Indeed, the last equalities are equivalent in the continuum theory to the equality of the integrals of curvature 2-forms over arbitrary two-dimensional surfaces, which means that the curvature tensor vanishes.

Equalities (10) are invariant with respect to the action of the gauge group (9). In accordance with (8) and (10), the transition amplitude depending on a concrete simplex $\Re$ is defined by

$$Z(\Re) = \int \prod_{e \in \mathscr{E}} dg_e \prod_{f \in \mathscr{F}} \delta(g_{e_1 f} g_{e_2 f} \ldots g_{e_N f}), \tag{11}$$

where $\mathscr{E}$ and $\mathscr{F}$ are the sets of indices enumerating the dual edges and the polygons, respectively; here, the integration is performed by the Haar measure on the group G. The Haar measure is considered to be normalized in the case of compact groups. The delta function in (11) is a delta function on the group G and is defined as follows:

$$\delta(g) = 0 \text{ for } g \neq 1, \quad \int dg \delta(g) = 1.$$

We can see that, in the present approach to the discrete *B*–*F*-theory, the field B corresponding to the $(D-2)$-form B in the continuous theory does not appear in the explicit form. In the case of a topological theory, this fact does not cause any difficulties; however, the transition amplitude in (11) cannot directly describe the transition amplitude in a discrete gravity of dimension higher than three. In the latter case, one should impose the constraints following from representation (7) (in the four-dimensional case) on the field B. It is clear that such constraints radically complicate the theory because there arise local degrees of freedom. The problem becomes even more complicated when matter is introduced into the theory. For example, in the case of action (2), the fermion part of this action is proportional to ω^a to the third power or to the field B^{ab} to the power one and a half. Hence, it is necessary to explicitly introduce the field ω^a into four-dimensional spaces. Therefore, the development of the quantum theory of gravity on the basis of the *B*–*F*-theory formalism seems to have no prospects.

The detailed description of the theory of quantum discrete gravity based on the *B*–*F* formalism can be found in [8–11].

In contrast to the multidimensional case, considerable computational progress has been made in two-dimensional discrete quantum gravity (see [12, 13]). We refer the reader to [14] for the relation between the three-dimensional quantum Yang–Mills theory on a lattice and three-dimensional gravity theory.

In the present paper, we propose a new version of discrete quantum gravity theory. This new theory differs both from the Regge theory and from the discrete variant of the *B*–*F*-theory. Just as in the *B*–*F*-theory, the connection in our theory is represented by the elements of a holonomy group. However, unlike the *B*–*F*-theory, all fundamental variables in the theory proposed here are defined directly on the elements of the simplicial complex. In particular, the holonomy group elements that play the role of connection 1-forms are defined on one-dimensional simplices. It is assumed that the holonomy group is a spinor group. In contrast to the *B*–*F*-theory, we explicitly introduce an analogue of a tetrad 1-form in our theory (see formulas (22)–(25) below). In the discrete variant, the elements of a tetrad 1-form are also defined on one-dimensional simplices and belong to the vector representation of the spinor group considered. The presence of a tetrad form in the theory allows us to introduce a Dirac field whose elements are transformed by a spinor representation and are defined at the vertices of a simplicial complex. Using these fields, one can easily construct a lattice action invariant with respect to gauge transformations, i.e., local orthogonal transformations of a tetrad and the corresponding transformations of other variables. In the naive continuum limit, this action is reduced to a simple action of the continuum theory of gravity.

In addition, in this paper we perform the quantization of discrete gravity. This means the determination of the fundamental statistical sum, which represents a functional gauge-invariant integral over the fields of matter, tetrad, and connection with the weight of an exponent of action. In this case, it turns out that the correct determination of the statistical sum requires that the gauge group should be compact, which is equivalent to a metric with Euclidean signature.

Although the theory considered is free of ultraviolet divergences, it contains infrared divergences. These divergences correspond to the increasing lengths of elementary one-dimensional simplices (edges) of a simplicial complex. Therefore, the infrared divergences

should be interpreted as the reflection of the birth of a macroscopic space–time, obeying the Einstein theory, with distributed matter from an original unstructured (from geometrical viewpoint) lattice space. The observation and interpretation of these infrared divergences is possibly the most interesting result of our study, whereas the existence of these divergences in the theory seems to make the latter very attractive.

In the last section of this paper, we consider the problem of quantum anomalies. We show that the theory considered is free of graviational, i.e., reparametrization, anomaly. We also discuss the problem of the doubling of fermion states on simplicial complexes.

2. DISCRETE GRAVITY

First, we present certain definitions and facts from the theory of simplicial complexes. This is necessary since they provide a basis for the definitions of the objects that are not available in the literature on combinatorial topology (see [3, 4]) but are used in our approach to discrete gravity.

Definition 1. A finite set $\mathfrak{K}$ of elements $a_0, a_1, \ldots, a_N$ is called a finite abstract simplicial complex with vertices $a_0, a_1, \ldots, a_N$ if the following conditions are satisfied:

1. Certain subsets of $\mathfrak{K}$ are marked and are referred to as abstract simplices of the complex $\mathfrak{K}$.

2. Among the marked subsets are those that contain a single element so that each vertex of the complex $\mathfrak{K}$ is its simplex.

3. If s is a certain simplex from $\mathfrak{K}$, then each of its parts, called a face of the simplex s, also is a simplex of $\mathfrak{K}$.

If an abstract simplex $s^r = (a_0, a_1, \ldots, a_r)$ has $(r + 1)$ vertices, then the number r is called its dimension. The maximum dimension of the simplices contained in the complex $\mathfrak{K}$ is called the dimension of the complex $\mathfrak{K}$. □

We will only focus on simplicial complexes (henceforth, we omit the adjective abstract) of finite dimension, or finite-dimensional complexes. An infinite complex $\mathfrak{K}$ differs from a finite one by the only fact that the number of its vertices is infinite. Let s be an arbitrary element of the complex $\mathfrak{K}$; the star of the simplex s, denoted by $\mathrm{st}_{\mathfrak{K}}(s)$, is a set of all simplices of the complex $\mathfrak{K}$ for which s is a face.

Definition 2. An infinite simplicial complex $\mathfrak{K}$ is called locally finite if the star $\mathrm{st}_{\mathfrak{K}}(s)$ of each simplex $s \in \mathfrak{K}$ consists of a finite number of simplices. □

Now, suppose that the vertices of a finite (infinite) complex $\mathfrak{K}$ are the points of a certain Euclidean (Hilbert) space and that the set of vertices of any simplex of dimension r does not belong to any plane of dimension $r - 1$. In this case, the complex $\mathfrak{K}$ is called a geometric simplicial complex. A geometric realization of the simplicial complex $\mathfrak{K}$ is the one-to-one mapping of the vertices of $\mathfrak{K}$ into the vertices of a geometric simplicial complex $\mathcal{K}$, such that each simplex is mapped into a simplex both under the direct and inverse mappings.

The following theorem is given without proof (see [3, 4]).

Theorem 1. A finite abstract D-dimensional complex $\mathfrak{K}$ can always be realized as a geometric complex $\mathcal{K}$ belonging to the Euclidean space $\mathbb{R}^{2D+1}$. An infinite abstract complex can be realized in a Hilbert space if and only if it is locally finite and consists of a countable set of simplices.

Any sequence $a, b, c, \ldots, f$ in terms of which one can represent a set of all vertices of a certain simplex is called the order of vertices of this simplex.

Definition 3. A simplex $(a_0, a_1, \ldots, a_r)$ has an orientation, or is oriented, if every order of its vertices is assigned a sign "+" or "–," so that orders differing by an odd permutation correspond to opposite signs. This fact can be expressed as the equality

$$s^r = \varepsilon(a_0, a_1, \ldots, a_r), \tag{12}$$

where ε stands for the sign ascribed to the sequence $a_0, a_1, \ldots, a_r$. If s^r is an oriented simplex, then $(-s^r)$ denotes the oppositely oriented simplex. A zero-dimensional simplex admits two opposite orientations: $+(a_0)$ and $-(a_0)$.

Let $(a_0, \ldots, a_{i-1}, a_{i+1}, \ldots, a_r)$ be the face of a simplex s^r obtained by eliminating one vertex a_i from the sequence $a_0, a_1, \ldots, a_r$. By definition, the orientation of this face, given by

$$B_i^{r-1} = (-1)^i \varepsilon(a_0, \ldots, a_{i-1}, a_{i+1}, \ldots, a_r), \tag{13}$$

is called an induced orientation of the simplex s^r. □

Formally, it is convenient to represent the orientation of simplex (12) as a product of its vertices:

$$s^r = \varepsilon a_0 a_1 \ldots a_r. \tag{14}$$

Here, the skew-commutativity rules are satisfied:

$$a_\alpha a_\beta = -a_\beta a_\alpha. \tag{15}$$

Although the definitions below involve known concepts, they are adjusted here to facilitate the presentation of the variant of discrete gravity considered in this paper.

In addition, we consider D-dimensional strongly connected simplicial complexes, i.e., those that satisfy the following condition: each D-simplex has at least one $(D-1)$-dimensional face that also belongs to a certain other D-dimensional simplex of the same complex.

Definition 4. Two oriented D-dimensional simplices s_1^D and s_2^D of a D-dimensional simplicial complex are called concordantly oriented if either the simplices s_1^D and s_2^D have no common $(D-1)$-dimensional faces or the orientation of their common $(D-1)$-dimen-

sional face B^{D-1} induced by the orientation of the simplex s_1^D is opposite to the orientation of the same face B^{D-1} induced by the orientation of the simplex s_2^D. □

Definition 5. A D-dimensional strongly connected simplicial complex $\mathfrak{K}$ is called orientable and oriented if there exists such an orientation for all its D-dimensional simplices that any pair of its D-dimensional simplices is concordantly oriented. The concordant orientation of D-dimensional simplices defines the orientation of the complex. □

Further, we will consider only oriented simplicial complexes.

Denote by $\{s_A^D\}$ a set of D-dimensional simplices of the complex and by $\{a_{A\alpha}\}$, $\alpha = 0, \ldots, D$, a set D of vertices of the simplex s_A^D. Suppose that the orientation of the simplex s_A^D is defined by the following order of its vertices:

$$s_A^D = a_{A\alpha_0}a_{A\alpha_1}\ldots a_{A\alpha_D}. \quad (16)$$

We also introduce the following notation for oriented 1-simplices having a common vertex $a_{A\alpha_0}$:

$$X^A_{\alpha_0\alpha_i} = a_{A\alpha_0}a_{A\alpha_i} = -X^A_{\alpha_i\alpha_0}, \quad i = 1, \ldots, D. \quad (17)$$

Definition 6. An oriented frame of a simplex s_A^D at a vertex $a_{A\alpha_0}$ is the set of oriented 1-simplices (17) expressed in a certain order such that an even permutation of these 1-simplices does not change the orientation while an odd permutation changes the orientation of the frame to the opposite. By definition, the frame

$$\mathcal{R}^{A\alpha_0} = (X^A_{\alpha_0\alpha_1}, X^A_{\alpha_0\alpha_2}, \ldots, X^A_{\alpha_0\alpha_D}) \quad (18)$$

is oriented positively. □

The correctness of these definitions, related to the orientations of complexes and frames, is verified by the fact that, in the case of a geometric realization of the complex and a subsequent smoothing of a piecewise flat surface, these definitions actually coincide with the definitions of an oriented smooth manifold and a mutual orientation of local coordinates.

Next, assume that $D = 4$; i.e., we will consider four-dimensional complexes.

Let γ^a, $a = 1, 2, 3, 4$, be four-dimensional Hermitian Dirac matrices such that

$$\gamma^a\gamma^b + \gamma^b\gamma^a = 2\delta^{ab}, \quad (\gamma^a)^\dagger = \gamma^a, \quad (19)$$

and the anti-Hermitian spin operators

$$\sigma^{ab} \equiv (1/4)[\gamma^a, \gamma^b]$$

satisfy ordinary commutation relations of the group SO(4). Denote by Spin(4) a spinor group in the spinor representation where each element can be expressed as

$$g = \exp\left(\frac{1}{2}\varepsilon_{ab}\sigma^{ab}\right), \quad g^\dagger = g^{-1}, \quad \varepsilon_{ab} = -\varepsilon_{ba}. \quad (20)$$

Let us assign to each one-dimensional oriented simplex $X^A_{\alpha_i\alpha_j} = a_{A\alpha_i}a_{A\alpha_j}$ $(\alpha_i \neq \alpha_j)$ an element of the group Spin(4), which we denote by $\Omega(X^A_{\alpha_i\alpha_j})$. By definition,

$$\Omega(X^A_{\alpha_i\alpha_j}) = \Omega^{-1}(X^A_{\alpha_j\alpha_i}) \subset \mathrm{Spin}(4). \quad (21)$$

Thus, the change of orientation of a one-dimensional simplex results in the inversion of the corresponding element Ω. We will refer to quantities (21) as elements of a connection holonomy group on a simplicial complex. We will also use the abbreviated notation

$$\Omega(X^A_{\alpha_i\alpha_j}) \equiv \Omega_{Aij}.$$

We stress that the value of the holonomy group element Ω_{Aij} is defined exclusively by the oriented one-dimensional simplex $a_{A\alpha_i}a_{A\alpha_j}$ and is independent of the frame that contains the one-dimensional simplex $X^A_{\alpha_i\alpha_j}$ as its element. This reservation is essential since every one-dimensional simplex is, in general, an element of several oriented frames.

Let V be a linear space with the basis γ^a:

$$v \in V \longleftrightarrow v = v^a\gamma^a.$$

On a simplicial complex, we define a 1-form e with the values in the space V by assigning the element

$$e(X^A_{\alpha_i\alpha_j}) \equiv e_{Aij} \in V \quad (22)$$

to the vertex $a_{A\alpha_i}$ and the oriented one-dimensional simplex $X^A_{\alpha_i\alpha_j}$. Just as in the case of the holonomy group elements, elements (22) depend only on the oriented one-dimensional simplex $X^A_{\alpha_i\alpha_j}$ but are independent of the frame that contains the simplex $X^A_{\alpha_i\alpha_j}$ as its element. Since e is a 1-form, there is the following relation between elements (22):

$$e_{Aij} = -\Omega_{Aij}e_{Aji}\Omega_{Aji}. \quad (23)$$

Here, the sign "–" in (23) is due to the fact that e_{aij} and e_{Aji} are the values of the 1-form on the vectors $X^A_{\alpha_i\alpha_j}$ and $X^A_{\alpha_j\alpha_i} = -X^A_{\alpha_i\alpha_j}$, respectively. The "facings" from the elements of a holonomy group on the right-hand side of Eq. (23) are necessary since the element e_{Aji} must be parallel-translated from the vertex $a_{A\alpha_j}$ to the vertex $a_{A\alpha_i}$ to compare this element with the element e_{Aij}.

Thus, according to the definition, the vector $v_{A\alpha_j} \in V$ at the vertex $a_{A\alpha_j}$ is obtained by the parallel translation of the vector $v_{A\alpha_i} \in V$ from the vertex $a_{A\alpha_i}$ along the simplex $X^A_{\alpha_i\alpha_j}$ if the following equality holds:

$$v_{A\alpha_j} = \Omega_{Aji} v_{A\alpha_i} \Omega_{Aij}.$$

It is natural to interpret the quantity

$$l^2_{Aij} \equiv \frac{1}{4}\mathrm{Tr}(e_{Aij})^2 \tag{24}$$

as the square of the length of the edge $X^A_{\alpha_i\alpha_j}$. Thus, the geometric properties of a simplicial complex prove to be completely defined. Therefore, simplicial complexes represent the most convenient, if not the only possible, lattice for constructing a discrete theory of gravity: the theory defined above represents a model of a piecewise flat Regge space. On the other hand, if we considered, for example, a cubic lattice, then it would be rather difficult to give a similar interpretation for quantity (24). Thus, the theory based on the cubic lattice would hardly pretend to play the role of geometrodynamics.

The introduced fields Ω and e describe the gravitational field. Now, let us introduce the matter field.

To each vertex $a_{A\alpha_i}$, we assign the Dirac spinors $\psi_{A\alpha_i} = \psi_{Ai}$ and $\bar{\psi}_{Ai}$ each of whose components assumes values in a complex Grassmann algebra. In the case of a metric with Euclidean signature, the spinors ψ_{Ai} and $\bar{\psi}_{Ai}$ are independent variables and are interchanged under the Hermitian conjugation. As in the case of gravitational fields, the spinors ψ_{Ai} and $\bar{\psi}_{Ai}$ are defined by the vertex itself rather than by the four-dimensional simplex to which this vertex belongs.

In the space of fields, there acts a gauge group according to the following rule. To each vertex $a_{A\alpha_i}$, let us assign an element of the group $S_{A\alpha_i} = S_{Ai} \in \mathrm{Spin}(4)$. According to the principle of gauge invariance, the fields Ω, e, ψ, and the transformed fields

$$\begin{aligned} \tilde{\Omega}_{Aij} &= S_{Ai}\Omega_{Aij}S^{-1}_{Aj}, \\ \tilde{e}_{Aij} &= S_{Ai}e_{Aij}S^{-1}_{Ai}, \\ \tilde{\psi}_{Ai} = S_{Ai}\psi_{Ai}, &\quad \tilde{\bar{\psi}}_{Ai} = \bar{\psi}_{Ai}S^{-1}_{Ai} \end{aligned} \tag{25}$$

are physically equivalent.

Our next goal is to construct a gauge-invariant action that reduces to the classical action in the limit of slowly varying fields.

Define a 1-form Θ on a complex with values in the space V, that is bilinear with respect to the fields $\bar{\psi}$ and ψ. To each vertex $a_{A\alpha_i}$ and the oriented one-dimensional simplex $X^A_{\alpha_i\alpha_j}$, we assign the element

$$\begin{aligned} \Theta(X^A_{\alpha_i\alpha_j}) &= \frac{i}{2}\gamma^a(\bar{\psi}_{Ai}\gamma^a\Omega_{Aij}\psi_{Aj} - \bar{\psi}_{Aj}\Omega_{Aji}\gamma^a\psi_{Ai}) \\ &\equiv \Theta_{Aij} \in V. \end{aligned} \tag{26}$$

The quantity Θ_{Aij} represents an Hermitian operator. One can easily verify that 1-form (26), just as the 1-form e, satisfies relation (23). This fact is established by the repeated application of the formula

$$S^{-1}\gamma^a S = S^a_b\gamma^b, \tag{27}$$

where

$$\begin{aligned} &S \equiv \exp\left(\frac{1}{2}\varepsilon_{ab}\sigma^{ab}\right), \quad \varepsilon_{ab} = -\varepsilon_{ba} = \varepsilon^a_b, \\ &S^a_b \equiv (\exp\varepsilon)^a_b = \delta^a_b + \varepsilon^a_b + \frac{1}{2}\varepsilon^a_c\varepsilon^c_b + \ldots. \end{aligned} \tag{28}$$

Under the gauge transformations (25), the 1-form Θ is transformed in the same way as the form e:

$$\tilde{\Theta}_{Aij} = S_{Ai}\Theta_{Aij}S^{-1}_{Ai}. \tag{29}$$

In the case of the Euclidean metric (19), the matrix γ^5 is defined by the formula

$$\gamma^5 = \gamma^1\gamma^2\gamma^3\gamma^4 = (\gamma^5)^\dagger.$$

The Euclidean action of the system of introduced fields can be expressed as follows:

$$\begin{aligned} I = \frac{1}{5}\times\frac{1}{24}\sum_{A,m}\sum_{\sigma(Am)}\varepsilon_{\sigma(Am)}\mathrm{Tr}\gamma^5\Bigg\{-\frac{1}{l_P^2}\Big[(\Omega_{Ami}\Omega_{Aij}\Omega_{Ajm}) \\ + \frac{1}{12}\Lambda e_{Ami}e_{Amj}\Big] + \frac{1}{24}\Theta_{Ami}e_{Amj}\Bigg\}e_{Amk}e_{Aml}. \end{aligned} \tag{30}$$

Here, $\sum_{A,m}$ denotes a summation over all vertices m of a four-dimensional simplex A and then over all simplices A. Since m runs over five values, we introduced the factor 1/5 for convenience. The expression $\sigma(Am)$ in (30) stands for a permutation in the frame

$$\mathcal{R}^{A\alpha_m} = (X^A_{\alpha_m\alpha_i}, X^A_{\alpha_m\alpha_j}, X^A_{\alpha_m\alpha_k}, X^A_{\alpha_m\alpha_l}), \tag{31}$$

so that $\varepsilon_{\sigma(Am)} = +1$ if the permutation yields a positively oriented frame and $\varepsilon_{\sigma(Am)} = -1$otherwise. The symbol $\sum_{\sigma(Am)}$ denotes the summation over all 24 such permutations.

Formulas (25) and (29) immediately imply the gauge invariance of the action (30). Note that l_P^2 in (30) is a dimensionless constant.

Now, let us show that, in the limit of slowly varying fields, the action (30) reduces to the continuum action of gravity, minimally connected with a Dirac field, in a four-dimensional Euclidean space.

Consider a certain subset of vertices from a simplicial complex and assign the coordinates (real numbers)

$$x^{\mu}_{A\alpha_i} \equiv x^{\mu}(a_{A\alpha_i}), \quad \mu = 1, 2, 3, 4 \tag{32}$$

to each vertex $a_{A\alpha_i}$ from this subset. We stress that these coordinates are defined only by their vertices rather than by the higher dimension simplices to whom these vertices belong; moreover, the correspondence between the vertices from the subset considered and the coordinates (32) is one-to-one.

Suppose that

$$\left|x^{\mu}_{A\alpha_i} - x^{\mu}_{A\alpha_j}\right| \ll 1. \tag{33}$$

Estimates (33) can easily be satisfied if the complex contains a very large (infinite) number of simplices and its geometric realization is an almost smooth four-dimensional surface.[2] Suppose also that the four 4-vectors

$$dx^{\mu}_{Aji} \equiv x^{\mu}_{A\alpha_i} - x^{\mu}_{A\alpha_j}, \quad \alpha_i \neq \alpha_j, \quad i = 1, 2, 3, 4, \tag{34}$$

are linearly independent and

$$\begin{vmatrix} dx^1_{Am1} & dx^2_{Am1} & \ldots & dx^4_{Am1} \\ \ldots & \ldots & \ldots & \ldots \\ dx^1_{Am4} & dx^2_{Am4} & \ldots & dx^4_{Am4} \end{vmatrix} > 0, \tag{35}$$

provided that the frame $(X^A_{\alpha_m\alpha_1}, \ldots, X^A_{\alpha_m\alpha_4})$ is positively oriented. Inequality (35) implies that positively oriented local coordinates are introduced on the almost flat surface considered. Here, the differentials of coordinates (34) correspond to one-dimensional simplices $a_{A\alpha_j}a_{A\alpha_i}$, so that, if the vertex $a_{A\alpha_j}$ has coordinates $x^{\mu}_{A\alpha_j}$, then the vertex $a_{A\alpha_i}$ has the coordinates $x^{\mu}_{A\alpha_j} + dx^{\mu}_{Aji}$.

In the continuum limit, the holonomy group elements (21) are close to the identity element. Let us represent these elements in the form convenient for passing to the continuum limit:

$$\Omega_{Aij} = \exp\omega_{Aij}, \quad \omega_{Aij} = \frac{1}{2}\omega^{ab}_{Aij}\sigma^{ab}. \tag{36}$$

[2] Here, by an almost smooth surface, we mean a piecewise smooth surface consisting of flat four-dimensional simplices, such that the angles between adjacent 4-simplices tend to zero and the dimensions of these simplices are commensurable.

Consider the following system of equations in $\omega_{Am\mu}$:

$$\omega_{Am\mu}dx^{\mu}_{Ami} = \omega_{Ami}. \tag{37}$$

In this system of linear equations, the indices A and m are fixed, the summation is carried out over the index μ, and index runs over all its values. Since the determinant (35) is positive, the quantities $\omega_{Am\mu}$ are defined uniquely. Suppose that a one-dimensional simplex $X^A_{\alpha_m\alpha_i}$ belongs to four-dimensional simplices with indices A_1, A_2, …, A_r. Introduce the quantity

$$\omega_{\mu}\left[\frac{1}{2}(x_{A\alpha_m} + x_{A\alpha_i})\right] \equiv \frac{1}{r}\{\omega_{A_1m\mu} + \ldots + \omega_{A_rm\mu}\}, \tag{38}$$

which is assumed to be related to the midpoint of the segment $[x^{\mu}_{A\alpha_m}, x^{\mu}_{A\alpha_i}]$. Recall that the coordinates $x^{\mu}_{A\alpha_i}$ just as the differentials (34), depend only on vertices but not on the higher dimensional simplices to which these vertices belong. According to the definition, we have the following chain of equalities:

$$\omega_{A_1mi} = \omega_{A_2mi} = \ldots = \omega_{A_rmi}. \tag{39}$$

It follows from (34) and (37)–(39) that

$$\omega_{\mu}\left(x_{A\alpha_m} + \frac{1}{2}dx_{Ami}\right)dx^{\mu}_{Ami} = \omega_{Ami}. \tag{40}$$

The value of the field ω_{μ} in (40) on each one-dimensional simplex is uniquely defined by this simplex.

Next, we assume that the fields ω_{μ} smoothly depend on the points belonging to the geometric realization of each four-dimensional simplex. In this case, the following formula is valid up to $O((dx)^2)$ inclusive:

$$\Omega_{Ami}\Omega_{Aij}\Omega_{Ajm} = \exp\left[\frac{1}{2}\Re_{Am\mu\nu}dx^{\mu}_{Ami}dx^{\nu}_{Amj}\right], \tag{41}$$

where

$$\Re_{Am\mu\nu} = \partial_{\mu}\omega_{Am\nu} - \partial_{\nu}\omega_{Am\mu} + [\omega_{Am\mu}, \omega_{Am\nu}]. \tag{42}$$

On the right-hand side of (41), as well as in equality (42), all fields are taken at the vertex $a_{A\alpha_m}$ of a four-dimensional simplex A as is indicated by the subscript in Am. When deriving formula (41), we used the Hausdorff formula.

In exact analogy with (37), let us write out the following relation for a tetrad field without explanations:

$$e_{Am\mu}dx^{\mu}_{Ami} = e_{Ami}. \tag{43}$$

Using (36) and (37), we can rewrite the 1-form (26) as

$$\Theta_{Aij} = \gamma^a\frac{i}{2}[\overline{\psi}_{Ai}\gamma^a\mathcal{D}_{\mu}\psi_{Ai} - \overline{\mathcal{D}_{\mu}\psi}_{Ai}\gamma^a\psi_{Ai}]dx^{\mu}_{Aij}, \tag{44}$$

to within $O(dx)$; here,

$$\mathcal{D}_\mu \psi_{Ai} = \partial_\mu \psi_{Ai} + \omega_{Ai\mu}\psi_{Ai}. \quad (45)$$

Before rewriting the action (30) in the continuum limit, we give the following obvious formula:

$$\sum_{\sigma(Am)} \varepsilon_{\sigma(Am)} dx^\mu_{Ami} dx^\nu_{Amj} dx^\lambda_{Amk} dx^\rho_{Aml} = 24\varepsilon^{\mu\nu\lambda\rho} v_{SA}. \quad (46)$$

Here, $\varepsilon^{\mu\nu\lambda\rho}$ is a completely antisymmetric symbol, which is equal to unity due to inequality (35) when $(\mu\nu\lambda\rho) = (1234)$, and v_{SA} is the volume of the geometric realization of simplex A in a four-dimensional Euclidean space when the Euclidean coordinates of the geometric realization of the simplex are equal to the corresponding coordinates of its vertices (32). The factor 24 in (46) is necessary since the volume v_{SA} of the four-dimensional simplex on the right-hand side is less than the volume of a four-dimensional parallelepiped constructed on the vectors $dx^\mu_{Ami}, \ldots, dx^\mu_{Aml}$ by a factor of 24.

Applying formulas (41)–(46) and changing the summation to integration, we obtain the following expression for the action (30) in the continuum limit:

$$I = \int \mathrm{Tr}\gamma^5 \times \left[-\frac{1}{4l_P^2}\left(\Re + \frac{1}{3}\Lambda e \wedge e\right) + \frac{1}{24}\Theta \wedge e\right] \wedge e \wedge e. \quad (47)$$

Here, the curvature 2-form (see (42)) and the 1-forms (see (43), (44)) are defined by

$$\Re \equiv \frac{1}{2}\sigma^{ab} R^{ab}_{\mu\nu} dx^\mu \wedge dx^\nu,$$
$$e = \gamma^a e^a_\mu dx^\mu, \quad (48)$$
$$\Theta = \gamma^a \frac{i}{2}[\bar\psi\gamma^a \mathcal{D}_\mu\psi - \overline{\mathcal{D}_\mu\psi}\gamma^a\psi]dx^\mu.$$

Thus, in the continuum limit, the action (30) proves to be equal to the action of gravity with a Λ-term and a metric with Euclidean signature that is minimally connected with a Dirac field.

3. QUANTIZATION OF DISCRETE GRAVITY

Let us determine the statistical sum Z for a discrete Euclidean gravity, which becomes the transition amplitude in discrete quantum gravity after passing to the Lorentzian signature. Let us enumerate the zero-dimensional (vertices) and one-dimensional (edges) simplices by indices $\mathcal{V}$ and $\mathcal{E}$, respectively, and denote by $\psi_\mathcal{V}$, $\Omega_\mathcal{E}$, etc. the corresponding variables. By definition,

$$Z = \mathrm{const}\left(\prod_\mathcal{E}\int d\Omega_\mathcal{E}\int de_\mathcal{E}\right) \times \left(\prod_\mathcal{V}\int d\bar\psi_\mathcal{V}\int d\psi_\mathcal{V}\right)\exp(-I). \quad (49)$$

Here, const is a certain normalizing factor, $d\Omega_\mathcal{E}$ is the Haar measure on the group Spin(4),

$$de_\mathcal{E} \equiv \prod_a d\omega^a_\mathcal{E}, \quad e_\mathcal{E} = \omega^a_\mathcal{E}\gamma^a, \quad (50)$$

and

$$d\bar\psi_\mathcal{V} d\psi_\mathcal{V} \equiv \prod_\nu d\bar\psi_{\mathcal{V}\nu} d\psi_{\mathcal{V}\nu}. \quad (51)$$

The index ν in (51) enumerates individual components of the spinors $\psi_\mathcal{V}$ and $\bar\psi_\mathcal{V}$, such that we have a product of the differentials of all independent generators of the Grassman algebra of Dirac spinors in (51). The action I in (49) is defined by formula (30).

Note that the measure (50) is actually calculated at one of the vertices of the edge $\mathcal{E}$ since the fields $e_\mathcal{E}$ are defined by vertices (see (22), (23)). However, in view of (23), it is of no importance to which vertex of the simplex $\mathcal{E}$ the element $e_\mathcal{E}$ is related since the measure (50) is invariant under transformation (23). Therefore, one can really assume that the measure (50) is related to the edge $\mathcal{E}$.

Obviously, all the measures used in the functional integral (49) are invariant under the gauge transformations (25). Since the action I in (49) also represents a gauge invariant, the statistical sum (49) is invariant under the action of the gauge group (25).

Let us show how the transition from the theory with the metric with Euclidean signature to the theory with Lorentzian signature occurs. For this purpose, we rewrite the continuum version of the action (47) as follows:

$$I = \int \varepsilon_{abcd}\left\{-\frac{1}{4l_P^2}\left[(d\omega^{ab} + \omega^a_c \wedge \omega^{cb}) + \frac{1}{3}\Lambda\omega^a \wedge \omega^b\right] + \frac{1}{6}\Theta^a \wedge \omega^b\right\} \wedge \omega^c \wedge \omega^d. \quad (52)$$

Note that the set of components $\omega^a_\mathcal{E}$ ($a = 1, 2, 3, 4$) of $e_\mathcal{E} = \omega^a_\mathcal{E}\gamma^a$ is transformed under the gauge transformations (25) as a vector representation of the group SO(4), which corresponds to a metric with Euclidean signature. To pass to Lorentzian signature, we have to change

$$\omega^4 = i\omega^0. \quad (53)$$

Indeed, in the case of a metric with Euclidean signature ($\omega^a = e^a_\mu dx^\mu$), we have the following expressions in the continuum limit:

$$g_{\mu\nu} = \sum_{\alpha=1}^{3} e^\alpha_\mu e^\alpha_\nu + e^4_\mu e^4_\nu,$$
$$e^a_\mu = (e^\alpha_\mu, e^4_\mu), \quad \alpha = 1, 2, 3. \tag{54}$$

The change (53) reduces the metric tensor (54) to the tensor

$$g_{\mu\nu} = -e^0_\mu e^0_\nu + \sum_{\alpha=1}^{3} e^\alpha_\mu e^\alpha_\nu, \tag{55}$$

which has a Lorentzian signature.

The components $\omega^a_{\mathscr{E}}$, where $a = 0, 1, 2, 3$, are transformed by the vector representation of the Lorentz group SO(3, 1). Now, we can see from the expression for action (52) that the following changes should be made together with (53):

$$\omega^{4a} = i\omega^{0\alpha}, \quad \Theta^4 = i\Theta^0, \quad \gamma^4 = i\gamma^0. \tag{56}$$

Otherwise, different terms in the action (52) would have different parities with respect to complex conjugation. Finally, the quantities $\Omega_{\mathscr{E}}$ become elements of the noncompact group Spin(3, 1).

It follows from formulas (2) and (52) that the changes (53) and (56) reduce the statistical sum (49) to the transition amplitude in the theory of gravity with the Lorentzian signature; here, the weights in the appropriate functional integrals are transformed by the rule

$$\exp(-I) \longrightarrow \exp(iS), \tag{57}$$

where S is the action of the system under consideration on a simplicial complex with the metric with Lorentzian signature and the noncompact gauge group Spin(3, 1).

Consider the statistical sum (49) with a zero Λ-term in the absence of fermions. In this case, the integral over the 1-form $e_{\mathscr{E}}$ becomes Gaussian:

$$Y\{\Omega\} = \int Dz \exp\left(\frac{1}{2} z_m \mathcal{M}_{mn} z_n\right). \tag{58}$$

Here, $\{z_m\}$, $m = 1, \ldots, Q$ denotes a set of real variables $\{\omega^a_{\mathscr{E}}\}$ and $\mathcal{M}_{mn}$ is a real symmetrical matrix depending on the elements of the holonomy group $\Omega_{\mathscr{E}}$. Thus,

$$\frac{1}{2} z_m \mathcal{M}_{mn} z_n \equiv \frac{1}{l_P^2}\frac{1}{5} \times \frac{1}{24} \sum_{A,m} \sum_{\sigma(Am)} \varepsilon_{\sigma(Am)}$$
$$\times \operatorname{Tr}(\gamma^5 \Omega_{Ami} \Omega_{Aij} \Omega_{Ajm} e_{Amk} e_{Aml}). \tag{59}$$

Denote by $\{\lambda_q\}$, where $q = 1, \ldots, Q$, a set of eigenvalues of the matrix $\mathcal{M}_{mn}$. Let $\varepsilon_q = \operatorname{sgn}\lambda_q$. Since, in general, there are both negative and positive eigenvalues among $\{\lambda_q\}$, the integral (58) should be redefined. This is done by passing to Lorentzian signature. Under this procedure, the eigenvalues are transformed by the rule

$$\lambda_q \longrightarrow e^{i\varphi}\lambda_q,$$

where $\varphi = 0$ in the Euclidean case and $\varphi = \pi/2$ in the case of the Minkowski signature. Thus, in the case of the Euclidean signature, the Gaussian integral

$$\mathcal{J}_E = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} dz \exp\left(\frac{1}{2}\lambda z^2\right) \tag{60}$$

reduces to the Fresnel integral in the case of the Minkowski signature:

$$\mathcal{J}_M = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} dz \exp\left(\frac{i}{2}\lambda z^2\right)$$
$$= \sqrt{\frac{i}{\lambda}} = (i)^{\varepsilon/2} \frac{1}{\sqrt{|\lambda|}}, \tag{61}$$

where $\varepsilon = \operatorname{sgn}\lambda$. Let us perform the analytic continuation

$$\lambda \longrightarrow e^{-i\varphi}\lambda$$

on the right-hand side of (61) and set $\varphi = \pi/2$. Thus, we recover the Euclidean signature of a metric and obtain the following value of integral (60):

$$\mathcal{J}_E = i^{(\varepsilon+1)/2} \frac{1}{\sqrt{|\lambda|}}. \tag{62}$$

Now, using (62), we redefine the integral (58) of interest:

$$Y\{\Omega\} = \text{const} \prod_q i^{(\varepsilon_q+1)/2} |\lambda_q|^{-1/2}. \tag{63}$$

If there are fermion fields in the theory, one should first calculate a functional integral over fermions. The subsequent integration over the 1-form e remains Gaussian and yields a contribution of the form (63) to the statistical sum. The remaining integral over the elements of the holonomy group Ω may prove to be divergent despite the compactness of this group. Indeed, certain eigenvalues λ_q may vanish under certain configurations of the field Ω. Since the expression under the integral sign depends on the negative powers of λ_q, the integral over the field Ω may prove to be divergent. From the physical point of view, these divergences are of great interest. Note that the tendency of eigenvalues λ_q to zero implies that the integral over the 1-form e^a is saturated when the absolute values of this field e^a (or its certain components) tend to infinity. On the other hand, as will be shown below, the fact that the field components have large values implies that the dynamics of these components is quasiclassical. Therefore, from the physical viewpoint, these divergences imply

that system (49) is essentially quasiclassical with action (52) or (2). Thus, we obtain a classical macroscopic space–time in which conditions for the existence of an observer may appear.

Concerning the problem under discussion, we note that the presence of Dirac fields in integral (49) only strengthens the divergence under the integration over the field e^a. Indeed, after the integration over the fermion field, the integral over the field e^a is rewritten as (cf. (60) and (61))

$$\mathcal{J} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} dz P_n(z)\exp\left(\frac{i}{2}\lambda z^2\right), \tag{64}$$

where $P_n(z)$ is a polynomial in z of degree n. For small λ, integral (64) is proportional to $|\lambda|^{-(n+1)/2}$.

A similar physical interpretation of divergences under the integration over the field e^a in the continuum quantum B–F-theory in a three-dimensional space–time was given by Witten in [15].

Let us notice another possible type of divergences in a discrete quantum gravity. If the statistical sum (49) was defined for a metric with Lorentzian signature, then the elements of the holonomy group would be the elements of the noncompact group Spin(3, 1). The gauge group (25) would also be noncompact, being a direct product $\mathcal{V}$ of the copies of the group Spin(3, 1). Since both the measure and action in the transition amplitude are gauge invariant, the functional integral in the transition amplitude would not be defined at all before the fixation (at least partial) of the gauge. However, the fixation of the gauge in the fundamental transition amplitude seems to be a so artificial procedure that the theory itself looses its sense. In our opinion, this means that the fundamental statistical sum for a discrete theory of gravity can be constructed only on the basis of a metric with Euclidean signature.

In their well-known paper [16], Hartle and Hawking made a hypothesis that the wave function of the universe must be calculated with the use of the functional integral on the basis of a metric with Euclidean signature. The reasoning was carried out by analogy with ordinary quantum theory with a positive definite Hamiltonian and Euclidean action. In the latter case, the transition to the metric with Euclidean signature makes it possible, in principle, to single out the fundamental state; this result was also assumed in the case of the gravity theory for which the Euclidean action is not positively definite. In our opinion, the arguments for a metric with Euclidean signature provided by the discrete theory of gravity are much more reliable than the arguments given in [16].

Since the gauge group is compact in the case of a metric with Euclidean signature, the eigenvalues of certain quantities prove to be discrete (quantized). In particular, so are elementary two-dimensional volumes and areas. Let us show this.

For fixed values of indices A, m, i, and j, we introduce the notation

$$\Omega_{Amij} = \Omega_{Ami}\Omega_{Aij}\Omega_{Ajm} = \exp\left(\frac{1}{2}\omega_{ab}\sigma^{ab}\right). \tag{65}$$

Element (65) of the group Spin(4) has the following representation in the group SO(4):

$$\Omega^{ab}_{Amij} = (\exp\omega)^{ab} = \delta^{ab} + \omega^{ab} + \frac{1}{2}\omega^{ac}\omega^{cb} + \ldots. \tag{66}$$

Next, we will consider (65) or (66) as a certain dynamical variable and all the other combinations of the variables Ω_{Bkl}, together with the variables e and ψ, as variables that are independent of variable (66) [although not necessarily commute with the selected variable (66)].

Suppose that the expression

$$s_{Amij} = a_{Am}a_{Ai}a_{Aj} \tag{67}$$

stands for the selected two-dimensional simplex and s_A, $s_{A'}$, … is a set of four-dimensional simplices belonging to $\mathrm{st}_{\mathcal{M}}(s_{Amij})$ (see Introduction).

Next, denote by

$$\begin{aligned}\mathcal{R}^{Am} &= (X^A_{mi}, X^A_{mj}, X^A_{mk}, X^A_{ml}),\\ \mathcal{R}^{A'm} &= (X^{A'}_{mi}, X^{A'}_{mj}, X^{A'}_{mk}, X^{A'}_{ml}),\end{aligned} \tag{68}$$

etc., positively oriented frames of the simplices s_A, $s_{A'}$, etc.; the first pairs of 1-simplices in the frames (68) are identical:

$$X^A_{mi} = X^{A'}_{mi} = \ldots, \quad X^A_{mj} = X^{A'}_{mj} = \ldots. \tag{69}$$

Therefore,

$$\Omega_{Amij} = \Omega_{A'mij} = \ldots. \tag{70}$$

Let us write out the contribution to the action (30), that is proportional to variable (70) and is defined on the simplex s_A. In the vector representation, this contribution is given by

$$\begin{aligned}\Delta I_A &= -\frac{1}{30l_P^2}\varepsilon_{abcd}\Omega^{ab}_{Amij}\sigma^{cd}_{Amkl},\\ \sigma^{cd}_{Amkl} &= \frac{1}{2}(e^c_{Amk}e^d_{Aml} - e^d_{Amk}e^c_{Aml}).\end{aligned} \tag{71}$$

Here, we used the following relations:

$$\Omega_{Amij} = \Omega^{-1}_{Amji},$$

$$\varepsilon_{abcd}\Omega^{ab}_{Amij} = -\varepsilon_{abcd}(\Omega^{-1}_{Amij})^{ab}.$$

It is natural to interpret the quantity σ^{cd}_{Amkl} in (71) as a projection, onto the plane cd, of the half area of the oriented parallelogram constructed on the 1-simplices (X^A_{mk}, X^A_{ml}) or as a projection of the area of the oriented 2-simplex $a_{Am}a_{Ak}a_{Al}$.

Introduce the dual of σ^{cd}_{Amkl}:

$$^*\sigma^{ab}_{Amkl} \equiv \frac{1}{2}\varepsilon_{abcd}\sigma^{cd}_{Amkl}. \tag{72}$$

Let us sum up the contributions of all the above 4-simplices s_A, $s_{A'}$, … to action (71) and denote the result by ΔI. We obtain

$$\Delta I = -\frac{1}{15l_P^2}\Omega^{ab}_{Amij}\,{}^*\sigma^{ab}_{mij},$$
$$^*\sigma^{ab}_{mij} = {}^*\sigma^{ab}_{Amkl} + {}^*\sigma^{ab}_{A'mkl} + \ldots. \tag{73}$$

Let l_{ab} be a set of left-invariant vector fields over the group SO(4) that are the generators of the Lie algebra of this group. It is convenient to interpret the quantities l_{ab} as first-order differential operators on the space of the group SO(4) that satisfy the commutation relations

$$[l_{ab}, l_{cd}] = \delta_{ad}l_{bc} - \delta_{bd}l_{ac} - \delta_{ac}l_{bd} + \delta_{bc}l_{ad}. \tag{74}$$

If Ω_{ab} is an element of the group SO(4) in the vector representation, then the commutation relations (74) are correlated with the following rule for the action of the operators l_{ab} on the element of the group SO(4):

$$\left(1 + \frac{1}{2}\varepsilon_{cd}l_{cd}\right)\Omega^{ab} = (\delta_{ac} + \varepsilon_{ac})\Omega^{cb}. \tag{75}$$

Here, the parameters ε_{cd} are assumed to be infinitesimal.

Now, assume that the operators l_{ab} act on the element Ω^{ab}_{Amij} in (73) so that formula (75) holds provided that the following change is made:

$$\Omega^{ab} \longrightarrow \Omega^{ab}_{Amij}.$$

Using Eq. (75), we obtain

$$l_{ab}(\Omega^{cd}_{Amij}\,{}^*\sigma^{cd}_{mij}) = {}^*\sigma^{ac}_{mij}(\Omega^{-1})^{cb}_{Amij} + (\Omega^{-1})^{ac}_{Amij}\,{}^*\sigma^{cb}_{mij}$$
$$\equiv 2\,{}^*\Sigma^{ab}_{Amij} \equiv \varepsilon_{abcd}\Sigma^{cd}_{Amij}. \tag{76}$$

The quantity Σ^{cd}_{Amij} in (76) is interpreted as a sum of projections of the areas of all oriented two-dimensional simplices $a_{Am}a_{Ak}a_{Aa}$, $a_{A'm}a_{A'k}a_{A'l}$, … dual to the simplex $a_{Am}a_{Ai}a_{Aj}$. In this case, the orientation of a simplex $a_{Am}a_{Ak}a_{Al}$ is assumed to be positive if the corresponding frame in (68), constructed on the vertices a_{Am}, a_{Ai}, a_{Aj}, a_{Ak}, and a_{Al}, is oriented positively.

Now, let us pass to the quantum-mechanical analysis. From the quantum-mechanical point of view, the functional integral (49) defines a transfer matrix (a transition amplitude in the case of Lorentzian signature) describing the evolution of quantum states. The quantum states $\Psi\{\Omega, \bar{\psi}\}$ are gauge invariant functionals of the elements of the holonomy group Ω and the Dirac field $\bar{\psi}$. Here, the field e (or, more precisely, its certain bilinear combinations) plays the role of the momentum variable of the field Ω, while the field ψ is the momentum variable for the field $\bar{\psi}$. The quantum states or the wave functions are defined over three-dimensional simplicial complexes. Suppose that the boundary of a four-dimensional simplicial complex $\mathfrak{K}$ consists of two nonconnected simplicial complexes $\partial_1\mathfrak{K}$ and $\partial_2\mathfrak{K}$ (the nonconnectedness of complexes means that they have no common simplex). Let Ψ_1 be the initial wave function defined over the complex $\partial_1\mathfrak{K}$ and Ψ_2 be the final wave function defined over the complex $\partial_2\mathfrak{K}$. Then, Ψ_2 is defined by the functional integral (49). In this case, the integration over the boundary $\partial_1\mathfrak{K}$ is performed with the weight $\Psi_1\{\Omega, \bar{\psi}\}$, the variables Ω and $\bar{\psi}$ are fixed on the boundary $\partial_2\mathfrak{K}$ (and only on this boundary), and the integration is performed with respect to all the other variables over the four-dimensional complex $\mathfrak{K}$. Such an integration yields the wave function Ψ_2, which we denote by Ψ.

It follows from (49), (73), and (76) that the quantity $^*\Sigma^{ab}_{Amij}$ corresponds to the operator $(15/2)l_P^2 l_{ab}$ in the quantum-mechanical approach. This means that the mean gauge-invariant quantity $(^*\Sigma^{ab})^2$, which is proportional to the square of the area of the two-dimensional simplex Σ^{cd}_{Amij} in (76), is determined by the formula

$$\left\langle (^*\Sigma^{ab}_{Amij})^2 \right\rangle = \left(\frac{15}{2}l_P^2\right)^2 \frac{\langle l_{ab}\Psi | l_{ab}\Psi\rangle}{\langle\Psi|\Psi\rangle}. \tag{77}$$

The symbol $\langle\ldots\rangle$ in (77) denotes the integration over the field Ω by the Haar measure. Since the operators l_{ab} in this measure are anti-Hermitian, formula (77) shows that the square of a surface element

$$\frac{1}{2}(^*\Sigma^{ab}_{Amij})^2 \tag{78}$$

corresponds to the operator

$$-\frac{1}{2}\left(\frac{15}{2}\right)^2 l_P^2 l_{ab}^2. \tag{79}$$

Now, from commutation relations (74), we obtain the following quantization rule for the area $\mathfrak{A}$ of the surface element:

$$\mathfrak{A} = \left(\frac{15}{2}\right)l_P^2\sqrt{2[j_1(j_1+1) + j_2(j_2+1)]},$$
$$j_1 = 0, 1, \ldots, \quad j_2 = 0, 1, \ldots. \tag{80}$$

The quantization rule (80) also remains valid for a metric with Lorentzian signature.

The question of quantizing surface elements in the lattice B–F-theory is well studied (see, for example, [8, 9] and references therein). Our result demonstrates that the application of the quantization rule for surface elements within the framework of the formalism considered is natural.

Finally, consider the following important fact. It is known from the theory of angular momenta that large values of the numbers j_1 and j_2 in (80) correspond to the quasiclassical limit. On the other hand, the limit as $j_1 \longrightarrow \infty$ and $j_2 \longrightarrow \infty$ implies that the elementary lengths tend to infinity (in the units of the Planck length l_P). This fact justifies the above statement that the divergences in the statistical sum (49) that may appear during integration over the field e as $|e| \longrightarrow \infty$ imply the generation of a macroscopic and quasiclassical Riemannian space from a completely quantum nonstructured original space. This fact gives rise to a continuum universe obeying the Einstein equation against the background of which quantum fluctuations exist.

4. DISCRETE GRAVITY AND QUANTUM ANOMALIES

We begin the study of quantum anomalies with a remark that there is no reparametrization anomaly in the approach to discrete quantum gravity considered here. Indeed, the local coordinates x^μ appear only as vertex markers (see (32)–(35)); they can be chosen arbitrarily (under the above nondegeneracy conditions) and are not involved in quantum calculations. Since, by definition, the version of discrete gravity considered here is regularized on small scales, the above statement implies that there are no quantum anomalies with respect to arbitrary transformations of local coordinates. As is known, this property of generally covariant field theories can be generally violated under quantization. For example, there is an anomaly (central charge) in the Virasoro quantum algebra, which generates generally covariant transformations on the space of two-dimensional gravity or on a world surface of a string. Another example of reparametrization anomaly under the quantization of four-dimensional gravity is given in [17]. In both examples, reparametrization anomalies arise during quantizing continuum theories. Here, it is important that there also exist approaches to the quantization of continuum generally covariant theories under which there are no reparametrization anomalies (see, for example, [1, 18–20]).

A more complicated problem in the lattice theory is the problem of axial anomaly and the problem of introducing a Weyl field. These problems are closely related to the so-called problem of doubling of fermion states on a lattice. It is well known that, on a periodic spatial lattice, when possible momenta are restricted to the Brillouin zone with the topology of the direct product D of copies of the circle S^1, the doubling of fermion states occurs [21–25].

Indeed, consider a regular cubic lattice embedded into $\mathbb{R}^4$ whose vertices have integer-valued coordinates. Each vertex has index $n = (n^a)$ consisting of four numbers $n^a \in \mathbb{Z}$, $a = 1, 2, 3, 4$. Let e^a be the generators of the lattice: $e^1 = (1, 0, 0, 0)$, $e^2 = (0, 1, 0, 0)$, etc. It is clear that, in the absence of gravitational and gauge fields, the fermion part of the action (30) has the form

$$I_\psi = \frac{i}{2}\sum_n \sum_{a=1}^4 \bar\psi_n \gamma^a (\psi_{n+e^a} - \psi_{n-e^a}). \tag{81}$$

The equation

$$\frac{i}{2}\sum_{a=1}^4 \gamma^a (\psi_{n+e^a} - \psi_{n-e^a}) = \epsilon \psi_n \tag{82}$$

determines the eigenmodes of action (81) that can easily be described due to the translation invariance of Eq. (82). Suppose that the momentum variables k^a fill the Brillouin zone B:

$$-\pi \le k^a \le \pi, \quad a = 1, 2, 3, 4. \tag{83}$$

Let us expand the fermion field into the Fourier integral

$$\psi_n = \int_B \frac{d^4k}{(2\pi)^4} e^{ikn} \tilde\psi(k), \quad kn = k^a n^a. \tag{84}$$

The spinors (84) are the eigenmodes of Eq. (82) if the spinors $\tilde\psi(k)$ are the eigenvectors of the matrix

$$\sum_{a=1}^4 \gamma^a \sin(k^a), \tag{85}$$

here, the eigenvalues of the modes are determined by the eigenvalues of matrix (85). Formula (85) implies that, if the momentum components k^a assume only one of the two values

$$k^a = 0, \pi, \tag{86}$$

independently of each other, then the eigenvalues of matrix (85) vanish. This means that the fermion field introduced contains several low-energy components. In particular, it was demonstrated that, if one introduces one right (left) Weyl field into the germ action of the theory, then there are pairs of Weyl fields in the low-energy limit that are combined into Dirac spinors.

In terms of the lattice variables ψ_n, the phenomenon described implies the following. Let us choose a certain subset of indices from the set of four indices a and denote these indices by α. The set chosen may either be empty or contain all four indices a. Let us divide all the

vertices $n = (n^a)$ of the lattice into even and odd ones depending on whether the number sum $\sum_a n^\alpha$ is even or odd. Let u be a certain nonzero spinor. Obviously, all zero modes of Eq. (82) are expressed as

$$\psi_n = (-1)^{\sum n^\alpha}_\alpha u. \tag{87}$$

The superposition of long-wave perturbations on the zero modes (87) gives rise to low-energy modes that survive in the limit as the lattice constant tends to zero. Therefore, the absence of axial anomaly in the gauge theory constructed by generalizing action (81) on a periodic lattice to the case of gauge interaction only implies that the anomalies in the divergence of axial current induced by various components of the spinor field cancel each other. On the other hand, any modification of the theory on a periodic lattice, such that only one component of the Dirac field remains in the long-wavelength limit, gives rise to the known axial anomaly.

An important question arises: Does the phenomenon of doubling of fermion states occur on irregular amorphous lattices and, in particular, in the theory of gravity on simplicial complexes? Our hypothesis is that the doubling phenomenon occurs on the simplicial complexes that form a periodic lattice under geometric realization. On the contrary, if the geometric realization of a simplicial complex results in an amorphous periodic lattice, then there is no doubling of fermion states.

Let us elucidate the situation by a simple example of a two-dimensional lattice. Consider the fermion part of the action of gravity on this lattice. In this case, the indices $a, b, \ldots$ assume two values 1 and 2. Formulas (19) remain valid, but the γ matrices have dimension 2×2. Then, the fermion part of the gravity action is expressed as

$$S_\psi = \frac{1}{6}\sum_{A,\, m}\sum_{\sigma(Am)} \varepsilon_{\sigma(Am)}\varepsilon_{ab}\Theta^a_{Ami}e^b_{Amj},$$
$$\Theta^a_{Aij} = \frac{i}{2}(\bar\psi_{Ai}\gamma^a\Omega_{Aij}\psi_{Ai} - \bar\psi_{Aj}\Omega^{-1}_{Aij}\gamma^a\psi_{Ai}). \tag{88}$$

Here, the notation is the same as before (see (18)–(31)). Recall that $\sigma(Am)$ denotes the permutation in the frame $(X^A_{\alpha_m\alpha_i}, X^A_{\alpha_m\alpha_j})$ and $\varepsilon_{\sigma(Am)} = \pm 1$ depending on if this permutation is even or odd.

Consider the case of a flat space, when

$$\Omega_{Aij} = 1, \quad (e^a_{ij} + e^a_{jk} + \ldots + e^a_{li}) = 0. \tag{89}$$

The geometrical sense of the second equality in (89) is as follows. The quantity e^a_{ij} should be considered as a vector in an orthonormal basis that starts at the vertex a_{α_i} and ends at the vertex a_{α_j}. Each subsequent vector in the sum (89) starts at the vertex at which the preceding vectors ends. Then, the second equality in (89) implies that we deal with a torsionless flat space, while the first equality implies that the curvature of this space is zero. Hence, we assume that a simplicial complex is realized in the space $\mathbb{R}^2$ and e^a_{ij} is a vector in $\mathbb{R}^2$ that connects appropriate vertices.

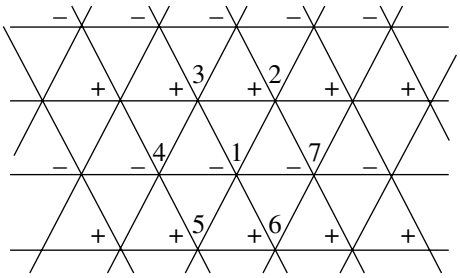


Fig. 1.

Now, consider a simplicial complex consisting of equilateral triangles (see Fig. 1). The vertices of interest for us are enumerated by numbers from 1 through 7.

The vectors connecting vertices 1 and 2, 7 and 3, etc. are denoted by e^a_{12}, e^a_{73}, etc., where $e^a_{12} = -e^a_{21}, \ldots$. Let us write out the equation for the zero modes of the action (88). For this purpose, we vary the action (88) with respect to $\bar\psi_1$ and equate the result to zero. Thus, we obtain the following equation at vertex 1:

$$\varepsilon_{ab}\gamma^a\{[(\psi_2 - \psi_5) + (\psi_3 - \psi_6)]e^b_{73} + [(\psi_4 - \psi_7) + (\psi_3 - \psi_6)]e^b_{35}\} = 0. \tag{90}$$

Equation (90), just as all the other equations of the zero mode, is satisfied if we set $\psi_{Ai} = u \neq 0$. It is also obvious that Eq. (90), together with all the other zero-mode equations, is satisfied if $\psi_{Ai} = \pm u \neq 0$, where the sign "+" or "–" corresponds to the arrangement of these signs at the vertices of the lattice in Fig. 1. In addition to this arrangement, there are other possible variants for arranging the signs "+" and "–" at the vertices of the lattice in Fig. 1 and assigning appropriate signs to the spinors ψ_{Ai} so that the mode becomes zero.

Thus, just as in the case of a square lattice, there exists a problem of doubling the fermion states on a regular simplicial complex.

Now, consider a simplicial complex that differs from that shown in Fig. 1 by one additional vertex (see Fig. 2). The zero-mode equations at vertices 1 and 8 in this case have the form

$$\varepsilon_{ab}\gamma^a[(\psi_2 - \psi_8)e^b_{83} + (\psi_4 - \psi_8)e^b_{35} + (\psi_5 - \psi_8)e^b_{46} + (\psi_7 - \psi_8)e^b_{68} + (\psi_3 - \psi_6)e^b_{24}] = 0,$$
$$\varepsilon_{ab}\gamma^a[(\psi_2 - \psi_1)e^b_{71} + (\psi_7 - \psi_1)e^b_{12}] = 0. \tag{91}$$

We can see that the addition of one vertex appreciably complicates the system of equations for the zero mode. Nevertheless, just as in Fig. 2, one can easily check that

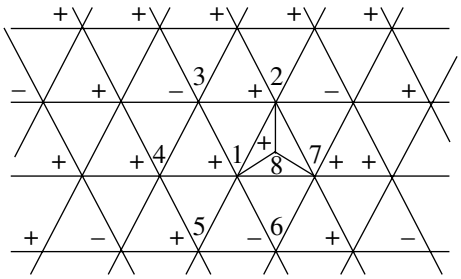


Fig. 2.

such an arrangement of signs yields a nontrivial zero mode.

Will the phenomenon of doubling of fermion states still be possible if a large number of additional vertices are chaotically introduced into the lattice of Fig. 1? The answer to this question seems to be negative. However, according to the analysis carried out, the absence of nontrivial zero modes may be characteristic only of simplicial complexes that are rather complicated from the viewpoint of periodicity.

In the general case of arbitrary dimension, the problem of doubling of fermion states should be formulated, in our opinion, as follows. Is it possible to find an (obviously nonperiodic) simplicial complex realized in a Cartesian space, such that the Dirac equation has the only zero mode? The existence of such a complex would imply the absence of the doubling of fermion states, as well as the absence of axial anomaly in the continuum limit of the theory.

Of course, in the continuum limit of this theory, the diagram technique (more generally, the theory of perturbations) would strongly differ (for example, by the traversal of poles) from the diagram technique used in quantum field theory. Possibly, the theory of perturbations in an anomaly-free continuum theory would be analogous to that developed in [1]. This problem requires separate analysis.

ACKNOWLEDGMENTS

I am grateful to V.E. Zakharov for his interest in the theory of discrete gravity that stimulated this paper. I am also grateful to S. Parkhomenko and S. Savchenko for fruitful discussions.

This work was supported by the Higher Scientific School Program (grant no. 00-1596579).

REFERENCES

1. S. N. Vergeles, Zh. Éksp. Teor. Fiz. **118**, 996 (2000) [JETP **91**, 859 (2000)].
2. T. Regge, Nuovo Cimento **19**, 558 (1961).
3. L. S. Pontryagin, *Foundations of Combinatorial Topology* (Nauka, Moscow, 1976).
4. P. J. Hilton and S. Wylie, *Homology Theory, an Introduction to Algebraic Topology* (Cambridge Univ. Press, Cambridge, 1960; Mir, Moscow, 1966).
5. R. Friedberg and T. D. Lee, Nucl. Phys. B **242**, 145 (1984).
6. J. Cheeger, W. Muller, and R. Schrader, Commun. Math. Phys. **92**, 405 (1984).
7. A. Bobenko and U. Pinkall, J. Reine Angew. Math. **475**, 187 (1996).
8. T. Regge and R. M. Williams, gr-qc/0012035.
9. J. C. Baez, gr-qc/9905087.
10. M. P. Reisenberger, gr-qc/9711052.
11. J. Iwasaki, gr-qc/9903112v2.
12. F. David, Nucl. Phys. B **257**, 543 (1985).
13. D. V. Boulatov, V. A. Kazakov, I. K. Kostov, and A. A. Migdal, Nucl. Phys. B **275**, 641 (1986).
14. D. I. Diakonov and V. Yu. Petrov, Zh. Éksp. Teor. Fiz. **118**, 1012 (2000) [JETP **91**, 873 (2000)].
15. E. Witten, Nucl. Phys. B **323**, 113 (1989).
16. J. B. Hartle and S. W. Hawking, Phys. Rev. D **28**, 2960 (1983).
17. A. Yu. Kamenshchik and S. L. Lyakhovich, hep-th/9608130.
18. M. Henneaux, Phys. Rev. Lett. **54**, 959 (1985).
19. E. Benedict, R. Jackiw, and H.-J. Lee, Phys. Rev. D **54**, 6213 (1996); D. Cangemi, R. Jackiw, and B. Zwiebach, Ann. Phys. (N.Y.) **245**, 408 (1996).
20. S. N. Vergeles, Zh. Éksp. Teor. Fiz. **117**, 5 (2000) [JETP **90**, 1 (2000)].
21. J. Kogut and L. Susskind, Phys. Rev. D **11**, 393 (1975).
22. L. Susskind, Phys. Rev. D **16**, 3031 (1977).
23. K. G. Wilson, Erice Lectures Notes (1975).
24. H. B. Nielsen and M. Ninomiya, Nucl. Phys. B **185**, 20 (1981); Nucl. Phys. B **193**, 173 (1981).
25. M. Luscher, hep-th/0102028.

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 93, No. 5, 2001, pp. 941–947.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 5, 2001, pp. 1086–1093.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Measurements of the Magnetic Moment of the Negative Muon in the Bound State in Various Atoms

T. N. Mamedov[a, *], D. Herlach[b], K. I. Gritsaj[a], O. Kormann[c], J. Major[c, d], A. V. Stoikov[a], and U. Zimmermann[b]

**e-mail: tmamedov@nu.jinr.ru*

[a]*Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia*

[b]*Paul Scherrer Institut, CH-5232 Villigen PSI, Switzerland*

[c]*Max-Planck-Institut für Metallforschung, D-70569 Stuttgart, Germany*

[d]*Universität Stuttgart, Institut für Theoretische und Angewandte Physik, D-70569 Stuttgart, Germany*

Received June 8, 2001

Abstract— The magnetic moment of the negative muon in the 1*s* state was measured in carbon, oxygen, magnesium, silicon, sulfur, and zinc. The attained precision of measurements allowed the dependence of the relativistic correction to the magnetic moment of the bound muon on the charge of the nucleus to be verified.

1. INTRODUCTION

Interest in measurements of the magnetic moment of the electron in the 1*s* state of various atoms has considerably increased in recent years because of the appearance of new experimental facilities. In 1928, Breit [1] showed that the magnetic moment of the electron in an atom in the 1*s* state should differ from the magnetic moment of the free electron because of its relativistic motion. This effect was considered in more detail in [2]. Measurements of the magnetic moment of the electron in an atom other than hydrogen in the 1*s* state, however, involve serious difficulties, because performing such experiments requires generating high-density flows of multiply ionized atoms (ions with one electron). Indeed, until recently, experimental *g* factors of the electron were only available for hydrogen in the 1*s* state [3, 4]. In 2000, the magnetic moment of the electron on the 1*s* level of an atom with a nucleus charge other than one was measured [5]. In [6], the *g* factors of the electron on the 1*s* level of atoms with nuclear charges up to $Z = 92$ were theoretically calculated.

In 1958, Hughes and Telegdi [7] noted that relativistic magnetic moment changes should also be characteristic of the negative muon in an atom. The relativistic correction to the magnetic moment of the negative muon in the 1*s* state can be measured for an arbitrary atom with a zero nuclear magnetic moment, which makes it possible to study its dependence on nuclear charge *Z* up to the lead atom.

The theoretical calculations performed in [6, 8] show that, apart from the relativistic correction discovered by Breit, there are additional radiative corrections to the magnetic moment of the electron in the 1*s* state. These corrections are caused by the occurrence of the electron in a strong Coulomb field of the nucleus. Accordingly, the *g* factor of the 1*s* electron in hydrogen and hydrogen-like ions can be written in the form

$$g_e^{1s} = 2(1 + a_e^{\text{free}} + a_e^{BS} + a_e^{\text{rel}}), \qquad (1)$$

where a_e^{free} is the radiative correction to the *g* factor of the free electron, $a_e^{BS} \equiv a_e^{BS}(\text{QED})$ is the additional radiation (quantum-electrodynamic) correction for a bound electron, and a_e^{rel} is the relativistic correction for the 1*s* electron.

The radiative correction to the magnetic moment of the free electron was measured with an accuracy close to that of theoretical calculations [9]; according to [10], $a_e^{\text{free}} = 0.001159652193(10)$.

The theoretical calculations [8] give the additional radiative correction for the 1*s* electron in the form

$$a_e^{BS}(QED) = \frac{(\alpha Z)^2}{4}\frac{\alpha}{\pi} + \ldots. \qquad (2)$$

The dependence of the relativistic correction on *Z* is described by the equation [1, 2, 6, 8]

$$a_e^{\text{rel}} = \frac{2}{3}(\sqrt{1-(\alpha Z)^2} - 1). \qquad (3)$$

Equation (3) shows that the relativistic correction becomes comparable with the radiative correction for

the free electron a_e^{free} at $Z \approx 6$ and exceeds the latter by approximately one order of magnitude at $Z = 25$.

Currently, the magnetic moment of the 1*s* electron has most accurately been measured for hydrogen [4]. The ratio between the magnetic moments (g factors) of the bound g_e^{1s} and free g_e^{free} electrons was found to be [4]

$$\frac{g_e^{1s}}{g_e^{\text{free}}} - 1 = -17.709(13) \times 10^{-6}.$$

This value closely agrees with the results of the calculations performed in [8], according to which

$$\frac{g_e^{1s}}{g_e^{\text{free}}} - 1 = -17.7051 \times 10^{-6}.$$

The error of measurements reported in [4] is, however, close to the expected a_e^{BS} value for hydrogen, which does not allow the a_e^{BS} value to be determined directly from this experiment.

Recently [5], the magnetic moment of the electron in the 1*s* state has been measured for the fivefold ionized carbon atom. For this purpose, a special unit was constructed which operated by the principle of the continuous Stern–Gerlach effect. C^{5+} ions were confined in a magnetic trap. A constant magnetic field in the middle part of the trap equaled 3.8 T. Parallel to the magnetic field, electric quadrupole and additional magnetic fields were applied. The latter quadratically varied along the axial axis. Transitions between states with electron spin projections ±1/2 were induced by a microwave field. The frequency of axial carbon ion motions in the field depended on the electron spin projection onto the direction of the magnetic field. Simultaneously changing the axial and cyclotron frequencies of ion motions allowed the magnetic moment of the electron in the C^{5+} ion to be determined,

$$g_e(C^{5+}) = 2.001042(2)$$

accordingly,

$$g_e(C^{5+}) - g_e^{\text{free}} = -0.001277(2),$$

$$g_e^{\text{free}} = 2.002319304386(20)$$

(see [10]). The result obtained in [5] was in close agreement with the relativistic correction to the magnetic moment of the electron in the 1*s* carbon state calculated by (3) (–0.001278), but the accuracy of measurements was insufficient for verifying the theoretical predictions concerning a_e^{BS}.

2. THE MAGNETIC MOMENT OF THE NEGATIVE MUON IN ATOMS

Corrections to the magnetic moment of the negative muon in the bound state in atoms with a zero moment of the electron shell were considered in [11, 12]. According to [11, 12], the g factor of the negative muon in the 1*s* state of an atom with a diamagnetic electron shell can be written as

$$g_\mu^{1s} = 2\left(1 + \sum_{i=1}^{7} a_\mu^{(i)}\right), \tag{4}$$

where g_μ^{1s} is the g factor of the muon in the 1*s* atom state and $a_\mu^{(1)}, \ldots, a_\mu^{(7)}$ are the corrections to the g factor: $a_\mu^{(1)}$ is the radiative correction for the free muon; $a_\mu^{(2)}$ is the radiative correction caused by the occurrence of the muon in the Coulomb field of the nucleus; $a_\mu^{(3)}$ is the relativistic correction; $a_\mu^{(4)}$ and $a_\mu^{(5)}$ are the corrections taking into account polarization of the nucleus and the electron shell, respectively; $a_\mu^{(6)}$ is the correction for diamagnetic screening of the external magnetic field by the electron shell; and $a_\mu^{(7)}$ is the correction for the center of mass. For the free muon,

$$g_\mu^{\text{free}} = 2(1 + a_\mu^{(1)}).$$

The $a_\mu^{(1)}$, $a_\mu^{(2)}$, and $a_\mu^{(3)}$ corrections are similar to the a_e^{free}, a_e^{BS}, and a_e^{rel} corrections for the 1*s* electron ($a_\mu^{(1)} \equiv a_\mu^{\text{free}}$, $a_\mu^{(2)} \equiv a_\mu^{BS}$, and $a_\mu^{(3)} \equiv a_\mu^{\text{rel}}$).

The radiative correction to the magnetic moment of the free muon is known to a high accuracy, $a_\mu^{(1)} = 0.0011659230(84)$ [10]. The radiative correction to the magnetic moment of a bound muon differs from $a_\mu^{(1)}$ by the $a_\mu^{(2)}$ value, which does not exceed 2% of the $a_\mu^{(3)}$ relativistic correction value even for large Z [12]. The correction for the center of mass is also much smaller than the relativistic correction and is related to the latter as $a_\mu^{(7)}/a_\mu^{(3)} \sim m_\mu/M$ [12], where m_μ and M are the masses of the muon and the nucleus, respectively.

The largest correction to the magnetic moment of a bound muon is caused by its relativistic motion in the Coulomb field of the nucleus [2],

$$a_\mu^{(3)} = -\frac{4}{3}\int F^2 dr, \tag{5}$$

where F is the small component of the radial wave function of the muon.

Calculations show [11, 12] that the relativistic correction to the magnetic moment of a bound muon is of

the order of 0.1, 1.1, and 3.2% for the oxygen, zinc, and lead atoms, respectively. It follows that the relativistic correction is comparable in magnitude with the radiative correction for oxygen and is larger than the latter approximately by one order of magnitude for zinc.

The magnetic moments of negative muons in the 1*s* state were measured experimentally for light (C, O, Mg, Si, and S) and heavy (Zn, Cd, and Pb) atoms in [13, 14] and [15], respectively. The accuracy of measurements performed in [13] for corrections to the *g* factor of negative muons in Mg, Si, and S was approximately 3%, which was close to the accuracy of theoretical calculations. Satisfactory agreement was attained in [13] between the experimental and calculated *g* factor values for the C, O, Mg, Si, and S atoms. The $(g_\mu^{\text{free}} - g_\mu^{1s})/g_\mu^{\text{free}}$ values obtained in [14] for negative muons in Mg, Si, and S were, however, smaller in magnitude than those reported in [13] by $(17 \pm 4) \times 10^{-4}$. According to [13],

$$\frac{g_\mu^{\text{free}} - g_\mu^{1s}}{g_\mu^{\text{free}}} = \begin{cases} (29.6 \pm 0.7) \times 10^{-4} & \text{for Mg} \\ (36.3 \pm 1.1) \times 10^{-4} & \text{for Si} \\ (48.2 \pm 1.6) \times 10^{-4} & \text{for S.} \end{cases}$$

In the case of heavy atoms, the accuracy of measurements [15] is about 50:

$$\frac{g_\mu^{\text{free}} - g_\mu^{1s}}{g_\mu^{\text{free}}} = \begin{cases} (120 \pm 62) \times 10^{-4} & \text{for Zn} \\ (201 \pm 140) \times 10^{-4} & \text{for Cd} \\ (468 \pm 220) \times 10^{-4} & \text{for Pb.} \end{cases}$$

It follows that the experimental data on heavy atoms do not contradict theoretical calculations but give no proof of changes in the magnetic moment of the Dirac particle caused by its relativistic motion in the Coulomb field of the atomic nucleus.

The purpose of this work was to check the existence of substantial discrepancies between theoretical calculations and the experimental data on the *g* factors of muons on the 1*s* level of the Mg, Si, and S atoms and to obtain statistically significant data on atoms with nuclear charges $Z \geq 30$. Our preliminary results for C, O(H_2O), Mg, and Si were reported in [16]. Currently, similar measurements are being performed by J. Brewer at TRIUMF (Canada).

The implantation of a negative muon into a medium causes its deceleration and capture by an atom of the medium. In a condensed medium, the muon reaches the 1*s* state in an atom in time shorter than 10^{-10} s. Because of its larger mass, the Bohr radius of the muon is approximately 200 times smaller that the radius of the *K*-electron orbit. The negative muon is an unstable particle and predominantly decays by the scheme

$$\mu^- \longrightarrow e^- + \nu_\mu + \bar{\nu}_e.$$

Because of parity violation in this process, the spatial distribution of decay electrons is asymmetric, and measurements of the magnetic moment of the muon are based on this circumstance. In a transverse magnetic field, the magnetic moment (and spin) of the muon precesses at the frequency

$$\omega = \frac{2\mu_\mu H}{\hbar} = \frac{g\mu_B^\mu H}{\hbar},$$

where μ_B^μ is the Bohr magneton for the muon. For polarized muons, the μSR-spectrum (distribution with respect to the time of muon capture in the sample) of decay electrons represents an exponent modulated by the cosine function with frequency ω. The modulation amplitude is proportional to the muon polarization in the 1*s* state. Measurements of the precession frequency of the muon allow its magnetic moment in the 1*s* atom state to be determined. The correction to the magnetic moment (*g* factor) of the bound negative muon can be found from the equation

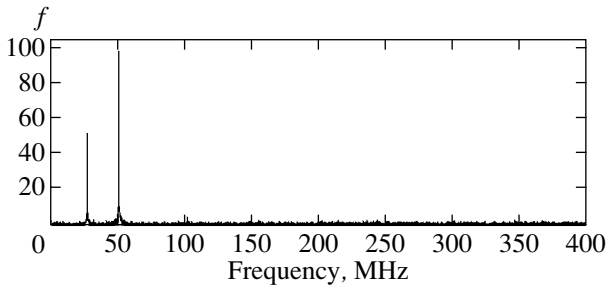
$$\frac{g_\mu^{\text{free}} - g_\mu^{1s}}{g_\mu^{\text{free}}} = \frac{\omega^{\text{free}} - \omega}{\omega^{\text{free}}}, \quad (6)$$

where ω^{free} and ω are the spin precession frequencies for the free muon and for μ^- in the 1*s* atom state, respectively.

3. MEASUREMENTS

Measurements were performed on a Stuttgart LFQ-spectrometer [17] using the μE4 muon channel of the proton accelerator at the Paul Scherrer Institute (PSI, Switzerland). The muon beam momentum was about 68 MeV/s. An 0.1–0.2 T external magnetic field transverse with respect to the muon spin was created by Helmholtz coils. Current stability in Helmholtz coils was controlled during measurements; it equaled $\Delta I/I \approx 2 \times 10^{-5}$. The mean diameter of Helmholtz coils was 510 mm, and the distance between coil centers was 240 mm. These dimensions were close to optimal for obtaining a magnetic field with a uniformity not worse than 10^{-5} in a $3 \times 3 \times 3$ cm^3 volume. The components of the magnetic field of the Earth and scattered fields of magnetic elements near the spectrometer were compensated by three pairs of additional coils to an accuracy not lower than 10^{-2} G. The residual magnetic field was measured by three permalloy sensors with mutually orthogonal orientations. The Helmholtz coils were positioned with respect to the beam (collimator) axis with the help of a laser.

The samples were prepared as cylinders 30 mm in diameter and 12, 18, 11, 10, 14, and 7 mm thick for carbon (reactor graphite), oxygen (water), magnesium, silicon, sulfur, and zinc, respectively. Water was packed into a cylindrical container made of a foam plastic with walls 2 mm thick. The weight of the container was 1.7 g.



The results of the Fourier analysis of the experimental data (μ*SR* histogram) for graphite measured during the standard exposure of about three hours to the beam of negative muons in a transverse magnetic field of 2000 G (*f* is the Fourier transform amplitude).

Crystalline silicon had specific resistance of about $10^4\ \Omega$ cm. The samples were mounted in such a way that sample axes coincided with the axis of the muon beam. The beam cross section diameter at the samples was about 16 mm. The position of the samples with respect to the beam axis was fixed accurate to 1 mm or less.

The intensity of muon stops in graphite (1 g/cm^2 thick) as a function of copper moderator thickness, that is, the curve of muon stops, was measured to determine the distribution of muons in the volume of the samples. A maximum of the curve of stops corresponded to an approximately 4 g/cm^2 moderator thickness, the total width at half-height equaled 0.8 g/cm^2, and the width at a 5% level of the maximum height was 1.4 g/cm^2. It follows that the volume of the region of muon stops in the samples did not exceed 6 cm^3.

Muon beams contained several percent electron impurities, which caused the appearance of background electrons in the experimental time distribution spectra (μSR-spectra) of electrons formed in the decay of muons stopped in the target. The time distribution of electrons in the beam and the background that they produced had a periodic structure with a frequency equal to that of the high-frequency accelerator field. The PSI accelerator field frequency was stabilized accurate to 10^{-8} and equaled 50.6330 MHz (e.g., see [18]). It follows that the experimental spectra contained a background with a periodic structure whose frequency was well known. This circumstance made it possible to control the time parameters of the unit as a whole including the characteristics of the time-to-code converter under operating conditions during the whole experiment (we used an EG&G' ORTEC Model 9308 time-to-code converter).

Figure shows the results of the Fourier analysis of the experimental data (μ*SR* histogram) for graphite obtained during the standard exposure of about three hours to a beam of negative muons. The Fourier spectrum contains the muon spin precession frequency in the external magnetic field and the F_{ac} frequency caused by the periodic background. Processing the μ*SR* histogram by the method of least squares shows that F_{ac} is determined accurate to 10^{-5} (0.5 kHz) and coincides with the accelerator frequency within the error of measurements. The F_{ac} values determined from the spectra recorded at various times during a session about 500 h long coincided with each other within the statistical error.

The data given above are evidence that the parameters of the μ*SR* unit that we used allowed the muon spin precession frequency to be measured accurately to 10^{-5} (0.5 kHz) at the required statistical level.

The free muon spin precession frequency was determined from the μ^+ precession frequency in copper as

$$\omega^{\text{free}} = \frac{\omega(\mu^+, \text{Cu})}{1 + K}. \qquad (7)$$

Here, K is the Knight shift for the positive muon in copper equal to $(60.0 \pm 2.5) \times 10^{-6}$ [19]. Accordingly, we first measured the positive muon spin precession frequency in copper and graphite. Next, the μ*E*4 muon channel was tuned to generating negative muon beams with the momentum equal to that of the positive beam, and μ^- spin precession frequencies were measured in C, O(H_2O), Mg, Si, S, and Zn. Measurements for O(H_2O), Mg, Si, S, and Zn alternated with measurements for graphite performed in the same magnetic fields as for the samples under study.

4. RESULTS AND DISCUSSION

The following spatial distribution asymmetry coefficients were obtained:

for positrons formed in the decay of μ^+ in

Cu—0.181 ± 0.001

C—0.218 ± 0.001;

for electrons formed in the decay of μ^- in

C—0.0486 ± 0.0003

O(H_2O)—0.0177 ± 0.0004

Mg—0.0324 ± 0.0004

Si—0.0304 ± 0.0004

S—0.0213 ± 0.0002

Zn—0.0107 ± 0.0005.

It was found from a comparison of the positive muon spin precession frequencies in graphite and copper that the paramagnetic shift for μ^+ in carbon was

$$\frac{(1 + K)\omega(\mu^+, \text{C}) - \omega(\mu^+, \text{Cu})}{\omega(\mu^+, \text{Cu})} = +(5.0 \pm 0.3) \times 10^{-4}.$$

The measured negative muon spin precession frequencies and the free muon spin precession frequencies in the corresponding fields (ω^{free}) are listed in Table 1

Table 1. Experimental free muon and μ^- spin precession frequencies in the samples in external magnetic field H

Sample	H, G	ω, rad/μs	ω^{free}, rad/μs	$\frac{\omega^{free}-\omega}{\omega^{free}}\times 10^4$
C	1000	85.048 ± 0.006	85.115 ± 0.002	7.9 ± 0.7
O(H_2O)	1000	127.455 ± 0.009	127.545 ± 0.011	7.0 ± 1.1
Mg	1000	127.264 ± 0.006	127.558 ± 0.010	23.1 ± 0.9
Si	1500	127.087 ± 0.009	127.545 ± 0.011	35.9 ± 1.1
S	1500	127.022 ± 0.025	127.563 ± 0.011	42.4 ± 2.1
Zn	2000	168.93 ± 0.38	170.243 ± 0.014	77 ± 22

Table 2. Corrections to the g factors of bound negative muons in carbon, oxygen (water), magnesium, silicon, sulfur, and zinc

Sample	$\frac{g_\mu^{free}-g_\mu^{1s}}{g_\mu^{free}}\times 10^4$ (this work)	$\frac{g_\mu^{free}-g_\mu^{1s}}{g_\mu^{free}}\times 10^4$ (data from [13, 15])	Theoretical values [12]	
			$\frac{g_\mu^{free}-g_\mu^{1s}}{g_\mu^{free}}\times 10^4$	$a_\mu^{(3)}\times 10^4$
C (graphite)	7.9 ± 0.7	7.6 ± 0.3	8.2 ± 0.1	6.29
		7.1 ± 0.6		
		8.0 ± 0.5		
O, in H_2O	7.0 ± 1.1	9.4 ± 1.0	14.3 ± 0.2	11.04 ± 0.01
Mg, metal	23.1 ± 0.9	26.4 ± 0.7	29.8 ± 0.6	23.79 ± 0.06
Mg, in MgH_2		29.6 ± 0.7		
Si, crystal	35.9 ± 1.1	36.3 ± 1.1	39.1 ± 1.0	31.70 ± 0.10
S, amorphous	42.4 ± 2.1	48.2 ± 1.6	49.1 ± 1.5	40.35 ± 0.15
Zn, metal	77 ± 22	130 ± 63	117.3	112.6 ± 1.0

for the samples that we studied. For carbon, ω^{free} was determined from the positive muon spin precession frequency in copper $\omega(\mu^+, Cu)$ measured at the same magnetic field as $\omega(\mu^-, C)$ for the negative muon in graphite (7). The same data were used to determine the ratio

$$R = \frac{\omega(\mu^+, Cu)}{\omega(\mu^-, C)},$$

which did not depend on the magnetic field value.

The ω^{free} values for O(H_2O), Mg, Si, S, and Zn were determined from the negative muon spin precession frequency in graphite measured in the corresponding magnetic field by the formula

$$\omega^{free} = R\frac{\omega(\mu^-, C)}{1+K}.$$

The corrections to the g factor of the negative muon in the 1s state of carbon, oxygen, magnesium, silicon, sulfur, and zinc atoms are compared in Table 2 with similar data obtained in [13, 15] and with the theoretical calculations [12]. The last column of Table 2 contains the relativistic corrections to the magnetic moment of the bound negative muon calculated in [12].

The corrections to the g factor (magnetic moment) of the negative muon in carbon, oxygen, magnesium, silicon, and sulfur in the 1s state determined in this work are close to the values reported in [13] and differ from the data obtained in [14], according to which the $(g_\mu^{free} - g_\mu^{1s})/g_\mu^{free}$ value is smaller by $(17 \pm 4) \times 10^{-4}$. The accuracy of our g_μ^{1s} measurements for light atoms (C, O, Mg, Si, and S) is close to the accuracy of measurements in [13] and approximately threefold higher for Mg and Si and 1.5 times higher for S than the accuracy of measurements in [14]. For Zn, the accuracy of measurements was improved three times compared with [15]. Within thrice the standard deviations (3σ), the experimental data on C, Si, S, and Zn obtained in this work agree with the results of theoretical calculations. With O(H_2O) and Mg, the discrepancies between the experimental and theoretical values are, however, of about 7σ. The reason for the discrepancies between the experimental data and the theoretical calculations [12] may be the neglect of possible Knight shifts and chemical shifts in the calculations performed in [12]. Recall that the capture of the negative muon in C, H_2O, Mg, Si, S, and Zn results in the formation in the media of sepa-

Table 3. Calculated a_e^{BS} and a_μ^{BS} corrections to the magnetic moments of electrons [8] and muons [12] and currently attainable accuracy (σ) of measuring corrections to the g factors of electrons and muons in the 1s state of several atoms

	e^-		μ^-	
	$a_e^{BS} \times 10^6$	$\sigma \times 10^6$	$a_\mu^{BS} \times 10^6$	$\sigma \times 10^6$
H	0.0102	0.013	–	–
C	0.4	1.0	8	30
Si	2.9	–	40	100
Zn	≈20	–	153	2000

rate atoms which are analogues of the B, N, Na, Al, P, and Cu atoms, respectively.

As water is a diamagnetic substance, the reason for the observed discrepancy between the experimental and calculated muon spin precession frequencies may be a chemical shift of the precession frequency. According to the data obtained in this work, the chemical shift of nitrogen in water equals $+(7.3 \pm 1.1) \times 10^{-4}$. This value does not contradict to the known NMR data on the chemical shifts of nitrogen in various compounds, which vary in a wide range from -400×10^{-6} to $+400 \times 10^{-6}$ (e.g., see [20]).

With Mg, we were unable to find data on the Knight shift in a Mg + Na alloy that might be directly taken into account in determining the frequency shift of muon spin precession. The NMR data on the $Mg_{17}Al_{12}$ alloy are, however, indicative of large Knight shifts for Mg and Al, of 1.3×10^{-3} for Mg and 1.7×10^{-3} for Al [21]. If the difference between the experimental and calculated g_μ^{1s} (Mg) values is caused by the Knight shift, then the Knight shift on Na in the Mg + Na alloy should be $(6.2 \pm 1.0) \times 10^{-4}$. This value is close in the order of magnitude to the Mg and Al Knight shift values measured for the $Mg_{17}Al_{12}$ alloy.

According to the NMR data on silicon with a 2.1×10^{19} cm^{-3} boron admixture, the Knight shift on boron amounts to $(0.65 \pm 0.05) \times 10^{-4}$ [22]. This value is in close agreement with the estimate based on the magnetic susceptibility of silicon containing boron in a 5.2×10^{19} cm^{-3} concentration [23]. In this work and [13], "pure" silicon samples with impurity concentrations not exceeding 10^{13} cm^{-3} were used. The concentration of free charge carriers in such samples is several orders of magnitude lower than in the samples used in the NMR [22] and magnetic susceptibility [23] measurements. The Knight shift in our samples should therefore be negligibly small (also see the estimates made in [13]).

The upper bound for the Knight shift contribution that should be taken into account in determining the g factor of the muon in Zn can be estimated from the NMR data on Cu_xZn_{1-x} alloys [24]. It follows from [24] that the Knight shift on Cu decreases approximately three times as x decreases from 1.0 to 0.25 and equals $(7 \pm 1) \times 10^{-4}$ at $x = 0.3$. This value is three times lower than the error of measurements performed in this work. It follows that the Knight shift can be ignored for Zn.

It would in our view be expedient to analyze the existing experimental data on the electron and muon magnetic moments in the 1s state of atoms with the purpose of estimating the a^{BS}(QED) quantum-electrodynamic correction to the electron and muon magnetic moments in the Coulomb field of the nucleus. The currently attainable accuracy of measurement and the expected values of this correction for several atoms are listed in Table 3. The calculated a^{BS}(QED) value for the muon is approximately one order of magnitude higher than that for the electron, although σ, that is, the accuracy of determining corrections to the g factor of the muon, is lower in comparison with measurements for the electron. We nevertheless believe that measurements for negative muons in atoms with nuclear charges $Z > 10$ offer no less promise than measurements for electrons. In addition, at such Z values, the muon in the 1s state occurs in a Coulomb field approximately two orders of magnitude higher compared with the electron. Accordingly, possible deviations of the a^{BS}(QED) values from theoretical predictions may be much larger for the muon than for the electron.

5. CONCLUSION

To summarize, the results of this work are in agreement with the experimental data obtained in [13]. These results show that the magnetic moment of the negative muon in the Coulomb field of the nucleus differs from the magnetic moment of the free muon. For carbon, silicon, sulfur, and zinc, our experimental results are in agreement with theoretical calculations within the error of measurements. This confirm the Z dependence of the relativistic correction to the magnetic moment of the negative muon in the 1s state of various atoms.

An analysis of the available experimental data shows that a further increase in the accuracy of measuring the g factor of the negative muon in silicon and zinc will allow a_μ^{BS} (QED) to be determined and, thereby, predictions of quantum electrodynamics for strong electric fields to be verified.

ACKNOWLEDGMENTS

The authors thank the Paul Scherrer Institute for the possibility of performing the measurements and I.A. Yutlandov for providing the magnesium sample.

REFERENCES

1. G. Breit, Nature **122**, 649 (1928).
2. H. Margenau, Phys. Rev. **57**, 383 (1940).

3. L. C. Balling and F. M. Pipkin, Phys. Rev. **139**, A19 (1965).
4. J. S. Tiedeman and H. G. Robinson, Phys. Rev. Lett. **39**, 602 (1977).
5. N. Hermanspahn, H. Häffner, H.-J. Kluge, *et al.*, Phys. Rev. Lett. **84**, 427 (2000).
6. H. Persson, S. Salomonson, P. Sunnergren, and I. Lindgren, Phys. Rev. A **56**, R2499 (1997).
7. V. W. Hughes and V. L. Telegdi, Bull. Am. Phys. Soc. **3**, 229 (1958).
8. H. Grotch and R. A. Hegstrom, Phys. Rev. A **4**, 59 (1971).
9. T. Kinoshita, Phys. Rev. Lett. **75**, 4728 (1995).
10. Particle Data Group, Eur. Phys. J. C **3**, 1 (1998).
11. K. W. Ford and J. G. Wills, Nucl. Phys. **35**, 295 (1962).
12. K. W. Ford, V. W. Hughes, and J. G. Wills, Phys. Rev. **129**, 194 (1963).
13. D. P. Hutchinson, J. Menes, G. Shapiro, and A. M. Patlach, Phys. Rev. **131**, 1362 (1963).
14. J. H. Brewer, Hyperfine Interact. **17–19**, 873 (1984).
15. T. Yamazaki, S. Nagamiya, O. Hashimoto, *et al.*, Phys. Lett. B **53**, 117 (1974).
16. T. N. Mamedov, V. N. Duginov, K. I. Gritsaj, *et al.*, Preprint No. E14-2000-158, JINR (Joint Inst. for Nuclear Research, Dubna, 2000).
17. R. Scheuermann, J. Schmidl, A. Seeger, *et al.*, Hyperfine Interact. **106**, 295 (1997).
18. E. Klempt, R. Schulze, H. Wolf, *et al.*, Phys. Rev. D **25**, 652 (1982).
19. A. Schenck, Helv. Phys. Acta **54**, 471 (1981).
20. *Nuclear Magnetic Resonance Spectroscopy of Nuclei Other than Protons*, Ed. by T. Axenrod and G. A. Webb (Wiley, New York, 1974).
21. T. J. Bastow and E. M. Smith, J. Phys.: Condens. Matter **7**, 4929 (1995).
22. R. K. Sunders and D. F. Holcomb, Phys. Rev. **136**, A810 (1964).
23. M. P. Sarachik, D. R. He, W. Li, and M. Levy, Phys. Rev. B **31**, 1469 (1985).
24. L. Bai-Qin and W. Ye-Ning, Phys. Rev. B **47**, 16 582 (1993).

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 93, No. 5, 2001, pp. 948–956.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 5, 2001, pp. 1094–1103.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Discrete Conversion of Gamma Rays in ^{229}Th and in Highly Ionized ^{125}Te^{Q} Ions

I. M. Band

Institute of Nuclear Physics, Russian Academy of Sciences, Gatchina, Leningradskaya oblast, 188300 Russia
e-mail: band@thd.pnpi.spb.ru

Received June 19, 2001

Abstract—The complex of problems associated with the discrete conversion of gamma rays is considered. The analogue of the internal conversion coefficient is determined for the case of discrete conversion, and specific computations are made. The calculated values of the half-life of the nuclear level $E_\gamma = 35.492$ keV in ^{125}Te^{Q} ions ($Q = 45$–48) are in good agreement with the measured quantities. Computations for the spectrum of optical photons accompanying the de-excitation of the nuclear level $E_\gamma = 3.5 \pm 1$ eV in ^{229}Th, carried out using the Dirac–Fock multiple configuration method, show that the highest-intensity spectral lines are situated in the region of 2.3–2.4 eV, which is in accord with the experimental data. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The discrete (subthreshold, resonance) internal conversion of gamma rays is currently attracting considerable attention. These are the cases in which the nuclear transition energy E_γ is slightly lower than the electron binding energy ε_b. For neutral atoms, this phenomenon has not virtually been observed in actual practice. Perhaps, the only exception to this is the de-excitation of the unique isomeric level $I = 3/2^+$, $E_\gamma = 3.5 \pm 1$ eV in ^{229}Th [1, 2].

Discrete conversion was confirmed in [3]. In these experiments, the half-life $T_{1/2}$ of the first excited nuclear level $I = 3/2^+$, $E_\gamma = 35.492$ keV was measured in highly ionized ^{125}Te^{Q}, $Q = 45$–48. The configuration of such ions has the form $1s^2 2s^2 2p^q$, $q = (3–0)$. The de-excitation of the level takes place through the $M1$ transition to the ground state $I_{gr} = 1/2^+$ of the nucleus.

The half-life of a nuclear level is connected with the internal conversion coefficient (ICC) through the relation

$$T_{1/2} = \frac{T_{1/2}^{\text{nucl}}}{1 + \alpha_{\text{tot}}}, \qquad (1)$$

where α_{tot} is the total ICC. For neutral tellurium, $\alpha_{\text{tot}} = 13.6$, $T_{1/2} = 1.49$ ns [4], whence it follows for that ^{125}Te^{Q} ions

$$T_{1/2}^{Q} = \frac{21.8}{1 + (\alpha_{\text{tot}})^{Q}} \text{ ns}. \qquad (2)$$

For Te^{Q} ions with $Q = 45$–48, the binding energy of K electrons is higher than the nuclear transitions energy and, hence, a "normal" conversion (for which the nuclear transition energy is higher than the binding energy of an atomic electron) at the K shell is impossible. For such ions, the value of ICC is determined by the conversion at the L subshells of the atom. Having calculated the ICC $\alpha_L(E_\gamma)$, we can use formula (2) to determine the half-life of a nuclear level under the assumption that only "normal" conversion takes place. Table 1 contains the half-lives $T_{1/2}^{Q}$ (calculated from formula (2) under the assumption that only "normal" α_L^{M1} conversion takes place) as well as the half-lives measured in [3]. It can be seen that the experimental data obtained in [3] can be interpreted only by assuming that de-excitation of the nuclear level in ^{125}Te^{Q} ions ($Q = 45$–48) is accompanied by a discrete conversion at the K-shell, which considerably increases α_{tot}^{M1} as compared to the "normal" ICC α_L^{M1}.

This is the case of de-excitation of the nuclear level $I = 3/2^+$, $E_\gamma = 35.492$ keV in ^{125}Te^{Q} which must be considered while comparing the results of the theory with the experimental data. It was assumed in [5, 6] that the experimental results obtained in these works serve as a proof of the discrete conversion in ^{229}Th. However, this provoked criticism from both experimenters [7, 8] and theoretical physicists [9, 10].

The phenomenon of discrete conversion has been studied by a number of authors [11–16]. The authors of [11–13] introduced the term discrete conversion factor (DCF), which is an analogue of the ICC in the case of discrete conversion:

$$\mathcal{R}_{i\to f}^{\tau L}(E_\gamma) = \frac{1}{2\pi}\frac{\Gamma_f}{(E_\gamma - \omega_f)^2 + (\Gamma_f/2)^2}\tilde{\alpha}_{i\to f}^{\tau L}. \qquad (3)$$

Here, E_γ is the nuclear transition energy and ω_f is the electron transition energy closest to E_γ. The authors of

Table 1. Half-lives $T_{1/2}^{Q}$, calculated under the assumption that only "normal" conversion can take place, as well as those measured in [3], in the case of de-excitation of the nuclear level $I = 3/2^+$, $E_\gamma = 35.492$ keV in $^{125}Te^{Q}$ ions

Q	$E_\gamma - \varepsilon_b$, eV	α_K^{M1}	α_L^{M1}	α_{tot}^{M1}	$T_{1/2}^{Q}$, ns	
					theory	experiment [3]
0	+3549	11.6	1.55	13.6		1.49 [4]
45	–206	–	1.63	1.63	8.25	<2
46	–521	–	1.65	1.65	8.20	<2
47	–870	–	1.60	1.60	8.33	6 ± 1
48	–1205	–	1.56	1.56	8.48	11 ± 2

[11–13] assume that the quantity $\tilde{\alpha}_{i\to f}^{\tau L}$ should be calculated using the same formulas that were used for calculating "normal" ICC's [17, 18]. Hence, they propose that the term discrete ICC should be preserved for the quantity $\tilde{\alpha}_{i\to f}^{\tau L}$. However, since the wave functions of bound and free electrons are normalized in different ways, the discrete ICC $\tilde{\alpha}_{i\to f}^{\tau L}$ has the dimensions of energy. It should be recalled that the wave functions of the discrete spectrum are normalized in such a way that

$$\int_0^\infty (g^2 + f^2) r^2 dr = 1, \qquad (4)$$

while the wave functions of free electrons are normalized to a unit energy interval in such a way that

$$r^2\left(g^2 + \frac{E+m}{E-m} f^2\right)_{r\to\infty} = \frac{E+m}{\pi(E^2 - m^2)^{1/2}}. \qquad (5)$$

In formulas (4) and (5), g and f are the large and small components of the relativistic radial wave function of an electron, m is the rest mass, and E is the total energy of an electron. Formula (3) was obtained in [11] from the condition that the atomic state f formed during conversion decays with a finite width Γ_f. In [11], the width Γ_f was defined as the total width of the excited atomic level, for which ω_f is closest the E_γ. In the present work, we assume that we can use for Γ_f the radiation width of the excited atomic level $\Gamma_f = \Gamma_f^{rad}$. In [13], the width Γ_f was defined for Te^{Q} ions as the radiation hole width determined by the transition $2p \longrightarrow 1s$ in an excited ion with configuration $1s2s^2p^q ns$. Note that the hole widths Γ_{ns}^{hole} for all levels of the excited Te^{Q} ion are practically constant and independent of ns.

It should be recalled that for $M1$ transitions at s shells of the atom, nearly 100% of the contribution to ICC comes from transitions to the states with the relativistic quantum number $\kappa = -1$ ($\kappa = (l - j)(2j + 1)$). Hence, while considering the discrete conversion during the nuclear transitions $M1$ at the s-shells of the atom, we confine the analysis to electronic transitions to the excited atomic levels containing the ns electrons.

Table 2 contains the half-lives $T_{1/2}^{Q}$ of the nuclear level $I = 3/2^+$, $E_\gamma = 35.492$ keV in $^{125}Te^{Q}$, which were calculated using formula (2) under the assumption that

$$\alpha_{tot}^{M1} = \mathcal{R}_{1s\to ns}^{M1} + \alpha_L^{M1}.$$

The DCF $\mathcal{R}_{1s\to ns}^{M1}$ was calculated from formula (3) in two cases: (a) using for Γ_f the radiation widths Γ_{ns}^{rad} calculated, as suggested in [11], for ns levels whose energy is closest to E_γ, and (b) using the widths Γ_{ns}^{hole} of the hole levels, as proposed in [13]. The computations presented in Table 2 were carried out using the Dirac–Fock

Table 2. Half-lives $T_{1/2}^{Q}$ of the nuclear level $I = 3/2^+$, $E_\gamma = 35.492$ keV in $^{125}Te^{Q}$ ions, calculated by using the DF method

Q	45	46	47
Configuration	$1s2s^22p^3ns$	$1s2s^22p^2ns$	$1s2s^22pns$
$T_{1/2}^{Q}$ (exp), ns	<2	<2	6 ± 1
ns	12s	8s	6s
$E_\gamma - \omega_{ns}$, eV	–1.4	–35	+44
Calculations based on the method proposed in [11]			
Γ_{ns}^{rad}, eV	1.1×10^{-3}	3.8×10^{-3}	7.5×10^{-3}
$T_{1/2}^{Q}$, ns	4.7	8.2	8.3
Calculations based on the method proposed in [13]			
Γ_{ns}^{hole}, eV	4.02	2.78	1.41
$T_{1/2}^{Q}$, ns	0.47	6.7	8.3

Note: n is the principal quantum number; s is the orbital quantum number of the electron; ω_{ns} the value of the energy of the excited level with configuration $1s2s^22p^q ns$, closest to E_γ; and $T_{1/2}^{Q}$ (exp) is the half-life measured in [3].

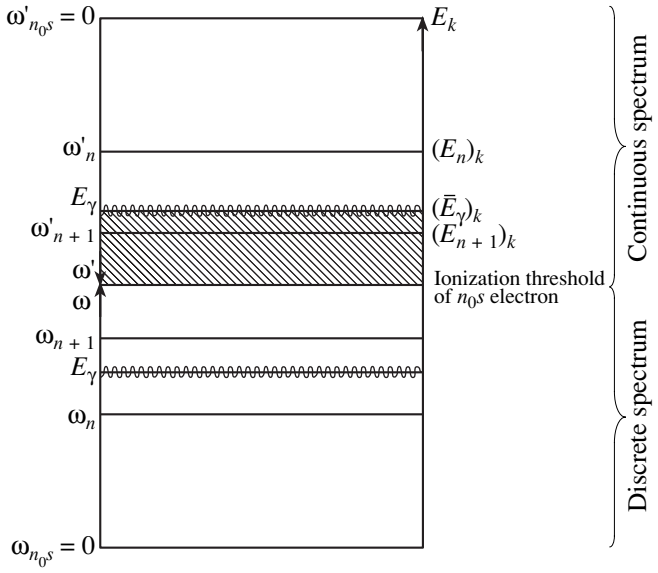


Fig. 1. Illustration to the model of DCF computation on the $n_0 s$ atomic shell: ω_n is the electron transition energy and E_γ is the nuclear transition energy. The kinetic energies of conversion electrons for "normal" conversion are indicated in the upper part of the plane on the E_k axis; $(E_n)_k$ is the kinetic energy of an electron for "normal" conversion; point $(E_n)_k = \varepsilon_b(ns)$ is the mirror reflection of point ω_n relative to the ionization threshold of an $n_0 s$ electron; $(\bar{E}_\gamma)_k$ is the mirror reflection of point E_γ, $(\bar{E}_\gamma)_k = \varepsilon_b(n_0 s) - E_\gamma$. The ω' axis is the mirror reflection of the ω axis. The hatched region is that where "normal" conversion is possible from the energy point of view.

(DF) method. It can be seen from the table that the results of computation of DCF using formula (3) do not agree with the experimental data [3] whether we take for Γ_f the radiative widths of the *ns* level or the hole widths.

Note that all the computations presented in this work were carried out with the help of methods and programs developed by us. This includes the computations of the electronic structures using the DF method [19, 20] and the multiple configuration Dirac-Fock (MCDF) method [15, 16, 21, 22], the computations of ICC and the electron transition probabilities using the DF method [23, 24] and the MCDF method [15,1 6]. Formulas for calculating ICC and the electron transition probabilities using the MCDF method are presented in [15, 16].

2. DISCRETE CONVERSION FACTORS NEAR THE RESONANCE POINTS

In this work, we propose the peculiarities of ICC [25, 26] be used to obtain a relation between the discrete and "normal" ICC near resonance points.

In the DF method, the radial components of the electronic wave functions for small r are proportional to the distance r from the center of the atom:

$$rg \approx \begin{cases} N_b(\sqrt{m^2 - E^2}\,r)^\gamma, \\ N_c(\sqrt{E^2 - m^2}\,r)^\gamma. \end{cases} \tag{6}$$

Here $\gamma = \sqrt{\kappa^2 - (\alpha Z)^2}$, α is the fine structure constant, N_c and N_b are the normalization factors for the wave functions of free (conduction) and bound electrons, respectively. For a free electron, we have

$$E = m + E_k, \tag{7}$$

while for a bound electron, we can write

$$E = m - \varepsilon_b, \tag{8}$$

where E_k is the kinetic energy of a free electron and ε_b is the binding energy of an electron in the atom. For low kinetic energies of the free electron and for low binding energies, the following relations hold:

$$\begin{aligned} m^2 - E^2 &\approx 2m\varepsilon_b, \\ E^2 - m^2 &\approx 2mE_k. \end{aligned} \tag{9}$$

If the kinetic energy E_k of a free electron is equal to the binding energy $\varepsilon_b(ns)$ of an atomic electron, the wave functions of the free and bound electrons with the same relativistic quantum number κ differ only in their normalization factors $N_b(\varepsilon_b(ns))$ and $N_c(E_k = \varepsilon_b(ns))$ for small r. The value of ICC is determined in the interior region of the atom [25, 26]. It was shown in [26] that the ICC formation region for $M1$ transitions at s shells of the atom is determined by the mean radius of the K shell. For low energies, the behavior of the wave functions within the ICC formation region is determined by formula (6).

We shall use the notation $E_k(ns)$ for $E_k = \varepsilon_b(ns)$. Let us compare the discrete ICC $\tilde{\alpha}^{M1}_{n_0 s \to ns}$ for the conversion transition $n_0 s$ of an electron to a higher *ns* orbit of the excited atom and the "normal" ICC $\alpha^{M1}_{n_0 s}(E_k(ns))$ in the case when the kinetic energy of a free electron is equal to the binding energy of an *ns* electron. The electron wave functions in the final state of the system for small r behave identically in the expressions for "normal" as well as discrete conversion, differing only in a constant factor equal to the ratio of the normalizing factors $N_c(E_k(ns))/N_b(\varepsilon_b(ns))$. Let us denote this constant (dimensional) factor by $2/\pi G_{ns}$. For close transition energy values, we obtain the following expression from the formulas [17, 18] defining the ICC:

$$\alpha^{M1}_{n_0 s}(E_k(ns)) \approx \frac{2}{\pi G_{ns}} \tilde{\alpha}^{M1}_{n_0 s \to ns}, \tag{10}$$

Table 3. Coefficients G_{ns} calculated by using formula (12), and the difference Δ_{ns} in total energies of the adjacent levels of the excited ions (atoms) (11)

Te^Q Transitions $1s \longrightarrow ns$						^{229}Th Transitions $7s \longrightarrow ns$		
$Q = 45$			$Q = 46$					
ns	G_{ns}, eV	Δ_{ns}, eV	ns	G_{ns}, eV	Δ_{ns}, eV	ns	G_{ns}, eV	Δ_{ns}, eV
11*s*	39	38	7*s*	141	139	8*s*	0.851	0.941
12*s*	30	30	8*s*	103	95	9*s*	0.347	0.381
13*s*	25	24	9*s*	73	68	10*s*	0.177	0.197
14*s*	19	19	10*s*	54	51	11*s*	0.103	0.112

where

$$\frac{2}{\pi G_{ns}} = \frac{N_c(E_k(ns))}{N_b(\varepsilon_b(ns))}.$$

It can be shown [27] that the following relation holds for the Coulomb field for electrons with quite large principal quantum numbers n:

$$\frac{N_c(E_k(ns))}{N_b(\varepsilon_b(ns))} \approx \frac{2}{\pi\Delta_{ns}},$$

where

$$\begin{aligned}\Delta_{ns} &= \varepsilon_b(ns) - \varepsilon_b((n+1)s) \\ &= E_{\text{tot}}((n+1)s) - E_{\text{tot}}(ns).\end{aligned} \tag{11}$$

It follows from relation (10) that

$$G_{ns} = \frac{2}{\pi}\frac{\tilde{\alpha}^{M1}_{n_0s\to ns}}{\alpha^{M1}_{n_0s}(E_k(ns))}. \tag{12}$$

Table 3 shows the coefficients G_{ns} and the differences Δ_{ns} between the energies of adjacent levels of the excited atoms, obtained as a result of calculations using formula (12). It can be seen from the table that for quite large values of n, the quantities G_{ns} are close to Δ_{ns}, the differences between the two being the smaller, the larger the value of n. As a result, formula (10) assumes the form

$$\alpha^{M1}_{n_0s}(E_k(ns)) \approx \frac{2}{\pi\Delta_{ns}}\tilde{\alpha}^{M1}_{n_0s\to ns}. \tag{13}$$

The situation described by formula (13) is illustrated in Fig. 1. The energies $E_k(ns)$ for which the "normal" ICC are calculated on the left-hand side of formula (13) are mirror images of the points $\varepsilon_b(ns)$ relative to the ionization threshold of the n_0s electron.

The following circumstance is worth noting. It follows from formula (3) that at the resonance points, the DCF is given by

$$\mathcal{R}^{M1}_{n_0s\to ns}(E_\gamma = \omega_{ns}) = \frac{2}{\pi\Gamma_{ns}}\tilde{\alpha}^{M1}_{n_0s\to ns}.$$

The analysis carried out by us shows that for quite large values of the principal quantum number n, the probability of discrete conversion near the resonance (see formula (13)) is defined mainly by the density $1/\Delta_{ns}$ of the levels in an atom, and not by the width of an individual level of an excited atom.

Table 2 shows the values of the radiative widths Γ^{rad}_{ns} and $\Gamma^{\text{hole}}_{ns}$ of $^{125}Te^Q$ ion for levels whose energies are closest to the nuclear transition energy $E_\gamma = 35.492$ keV. For ^{229}Th in the excited state $7s8s6d^2_{3/2}$, the width of the level with energy $\omega_{8s} = 3.45$ eV closest to the nuclear transition energy $E_\gamma = 3.5$ eV is $\Gamma^{\text{rad}}_{8s} = 2.08\times10^{-8}$ eV. Table 3 contains the energy differences Δ_{ns} between adjacent atomic levels. The data presented in Tables 2 and 3 can be used to determine the ratios Δ_{ns}/Γ_{ns} for ns levels with energies closest to the nuclear transition energies in $^{125}Te^Q$ and ^{229}Th, i.e., to estimate the amount by which the DCF at the resonance points, determined from formula (3), exceeds the quantity

$$\frac{2}{\pi\Delta_{ns}}\tilde{\alpha}^{M1}_{n_0s\to ns},$$

calculated from formula (13) at the resonance points. For $^{125}Te^Q$ ions with $Q = 45$–47, the ratio Δ_{ns}/Γ_{ns} lies within the limits

$$2.5\times10^4 \leqslant \frac{\Delta_{ns}}{\Gamma^{\text{rad}}_{ns}} \leqslant 3.2\times10^4,$$

$$7.5 \leqslant \frac{\Delta_{ns}}{\Gamma^{\text{hole}}_{ns}} \leqslant 1.7\times10^2,$$

for ^{229}Th, the ratio is

$$\Delta_{8s}/\Gamma^{\text{rad}}_{8s} \approx 4.5\times10^7.$$

The estimates presented here, as well as the results of computations presented in Table 2, show that the conversion width of the process differs considerably from the width of an individual level of an excited atom.

The ratio (13) can be used as a model for describing the discrete conversion process. In this model, the analogue (factor) of discrete conversion at the resonance point $E_\gamma = \omega_{ns}$ is defined by us as the "normal" ICC calculated for the kinetic energy of a free electron, which is equal to the binding energy $\varepsilon_b(ns)$ of an ns electron in the atom. All computations presented below are made using this model.

3. DISCRETE CONVERSION FACTORS AWAY FROM THE RESONANCE POINT

Let us now consider the definition and the properties of the δ-function. We consider the representation of the δ-function in the form

$$\delta(x, \Delta)_{\Delta \to 0} = \frac{1}{\pi}\frac{\Delta}{x^2 + (\Delta/2)^2}. \tag{14}$$

For the parameter Δ, we choose the variable quantity Δ_{ns} (11), such that $\Delta_{ns} \longrightarrow 0$ as $n \longrightarrow \infty$. We use the well-known property of the δ-function

$$f(\xi) = \int_0^\infty f(x)\delta(x-\xi)dx \tag{15}$$

and approximate integral (15) by the sum over known values of the function $f(x_n)$ at the points x_n:

$$f(\xi) \approx \sum_n \delta(x_n - \xi) f(x_n) dx_n. \tag{16}$$

Substituting into relation (16) expression (14) for the δ-function and the values of the model DCF (13), i.e., the "normal" ICC, $\alpha^{M1}_{n_0 s}(E_k(ns))$, we obtain the "total" DCF in the form

$$[\mathcal{F}^{M1}_{n_0 s}(E_\gamma)]^{\text{tot}} = \frac{1}{N}\sum_n \frac{\Delta^2_{ns}\alpha^{M1}_{n_0 s}(E_k(ns))}{(E_k(ns) - (\bar{E}_\gamma)_k)^2 + (\Delta_{ns}/2)^2}, \tag{17}$$

$$N = \sum_n \frac{\Delta^2_{ns}}{E^2_k(ns) + (\Delta_{ns}/2)^2}. \tag{18}$$

The parameter $(\bar{E}_\gamma)_k$ in (17) stands for the kinetic energy corresponding to the energy E_γ in the specularly reflected plane (see Fig. 1). We defined the "total" DCF keeping in mind the fact that the summation in (17) is carried out over all levels of the excited atom. From the energy point of view, "normal" conversion is possible for $E_k(ns) \le (\bar{E}_\gamma)_k$ or $\varepsilon_b(ns) \le \bar{E}_\gamma$. Otherwise, the ICC $\alpha^{M1}_{n_0 s}(E_k(ns)) = 0$. Hence, the DCF can be written in the form

$$\mathcal{F}^{M1}_{n_0 s}(E_\gamma) = \frac{1}{N} \times \sum_{\varepsilon_b(ns) \le \bar{E}_\gamma} \frac{\Delta^2_{ns}\alpha^{M1}_{n_0 s}(E_k(ns))}{(E_k(ns) - (\bar{E}_\gamma)_k)^2 + (\Delta_{ns}/2)^2}. \tag{19}$$

Finally, we transform formulas (17)–(19), presenting them in terms of ω_{ns}, E_γ, and $\varepsilon_b(ns)$. Note that

$$\begin{aligned} E_k(ns) &= \varepsilon_b(n_0 s) - \omega_{ns} = \varepsilon_b(ns), \\ \bar{E}_{\bar{\gamma}}(\bar{E}_\gamma)_k &= \varepsilon_b(n_0 s) - E_\gamma, \\ E_k(ns) - (\bar{E}_\gamma)_k &= E_\gamma - \omega_{ns}. \end{aligned} \tag{20}$$

It follows from Eqs. (20) that formulas (17)–(19) assume the form

$$[\mathcal{F}^{M1}_{n_0 s}(E_\gamma)]^{\text{tot}} = \frac{1}{N}\sum_{ns} \frac{\Delta^2_{ns}\alpha^{M1}_{n_0 s}(E_k(ns))}{(E_\gamma - \omega_{ns})^2 + (\Delta_{ns}/2)^2}, \tag{21}$$

$$N = \sum_{ns} \frac{\Delta^2_{ns}}{\varepsilon^2_b(ns) + (\Delta_{ns}/2)^2}, \tag{22}$$

$$\mathcal{F}^{M1}_{n_0 s}(E_\gamma) = \frac{1}{N}\sum_{\varepsilon_b(ns) \le \bar{E}_\gamma} \frac{\Delta^2_{ns}\alpha^{M1}_{n_0 s}(E_k(ns))}{(E_\gamma - \omega_{ns})^2 + (\Delta_{ns}/2)^2}. \tag{23}$$

Formula (23) allows the following interpretation of the discrete conversion process. The de-excitation of a nuclear level causes the excitation of the atom, which is accompanied by "normal" conversion at the excited atom levels for which the electron binding energy $\varepsilon_b(ns) \le \bar{E}_\gamma$.

Since formula (16) for $f(\zeta)$ approximates integral (15) with a certain error, the DCF $\mathcal{F}^{M1}_{n_0 s}(E_\gamma = \omega_{ns})$ calculated using formulas (22), (23) at the resonance points match only to a certain extent with the discrete conversion analogue $\alpha^{M1}_{n_0 s}(E_k(ns))$ defined by formula (13).

We compared the values of $\mathcal{F}^{M1}_{1s}(E_\gamma = \omega_{ns})$ (23) calculated for $^{125}\text{Te}^Q$ at the resonance point $E_\gamma = \omega_{ns}$ with the quantities $\alpha^{M1}_{1s}(E_k(ns))$. In other words, we calculated the error connected with the approximation of integral (15) by sum (16). We obtained the following result:

$$\frac{\left|\mathcal{F}^{M1}_{1s}(\omega_{ns}) - \alpha^{M1}_{1s}(E_k(ns))\right|}{\alpha^{M1}_{1s}(E_k(ns))} \le (0.16 \times 100)\%. \tag{24}$$

While indicating the computational error in the following, we will mean just this error of 16%.

Table 4. "Normal" ICC $\alpha_L^{M1}(E_\gamma)$, total DCF $[\mathcal{F}_{1s}^{M1}(E_\gamma)]^{\text{tot}}$ (21), DCF $\mathcal{F}_{1s}^{M1}(E_\gamma)$ (23), and half-lives $T_{1/2}^Q$ calculated by using formula (2) for $\alpha_{\text{tot}}^{M1} = \mathcal{F}_{1s}^{M1}(E_\gamma) + \alpha_L^{M1}(E_\gamma)$ for the isomeric nuclear level $I = 3/2^+$, $E_\gamma = 35.492$ keV in $^{125}\text{Te}^Q$

Q	$E_\gamma - \varepsilon_b(1s)$, eV	$\alpha_{1s}^{M1}(E_\gamma)$	$[\mathcal{F}_{1s}^{M1}(E_\gamma)]^{\text{tot}}$ (21)	$\mathcal{F}_{1s}^{M1}(E_\gamma)$ (23)	$T_{1/2}^Q$, ns	
					theory	experiment [3]
0	3494	11.7				1.49 [4]
24	2725	11.3				
34	1685	10.9				
40	94	10.4				
44	89	7.18		1.3 ± 0.20		
45	–125	–	13.8 ± 2.2	9.12 ± 1.50	1.84 ± 0.29	<2
46	–454	–	11.4 ± 1.8	6.41 ± 1.00	2.41 ± 0.39	<2
47	–797	–	11.6 ± 1.9	1.40 ± 0.20	5.44 ± 0.87	6 ± 1
48	–1201	–	8.91 ± 1.40	1.78 ± 0.30	5.12 ± 0.80	11 ± 2

Table 4 contains the DCF (23) and the half-lives $T_{1/2}^Q$ in $^{125}\text{Te}^Q$ ions for $Q = 45$–48, calculated from formula (2) for

$$\alpha_{\text{tot}}^{M1}(E_\gamma) = \mathcal{F}_{1s}^{M1}(E_\gamma) + \alpha_L^{M1}(E_\gamma).$$

The table shows that a good agreement is observed between the half-lives $T_{1/2}^Q$ calculated in the present work and those measured in [3]. It also contains the "normal" ICC calculated for ions with charge number $Q = 0$, 24, 34, 40, and 44 in $^{125}\text{Te}^Q$, as well as total DCF calculated using formula (21) for ions with $Q = 45$–48. Note that the values of "normal" ICC (in cases when "normal" conversion is possible) are close to those of DCF calculated in cases when "normal" conversion is forbidden due to energy considerations. This circumstance confirms the fact that the processes occurring in a nucleus do not depend significantly on the atomic electron shell surrounding the nucleus. The closeness of the calculated total DCF $[\mathcal{F}_{1s}^{M1}(E_\gamma)]^{\text{tot}}$ and the "normal" ICC $\alpha_{1s}^{M1}(E_\gamma)$ for all ions with $Q = 0$–48 serves as an additional argument in favor of the method proposed here for analyzing discrete conversion.

The discrepancy between the experimental and theoretical half-lives for $Q = 48$ is due to the fact that formula (16) is practically inapplicable in this case since the points n in the spectrum of the excited ion are separated from one another by large distances.

4. DISCRETE CONVERSION FOR DE-EXCITATION OF THE LEVEL $I = 3/2^+$, $E_\gamma = 3.5 \pm 1$ eV IN ^{229}Th

Let us consider the case of ^{229}Th. According to the authors of the experimental works [5, 6], the optical emission spectra measured by them are associated with the de-excitation of the isomer $I = 3/2^+$, $E_\gamma = 3.5 \pm 1$ eV in ^{229}Th [1, 2]. The de-excitation of this level occurs through the $M1$ transition to the ground state $I_{\text{gr}} = 5/2^+$ of the nucleus. It should be recalled that the configuration of the ground state of the thorium atom has the form $7s^2 6d_{3/2}^2$. In the course of $M1$ conversion transitions at the $7s$ shell of the atom (the maximum value of ICC is observed precisely at this shell), the main contribution to the ICC comes from transitions of a $7s$ electron to the excited states of the atom with configurations $7sns6d_{3/2}^2$. The spectrum of secondary photons is determined by the $E1$ transitions to the states of the atom with configurations $7sn'p6d_{3/2}^2$. The highest intensity spectral lines were observed in the energy interval 2.3–2.4 eV and around 3.5 eV in the experiments conducted in both [5] and [6].The authors of [5] attribute the peaks in the region of 3.5 eV to γ radiation, while the peaks around 2.4 eV were attributed to the radiation connected with the atomic excitation during the discrete $M1$ conversion. Experiments in [5] were carried out on metallic samples containing ^{233}U, while experiments in [6] were carried out on liquid samples of compounds containing ^{233}U. In both cases, measurements were not made in a vacuum. Analogous measurements in vacuum were made by the authors of [7, 8] and no radiation was detected. The authors of these works believe that the peaks in the region of 3.5 eV observed in [5, 6] are connected with the luminescence of air under the action of α-particles. However, they could not explain the emergence of peaks observed in the region 2.3–2.4 eV by the authors of [5, 6]. In his theoretical work, Tkalya [9] explains the situation observed in the experiments carried out in [7, 8] by considering a new channel of radiationless decay of this level in metallic thorium through conduction electrons.

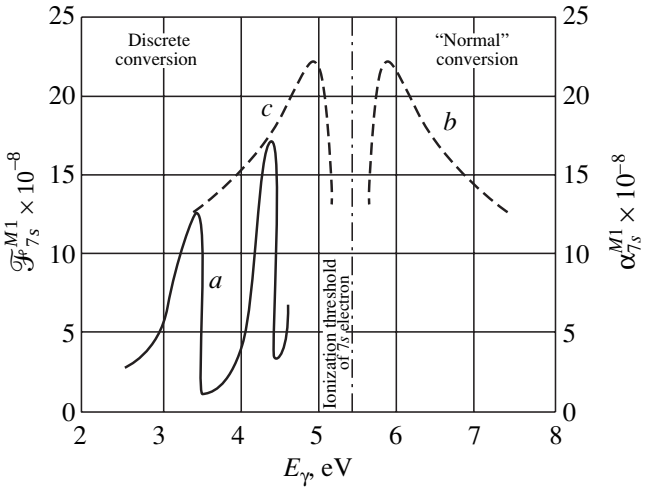


Fig. 2. Discrete conversion factors (DCF) and "normal" internal conversion coefficients (ICC) $\alpha_{7s}^{M1}(E_\gamma)$ in ^{229}Th. (a) $\mathcal{F}_{7s}^{M1}(E_\gamma)$, DCFs are calculated by formula (23) for $E_\gamma < \varepsilon_b(7s)$; (b) $\alpha_{7s}^{M1}(E_\gamma)$; "normal" ICC, $E_\gamma > \varepsilon_b(7s)$; (c) envelope of the curve $\mathcal{F}_{7s}^{M1}(E_\gamma)$.

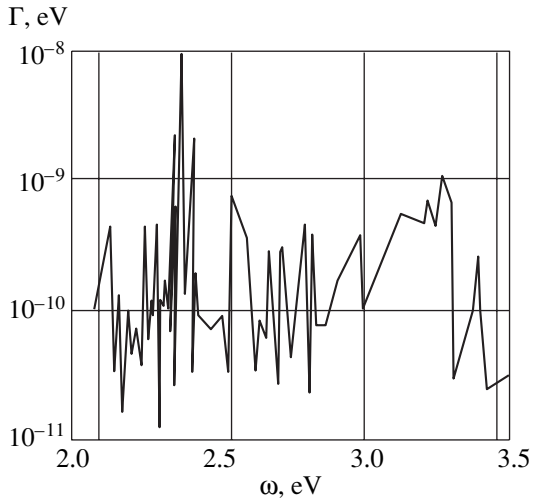


Fig. 3. Secondary optical photon spectrum associated with the decay of the isomeric level $I = 3/2^+$, $E_\gamma = 3.5 \pm 1$ eV in ^{229}Th; Γ(ω) are radiation widths of energy levels in the spectrum, which are calculated in the velocity calibration. The calculations are made in the framework of the MCDF method.

Let us consider the results obtained in [10], where the energies of spectral lines connected with $E1$ transitions from the excited levels of the type $7s8s6d_{3/2}^2$ in thorium to levels with configurations $7s7p_{1/2}6d_{3/2}^2$ and $7s7p_{3/2}6d_{3/2}^2$ were calculated. Calculations were made by using the DF method. The energies for these transitions were found to be 1.08 and 1.8 eV respectively. This leads to the conclusion that the peaks observed in the region of 2.3–2.4 eV in [5, 6] are not connected with conversion in ^{229}Th. We used the MCDF method to calculate the energy spectrum for the splitting of the $7s8s6d_{3/2}^2(J)$ configuration in moments J, and of the $7s7p6d_{3/2}^2(J')$ configuration in moments J'. It was found that the $7s8s6d_{3/2}^2(J)$ configuration is split into 11 levels with moments J = 0–3. Since the ground state $7s^26d_{3/2}^2$ of the thorium atom has a moment $J_{\text{ground}} = 2$, a $7s$ electron may pass over in the course of $M1$ conversion transitions to the orbits of an excited atom with configuration $7s8s6d_{3/2}^2(J)$ having moments J = 1–3. Transition to the orbit $7s8s6d_{3/2}^2$ with moment $J = 0$ is forbidden. However, our computations based on the MCDF method show that the values 1.08 and 1.8 eV of the transition energy correspond precisely to the $E1$ transitions from level $7s8s6d_{3/2}^2(J = 0)$ to levels $7s7p6d_{3/2}^2(J' = 1)$. Computations [15] based on the MCDF method for the $E1$ transitions from levels $7s8s6d_{3/2}^2(J$ = 1–3) to levels $7s7p6d_{3/2}^2(J')$ show that the highest intensity spectral lines correspond to the energies 2.3–2.4 eV. Thus, the conclusions [10] based on the computations made by using the averaged DF method [19, 20] were not confirmed in this case.

In the present work, we have used the model described above to calculate the DCF during de-excitation of the level $I = 3/2^+$, $E_\gamma = 3.5 \pm 1$ eV in gaseous ^{229}Th. The left-hand side of Fig. 2 shows the DCF calculated by formula (23) for energies E_γ varying between 2.5 and 4.5 eV. The curves $\mathcal{F}_{7s}^{M1}(E_\gamma)$ are of resonance type and the values of the factor vary in the interval $10^8 \lesssim \mathcal{F}_{7s}^{M1}(E_\gamma) \lesssim 2 \times 10^9$ for $2.5 \le E_\gamma \le 4.5$ eV. The dashed curve shows the envelope of the curve $\mathcal{F}_{7s}^{M1}(E_\gamma)$ passing through the points $E_\gamma = \omega_{ns}$, where ω_{ns} are the energies of the levels of the excited atoms with configurations $7sns6d_{3/2}^2$ ($n \ge 8$). The right-hand side of the figure shows the curve $\alpha_{7s}^{M1}(E_\gamma)$ calculated for energies E_γ higher than the ionization threshold for a $7s$ electron. It can be seen that the curve $\alpha_{7s}^{M1}(E_\gamma)$ is a mirror image of the envelope of the curve $\mathcal{F}_{7s}^{M1}(E_\gamma)$ relative to the ionization threshold of the $7s$ electron.

We calculated the optical spectra in the framework of the MCDF method [15]. In the MCDF computations of the excited states of an atom with configuration $7sns6d^2$, we considered the superposition of orbitals including the orbitals with n = 8–11, while for states of the type $7sn'pd^2$, we considered the superposition of orbitals including the orbitals with n' = 7–10. The opti-

Table 5. Errors $\beta_{ns} = \dfrac{2\tilde{\alpha}_{n_0s\to ns}(k_2)/\pi\Delta_{ns}}{\alpha^{M1}_{n_0s}(k_1)}$ associated with the use of approximation (13)

	$^{125}Te^{45}$ ($n_0s \equiv 1s$); k_1 and k_2 in keV				^{229}Th ($n_0s \equiv 7s$); k_1 and k_2 in eV			
ns	12s	13s	14s	15s	8s	9s	10s	11s
k_1	35.917	35.886	35.861	35.842	7.53	6.61	6.23	6.04
k_2	35.493	35.524	35.548	35.567	3.45	4.37	4.75	4.94
$\tilde{\alpha}^{M1}_{n_0s\to ns}(k_2)$, eV	3.99(2)	3.02(2)	2.40(2)	1.87(2)	1.73(10)	3.31(9)	1.27(9)	6.31(8)
Δ_{ns}, eV	30.6	24.2	19.5	16.0	0.914	0.381	0.196	0.115
$\dfrac{2\tilde{\alpha}^{M1}_{n_0s\to ns}(k_2)}{\pi\Delta_{ns}}$	8.30	8.10	7.83	7.44	1.20(10)	8.69(9)	4.12(9)	3.49(9)
$\alpha^{M1}_{n_0s}(k_1)$	8.30	8.07	7.83	7.60	1.24(9)	1.75(9)	2.01(9)	2.14(9)
β_{ns}	0%	0.4%	0%	2%	9.68	4.96	2.05	1.63

cal spectrum in the energy interval $2 \le \omega \le 4$ eV, including transitions from all atomic levels with an excitation energy $\omega_{ns} \lesssim 4.5$ eV, is presented in Fig. 3. The radiation widths of the $E1$ transitions were calculated in the velocity calibration [15]. It can be seen from the figure that the spectral line with the highest intensity lies around $\omega \approx 2.3$–2.4 eV. The radiation widths of the highest-intensity levels lie in the interval $10^{-9} < \Gamma(\omega) < 10^{-8}$ eV.

It cannot be affirmed that the agreement between our calculations of the optical photon spectra for $E1$ transitions of the type $7sns6d^2_{3/2}(J) \longrightarrow 7sn'p6d^2(J')$ ($J \neq 0$) and the experimental results [5, 6] is a convincing proof linking the observed spectrum with the discrete conversion in ^{229}Th. However, if there are no other factors responsible for the excitation of the thorium atom, our computations based on the MCDF method show that the spectrum of secondary photons for discrete conversion must be exactly of the same type as that observed in the experiments [5, 6].

5. COMPUTATIONAL ERRORS CONNECTED WITH THE MODEL REPRESENTATION OF DCF DEFINED BY FORMULA (13)

Let us estimate for the $^{125}Te^Q$ ions and for ^{229}Th the order of magnitude of the error associated with the model DCF representation based on formula (13). For a kinetic energy of a conversion electron equal to $E_k(ns)$, the nuclear transition energy $\alpha^{M1}_{n_0s}(E_k(ns))$ for "normal" conversion must be equal to $k_1 = \varepsilon_b(n_0s) + \varepsilon_b(ns)$. The discrete ICC $\tilde{\alpha}^{M1}_{n_0s\to ns}$ must be calculated for the nuclear transition energy $k_2 = \varepsilon_b(n_0s) - \varepsilon_b(ns)$. Let us compare the left- and right-hand sides of formula (13) by calculating $\alpha^{M1}_{n_0s}(E_k(ns))$ and $2\tilde{\alpha}^{M1}_{n_0s\to ns}/\pi\Delta_{ns}$ for transition energies k_1 and k_2 respectively. Table 5 contains the results of such calculations for $^{125}Te^Q$ and ^{229}Th and the quantities

$$\beta_{ns} = \frac{2\tilde{\alpha}^{M1}_{n_0s\to ns}(k_2)/\pi\Delta_{ns}}{\alpha^{M1}_{n_0s}(k_1)}$$

for conversion at the 1s shell in $^{125}Te^Q$ and at 7s shell in ^{229}Th. It can be seen from the table that the discrepancy in β_{ns} in the case of $^{125}Te^{45}$ is just a fraction of one percent. In this case, our calculations provide the true value of DCF with a fairly high degree of accuracy. For ^{229}Th, the quantities $\alpha^{M1}_{7s}(E_k(ns))$ and $2\tilde{\alpha}^{M1}_{7s\to ns}/\pi\Delta_{ns}$ may differ by an order of magnitude. Hence, our computations provide just an order-of-magnitude estimate for DCF: the value of DCF is found to lie in the range 10^8–10^{10} for $2.5 \le E_\gamma \le 4.5$ eV. Note that the DCF $\mathcal{R}^{M1}_{7s\to ns}(E_\gamma)$ (3), calculated in accordance with the technique proposed in [11], vary in the range $10 \lesssim \mathcal{R}^{M1}_{7s\to ns}(E_\gamma) \lesssim 10^{18}$ for $2.5 \le E_\gamma \le 4.5$ eV.

6. CONCLUSIONS

1. The model proposed by us for nuclear $M1$ transitions through conversion at n_0s atomic shells allows us to interpret the discrete conversion process as follows. The de-excitation of a nuclear level causes the excitation of the atom, accompanied by "normal" conversion at the excited atom levels for which the electron binding energy $\varepsilon_b(ns) \le \bar{E}_\gamma$. This model was used by us to derive formulas for calculating DCF and for making specific computations.

2. The computations of the half-lives $T_{1/2}^{Q}$ of the nuclear level $I = 3/2^+$, $E_\gamma = 35.492$ keV in $^{125}\mathrm{Te}^{Q}$ ions for $Q = 45$–48, carried out by using the technique proposed in this work, are in good agreement with the experimental data [3], thus justifying the use of this model for an analysis of the phenomena associated with a discrete conversion of gamma rays.

3. The DCF $\mathcal{F}_{7s}^{M1}(E_\gamma)$ calculated for gaseous ^{229}Th by using the technique proposed here during the deexcitation of the nuclear level $I = 3/2^+$, $E_\gamma = 3.5 \pm 1$ eV are of resonance type and lie in the interval $10^8 \lesssim \mathcal{F}_{7s}^{M1}(E_\gamma) \lesssim 1.7 \times 10^9$. The highest-intensity lines of the spectrum of $E1$ electronic transitions

$$7sns6d_{3/2}^2(J = 1\text{–}3) \longrightarrow 7sn'p6d_{3/2}^2(J'),$$
$$n \geq 8, \quad n' \geq 7,$$

calculated in the framework of the multiple configuration Dirac–Fock method in the velocity calibration lie in the interval 2.3–2.4 eV, which is in accord with the experimental data [5, 6]. The radiation width of the highest-intensity spectral lines lies in the interval $10^{-9} \lesssim \Gamma(\omega) \lesssim 10^{-8}$ eV.

4. It is shown that the probabilities of the $E1$ electron transitions in atoms are determined to a considerable extent by the moments of atomic energy levels between which a transition occurs.

ACKNOWLEDGMENTS

The authors are sincerely obliged to M.G. Kozlov, L.N. Labzovskii, and V.S. Polikanov for a fruitful discussion of the results and for valuable remarks which were taken into consideration while preparing this work for publication. This research was financed by the Russian Foundation for Basic Research (project no. 99-03-32640).

REFERENCES

1. C. W. Reich and R. G. Helmer, Phys. Rev. Lett. **64**, 271 (1990).
2. R. G. Helmer and C. W. Reich, Phys. Rev. C **49**, 1845 (1994).
3. F. Attalah, M. Aiche, J. F. Chemin, *et al.*, Phys. Rev. Lett. **75**, 1715 (1995).
4. É. E. Berlovich, V. V. Lukashevich, A. V. Popov, and V. M. Romanov, Yad. Fiz. **12**, 217 (1970) [Sov. J. Nucl. Phys. **12**, 117 (1971)].
5. G. M. Irwin and K. H. Kim, Phys. Rev. Lett. **79**, 990 (1997).
6. D. S. Richardson, D. M. Benton, D. E. Evans, *et al.*, Phys. Rev. Lett. **80**, 3206 (1998).
7. S. B. Utter, P. Beiersdorfer, A. Barnes, *et al.*, Phys. Rev. Lett. **82**, 505 (1999).
8. R. W. Shaw, J. P. Young, S. P. Cooper, and O. F. Webb, Phys. Rev. Lett. **82**, 1109 (1999).
9. E. V. Tkalya, Pis'ma Zh. Éksp. Teor. Fiz. **70**, 367 (1999) [JETP Lett. **70**, 371 (1999)].
10. F. F. Karpeshin, I. M. Band, M. B. Trzhaskovskaya, and A. A. Pastor, Phys. Rev. Lett. **83**, 1072 (1999); F. F. Karpeshin, I. M. Band, M. B. Trzhaskovskaya, and A. A. Pastor, Izv. Akad. Nauk, Ser. Fiz. **63**, 38 (1999).
11. B. A. Zon and F. F. Karpeshin, Zh. Éksp. Teor. Fiz. **97**, 401 (1990) [Sov. Phys. JETP **70**, 224 (1990)].
12. V. F. Strizhov and E. V. Tkalya, Zh. Éksp. Teor. Fiz. **99**, 697 (1991) [Sov. Phys. JETP **72**, 387 (1991)].
13. F. F. Karpeshin, M. R. Harston, F. Attalah, *et al.*, Phys. Rev. C **53**, 1640 (1996).
14. F. F. Karpeshin, I. M. Band, and M. B. Trzhaskovskaya, Zh. Éksp. Teor. Fiz. **116**, 1565 (1999) [JETP **89**, 845 (1999)].
15. I. M. Band, Izv. Akad. Nauk, Ser. Fiz. **65**, 140 (2001).
16. I. M. Band, Izv. Akad. Nauk, Ser. Fiz. **65**, 624 (2001).
17. I. M. Band, M. A. Listengarten, and A. P. Feresin, *Anomalies in Coefficients of Gamma Ray Internal Conversion* (Nauka, Leningrad, 1976).
18. I. M. Band, M. A. Listengarten, M. B. Trzhaskovskaya, and V. I. Fomichev, Preprint LIYaF-289 (Inst. of Nuclear Physics, Leningrad, 1976).
19. I. P. Grant, Adv. Phys. **19**, 747 (1970).
20. I. M. Band and V. I. Fomichev, Preprint LIYaF-498 (Inst. of Nuclear Physics, Leningrad, 1979).
21. I. P. Grant, B. Y. McKenzie, P. H. Norrington, *et al.*, Comput. Phys. Commun. **21**, 207 (1980).
22. I. P. Grant and N. C. Paper, J. Phys. B **9**, 761 (1976).
23. I. M. Band and M. B. Trzhaskovskaya, At. Data Nucl. Data Tables **55**, 43 (1993).
24. I. M. Band, M. A. Listengarten, and M. B. Trzhaskovskaya, Opt. Spektrosk. **71**, 225 (1991) [Opt. Spectrosc. **71**, 130 (1991)].
25. V. B. Berestetskiĭ, Zh. Éksp. Teor. Fiz. **18**, 1057 (1948).
26. I. M. Band, L. A. Sliv, and M. B. Trzhaskovskaya, Nucl. Phys. A **156**, 170 (1970).
27. A. I. Akhiezer and V. B. Berestetskiĭ, *Quantum Electrodynamics* (Fizmatgiz, Moscow, 1959; Wiley, New York, 1965).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 93, No. 5, 2001, pp. 957–968.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 5, 2001, pp. 1104–1116.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Color Dipole BFKL–Regge Factorization and High-Energy Photon–Photon Scattering¶

N. N. Nikolaev[a, b, *], J. Speth[a], and V. R. Zoller[c, **]

[a]*Institut für Kernphysik, Forschungszentrum Jülich, D-52425 Jülich, Germany*
[b]*Landau Institute for Theoretical Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia*
[c]*Institute for Theoretical and Experimental Physics, Russian Academy of Sciences, Moscow, 117218 Russia*
**e-mail: kph154@ikp301.ikp.kfa-juelich.de*
***e-mail: zoller@heron.itep.ru*

Received June 25, 2001

Abstract—Based on the color dipole representation, we investigate consequences for the $\gamma^*\gamma^*$, $\gamma^*\gamma$ scattering of the finding by Fadin, Kuraev, and Lipatov that the incorporation of asymptotic freedom into the BFKL equation makes the QCD pomeron a series of isolated poles in the angular momentum plane. The emerging color dipole BFKL–Regge factorization allows us to relate in a model-independent way the contributions of each BFKL pole to the $\gamma^*\gamma^*$, $\gamma^*\gamma$ scattering and the deep inelastic scattering on protons. Numerical predictions based on our early work on the color dipole BFKL phenomenology of the deep inelastic scattering on protons gives a good agreement with the recent experimental data from OPAL and L3 experiments at LEP200. We discuss the role of nonperturbative dynamics and predict a pronounced effect of the Regge-factorization breaking due to large unfactorizable nonperturbative corrections to the perturbative vacuum exchange. We comment on the salient features of the BFKL–Regge expansion for the $\gamma^*\gamma^*$, $\gamma^*\gamma$ scattering including the issue of the decoupling of subleading BFKL poles and the soft plus rightmost hard BFKL pole dominance. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

We study the scattering of virtual and real photons,

$$\gamma^*(q) + \gamma^*(p) \longrightarrow X, \tag{1}$$

in the high-energy regime of a large Regge parameter $1/x$; this parameter depends on virtualities of photons as

$$\frac{1}{x} = \frac{W^2 + Q^2 + P^2}{Q^2 + P^2 + \mu^2} \gg 1 \tag{2}$$

and has the correct parton model limit if either $Q^2 \ll P^2$ or $P^2 \ll Q^2$. In Eq. (2), $W^2 = (q + p)^2$ is the center-of-mass energy squared of colliding spacelike photons $\gamma^*(q)$ and $\gamma^*(p)$ with the respective virtualities $q^2 = -Q^2$ and $p^2 = -P^2$.

The high-energy virtual photon–virtual photon scattering can be viewed as an interaction of small size color dipoles from the beam and target photons. The recent strong theoretical [1–5] and experimental [1, 6–9] (see also a compilation in [10]) interest in the high-energy $\gamma^*\gamma^*$, $\gamma^*\gamma$, and $\gamma\gamma$ scattering stems from the fact that virtualities of photons give a handle on the size of color dipoles in the beam and target photons and, eventually, the short-distance properties of the QCD pomeron exchange. For earlier development of the subject, see the pioneering paper [11].

As noticed by Fadin, Kuraev, and Lipatov in 1975 [12] and discussed in more detail by Lipatov in [13], the incorporation of the asymptotic freedom into the BFKL equation [14] makes the QCD pomeron a series of isolated poles in the angular–momentum plane. The contribution of each isolated pole to the high-energy scattering amplitude satisfies the familiar Regge factorization [15]. In [16], we reformulated the consequences of the Regge factorization in our color dipole (CD) approach to the BFKL pomeron. In this paper, we address several closely related issues in the photon–photon scattering in Regge regime (2) which can be tested at LEP200 and Next Linear Collider (NLC).

First, following our early work [16–18], we discuss how the CD BFKL–Regge factorization leads to parameter-free predictions for the total cross sections of the $\gamma^*\gamma^*$, $\gamma^*\gamma$, and $\gamma\gamma$ scattering. We find good agreement with the recent experimental data from the L3 and OPAL experiments at LEP [6–9].

Second, we discuss the interplay of soft and hard dynamics of the vacuum exchange and comment on the onset of the soft plus rightmost hard BFKL-pole dominance in the $\gamma^*\gamma^*$ diffractive scattering. The nodal properties of eigenfunctions of the color dipole BFKL equation suggest an interesting possibility of the decoupling of subleading BFKL singularities when the virtuality of one or both of photons is in the broad vicinity of $Q^2 \sim 20\ \mathrm{GeV}^2$. This makes the leading hard plus soft approx-

¶This article was submitted by the authors in English.

imation (LHSA) previously advocated by us in [18] very efficient.

Third, we discuss the impact of the CD BFKL expansion on the contentious issue of testing the factorization properties of photon–photon scattering in the Q^2, P^2-plane which was previously discussed [4] only in the α_S = const approximation to the BFKL equation. Our result is that the nonperturbative corrections break down the Regge factorization. The experimental observation of this phenomenon would contribute to better understanding of the nonperturbative dynamics of high-energy processes.

2. OVERVIEW OF THE COLOR DIPOLE BFKL–REGGE FACTORIZATION

In the color dipole basis, the beam–target scattering is viewed as a transition of γ^* into a quark–antiquark pair and the interaction of the beam (b) and target (t) color dipoles of the flavor A, $B = u, d, s, c$. As a fundamental quantity, we use the forward dipole scattering amplitude and/or the dipole–dipole cross section $\sigma(x, \mathbf{r}, \mathbf{r}')$. Once $\sigma(x, \mathbf{r}, \mathbf{r}')$ is known, the total cross section of scattering of an $A\bar{A}$ color dipole in the beam on a $B\bar{B}$ color dipole in the target, $\sigma^{AB}(x)$, is calculated as

$$\sigma^{AB}(x) = \int dz d^2\mathbf{r} dz' d^2\mathbf{r}' |\Psi_A(z, \mathbf{r})|^2 \times |\Psi_B(z', \mathbf{r}')|^2 \sigma(x, \mathbf{r}, \mathbf{r}'), \quad (3)$$

where $\mathbf{r}$ and $\mathbf{r}'$ are the two-dimensional vectors in the impact parameter plane. In the color dipole factorization formula (3), the dipole–dipole cross section $\sigma(x, \mathbf{r}, \mathbf{r}')$ is beam–target symmetric and universal for all beams and targets, the beam and target dependence is concentrated in the probabilities $|\Psi_A(z, \mathbf{r})|^2$ and $|\Psi_B(z', \mathbf{r}')|^2$ to find an $A\bar{A}$ color dipole with $\mathbf{r}$ in the beam and a $B\bar{B}$ color dipole with $\mathbf{r}'$ in the target, respectively. Hereafter, we focus on cross sections averaged over polarizations of the beam and target photons, in which case only the $n = 0$ term of the Fourier series,

$$\sigma(x, \mathbf{r}, \mathbf{r}') = \sum_{n=0}^{\infty} \sigma_n(x, r, r') \exp(in\varphi), \quad (4)$$

where φ is the azimuthal angle between $\mathbf{r}$ and $\mathbf{r}'$, contributes in (3).

In 1975, Fadin, Kuraev, and Lipatov observed [12] (see also Lipatov's extensive discussion [13]) that the incorporation of asymptotic freedom into the BFKL equation makes the QCD pomeron a series of isolated poles in the angular momentum plane. The contribution of each pole to scattering amplitudes satisfies the standard Regge factorization [15], which in the CD basis implies the CD BFKL–Regge expansion for the vacuum exchange dipole–dipole cross section:

$$\sigma_n(x, r, r') = \sum_m C_m \sigma_m(r) \sigma_m(r') \left(\frac{x_0}{x}\right)^{\Delta_m}. \quad (5)$$

Here, the dipole cross section $\sigma_m(r)$ is an eigenfunction of the CD BFKL equation [16, 17, 19–21]

$$\frac{\partial \sigma_m(x, r)}{\partial \ln(1/x)} = \mathcal{H} \otimes \sigma_m(x, r) = \Delta_m \sigma_m(x, r), \quad (6)$$

with the eigenvalue (intercept) Δ_m where $\mathcal{H}$ is the kernel of the BFKL equation in the CD representation [16]. Arguably, for the transition of γ^* into heavy flavors, with $A = c, b, \ldots$, the hardness scale is set by $Q^2 + 4m_A^2$, and for light flavors, $Q^2 + m_\rho^2$ is a sensible choice. Hence, for the light-light transition, we evaluate Regge parameter (2) with $\mu^2 = m_\rho^2$, for the light–charm contribution, we take $\mu^2 = 4m_c^2$, and for the charm–charm contribution, we take $\mu^2 = 8m_c^2$.

We refer to our early works [16, 17, 21] for the details on the CD formulation of the BFKL equation, the infrared regularization by a finite propagation radius R_c for perturbative gluons, the freezing of strong coupling at large distances, the choice of the physically motivated boundary condition for the hard BFKL evolution, and for the description of eigenfunctions. The successful application of the CD BFKL–Regge expansion to the proton and pion structure functions and the evaluation of the hard-pomeron contribution to the rise of hadronic and real photoabsorption cross sections can be found in [16–18, 21]. We only recapitulate the salient features of the formalism that are essential for the present discussion.

There is a useful analogy between the intercept $\Delta = \alpha(0) - 1$ and the binding energy for the bound state problem for the Schrödinger equation. The eigenfunction $\sigma_0(r)$ for the rightmost hard BFKL pole (the ground state) corresponding to the largest intercept $\Delta_0 \equiv \Delta_{\mathbb{P}}$ is node-free. The eigenfunctions $\sigma_m(r)$ for excited states with m radial nodes have the intercept $\Delta_m < \Delta_{\mathbb{P}}$. Our choice of $R_c = 0.27$ fm yields the intercept $\Delta_{\mathbb{P}} = 0.4$ for the rightmost hard BFKL pole and $\Delta_m \approx \Delta_0/(m+1)$ for subleading hard poles. The node of $\sigma_1(r)$ is located at $r = r_1 \approx 0.05$–0.06 fm, for larger m the rightmost nodes move to a somewhat larger r and accumulate at $r \sim 0.1$ fm; see [16, 17] for a more detailed description of the nodal structure of $\sigma_m(r)$. Here we only emphasize that for solutions with $m \geq 3$, the third and higher nodes are located at a very small r far beyond the resolution scale $1/\sqrt{Q^2}$ of foreseeable deep inelastic scattering (DIS) experiments. We note that the Regge cut in the complex angular momentum plane

found in the much discussed approximation α_S = const resembles an infinite continuous sequence of poles. In a counterpart of our CD BFKL–Regge expansion (5) for the approximation α_S = const, the intercept Δ_m would be a continuous parameter in contrast to the discrete spectrum for the standard running α_S parameter.

Because the BFKL equation sums cross sections of the production of multigluon final states, the perturbative two-gluon Born approximation is an arguably natural boundary condition. This leaves the starting point x_0 as the only free parameter that completely fixes the result of the hard BFKL evolution for the dipole–dipole cross section. We follow the choice $x_0 = 0.03$ made in [21]. The very ambitious program for describing $F_{2p}(x, Q^2)$ starting from this, perhaps excessively restrictive, perturbative two-gluon boundary condition has been launched by us in [16] and met with remarkable phenomenological success [17, 18].

Because the subleading solutions with $m \geq 3$ cannot be resolved in the attainable region of r and all these solutions have similar intercepts $\Delta_m \ll 1$, in practical evaluation of σ^{AB} we can truncate expansion (5) at $m = 3$, thereby lumping the contributions of all singularities with $m \geq 3$ in the $m = 3$ term. Specifically, if we endow

$$\sigma_3(r) = \sigma_{\text{Born}}(r) - \sum_{m=0}^{2} \sigma_m(r) \tag{7}$$

with the effective intercept $\Delta_3 = 0.06$, the truncated expansion reproduces the numerical solution $\sigma(x, r)$ of our CD BFKL equation in the wide range of dipole sizes $10^{-3} \lesssim r \lesssim 10$ fm with an accuracy of approximately 10% even at moderately small x. This truncation can be justified *a posteriori* if the contribution from $m \geq 3$ turns out to be a small correction, which is indeed the case at small x.

Whereas the scattering of small dipoles $r \lesssim R_c$ is dominated by the exchange of perturbative gluons, the interaction of large dipoles with the proton target has been modeled in [17, 18, 21] by the nonperturbative soft pomeron with the intercept $\alpha_{\text{soft}}(0) - 1 = \Delta_{\text{soft}} = 0$. An extra term $\sigma_{\text{soft}}(r, r')$ must then be added to the right-hand side of expansion (5).

From the early phenomenology of DIS and the diffractive vector meson production off the proton target, we only know the parameterization of $\sigma_{\text{soft}}(r, r')$ when one of the dipoles is definitely large, of the order of the proton size. Evaluation of the soft contribution to the $\gamma^*\gamma^*$ scattering inevitably introduces model dependence if both dipoles are small. Modeling the soft contribution by the exchange of two nonperturbative gluons suggests [22]

$$\sigma_{\text{soft}}(r, r') \propto \frac{r^2 r'^2}{r^2 + r'^2}$$

and the nonfactorizable cross section in the form

$$\sigma_{\text{soft}}^{\gamma^*\gamma^*}(Q^2, P^2) \propto \frac{1}{Q^2 + P^2}.$$

A similar nonperturbative cross section σ_{soft} is found in the soft pomeron models [23]. The explicit parameterization is given in Appendix.

Finally, at moderately small values of x, the t-channel gluon tower exchange described above must be complemented by the t-channel $q\bar{q}$ exchange often associated with the DIS off vector mesons (hadronic component) and off the perturbative (point-like) $q\bar{q}$-component of the target photon wave function. We add the corresponding corrections only to the real photon structure function $F_{2\gamma}(x, Q^2)$ to estimate the interplay of the vacuum and non-vacuum exchanges in the currently accessible kinematical region of not very small x. In all other cases of interest, we concentrate on the pure vacuum exchange at $x \lesssim x_0$, where the nonvacuum corrections are negligibly small.

In our evaluation of the box diagram contribution to $F_{2\gamma}^{pl}(x, Q^2)$, which is [24]

$$F_{2\gamma}^{pl}(x, Q^2) = \frac{3\alpha_{em}}{\pi} \sum_{q = u, d, s, c} e_q^4 x \left\{ [x^2 + (1-x)^2] \times \ln \frac{Q^2(1-x)}{xQ_q^2} + 8x(1-x) - 1 \right\}, \tag{8}$$

we take the ρ-meson mass as the lower cut-off for the light-flavor-loop integral, $Q_q^2 = m_\rho^2$ for $q = u, d, s$, and the charm quark mass for the c-loop, $Q_c^2 = m_c^2$. In Eq. (8), e_q is the quark charge.

To describe the hadronic component of $F_{2\gamma}$ we take the coherent mixture of the ρ^0 and ω mesons [25]. Supplemented with the standard assumptions on the vector meson valence quark density, this gives

$$F_{2\gamma}^{\text{had}}(x) = \frac{\alpha_{em}}{12}[4(g_\rho + q_\omega)^2 + (g_\rho - g_\omega)^2]\sqrt{x}(1-x), \tag{9}$$

where the coupling constants $q_V^2 = 4\pi/f_V^2$ entering the Fock state expansion

$$|\gamma\rangle^{\text{had}} = \frac{e}{f_\rho}|\rho\rangle + \frac{e}{f_\omega}|\omega\rangle + \ldots$$

are given by $g_\rho^2 = 0.5$ and $g_\omega^2 = 0.043$ [10]. We neglect the Q^2 evolution which, at reasonable values of the lower scale, is a small correction on the interval $1.9 \leq Q^2 \leq 5$ GeV^2 where the small-x data on $F_{2\gamma}$ were taken.

Combining Eqs. (5) and (3) and adding the soft and nonvacuum components, we obtain our principal result

Table 1. CD BFKL–Regge expansion parameters

m	Δ_m	σ_m^p, mb	C_m, mb^{-1}	A_m^γ/α_{em}	σ_m^γ, μb	$\sigma_m^{\gamma\gamma}$, nb
0	0.402	1.243	0.804	0.746	6.767	36.84
1	0.220	0.462	2.166	0.559	1.885	7.69
2	0.148	0.374	2.674	0.484	1.320	4.65
3	0.06	3.028	0.330	0.428	9.456	29.53
soft	0	31.19	0.0321	0.351	79.81	204.2

for the virtual–virtual scattering ($m = 0, 1, 2, 3$, $A, B = u, d, s, c$)

$$\sigma_{\rm vac}^{\gamma^*\gamma^*}(x, Q^2, P^2) = \frac{(4\pi^2\alpha_{em})^2}{Q^2P^2} \times \sum_m C_m \sum_{A,B} f_m^A(Q^2) f_m^B(P^2) \left(\frac{3x_0}{2x_{AB}}\right)^{\Delta_m} + \sigma_{\rm soft}^{\gamma^*\gamma^*}(x, Q^2, P^2). \quad (10)$$

To make the scale dependence discussed in Introduction explicit, we provide μ and x defined by Eq. (2) with two indices, A and B, indicating the flavor of the beam and the target dipoles: $\mu_{AB}^2 = m_\rho^2$ for $A, B = u, d, s$ while $\mu_{AB}^2 = 4m_c^2$ if either $A = c$ or $B = c$ and the second dipole is made of light quarks and $\mu_{AB}^2 = 8m_c^2$ if $A = B = c$.

For the DIS off real (quasi-real) photons with $P^2 \approx 0$ we have ($A = u, d, s, c$)

$$F_{2\gamma}(x, Q^2) = \sum_m A_m^\gamma \sum_A f_m^A(Q^2) \left(\frac{3}{2}\frac{x_0}{x_A}\right)^{\Delta_m} + F_{2\gamma}^{\rm soft}(x, Q^2) + F_{2\gamma}^{n\rm vac}(x, Q^2), \quad (11)$$

where

$$x_A = \frac{Q^2 + \mu_A^2}{W^2 + Q^2}$$

and $\mu_A^2 = m_\rho^2$ for $A = u, d, s$, while $\mu_A^2 = 4m_c^2$ for $A = c$. The $c\bar{c}$ component of the target photon wave function is strongly suppressed at $P^2 \approx 0$ and can be neglected for all the practical purposes as well as the $c\bar{c}$ content of the target proton. This observation simplifies factorization relation (11) for the real photon structure function. In Eq. (11), the nonvacuum component denoted by $F_{2\gamma}^{n\rm vac}$ is

$$F_{2\gamma}^{n\rm vac}(x, Q^2) = F_{2\gamma}^{\rm had}(x, Q^2) + F_{2\gamma}^{pl}(x, Q^2) \quad (12)$$

and the cross sections

$$\sigma_m^{\gamma^*}(Q^2) = \langle\gamma_T^*|\sigma_m(r)|\gamma_T^*\rangle + \langle\gamma_L^*|\sigma_m(r)|\gamma_L^*\rangle \quad (13)$$

are calculated with the well-known color dipole distributions in the transverse (T) and longitudinal (L) photon of the virtuality Q^2 derived in [26], and the eigenstructure functions are defined as usual,

$$f_m(Q^2) = \frac{Q^2}{4\pi^2\alpha_{em}}\sigma_m^{\gamma^*}(Q^2). \quad (14)$$

The Regge parameter involves the factor 3/2 because in the scattering of a color dipole on the photon, the effective dipole–dipole collision energy is 3/2 of that in the reference scattering of the color dipole on the three-quark nucleon at the same total c.m.s. energy W. The analytical formulas for the eigen-structure functions $f_m(Q^2)$ and $f_m^c(Q^2)$ are given in Appendix. Here, as in all our previous calculations, we put $m_c = 1.5$ GeV. We do not need any new parameters in addition to those used in the description of the DIS and real photoabsorption on protons [16–18]; the results for the expansion parameters A_m^γ and $\sigma_m^\gamma(0)$ are summarized in Table 1.

We recall that because of the diffusion in the color dipole space, the exchange of perturbative gluons also contributes to the interaction of large dipoles $r > R_c$ [20]. However, this hard interaction-driven effect is still small at a moderately large Regge parameter. For this reason, we refer to the terms with $m = 0, 1, 2, 3$ as the hard contribution as opposed to the genuine soft interaction.

3. ISOLATING THE SOFT PLUS RIGHTMOST HARD BFKL POLE IN HIGHLY VIRTUAL–VIRTUAL γ*γ* SCATTERING

We start with the theoretically purest case of the highly virtual photons, P^2, $Q^2 \gg 1$ GeV^2, and focus on the vacuum exchange component of the total cross section. The CD BFKL approach with asymptotic freedom uniquely predicts that subleading eigenstructure functions have a node at $Q^2 \sim 20$ GeV^2; in this region of Q^2, the rightmost hard pole contribution dominates. This suppression of the subleading hard background is

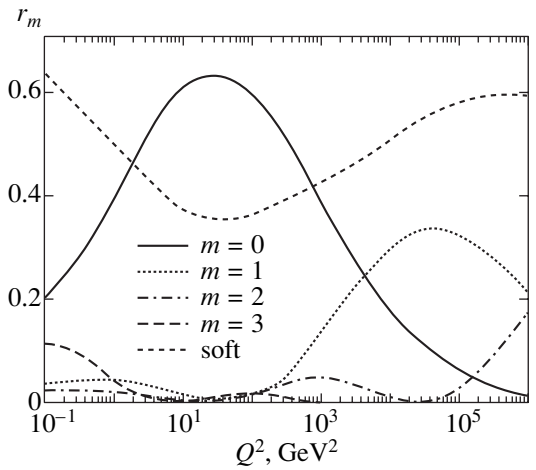


Fig. 1. The normalized ratio of the soft-to-rightmost-hard and subleading hard-to-rightmost hard expansion coefficients (m = 0, 1, 2, 3, soft) $r_m(Q^2) = \sigma_m^{\gamma^*\gamma^*}/\sigma_{\rm vac}^{\gamma^*\gamma^*}$ of the BFKL–Regge expansion for the $\gamma^*\gamma^*$ scattering at $x = x_0$.

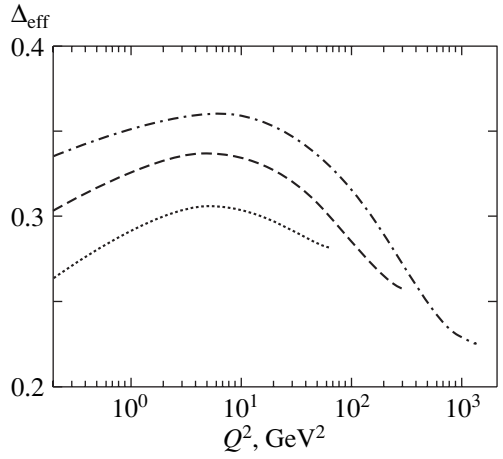


Fig. 2. Predictions from the CD BFKL–Regge expansion for the effective intercept $\Delta_{\rm eff}$, Eq. (15), for the diagonal case $Q^2 = P^2$ and W = 50 GeV (dotted curve), 100 GeV (dashed curve), and 200 GeV (dot-dashed curve).

shown in Fig. 1, where we plot the ratio (m = 0, 1, 2, 3, soft)

$$r_m(Q^2) = \frac{\sigma_m^{\gamma^*\gamma^*}\left(\frac{3}{2}x_0, Q^2, Q^2\right)}{\sigma_{\rm vac}^{\gamma^*\gamma^*}\left(\frac{3}{2}x_0, Q^2, Q^2\right)}$$

that defines the relative size of different contributions to $\sigma_{\rm vac}^{\gamma^*\gamma^*}$ at $x = 3x_0/2$. At this value of x, the contribution of the subleading hard BFKL poles remains marginal in a broad range of Q^2, although the contribution of the single-node component $m = 1$ becomes substantial at $Q^2 \gtrsim 10^3$ GeV^2.

The soft-pomeron exchange contributes substantially over all Q^2 and dominates at $Q^2 \lesssim 1$ GeV^2. However, at very large $W \sim 100$ GeV, which is of the practical interest at LEP and LHC, such small values of Q^2 correspond to very small x, where the soft and subleading hard contributions are Regge suppressed by the factor $(x/x_0)^{\Delta_{\mathbb{P}}}$ and $(x/x_0)^{0.5\Delta_{\mathbb{P}}}$, respectively. The latter is clearly seen in Fig. 2, where the effective pomeron intercept

$$\Delta_{\rm eff} = -\frac{\partial \ln \sigma_{\rm vac}^{\gamma^*\gamma^*}}{\partial \ln x} \qquad (15)$$

is presented for the diagonal case $Q^2 = P^2$ at three different values of W.

According to the results shown in Fig. 1, the dominance of the soft plus rightmost hard BFKL pomeron exchange in the virtual–virtual $\gamma^*\gamma^*$ scattering holds in a very broad range of Q^2, $P^2 \lesssim 500$ GeV^2, which nearly exhausts the interesting kinematical region at LEP200 and NLC. The quality of the leading hard pole plus soft approximation (LHSA) can also be judged from Fig. 3 for the diagonal case where $Q^2 = P^2$; in Fig. 3 we show the soft component of the cross section separately (the dashed curve). That the contribution of the subleading hard BFKL exchange is marginal is clear from the finding that the approximation of a soft pomeron plus the rightmost hard BFKL exchange (LHSA) shown by long-dash curve nearly exhausts the result of the complete CD BFKL–Regge expansion for the vacuum exchange.

Recently, the L3 collaboration [8] reported the first experimental evaluation of the vacuum exchange in the

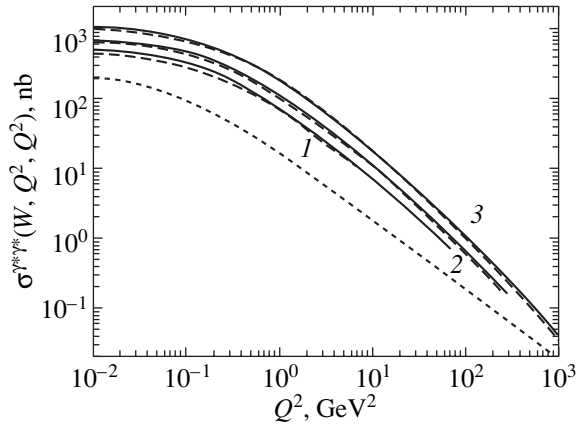


Fig. 3. Predictions from the CD BFKL–Regge expansion for the vacuum exchange component of the virtual–virtual $\gamma^*\gamma^*$ cross section for the diagonal case $Q^2 = P^2$ and for the cms collision energy W = 50, 100, and 200 GeV (solid curves *1*, *2*, and *3*, respectively). The leading hard BFKL exchange plus soft-pomeron exchange approximation (LHSA) is shown by the long-dash curve. The soft pomeron component of the cross section is shown separately by the dashed curve.

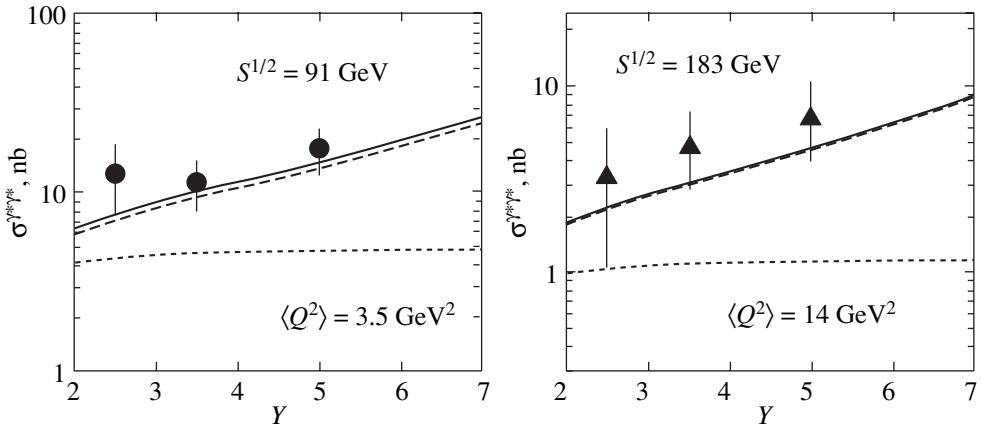


Fig. 4. Predictions from the CD BFKL–Regge expansion for the vacuum exchange component of the virtual–virtial $\gamma^*\gamma^*$ cross section for the diagonal case of $\langle Q^2 \rangle = \langle P^2 \rangle$ are confronted with the experimental data by the L3 Collaboration [10]. The experimental data and theoretical curves are shown vs. the variable $Y = \ln(W^2/\sqrt{Q^2P^2})$.

equal virtuality $\gamma^*\gamma^*$ scattering. Their procedure of subtracting the nonvacuum reggeon and/or the quark parton model contribution is described in [8], arguably the subtraction uncertainties are marginal within the present error bars. In Fig. 4, we compare our predictions to the L3 data. The experimental data and theoretical curves are shown vs. the variable $Y = \ln(W^2/\sqrt{Q^2P^2})$. The virtuality of two photons varies in the range of 1.2 GeV^2 < Q^2, P^2 < 9 GeV^2 ($\langle Q^2, P^2 \rangle$ = 3.5 GeV^2) at $\sqrt{s} \approx 91$ GeV and 2.5 GeV^2 < Q^2, P^2 < 35 GeV^2 at $\sqrt{s} \approx 183$ GeV ($\langle Q^2, P^2 \rangle$ = 14 GeV^2). We applied the averaging procedure described in [8] to the theoretical cross sections. The solid curve is a result of the complete BFKL–Regge expansion for the vacuum exchange, the long-dash curve is a sum of the rightmost hard BFKL exchange and the soft-pomeron exchange. The soft pomeron contribution is shown by the dashed line. The agreement of our estimates with the experiment is good, the contribution of the subleading hard BFKL exchange is negligible within the experimental error bars.

The early calculations [2–4] of the perturbative vacuum component of $\sigma^{\gamma^*\gamma^*}$ used the approximation α_S = const that predicts the P^2, Q^2-dependence different from our result for the CD BFKL approach with running α_S. A detailed comparison with numerical results by Brodsky, Hautmann, and Soper (BHS) [4] has been reported by the L3 Collaboration [8], which found that BHS formulas substantially overestimate $\sigma_{\rm vac}^{\gamma^*\gamma^*}$. In [3], the same perturbative fixed-α_S BFKL model with a massive *c*-quark was considered. At $\langle Q^2 \rangle$ = 14 GeV^2 and moderately small x ($x \gtrsim 3 \times 10^{-2}$), the model is in agreement with the L3 data, but at smaller x, already at $x \approx 7 \times 10^{-3}$, it substantially overestimates $\sigma_{\rm vac}^{\gamma^*\gamma^*}$. At $\langle Q^2 \rangle$ = 3.5 GeV^2 the results in [3] are substantially above the L3 data for all x.

4. VIRTUAL–REAL $\gamma^*\gamma$ SCATTERING: THE RIGHTMOST HARD BFKL POLE IN THE PHOTON STRUCTURE FUNCTION

The discussion of the photon structure function closely follows that of the proton and pion structure functions in [16–18]. Our normalization of eigenfunctions is such that the vacuum (sea) contribution to the proton structure function (m = soft, 0, 1, …, 3)

$$F_{2p}(x, Q^2) = \sum_m f_m(Q^2)\left(\frac{x_0}{x}\right)^{\Delta_m}, \qquad (16)$$

has the CD BFKL–Regge expansion coefficients $A_m^p = 1$. There is a fundamental point that the distribution of small-size color dipoles in the photon is enhanced compared to that in the proton [18]; this enhances the importance of the rightmost hard BFKL exchange. Indeed, closer inspection of the expansion coefficients A_m^γ shown in Table 1 reveals that subleading hard BFKL exchanges are suppressed by the factor ~1.5 to 2, whereas the soft-pomeron exchange contribution is suppressed by the factor ~3.

Our predictions for the photon structure function are parameter-free and are presented in Fig. 5. At moderately small $x \sim 0.1$, there is a substantial non-vacuum Regge exchange contribution from the DIS off the hadronic ($q\bar{q}$) component of the target photon wave function, which can be regarded as well constrained by the large x data. We use here the parameterizations presented above (Eqs. (8), (9), and (12)). The solid curve

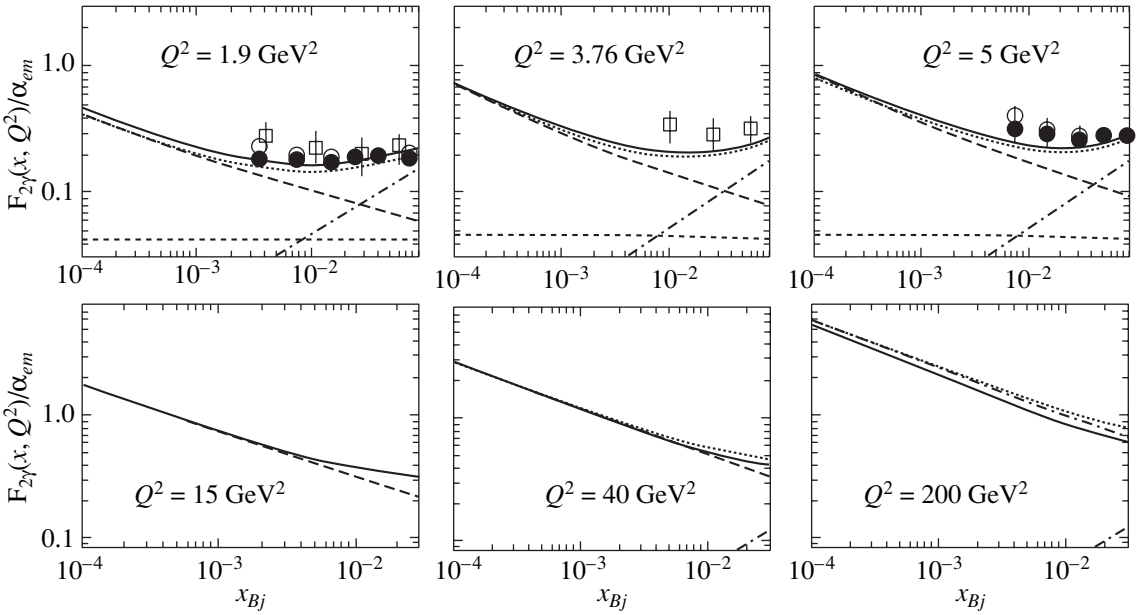


Fig. 5. Predictions from the CD the BFKL–Regge expansion for the photon structure function. The solid curve shows the result from complete BFKL–Regge expansion with the soft-pomeron (the dashed curve) and valence (the dot-dashed curve) components included, the dotted curve shows the rightmost hard BFKL (LH) plus soft-pomeron (S) plus non-vacuum (NV) approximation (LHSNVA). The long-dash line corresponds to the LH plus S approximation (LHSA). Data points are from [6, 9].

shows the result of the complete BFKL–Regge expansion with the soft-pomeron (the dashed curve) and quasi-valence (the dot-dashed curve) components included, the dotted curve shows the right-most hard BFKL (LH) plus soft-pomeron (S) plus non-vacuum (NV) approximation (LHSNVA). A comparison of the solid and dotted curves clearly shows that the subleading hard BFKL exchanges are numerically small in the experimentally interesting region of Q^2, the rightmost hard BFKL pole exhausts the hard vacuum contribution for $2 \lesssim Q^2 \lesssim 100$ GeV^2. The nodal properties of subleading hard BFKL structure functions are clearly seen: LHSNVA slightly underestimates $F_{2\gamma}$ at $Q^2 \lesssim 10$ GeV^2 and overestimates $F_{2\gamma}$ at $Q^2 \gtrsim 50$ GeV^2. For another illustration of the same modal property of the subleading hard components, see Fig. 6, where we show the vacuum component of the virtual–real total cross section $\sigma_{\mathrm{tot}}^{\gamma^*\gamma}$ as a function of Q^2 at fixed W. As seen from Fig. 1, the soft contribution rises towards small Q^2, but this rise is compensated to a large extent by the small-x enhancement of the rightmost hard BFKL contribution by the large Regge factor $(x_0/x)^{\Delta_{\mathbb{P}}}$. For this region, the soft background (the dashed curve) remains marginal over the entire range of Q^2. Because of the node effect, the $m = 1$ subleading component changes the sign and becomes quite substantial at very large Q^2 and moderately small x.

Recently, the L3 and OPAL collaborations reported the first experimental data on the photon structure function at sufficiently small x [6, 9]. These data are shown in Fig. 5 and are in good agreement with the predictions from the CD BFKL–Regge expansion. A comparison with the long-dash curve, which is the sum of the rightmost hard BFKL and soft exchanges, shows that the experimental data are in the region of x and Q^2 still affected by the nonvacuum Reggeon (quasivalence) exchange; going to smaller x and larger Q^2 would greatly improve the sensitivity to the pure vacuum exchange.

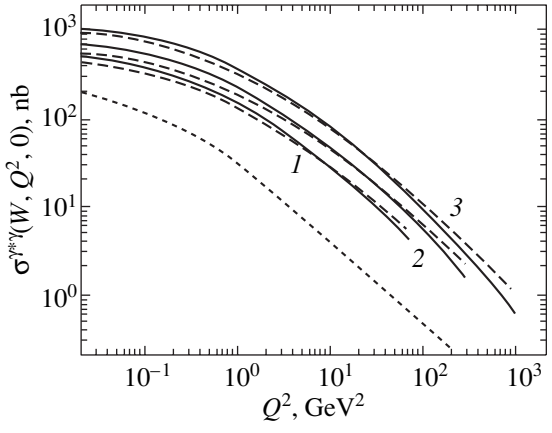


Fig. 6. Predictions from the CD BFKL–Regge expansion for the vacuum exchange component of the virtual–real $\gamma^*\gamma$ total cross section and for the cms collision energy W = 50, 100, and 200 GeV (solid curves, *1*, *2*, and *3*, respectively). The result from the rightmost hard BFKL (LH) plus soft-pomeron (S) approximation (LHSA) is shown by the long-dash curve. The soft-pomeron exchange component of the cross section is shown separately by the dashed curve.

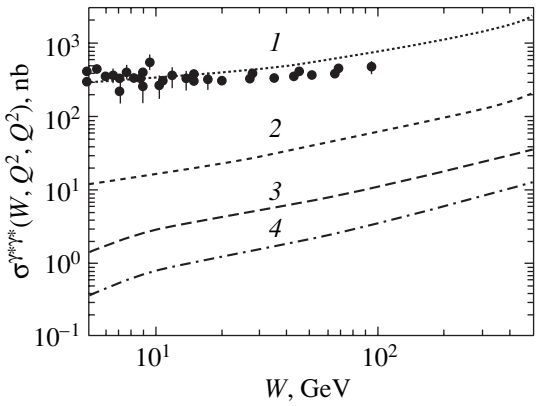


Fig. 7. Our predictions from the CD BFKL–Regge factorization for the single-vacuum exchange contribution to the real–real $\gamma\gamma$ scattering are compared with the recent experimental data from the OPAL collaboration [7, 10] for the cases of $Q^2 = 0$ (curve *1*), 2 (curve *2*), 10 (curve *3*), and 25 GeV^2 (curve *4*).

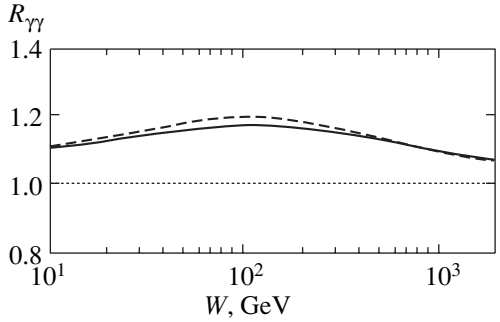


Fig. 8. Our evaluation of $R_{\gamma\gamma}$ for the single-vacuum component of total cross sections using the CD BFKL approach (solid curve) and the rightmost hard BFKL plus soft-pomeron approximation (LHSA, dashed curve).

In order to give a crude idea on finite-energy effects at large x and not so large values of the Regge parameter, we stretch the theoretical curves slightly to $x \gtrsim x_0$ by multiplying the BFKL–Regge expansion result with the purely phenomenological factor $1 - x$ motivated by the familiar behavior of the gluon structure function of the photon $\sim(1 - x)^n$ with the exponent $n \sim 1$.

5. THE REAL–REAL $\gamma\gamma$ SCATTERING

We recall that because of the well-known BFKL diffusion in the color dipole space, the exchange by perturbative gluons also contributes to the interaction of large dipoles $r > R_c$ [20]. As discussed in [18], this gives rise to a substantially rising component of the hadronic and real photoabsorption cross sections and a scenario in which the observed rise of hadronic and real photon cross sections are entirely due to this intrusion of hard scattering. This is a motivation behind our choice of the intercept $\Delta_{\rm soft} = 0$ for the soft pomeron exchange. Furthermore, in order to make this picture quantitative, one must invoke strong absorption/unitarization to tame the too rapid growth of the large dipole component of the hard BFKL dipole cross section. The case of the real–real $\gamma\gamma$ scattering is not an exception and the above enhancement of small dipole configurations in photons compared to hadrons uniquely predicts that the hard BFKL exchange component of the real–real $\gamma\gamma$ scattering is enhanced compared to the proton–proton and/or pion–proton scattering. This is clearly seen from Table 1, where we show the coefficients

$$\sigma_m^{\gamma\gamma} = \sigma_m^{\gamma}\sigma_m^{\gamma}C_m \tag{17}$$

of the expansion for the vacuum exchange component of the total $\gamma\gamma$ cross section ($m = 0, 1, 2, 3$, soft)

$$\sigma_{\rm vac}^{\gamma\gamma} = \sum_m \sigma_m^{\gamma\gamma}\left(\frac{W^2 x_0}{m_\rho^2}\right)^{\Delta_m}. \tag{18}$$

One must look at the soft–hard hierarchy of $\sigma_m^{\gamma\gamma}$ and σ_m^{γ}, σ_m^{p} in the counterparts of (18) for the γp and pp scattering. This enhancement of the hard BFKL exchange is confirmed by a simplified vacuum pole plus nonvacuum reggeon exchange fits to the real–real $\gamma\gamma$ total cross section: the found intercept of the effective vacuum pole, $\epsilon^{\gamma\gamma} \approx 0.21$, is much larger than $\epsilon \approx 0.095$ from similar fits to the hadronic cross section data. In Fig. 7, we compare our predictions from the CD BFKL–Regge factorization for the single-vacuum exchange contribution to the real–real $\gamma\gamma$ scattering with the recent experimental data from the OPAL collaboration [7, 10]. The theoretical curves are in the right ballpark, but the truly quantitative discussion of total cross sections of soft processes requires better understanding the absorption/unitarization effects.

6. REGGE FACTORIZATION IN THE $\gamma^*\gamma^*$ AND $\gamma\gamma$ SCATTERING

If the vacuum exchange were an isolated Regge pole, the well-known Regge factorization would hold for asymptotic cross sections [15],

$$\sigma_{\rm tot}^{bb}\sigma_{\rm tot}^{aa} = \sigma_{\rm tot}^{ab}\sigma_{\rm tot}^{ab}. \tag{19}$$

In the CD BFKL approach, this Regge factorization holds for each term in the BFKL–Regge expansion for the vacuum exchange, but evidently the sum of the factorized terms does not satisfy factorization (19). One can hope for an approximate factorization if a single term dominates the BFKL–Regge expansion. The corrections to the exact factorization still exist even for the single pole exchange because of the light $q\bar{q}$ and charm $c\bar{c}$ mass scale difference discussed above.

One such case is the real–real $\gamma\gamma$ scattering dominated by the soft-pomeron exchange (although the factorization of the soft on-shell amplitudes never gained

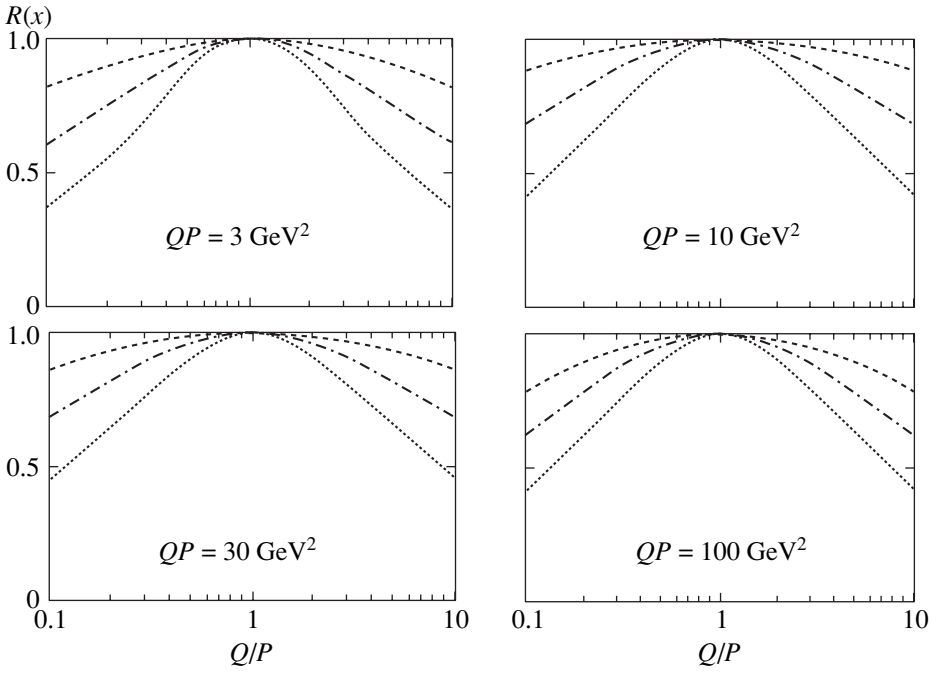


Fig. 9. The factorization cross section ratio $R_{\gamma^*\gamma^*}(x)$ at fixed x and QP as a function of Q/P for $x = 10^{-2}$ (dotted line), 10^{-3} (long-dash line), and 10^{-4} (dashed line).

strong support from the high-energy Regge phenomenology). For this reason, the CD BFKL–Regge expansion, that reproduces well the vacuum exchange components of the pp and γp scattering cannot fail for the vacuum component in real–real $\gamma\gamma$ scattering. The rise of the contribution of the hard-BFKL exchange breaks the Regge factorization relation

$$R_{\gamma\gamma} = \frac{\sigma_{\rm vac}^{\gamma\gamma}\sigma^{pp}}{\sigma_{\rm vac}^{\gamma p}\sigma_{\rm vac}^{\gamma p}} = 1, \qquad (20)$$

which would restore at extremely high energies such that the rightmost hard BFKL exchange dominates. This property is illustrated in Fig. 8, where we show our evaluation of R for the single-vacuum component of the total cross sections entering (19). At moderately high energies, the naive factorization breaks but the expected breaking is still weak, $\lesssim$20%. This curve must not be taken at face value for $W \gtrsim 0.1$–1 TeV because of likely strong absorption effects, but the trend of R, being larger than unity and rising with energy, should withstand the unitarity effects. The second case is the highly virtual–virtual $\gamma^*\gamma^*$ scattering. As we emphasized in Section 3, the CD BFKL approach here uniquely predicts that because of the nodal property of the subleading eigen-structure functions, the superposition of soft and rightmost hard BFKL poles dominate the vacuum exchange in a broad range of Q^2, $P^2 \lesssim 10^3$ GeV^2.

The above discussion clearly suggests that different cross sections must be taken at the same value of $x^{-1} = W^2/(Q^2 + P^2)$, in which case the vacuum components of the $\gamma^*\gamma^*$ scattering at Q^2, $P^2 \gg 4m_c^2$ and $x \ll x_0$ would satisfy

$$R_{\gamma^*\gamma^*}(x) = \frac{[\sigma^{\gamma^*\gamma^*}(x, Q^2, P^2)]^2}{\sigma^{\gamma^*\gamma^*}(x, Q^2, Q^2)\sigma^{\gamma^*\gamma^*}(x, P^2, P^2)} = 1. \qquad (21)$$

In accordance with the results shown in Fig. 1, the soft exchanges break factorization relation (21). The breaking is quite substantial at moderate $x = 0.01$ (dotted line in Fig. 9), and breaking effects disappear rapidly (as x^{Δ_0}) when $x \longrightarrow 0$. If the vacuum singularity were the Regge cut, as is the case with the approximation $\alpha_S = \text{const}$, the restoration of factorization is much slower, cf. our Fig. 9 and Fig. 9 in [4].

For the obvious reason that the soft-pomeron exchange is so predominant in the real photon scattering, whereas the soft plus rightmost hard BFKL exchange is outstanding in the virtual–virtual and real–virtual photon–photon scattering, it is ill advised to look at factorization ratio $R_{\gamma^*\gamma^*}(W)$ when one of the photons is quasi-real, $P^2 \sim 0$. In this limit, one would find strong departures of $R_{\gamma^*\gamma^*}(W)$ from unity. For precisely the same reason of predominance of the soft-pomeron exchange in the pp scattering vs. a nearly dominant rightmost hard BFKL pole exchange in DIS

Table 2. CD BFKL–Regge the all-flavor structure function parameters

m	a_m	c_m	r_m^2, GeV^{-2}	R_m^2, GeV^{-2}	$z_m^{(1)}$	$z_m^{(2)}$	$z_m^{(3)}$	δ_m
0	0.0232	0.3261	1.1204	2.6018				1.
1	0.2788	0.1113	0.8755	3.4648	2.4773			1.0915
2	0.1953	0.0833	1.5682	3.4824	1.7706	12.991		1.2450
3	1.4000	0.04119	3.9567	2.7706	0.23585	0.72853	1.13044	0.5007
soft	0.1077	0.0673	7.0332	6.6447				

at small x and 5–10 $\lesssim Q^2 \lesssim$ 100 GeV^2, see [18], the naive factorization estimate

$$\sigma^{\gamma^*\gamma^*}(W, Q^2, P^2) \approx \frac{\sigma^{\gamma^* p}(W, Q^2)\sigma^{\gamma^* p}(W, P^2)}{\sigma^{pp}(W)} \quad (22)$$

would not make much sense.

7. CONCLUSIONS

We explored the consequences for small-x photon structure functions $F_{2\gamma}(x, Q^2)$ and high-energy two-photon cross sections $\sigma^{\gamma^*\gamma^*}$ and $\sigma^{\gamma\gamma}$ from the color dipole BFKL–Regge factorization. Because of the nodal properties of eigenstructure functions of subleading hard BFKL exchanges, the CD BFKL approach uniquely predicts that the vacuum exchange is strongly dominated by the combination of soft plus rightmost hard BFKL pole exchanges in a very broad range of photon virtualities Q^2, P^2 which includes much of the kinematical domain attainable at LEP200 and NLC. Starting with a very restrictive perturbative two-gluon exchange as a boundary condition for the BFKL evolution in the color dipole basis and having fixed the staring point of the BFKL evolution in the early resulting CD BFKL–Regge phenomenology of the proton structure function, we presented parameter-free predictions for the vacuum exchange contribution to the photon structure function that agree well with OPAL and L3 determinations. A good agreement is found between our predictions for the energy and photon virtuality dependence of the photon–photon cross section $\sigma^{\gamma^*\gamma^*}(W, Q^2, P^2)$ and the recent data taken by the L3 Collaboration [8]. We commented on the utility of Regge factorization tests of the CD BFKL–Regge expansion.

APPENDIX

A.1. CD BFKL All-Flavor Eigen-Structure Functions

In the early discussion of DIS off protons, the results of numerical solutions of the CD BFKL equation for the all-flavor ($u + d + s + c$) eigen-structure functions $f_m(Q^2)$ were parameterized as

$$f_0(Q^2) = a_0 \frac{R_0^2 Q^2}{1 + R_0^2 Q^2}[1 + c_0 \ln(1 + r_0^2 Q^2)]^{\gamma_0}, \quad (23)$$

$$f_m(Q^2) = a_m f_0(Q^2) \frac{1 + R_0^2 Q^2}{1 + R_m^2 Q^2} \prod_{i=1}^{m} \left(1 - \frac{z}{z_m^{(i)}}\right), \quad (24)$$

$$m \geq 1,$$

where $\gamma_0 = 4/3\Delta_0$ and

$$z = [1 + c_m \ln(1 + r_m^2 Q^2)]^{\gamma_m} - 1, \quad \gamma_m = \gamma_0 \delta_m. \quad (25)$$

The parameters tuned to reproduce the numerical results for $f_m(Q^2)$ at $Q^2 \lesssim 10^5$ GeV^2 are listed in Table 2.

The soft component of the proton structure function as derived from $\sigma_{\rm soft}(r)$ taken from [27] is parameterized as

$$f_{\rm soft}(Q^2) = \frac{a_{\rm soft} R_{\rm soft}^2 Q^2}{1 + R_{\rm soft}^2 Q^2}[1 + c_{\rm soft} \ln(1 + r_{\rm soft}^2 Q^2)], \quad (26)$$

with the parameters given in Table 2.

The cross section $\sigma_{\rm soft}^{\gamma^*\gamma^*}(Q^2, P^2)$ obtained by the continuation of the above,

$$\sigma_{\rm soft}^{\gamma^* p} = \frac{4\pi^2 \alpha_{em}}{Q^2} f_{\rm soft}(Q^2),$$

into the Q^2, P^2-plane reads

$$\sigma_{\rm soft}^{\gamma^*\gamma^*}(Q^2, P^2) = \frac{\sigma_{\rm soft}^{\gamma\gamma}}{1 + R_{\rm soft}^2(Q^2 + P^2)} \times \left[1 + c_{\rm soft} \ln\left(1 + \frac{r_{\rm soft}^2 Q^2}{1 + r_{\rm soft}^2 P^2} + \frac{r_{\rm soft}^2 P^2}{1 + r_{\rm soft}^2 Q^2}\right)\right], \quad (27)$$

Table 3. CD BFKL–Regge charm structure function parameters

m	a_m	c_m	r_m^2, GeV^{-2}	R_m^2, GeV^{-2}	K_m^2, GeV^{-2}	$z_m^{(1)}$	$z_m^{(2)}$	δ_m
0	0.02140	0.2619	0.3239	0.2846				1.
1	0.0782	0.03517	0.0793	0.2958	0.2846	0.2499		1.9249
2	0.00438	0.03625	0.0884	0.2896	0.2846	0.0175	3.447	1.7985
3	−0.26313	2.1431	3.7424×10^{-2}	8.1639×10^{-2}	0.13087	158.52	559.50	0.62563
soft	0.01105	0.3044	0.09145	0.1303				

with the parameters given in Table 2 and the on-shell cross section

$$\sigma_{\text{soft}}^{\gamma\gamma} = \frac{(4\pi^2\alpha_{em}a_{\text{soft}}R_{\text{soft}}^2)^2}{\sigma_{\text{soft}}^{pp}}. \quad (28)$$

A.2. CD BFKL Charm Eigen-Structure Function

In practical evaluations, one needs the charm eigenstructure function $f_m^c(Q^2)$. For the rightmost hard BFKL pole, it is given by

$$f_0^2(Q^2) = a_0\frac{R_0^2Q^2}{1+R_0^2Q^2}[1+c_0\ln(1+r_0^2Q^2)]^{\gamma_0}, \quad (29)$$

where $\gamma_0 = 4/3\Delta_0$, while for the subleading hard BFKL poles we have

$$f_m^c(Q^2) = a_m f_0(Q^2)\frac{1+K_m^2Q^2}{1+R_m^2Q^2}\prod_{i=1}^{m_{\max}}\left(1-\frac{z}{z_m^{(i)}}\right), \quad (30)$$

$$m \geq 1,$$

where $m_{\max} = \min\{m, 2\}$ and

$$z = [1+c_m\ln(1+r_m^2Q^2)]^{\gamma_m} - 1, \quad \gamma_m = \gamma_0\delta_m. \quad (31)$$

The parameters tuned to reproduce the numerical results for $f_m^c(Q^2)$ at $(Q^2) \lesssim 10^5$ GeV^2 are listed in Table 3.

The soft component of the charm structure function is parameterized as

$$f_{\text{soft}}^c(Q^2) = \frac{a_{\text{soft}}R_{\text{soft}}^2Q^2}{1+R_{\text{soft}}^2Q^2}[1+c_{\text{soft}}\ln(1+r_{\text{soft}}^2Q^2)], \quad (32)$$

with parameters given by in Table 3.

ACKNOWLEDGMENTS

This work was supported in part by the grants INTAS-96-597, INTAS-97-30494, and DFG 436RUS17/11/99.

REFERENCES

1. R. Nisius, Phys. Rep. **332**, 165 (2000).
2. J. Bartels, A. De Roeck, and H. Lotter, Phys. Lett. B **389**, 742 (1996).
3. J. Bartels, C. Ewerz, and R. Staritzbichler, hep-ph/0004029.
4. S. J. Brodsky, F. Hautmann, and D. E. Soper, Phys. Rev. D **56**, 6957 (1997).
5. M. Boonekamp, A. De Roek, Ch. Royon, and S. Wallon, Nucl. Phys. B **555**, 540 (1999).
6. M. Acciarri *et al.* (L3 Collab.), Phys. Lett. B **436**, 403 (1998).
7. G. Abbiendi *et al.* (OPAL Collab.), Eur. Phys. J. C **14**, 51 (2000); **14**, 73 (2000); **14**, 187 (2000); **14**, 199 (2000); **14**, 373 (2000).
8. M. Acciarri *et al.* (L3 Collab.), Phys. Lett. B **453**, 333 (1999).
9. K. Ackerstaff *et al.* (OPAL Collab.), Phys. Lett. B **412**, 225 (1997).
10. Particle Data Group, Eur. Phys. J. C **3**, 1 (1998).
11. V. M. Budnev, I. F. Ginzburg, G. V. Meledin, and V. G. Serbo, Phys. Rep. **15**, 181 (1975).
12. V. S. Fadin, E. A. Kuraev, and L. N. Lipatov, Phys. Lett. B **60**, 50 (1975); E. A. Kuraev, L. N. Lipatov, and V. S. Fadin, Zh. Éksp. Teor. Fiz. **71**, 840 (1976) [Sov. Phys. JETP **44**, 443 (1976)]; Zh. Éksp. Teor. Fiz. **72**, 377 (1977) [Sov. Phys. JETP **45**, 199 (1977)].
13. L. N. Lipatov, Zh. Éksp. Teor. Fiz. **90**, 1536 (1986) [Sov. Phys. JETP **63**, 904 (1986)].
14. Ya. Ya. Balitsky and L. N. Lipatov, Yad. Fiz. **28**, 1597 (1978) [Sov. J. Nucl. Phys. **28**, 822 (1978)].
15. V. N. Gribov, B. L. Ioffe, I. Ya. Pomeranchuk, and A. P. Rudik, Zh. Éksp. Teor. Fiz. **42**, 1419 (1962) [Sov. Phys. JETP **16**, 220 (1963)].
16. N. N. Nikolaev, B. G. Zakharov, and V. R. Zoller, Pis'ma Zh. Éksp. Teor. Fiz. **66**, 134 (1997) [JETP Lett. **66**, 138 (1997)].
17. N. N. Nikolaev and V. R. Zoller, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 91 (1999) [JETP Lett. **69**, 103 (1999)]; Pis'ma Zh. Éksp. Teor. Fiz. **69**, 176 (1999) [JETP Lett. **69**, 187 (1999)].
18. N. N. Nikolaev, J. Speth, and V. R. Zoller, Phys. Lett. B **473**, 157 (2000).

19. N. N. Nikolaev, B. G. Zakharov, and V. R. Zoller, Pis'ma Zh. Éksp. Teor. Fiz. **59**, 8 (1994) [JETP Lett. **59**, 6 (1994)].
20. N. N. Nikolaev, B. G. Zakharov, and V. R. Zoller, Zh. Éksp. Teor. Fiz. **105**, 1498 (1994) [JETP **78**, 806 (1994)].
21. N. N. Nikolaev and B. G. Zakharov, Phys. Lett. B **327**, 149 (1994); **333**, 250 (1994); **327**, 157 (1994).
22. E. M. Levin and M. G. Ryskin, Yad. Fiz. **34**, 1154 (1981) [Sov. J. Nucl. Phys. **34**, 619 (1981)].
23. O. Nachtmann, Ann. Phys. **209**, 436 (1991); H. Dosch and Yu. A. Simonov, Phys. Lett. B **205**, 339 (1988); Yu. A. Simonov, Nucl. Phys. B **324**, 67 (1989); H. Dosch, E. Ferreira, and A. Krämer, Phys. Rev. D **50**, 1992 (1994); P. V. Landshoff and O. Nachtmann, Z. Phys. C **35**, 405 (1987); H. G. Dosch, T. Gousset, G. Kulzinger, and H. J. Pirner, Phys. Rev. D **55**, 2602 (1997); A. Donnachie and P. V. Landshoff, Phys. Lett. B **437**, 408 (1998); K. Golec-Biernat and M. Wüsthoff, Phys. Rev. D **59**, 014017 (1999); U. D'Alesio, A. Metze, and H. J. Pirner, Eur. Phys. J. C **9**, 601 (1999); G. Kulzinger, H. G. Dosch, and H. J. Pirner, Eur. Phys. J. C **7**, 73 (1999).
24. T. F. Walsh and P. M. Zerwas, Phys. Lett. B **44**, 195 (1973); R. L. Kingsley, Nucl. Phys. B **60**, 45 (1973).
25. G. A. Schuler and T. Sjöstrand, Phys. Lett. B **376**, 193 (1996).
26. N. N. Nikolaev and B. G. Zakharov, Z. Phys. C **49**, 607 (1991).
27. J. Nemchik, N. N. Nikolaev, E. Predazzi, *et al.*, Zh. Éksp. Teor. Fiz. **113**, 1930 (1998) [JETP **86**, 1054 (1998)].

Journal of Experimental and Theoretical Physics, Vol. 93, No. 5, 2001, pp. 969–976.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 5, 2001, pp. 1117–1125.

ATOMS, SPECTRA, RADIATION

On the Theory of Superradiant Rayleigh Scattering from a Bose–Einstein Condensate

E. D. Trifonov
Gertsen State Pedagogical University, St. Petersburg, 191186 Russia
e-mail: thphys@herzen.spb.ru
Received November 27, 2000

Abstract—A theory of the superradiant Rayleigh scattering by a Bose–Einstein condensate is suggested. The theory is based on a semiclassical approach. Atomic states with definite momenta are used as basis functions. The Maxwell–Bloch equations are derived and solved to describe the intensity of scattered radiation and the evolution of the population of coherent atomic states. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Superradiance predicted by Dicke [1] in 1953 (also see books [2–4] and the references therein) has recently become the focus of interest of researchers in view of the possibility of creating Bose–Einstein condensates by laser and evaporative cooling [5–8]. First and foremost, this refers to the recent experiments described in [9]. In these experiments, superradiant Rayleigh scattering by sodium atoms in the Bose–Einstein condensation state was observed. Under certain excitation conditions, scattered radiation was narrowly directed along the maximum condensate extension and was much stronger than usual (noncooperative) scattering. Recall that superradiant scattering appears under the conditions of phase memory retention by the atomic system [10–12]. For this reason, the Bose–Einstein condensation state, which is a coherent state of atomic systems, is an ideal object for studying cooperative effects. The distinguishing feature of cooperative effects is the impossibility in principle of separating the process into elementary events (absorption, radiation, scattering, etc. events) that occur in different atoms. This distinguishes superradiance not only from spontaneous radiation but also from laser generation or induced scattering.

The quantum-electrodynamic approach to the problem of superradiant scattering by a Bose–Einstein condensate was described in [13]. In this work, we suggest a simpler model of the phenomenon based on a semiclassical approach. We derive the Maxwell–Bloch equations for the system under consideration. The obtained solutions are in qualitative agreement with the experimental results reported in [9].

2. REMARKS ON EXPERIMENT

Theoretically estimating the time dependence of the intensity of superradiant scattering and the formation of coherent atomic state waves requires the use of the characteristic parameters that determine the conditions of experiments [9]. We will therefore shortly describe these experiments.

The Bose–Einstein condensate of sodium atoms was prepared in a magnetic trap by laser cooling. The sodium atoms were in the ground superfine structure state $F = 1$, $3S_{1/2}$. The magnetic trap had an extended form; it was 20 μm in diameter and 200 μm long. The number of atoms in the trap was $N \sim 10^7$. The sample was irradiated by a laser pulse whose frequency was shifted by 1.7 GHz to the red from the $3S_{1/2}$ ($F = 1$) $\longrightarrow$ $3P_{3/2}$ ($F = 0, 1, 2$) transition frequency. The wavelength of this transition was $\lambda = 589$ nm. The laser beam crossed the condensate normally to its axis. A typical laser radiation intensity was 1–100 mW/cm^2, and the pulse width was 10–800 μs.

Not only the intensity of the scattered radiation but also the distribution of the momenta of atoms which they acquired in interactions with laser and scattered photons was studied. For this purpose, the magnetic trap was switched off immediately after laser irradiation, and the spatial positions of the atoms were photographed by a resonance probe beam in 20–50 ms.

The authors noted that the effect depended on laser beam polarization. If the beam was polarized parallel to the sample axis, motions of atoms corresponded to recoil momenta under usual (incoherent) Rayleigh scattering with a $\sin^2\theta$ radiation pattern, where θ is the angle made by the polarization and observation directions. However, if laser beam polarization was perpendicular to the sample axis, intense scattering along the sample was observed. An atom participating in scattering obtained the laser photon momentum directed along the laser beam; when this atom emitted a scattered photon, it obtained the same momentum, however, directed along the sample. The angle between the resultant atomic momentum and the sample axis therefore equaled 45°. Some atoms succeeded in participat-

ing in subsequent scattering events and, accordingly, obtained additional momenta. Switching the trap off allowed further motions of atoms determined by the momenta that they obtained to be traced (see Fig. 1a). Apart from the dependence of scattering on laser beam polarization, the authors note that there exists an intensity threshold below which only usual spontaneous scattering is observed. Such scattering pulses have the intensity proportional to that of the excitation pulses.

In conclusion, consider the values of some parameters that characterize the conditions of these experiments. The concentration of atoms in the Bose–Einstein condensate was $N_0 = 10^{14}$ cm^{-3}. The electric field amplitude of the laser beam was $E_0 = 1$–10 V/cm. The amplitude was estimated by the formula

$$E[\text{V/cm}] = 27.46\sqrt{I_0[\text{W/cm}^2]},$$

which follows from $I_0 = cE_0^2/8\pi$. The transition dipole moment was $d \approx 10^{-17}$cgs units. The detuning was $\Delta = 2\pi 1.7$ GHz $\approx 10^{10}$ s^{-1}. The radiation width of the transition was

$$\gamma = \frac{4}{3}\frac{d^2(2\pi)^3}{\hbar\lambda^3} \approx 10^8 \text{ s}^{-1}.$$

Further, we use the time scale parameter called superradiance time,

$$\tau_R = \frac{2\pi}{\gamma N_0 \lambda^2 L},$$

where L is the magnetic trap length. Under the experimental conditions studied, $\tau_R \approx 10^{-11}$ s.

3. INCOHERENT RAYLEIGH SCATTERING

Incoherent Rayleigh scattering is scattering by separate atoms, and its intensity is an additive value. In gases, incoherent scattering arises because of local molecular concentration fluctuations. Incoherent scattering by Bose–Einstein condensates is caused by quantum fluctuations of the particle density distribution. The effects of atomic scattering fields on other atoms are ignored. Recall that the coherent part of Rayleigh scattering is scattering forward. An atom participating in coherent scattering does not change its momentum and remains in the Bose–Einstein condensation state.

Clearly, incoherent scattering by a system of atoms reduces to the problem of scattering by a single atom. The incident field induces a dipole moment of the atom. The intensity of radiation (radiation energy per unit time) of an oscillating dipole with amplitude P_0 and frequency ω is

$$W = \frac{P_0^2\omega^4}{3c^3}, \tag{1}$$

where c is the velocity of light in the vacuum. Let us estimate dipole moment P_0. The atom is treated as a two-level system (as in quasi-resonance scattering). The equation for the density matrix of the atom ρ has the following form:

$$i\hbar\dot{\rho} = [\hat{H}\rho]. \tag{2}$$

After separating rapidly oscillating multipliers, the density matrix amplitudes R_{ik} are obtained in the rotating wave approximation in the form

$$\begin{aligned} i\hbar\dot{R}_{aa} &= i\hbar\gamma R_{bb} + H_{ab}R_{ba} - R_{ab}H_{ba}, \\ i\hbar\dot{R}_{bb} &= -i\hbar\gamma R_{bb} + H_{ba}R_{ab} - R_{ba}H_{ab}, \\ i\hbar\dot{R}_{ab} &= \hbar\left(-i\hbar\frac{\gamma}{2} + \Delta\right)R_{ab} + H_{ab}R_{bb} - R_{aa}H_{bb}. \end{aligned} \tag{3}$$

Here, $H_{ab} = -dE_0$ is the off-diagonal matrix element of the interaction operator with the incident field and indices a and b refer to the ground and excited atom states, respectively. Under stationary conditions ($\dot{R}_{ik} = 0$), the last equation yields

$$\hbar\left(\Delta - i\frac{\gamma}{2}\right)R_{ab} - dE_0(R_{bb} - R_{aa}) = 0. \tag{4}$$

Ignoring population changes ($R_{aa} = 1$, $R_{bb} = 0$) and assuming that $\Delta \gg \gamma$, we obtain

$$R_{ab} = \frac{dE_0}{\Delta\hbar}. \tag{5}$$

Therefore,

$$P_0 = d|R_{ab}| = \frac{d^2E_0}{\Delta\hbar}. \tag{6}$$

This gives the mean intensity of light scattering by one atom in the form

$$W_s = \frac{1}{3}\left(\frac{d^2E_0}{\Delta\hbar}\right)^2\frac{\omega^4}{c^3}. \tag{7}$$

The $\tilde{W}_s = W_s/\hbar\omega$ value is called the rate of Rayleigh scattering. As the flux density of incident radiation is

$$I_0 = \frac{cE_0^2}{8\pi},$$

we have

$$\tilde{W}_s = \frac{8\pi}{3}\frac{d^4\omega^3}{\hbar^3c^4\Delta^2}I_0. \tag{8}$$

According to the authors of the experiments described in [9], Rayleigh scattering rates corresponding to 1–100 mW/cm^2 laser radiation intensities are 4.5×10–4.5×10^3 s^{-1}, which is in agreement with (8) with the parameter values specified in the end of Section 2.

To conclude this section, consider the equation for the number of photons scattered over time τ_R,

$$\tilde{W}_s\tau_R = \frac{P_0^2\omega^3}{3\hbar c^3}\frac{2\pi}{\gamma N_0\lambda^2 L} = \frac{2\pi|R_{ab}|^2}{N_0\lambda^2 L}. \quad (9)$$

According to the estimates made in the end of Section 2, we have

$$\frac{2\pi}{N_0\lambda^2 L} \sim 10^{-3}.$$

4. COOPERATIVE (SUPERRADIANT) LIGHT SCATTERING

Generally, light scattering cannot be treated as an atomically additive process, in particular, when atomic scattering field effects on other atoms are taken into account. Under the Bose–Einstein condensate conditions, these effects take place with phase memory retention (coherence) of atomic states and are therefore superradiant in character. A more general statement of the problem which implies considering light scattering by the multiatomic system as a whole is required to describe the observed effect. We assume that, initially, all atoms occur in the same electronic state with an almost zero translational motion momentum. Zero-point vibrations of atoms in a magnetic harmonic trap will be ignored. Incident laser radiation is in quasi-resonance with one of the atom electronic transitions. Each atom is therefore treated as a two-level electronic system. An atom will be characterized not only by its electronic state but also by the translational motion state. Interactions between atoms and radiation obey the law of conservation of momentum. Formally, this is ensured by the selection rules. The matrix element of the interaction between an atom (in the state with a certain momentum) and electromagnetic wave with wave vector $\mathbf{k}$ is nonzero for the final state of the atom whose momentum differs from the initial momentum value by $\pm\hbar\mathbf{k}$. The electronic states of the atom (ground and excited) will be denoted by indices a and b, ϕ_a and ϕ_b. Translational motion is described by momentum eigenfunctions. The wave functions of the atom are therefore written in the form

$$|a; \mathbf{k}\rangle = \frac{1}{\sqrt{V}}\exp(i\mathbf{k}\cdot\mathbf{r})\phi_a,$$

$$|b; \mathbf{k}'\rangle = \frac{1}{\sqrt{V}}\exp(i\mathbf{k}'\cdot\mathbf{r})\phi_b,$$

where V is the system volume. In particular, the wave function of the initial state is written as $|a; 0\rangle$, and the wave function of the excited state, which arises after absorption of one laser photon, is written as $|b; \mathbf{k}_0\rangle$. If the transition from this excited to the ground state occurs with emitting momentum $\mathbf{k}$, the new state is $|a; \mathbf{k}_0 - \mathbf{k}\rangle$.

In conformity with the experimental conditions (see Section 2), we assume that the exciting laser beam is directed normally to the condensate axis, and that photon scattering occurs along this axis. It follows that specifying the momentum of an atom requires knowledge of two momentum components. Let the laser beam direction be axis x and the condensate extension direction be y. The system of atomic states necessary for describing scattering then has the form $|a; 0, 0\rangle$, $|a; k_0, \pm k_0\rangle$, $|a; 2k_0, 0\rangle$, $|b; k_0, 0\rangle$, $|a; 2k_0, \pm k_0\rangle$, $|a; 2k_0, \pm 2k_0\rangle$, and so on. We will only consider scattering events not higher than twofold and number the above states by one index as follows:

$$\begin{aligned}
\psi_1 &= |a; 0, 0\rangle,\\
\psi_2 &= |a; k_0, -k_0\rangle,\\
\psi_3 &= |b; k_0, 0\rangle,\\
\psi_4 &= |a; k_0, k_0\rangle,\\
\psi_5 &= |b; 2k_0, -k_0\rangle,\\
\psi_6 &= |b; 2k_0, k_0\rangle,\\
\psi_7 &= |a; -2k_0, -2k_0\rangle,\\
\psi_8 &= |a; 2k_0, 0\rangle,\\
\psi_9 &= |a; 2k_0, 2k_0\rangle.
\end{aligned} \quad (10)$$

Transitions between these states are shown in the diagram, Fig. 1b.

We will consider interactions between these states and two fields, namely, the external laser field and the field scattered along the condensate. The latter field is described by two waves propagating in the opposite directions. These fields will be denoted by E_L, E_+, and E_-. Clearly, interaction with field E_L connects the states $1 \longleftrightarrow 3$, $2 \longleftrightarrow 5$, and $4 \longleftrightarrow 6$. Interaction with field E_- connects the states $3 \longleftrightarrow 2$, $5 \longleftrightarrow 7$, and $6 \longleftrightarrow 8$. Lastly, interaction with field E_+ connects the states $3 \longleftrightarrow 4$, $5 \longleftrightarrow 8$, and $6 \longleftrightarrow 9$.

Fields E_- and E_+ are caused by medium polarization. In the one-dimensional case, the field can be written through the polarization as (e.g., see [4])

$$E(y, t) = \frac{2\pi i\omega_0}{c}\int_0^L P\left(y', t - \frac{|y - y'|}{c}\right) \times \exp(ik_0|y - y'|)\exp(-i\omega_0 t)dy' + \text{c.c.}, \quad (11)$$

where ω_0 is the laser radiation frequency, $P(y', t)$ is the complex amplitude of medium polarization, and $k_0 = \omega_0/c$. This polarization amplitude is in turn expressed through off-diagonal density matrix elements as a quantum-mechanical average.

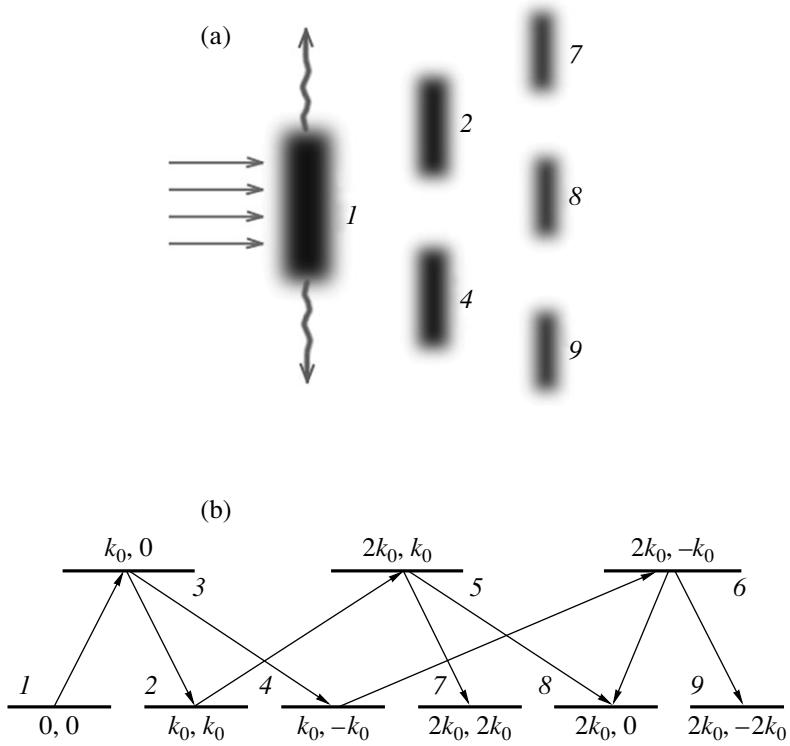


Fig. 1. (a) Scheme of formation of coherent atomic states with nonzero momenta as a result of superradiant light scattering. The original photograph of these coherent atomic "clouds" was given in [9]. (b) Scheme of transitions between coherent atomic states during superradiant Rayleigh scattering.

Let $\rho_{b,k;a,k'}$ be an arbitrary off-diagonal element of the density matrix of an atom. (We only specify the y component of the wave vector in density matrix indices.) The partial complex polarization amplitude P' determined by this element is found from the equality

$$P'(y', t) = \exp(i\omega_0 t) N P_{a,k';b,k}(y', t)\rho_{b,k;a,k'}. \quad (12)$$

Here, N is the total number of atoms in the system and $P_{a,k';b,k}(y', t)$ is the matrix element of the polarization operator (referred to one atom) $\hat{d}\,\delta(\mathbf{r}' - \mathbf{r})$,

$$P_{a,k';b,k}(y', t) = \exp(-i\omega_{ba} t)\frac{d}{V} \times \int \exp(-ik'y')\delta(\mathbf{r}' - \mathbf{r})\exp(iky')d\mathbf{r} = \frac{d}{V}\exp(-i\omega_{ba} t)\exp(i(k - k')y'), \quad (13)$$

where ω_{ba} is the transition frequency and $\hat{d}$ is the dipole moment operator. Further, let us introduce the amplitudes of density matrix elements,

$$R_{b,k;a,k'} = \exp(i(\omega_0 - \omega_{ba})t)\rho_{b,k;a,k'}. \quad (14)$$

We then have

$$P'(y', t) = dN_0 \exp(i(k - k')y')R_{b,k;a,k'}. \quad (15)$$

The field caused by this polarization has the form

$$E'(y, t) = \frac{2\pi i\omega_0}{c}\exp(-i\omega_0 t) \times \int_0^L P'(y', t)\exp(ik_0|y - y'|)dy' + \text{c.c.} \quad (16)$$

Ignoring the delay of the polarization amplitude and rapidly oscillating terms in the integration, we obtain

$$E'(y, t) = i\frac{2\pi\omega d N_0 y}{c}R_{b,k;a,k'} \times \exp(-i\omega_0 t)\exp(ik_0 y) + \text{c.c.} \quad (17)$$

We find that the amplitude of the field created by uniform polarization is linear with respect to the y coordinate. Further, we use the mean field approximation and set $y = L/2$. In addition, we will use the $\hbar/\tau_R d$ value as the electric field strength unit. This gives

$$E'(y, t) = iR_{b,k';a,k}\exp(ik_0 y)\exp(-i\omega_0 t) + \text{c.c.}, \quad k' - k = k_0. \quad (18)$$

The interaction operator between an atom and this field has the form

$$\hat{H}' = -\hat{d}E'. \quad (19)$$

Ignoring rapidly oscillating terms in the matrix elements of this operator, we obtain

$$H'_{b,k';a,k} = -iR_{b,k';a,k}, \quad H'_{a,k;b,k'} = iR_{a,k;b,k'}. \quad (20)$$

We considered partial field E' caused by the $R_{b,k';a,k}$ polarization amplitude. The total $E_\pm$ fields can be written as

$$\begin{aligned} E_+ &= iE_+^0 \exp(ik_0 y)\exp(-i\omega_0 t) + \text{c.c.},\\ E_- &= iE_-^0 \exp(-ik_0 y)\exp(-i\omega_0 t) + \text{c.c.}, \end{aligned} \quad (21)$$

where

$$\begin{aligned} E_+^0 &= \sum_{k'-k=k_0} R_{b,k';a,k},\\ E_-^0 &= \sum_{k'-k=-k_0} R_{b,k';a,k}. \end{aligned} \quad (22)$$

The interaction operator matrix elements in the approximation that we use therefore take the form

$$\begin{aligned} H_{b,k';a,k} &= -i\sum_{\tilde{k}'-\tilde{k}=k_0} R_{b,\tilde{k}';a,\tilde{k}},\\ H_{a,k;b,k'} &= i\sum_{\tilde{k}'-\tilde{k}=-k_0} R_{a,\tilde{k};b,\tilde{k}'}. \end{aligned} \quad (23)$$

Recall that the squares of the moduli of the amplitudes of these fields have the meaning of the numbers of photons (per atom) emitted through sample ends in time τ_R (e.g., see [4]).

We will not write the equation for the density matrix in the general form and will only describe the algorithm for constructing this equation. Above, we determined the form of the matrix of the interaction operator with the laser field and the fields created by polarization waves propagating along the sample. We introduced new amplitudes for the density matrix [Eq. (14)], and the equations for the off-diagonal elements (over electronic states) should therefore contain terms with detuning,

$$\dot{R}_{b,k';a,k} = i(\omega_0 - \omega_{ba})R_{b,k';a,k} + \ldots. \quad (24)$$

Apart from interactions with fields E_L and $E_\pm$, there is incoherent scattering in an arbitrary direction considered in Section 3. Incoherent scattering is not only an alternative process but also an initiating factor for superradiant scattering. Superradiant scattering arises as a consequence of quantum fluctuations of the field propagating along the sample. In our model, such a fluctuation is created as a result of incoherent (spontaneous) scattering. Let us estimate the intensity (the number of photons per time τ_R) of spontaneous scattering (by one atom) in the body angle of the sample. According to (9), we have

$$I_{\text{incoh}} = \frac{2\pi|R_{a,b}|^2}{N_0\lambda^2 L}\frac{D^2}{L^2}. \quad (25)$$

It follows that the amplitude of the corresponding fluctuation field is

$$E_{fl} = |R_{a,b}|s, \quad (26)$$

where

$$s = \left(\frac{2\pi}{N_0\lambda^2 L}\right)^{1/2}\frac{D}{L}. \quad (27)$$

In the equation for the density matrix, this fluctuation field is added to fields $E_\pm$.

The interaction operator with the laser field has the form $-\hat{d}E_L$. To exclude imaginary unit, its is expedient to select the laser wave field in the form $E_L = iE$ in the equation for the density matrix.

Spontaneous scattering is taken into account by introducing relaxation terms into the equations for the off-diagonal density matrix elements corresponding to radiative transitions and into the equations for the corresponding diagonal elements.

The algorithm described above will be used to obtain equations for the density matrix that take into account single scattering events only. This requires considering the states $\psi_1 = |a; 0, 0\rangle$, $\psi_2 = |a; k_0, -k_0\rangle$, $\psi_3 = |b; k_0, 0\rangle$, and $\psi_4 = |a; k_0, k_0\rangle$.

If the superradiance time is selected as the time unit, the equation for the density matrix takes the form

$$\begin{aligned} \dot{R}_{11} &= -E(R_{13} + \bar{R}_{13}),\\ \dot{R}_{22} &= E_+^0 R_{23} + \bar{E}_+^0 \bar{R}_{23},\\ \dot{R}_{33} &= E(R_{13} + \bar{R}_{13}) - E_+^0 R_{23} - E_-^0 \bar{R}_{34} - \bar{E}_-^0 R_{34} - \gamma R_{33},\\ \dot{R}_{44} &= \bar{E}_-^0 R_{34} + E_-^0 \bar{R}_{34},\\ \dot{R}_{12} &= -E\bar{R}_{23} + E_+^0 R_{13},\\ \dot{R}_{13} &= E(R_{11} - R_{33}) - \bar{E}_+^0 R_{12} - \bar{E}_-^0 R_{14} + \left(i\Delta - \frac{\gamma}{2}\right)R_{13},\\ \dot{R}_{14} &= -ER_{34} + E_-^0 R_{13},\\ \dot{R}_{23} &= E\bar{R}_{12} + \bar{E}_+^0(R_{33} - R_{22}) - \bar{E}_-^0 R_{24} + \left(i\Delta - \frac{\gamma}{2}\right)R_{23},\\ \dot{R}_{24} &= \bar{E}_+^0 R_{34} + E_-^0 R_{23},\\ \dot{R}_{34} &= ER_{14} + E_-^0(R_{33} - R_{44}) - E_+^0 R_{24} + \left(-i\Delta - \frac{\gamma}{2}\right)R_{34}, \end{aligned} \quad (28)$$

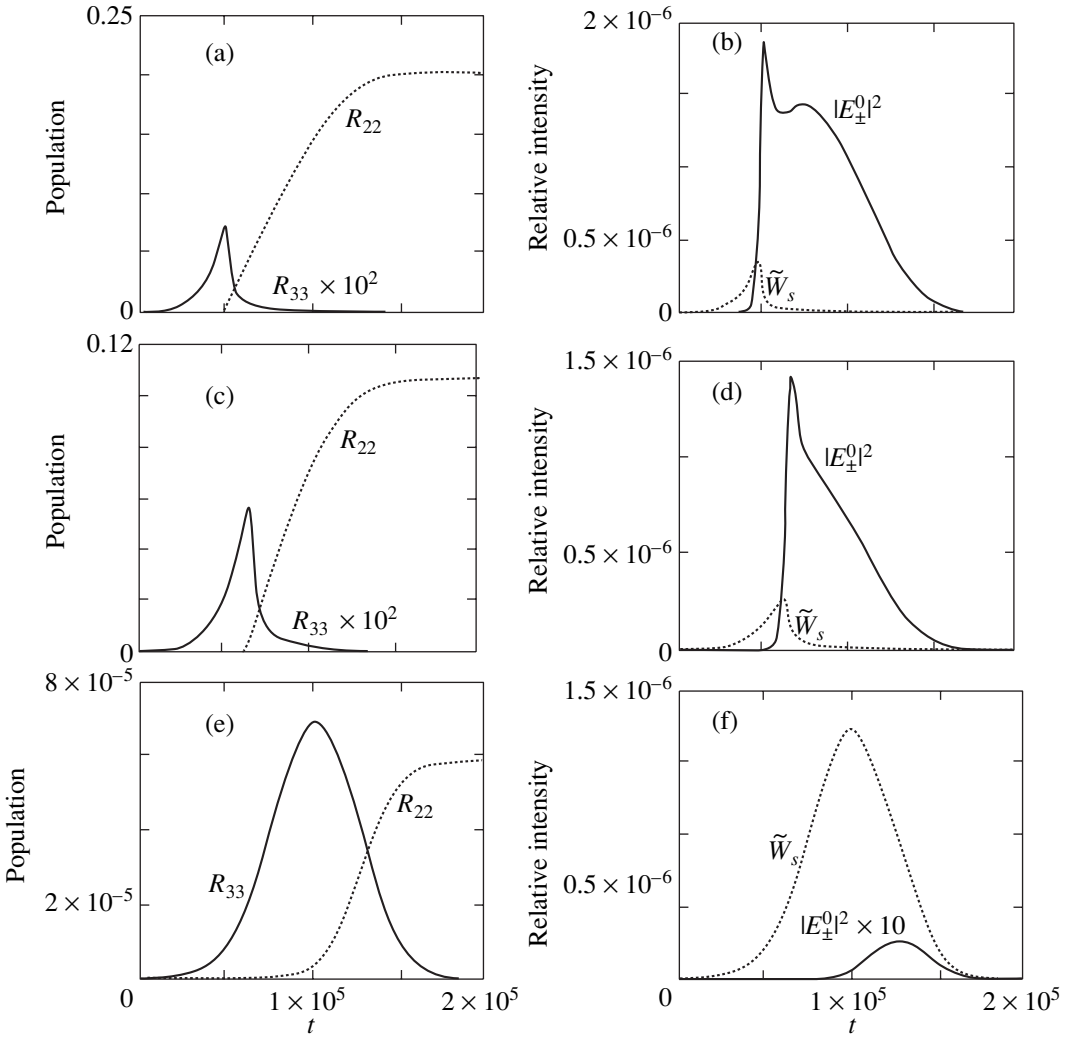


Fig. 2. (b, d, f) Pulses of light scattered along the condensate and (a, c, e) time dependence of the population of atomic states at various laser beam intensities for the model of single scattering events; E = (a, b) 10^{-3}, (c, d) 5×10^{-3}, and (e, f) 10^{-4}. Fields E are given in $\hbar/d\tau_R$ units. The rate of incoherent Rayleigh scattering is shown by the dotted lines in (b, d, f). Time is measured in τ_R units.

where E^0_+ and E^0_- are the amplitudes of fields (22),

$$E^0_+ = \bar{R}_{23} + R_{13}s, \quad E^0_- = R_{34} + R_{13}s, \tag{29}$$

and $\bar{E}$ and $\bar{R}$ are the complex conjugate values.

A similar equation for the density matrix that takes into account recurring scattering events is fairly cumbersome. Note that if we do not aim at accurately describing relaxation processes, we may replace the von Neumann equation by the corresponding Schrödinger equation, which has a more compact form,

$$\begin{aligned}
\dot{C}_1 &= -EC_3,\\
\dot{C}_2 &= \bar{E}^0_+ C_3 - EC_5,\\
\dot{C}_3 &= \left(i\Delta - \frac{\gamma}{2}\right)C_3 - \bar{E}^0_+ C_2 - E^0_- C_4 + EC_1,\\
\dot{C}_4 &= \bar{E}^0_- C_3 - EC_6,\\
\dot{C}_5 &= \left(i\Delta - \frac{\gamma}{2}\right)C_5 - E^0_+ C_7 - E^0_- C_8 + EC_2,\\
\dot{C}_6 &= \left(i\Delta - \frac{\gamma}{2}\right)C_6 - E^0_+ C_8 - E^0_- C_9 + EC_4,\\
\dot{C}_7 &= \bar{E}^0_+ C_5,\\
\dot{C}_8 &= \bar{E}^0_+ C_6 + \bar{E}^0_- C_5,\\
\dot{C}_9 &= \bar{E}^0_- C_6.
\end{aligned} \tag{30}$$

Here, C_i are the coefficients of the expansion of the wave function in basis functions (10) selected above, and

$$\begin{aligned}
E^0_+ &= \bar{C}_2 C_3 + C_5 \bar{C}_7 + C_6 \bar{C}_8 + C_1 \bar{C}_3 s,\\
E^0_- &= \bar{C}_4 C_3 + C_5 \bar{C}_8 + C_6 \bar{C}_9 + \bar{C}_1 C_3 s
\end{aligned} \tag{31}$$

in conformity with (22). For the problem stated above, the only nonzero initial condition is $R_{11}(0) = 1$ or $C_1(0) = 1$.

We use superradiance time τ_R as time unit. Under the conditions of experiments performed in [9], this time is of the order of 10^{-11} s. At the same time, the exciting pulse width is of the order of 1 ms. The time scale might be selected commensurate with the pulse width. How would the equations change as a result of such a time scale selection? Clearly, if new time $t' = \alpha t$ and new parameters $E' = \alpha^{-1}E$, $\Delta' = \alpha^{-1}\Delta$, and $\gamma' = \alpha^{-1}\gamma$ are used, the change of variables $R'_{i,k} = \alpha^{-1}R_{i,k}$ and $C'_i = \alpha^{-1/2}C_i$ leaves Eqs. (28) and (30) unaltered. We cannot, however, claim invariance of the solution with respect to such a transformation, because the initial condition then also changes, $R'_{i,k}(0) = \alpha^{-1}$ and $C'_i(0) = \alpha^{-1/2}C_i$.

5. COMPUTER SIMULATION OF SUPERRADIANT SCATTERING

In the end of Section 2, we gave the orders of parameter values that characterize the effect in the experiment [9]. Recall that we selected the time unit equal to superradiance time τ_R, which was of the order of 10^{-11} s under the experimental conditions. The radiation constant of the transition in (τ_R^{-1}) units is of the order of 10^{-3}, detuning $\Delta = 10^{-2}$, laser beam electric field strength (in $\hbar/\tau_R d$ units) $E = 10^{-4}$–10^{-3}, and pulse width $T_p = 10^7$. To guarantee accuracy of calculations, we had to decrease the detuning and the pulse width compared with their experimental values, namely, we set Δ equal to 2×10^{-3} and $T_p = 2 \times 10^5$. These limitations are not of a fundamental character and can be removed by using a more powerful computer or a more rational method for integrating equations. It was assumed that the laser pulse had a Gaussian form.

The results of solving system (28) are given in Fig. 2, where the form and intensity of the scattered pulse and the dynamics of population of coherent atomic states with momenta $|a; k_0, \pm k_0\rangle$ for various laser pulse intensities are shown. At a laser field amplitude of $E_0 = 10^{-3}$, the fraction of atoms in the states $|a; k_0, \pm k_0\rangle$ after switching laser irradiation off is about 0.2 for each of the two scattering directions. The form of the scattering pulse changes (time lag increases) as E_0 decreases, and, when $E_0 = 10^{-4}$ is attained, the intensity of the scattered pulse and the population of the coherent atomic wave sharply decrease (by four orders of magnitude). The $E_0 = 10^{-4}$ value can be considered a threshold of superradiant scattering. Usual coherent scattering, whose intensity is proportional to the intensity of the laser field, then remains (see Fig. 2c).

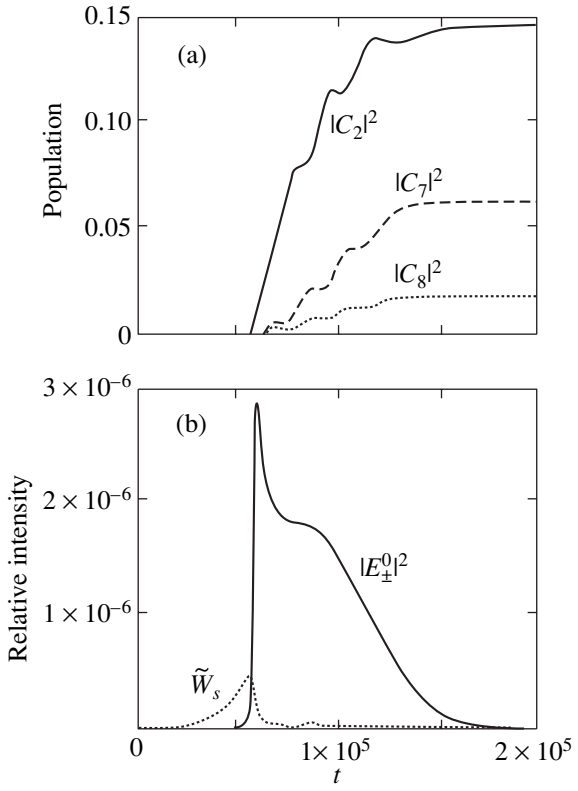


Fig. 3. (a) Time dependence of the population of coherent atomic states and (b) scattered light pulse shape for the model taking into account double scattering events, $E = 10^{-3}\hbar/d\tau_R$; dotted line in (a) shows the rate of incoherent Rayleigh scattering. Time is measured in τ_R units.

The results of solving system (30) for the occupancies of the $|a; k_0, \pm k_0\rangle$ and $|a; 2k_0, \pm 2k_0\rangle$ coherent atomic states calculated taking into account recurring scattering events are shown in Fig. 3. These calculations were performed for pulses $T_p = 10^5$ wide and detuning $\Delta = 10^{-3}$.

An analysis of these results shows that the effect depends on the area of the exciting laser pulse. At a fixed area, the effect (generation of coherent waves of atomic states) is stronger at a higher field and, accordingly, narrower laser pulses.

ACKNOWLEDGMENTS

I thank M.G. Benedikt, who called my attention to this problem, V.I. Perel', V.L. Gurevich, and V. Ketterle for stimulating interest in this work, and A.S. Troshin, I.V. Sokolov, A.I. Zaitsev, and D.S. Bulyanitsa for discussions and valuable comments.

This work was financially supported by the Ministry of Education of the Russian Federation and the Russian Foundation for Basic Research (project no. 01-02-17075).

REFERENCES

1. R. H. Dicke, Phys. Rev. **93**, 99 (1954).
2. L. Allen and J. H. Eberly, *Optical Resonance and Two-Level Atoms* (Wiley, New York, 1975; Mir, Moscow, 1978).
3. A. V. Andreev, V. I. Emel'yanov, and Yu. A. Il'inskiĭ, *Cooperative Effects in Optics* (Nauka, Moscow, 1988; Inst. of Physics Publ., Bristol, 1993).
4. M. G. Benedict, A. M. Ermolaev, V. A. Malyshev, I. V. Sokolov, and E. D. Trifonov, *Superradiance: Multiatomic Coherent Emission* (Inst. of Physics Publ., Bristol, 1996).
5. M. H. Anderson, J. R. Ensher, M. R. Matthews, *et al.*, Science **269**, 198 (1995).
6. K. B. Davis, M.-O. Mewes, M. R. Andrews, *et al.*, Phys. Rev. Lett. **75**, 3969 (1995).
7. C. C. Bradly, C. A. Sackett, and R. G. Hulet, Phys. Rev. Lett. **78**, 985 (1997).
8. D. G. Fried, T. C. Killian, L. Willmann, *et al.*, Phys. Rev. Lett. **81**, 3811 (1998).
9. S. Inouye, A. P. Chikkatur, D. M. Stamper-Kurn, *et al.*, Science **285**, 571 (1999).
10. S. G. Rautian and B. M. Chernobrod, Kvantovaya Élektron. (Moscow) **6**, 2645 (1979).
11. S. G. Rautian and B. M. Chernobrod, Zh. Éksp. Teor. Fiz. **78**, 1365 (1980) [Sov. Phys. JETP **51**, 687 (1980)].
12. E. D. Trifonov, A. S. Troshin, and N. I. Shamrov, Opt. Spektrosk. **48**, 1036 (1980) [Opt. Spectrosc. **48**, 567 (1980)].
13. M. G. Moore and P. Meystre, Phys. Rev. Lett. **83**, 5202 (1999).

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 93, No. 5, 2001, pp. 977–984.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 5, 2001, pp. 1126–1134.

ATOMS, SPECTRA, RADIATION

Localized Electromagnetic Modes and the Transmission Spectrum of a One-Dimensional Photonic Crystal with Lattice Defects

S. Ya. Vetrov[a, *] and A. V. Shabanov[b]

[a]*Krasnoyarsk State Technical University, Krasnoyarsk, 660074 Russia*
[b]*Kirenskii Institute of Physics, Siberian Division, Russian Academy of Sciences, Akademgorodok, Krasnoyarsk, 660036 Russia*
**e-mail: chery@escape.akadem.ru*

Received December 21, 2000

Abstract—The properties of localized electromagnetic modes in a one-dimensional photonic crystal with a structural defect layer were studied. The role of the defect was played by an anisotropic nematic liquid crystal layer. The frequency and the damping factor of defect modes were shown to substantially depend on the defect layer thickness and the orientation of the optical axis of the nematic. The transmission spectra of photonic crystals with one and two lattice defects were studied. Taking into account the special feature of nematic liquid crystals distinguishing them from usual crystals, namely, large permittivity anisotropy, it was shown that the transmission spectrum of the photonic crystal could be controlled by varying the orientation of the optical axis of the nematic, for instance, under the action of an external electric field. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In recent years, much attention has been given to photonic crystals [1], which are a special class of artificial structures with periodic dielectric property variations on a spatial scale of the order of the optical wave length. Depending on the periodicity dimension, photonic crystals are classified as one-, two-, and three-dimensional. The value of the concept of photonic structures lies in the possibility of studying physical phenomena from a new standpoint based on the traditional ideas of the physics of solids and electromagnetism. There is a close formal analogy between the theory of electromagnetic radiation propagation in periodic media and the quantum theory of electrons in crystals. The band structure of the energy spectrum of electrons in crystals caused by Bragg reflection of electrons is similar to the structure of the spectrum of a photonic crystal. Many interesting and potentially useful phenomena are related to the presence of photonic band gaps in photonic crystals and their unusual dispersion properties. Photonic crystalline structures make it possible to solve some fundamental problems, for instance, problems related to controlling spontaneous light emission from atoms and molecules [2–4] and localizing and channeling light [1, 5–11]. In practical applications, photonic band gap structures are extensively used in creating photonic crystal waveguides [8, 12, 13], superfast optical switches [14, 15], Bragg reflectors [16], detectors [17], and optical schemes [18, 19]. Note that the ability to use electrooptic effects for controlling photonic band gap structures is of importance for many practical applications. It was shown in [20] for three-dimensional photonic band gap structures formed with nematic liquid crystal inclusions that the transmission spectrum of a photonic crystal could be effectively controlled by varying the orientation of the optical axis of the nematic. Earlier [21], surface electromagnetic waves at the interface between an isotropic medium and a superlattice of alternating isotropic and nematic liquid crystal layers were studied. It was shown that the characteristics of surface waves could be effectively controlled. Volume electromagnetic waves in infinite and bounded superlattices comprising alternating isotropic and nematic layers were studied in [22]. It was shown that the spectrum of electromagnetic waves in superlattices could be considerably modified by changing the orientation of the optical axis of nematic liquid crystals. The possibility of effectively controlling the electromagnetic wave transmission coefficient by changing the orientation of the director of nematic liquid crystals was also noted.

In this work, we study the properties of localized electromagnetic modes in a one-dimensional photonic crystal with a defect structural layer. The role of the defect layer is played by an anisotropic nematic liquid crystal layer. We also study the transmission spectra of photonic crystals with one and two lattice defects. We show that the characteristics of localized modes and the transmission spectrum of a photonic crystal can be controlled by changing the orientation of the optical axis of the nematic.

Sections 2 and 3 contain a description of the theory used to perform the numerical simulation whose results are given in Section 4.

1063-7761/01/9305-0977$21.00

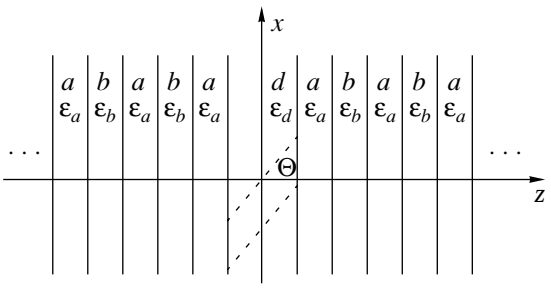


Fig. 1. One-dimensional photonic band gap structure with a lattice defect.

2. EQUATIONS FOR LOCALIZED MODE FREQUENCIES AND FORMS

The photonic band gap structure that we consider is an infinite layered medium comprising alternating isotropic layers of two materials with a structural defect (Fig. 1). The role of the defect layer is played by a nematic liquid crystal layer denoted by "*d*." The thickness of this layer is W_d, and θ is the angle between the optical axis of the nematic and the *z* axis. The nematic liquid crystal layer is inserted between two semibounded superlattices whose unit cell consists of materials *a* and *b* with layer thicknesses W_a and W_b, respectively.

The structure under consideration is characterized by permittivities ε_a and ε_b of layers *a* and *b* and the permittivity tensor of the nematic liquid crystal layer

$$\hat{\varepsilon} = \begin{bmatrix} \varepsilon_\perp \cos^2\theta + \varepsilon_\parallel \sin^2\theta & 0 & \frac{1}{2}\sin(2\theta)\Delta\varepsilon \\ 0 & \varepsilon_\perp & 0 \\ \frac{1}{2}\sin(2\theta)\Delta\varepsilon & 0 & \varepsilon_\perp \sin^2\theta + \varepsilon_\parallel \cos^2\theta \end{bmatrix}, \quad (1)$$

where $\varepsilon_\perp = \varepsilon_{x'x'} = \varepsilon_{y'y'}$, $\varepsilon_\parallel = \varepsilon_{z'z'}$ are the components of the permittivity tensor in principal axes and $\Delta\varepsilon = \varepsilon_\parallel - \varepsilon_\perp$. Further, we only take into account the diagonal components of tensor (1). This is justified if $\theta = 0$ or $\theta = \pi/2$. The Maxwell equations for the anisotropic defect photonic crystal layer on the class of *H*-type fields with frequency ω,

$$\{E_x, H_y\} = \{E_x(z), H_y(z)\}e^{-i\omega t} \quad (2)$$

have the form

$$\left(\frac{d^2}{dz^2} + \frac{\varepsilon_{xx}\omega^2}{c^2}\right)E_x(z) = 0,$$
$$H_y(z) = -\frac{ic}{\omega}\frac{dE_x}{dz}, \quad (3)$$

where *c* is the velocity of light in the vacuum. The equations for *E*-type fields are obtained from (3) by the replacements $E_x(z) \longrightarrow E_y(z)$, $H_y(z) \longrightarrow -H_x(z)$, and $\varepsilon_{xx} \longrightarrow \varepsilon_{yy} = \varepsilon_\perp$. The Maxwell equations for isotropic photonic crystal layers are obtained from (3) by the replacement $\varepsilon_{xx} \longrightarrow \varepsilon_a$ or $\varepsilon_{xx} \longrightarrow \varepsilon_b$.

The geometry of the problem set above makes it possible to adapt the method of studying localized acoustic phonon modes in a superlattice with an isotropic defect layer [23] to our purposes. The solution to the Maxwell equations for an electric field localized in a defect *L* mode can be written in the form

$$E_x(z, t) = E_L(z)\exp(-i\omega_L t), \quad (4)$$

where $E_L(z) \equiv E_x(z)$ is the electric field strength for the localized *L* mode and ω_L is the localized mode frequency. For a photonic band gap structure with a lattice defect, the electric field strength in various layers can be written taking into account general solution (3) to the Maxwell equation for a field in a layer in the form

$$E_L(z) = \begin{cases} f(A^r_{Lj}, B^r_{Lj}, k_{Lj}, z - z^r_{mj})\exp[iq_z(m-1)W] \\ -\frac{W_j}{2} \le z - z^r_{mj} \le \frac{W_j}{2}; \\ f(A_{Ld}, B_{Ld}, k_{Ld}, z - z_d) \\ -\frac{W_d}{2} \le z - z_d \le \frac{W_d}{2}; \\ f(A^l_{Lj}, B^l_{Lj}, k_{Lj}, z - z^l_{mj})\exp[iq_z(m-1)W] \\ -\frac{W_j}{2} \le z - z^l_{mj} \le \frac{W_j}{2}. \end{cases} \quad (5)$$

Here, the $f(A, B, k, z)$ function is defined by the equality

$$f(A, B, k, z) = Ae^{ikz} + Be^{-ikz}, \quad (6)$$

z_d is the coordinate of the center of the defect layer; z^l_{mj} and z^r_{mj} are the coordinates of the centers of the *j*th (*j* = *a*, *b*) layer in the *m*th (*m* = 1, 2, 3, …) period of semiinfinite superlattices from the left and from the right, respectively; $k_{L\mu}$ is given by the equality

$$k_{L\mu} = \frac{\omega_L n_\mu}{c}, \quad \mu = a, b, d, \quad (7)$$

where

$$n_\mu = \sqrt{\varepsilon_\mu}, \quad \varepsilon_d = \varepsilon_{xx} = \varepsilon_\perp \cos^2\theta + \varepsilon_\parallel \sin^2\theta,$$

and Bloch wave number q_z should be complex,

$$q_z = \frac{n\pi}{W} + iq, \quad q > 0, \quad n = 1, 2, 3, \ldots. \quad (8)$$

Here, $W = W_a + W_b$ is the period of the ideal layered medium. The continuity condition for E_x and H_y at the interfaces gives equations for determining the ω_L fre-

quency and the q damping factor of localized modes. These equations have the form

$$\cosh(qW) = (-1)^n\left(\cos\alpha\cos\beta - \frac{1}{2}n_{ba}^{+}\sin\alpha\sin\beta\right), \quad (9)$$

$$(-1)^n\sinh(qW)(2\cos\alpha\cos\beta - n_{da}^{+}\sin\alpha\sin\gamma) - \left(\sin\alpha\cos\beta + \frac{1}{2}n_{ba}^{+}\sin\beta\cos\alpha\right) \quad (10)$$

$$\times(2\sin\alpha\cos\gamma + n_{da}^{+}\sin\gamma\cos\alpha) + \frac{1}{2}n_{ba}^{-}n_{da}^{-}\sin\beta\sin\gamma = 0,$$

where

$$\alpha = k_{La}W_a, \quad \beta = k_{Lb}W_b, \quad \gamma = k_{Ld}W_d,$$
$$n_{ba}^{\pm} = \frac{n_b}{n_a} \pm \frac{n_a}{n_b}, \quad n_{da}^{\pm} = \frac{n_d}{n_a} \pm \frac{n_a}{n_d}.$$

The electric field distribution in localized modes of the photonic crystal structure can be written as

$$E_L(z) = \begin{cases} A_{La}^r f(A_{Ld}, B_{Ld}, k_{Ld}, z - z_d), \\ -\dfrac{W_d}{2} \le z - z_d \le \dfrac{W_d}{2}; \\ A_{La}^r f\left(1, \dfrac{\exp(-iq_zW) - \varphi_1}{\varphi_1}, k_{La}, z - z_{ma}^r\right) \\ \times \exp[iq_z(m-1)W], \quad -\dfrac{W_a}{2} \le z - z_{ma}^r \le \dfrac{W_a}{2}; \\ A_{La}^r f(A_{Lb}^r, B_{Lb}^r, k_{Lb}, z - z_{mb}^r)\exp[iq_z(m-1)W], \\ -\dfrac{W_b}{2} \le z - z_{mb}^r \le \dfrac{W_b}{2}; \\ A_{La}^r f(A_{L\nu}^l, B_{L\nu}^l, k_{L\nu}, z - z_{m\nu}^l)\exp[iq_z(m-1)W], \\ \nu = a, b, \end{cases} \quad (11)$$

where

$$\varphi_1 = \cos\alpha\cos\beta - i\sin\alpha\cos\beta - \frac{1}{2}n_{ba}^{-}(\sin\alpha\sin\beta + i\cos\alpha\sin\beta),$$

$$\varphi_2 = -\frac{i}{2}n_{ba}^{-}\sin\beta,$$

A_{La}^r is the normalization constant, and q_z is given by (9) and (10). Here,

$$\begin{pmatrix} A_{Lb}^r \\ B_{Lb}^r \end{pmatrix} = \hat{\Lambda}^{-1}(c_1, k_{Lb}, -W_b)\hat{\Lambda}(c_1, k_{La}, W_a)\hat{C},$$

$$\begin{pmatrix} A_{Ld} \\ B_{Ld} \end{pmatrix} = \hat{\Lambda}^{-1}(c_1, k_{Ld}, W_d)\hat{\Lambda}(c_1, k_{La}, -W_a)\hat{C},$$

$$\begin{pmatrix} A_{La}^l \\ B_{La}^l \end{pmatrix} = \hat{\Lambda}^{-1}(c_1, k_{La}, W_a)\hat{\Lambda}(c_1, k_{Ld}, -W_d)\begin{pmatrix} A_{Ld} \\ B_{Ld} \end{pmatrix}, \quad (12)$$

$$\begin{pmatrix} A_{Lb}^l \\ B_{Lb}^l \end{pmatrix} = \hat{\Lambda}^{-1}(c_1, k_{Lb}, W_b)\hat{\Lambda}(c_1, k_{La}, -W_a)\begin{pmatrix} A_{La}^l \\ B_{La}^l \end{pmatrix},$$

where

$$\hat{\Lambda}(c_1, k, z) = \begin{pmatrix} e^{ikz/2} & e^{-ikz/2} \\ ic_1ke^{ikz/2} & -ic_1ke^{-ikz/2} \end{pmatrix}, \quad (13)$$

$$\hat{C} = \begin{pmatrix} 1 \\ \dfrac{\exp(-iq_zW) - \varphi_1}{\varphi_2} \end{pmatrix}. \quad (14)$$

3. TRANSMISSION SPECTRUM

The transmission spectrum of a bounded photonic crystal with lattice defects will be studied by the transfer matrix method [24]. Let the permittivities of the layers be written as

$$\varepsilon = \begin{cases} \varepsilon(0) = 1, \quad z < z_0 \\ \varepsilon(1) = \varepsilon_b, \quad z_0 < z < z_1 \\ \varepsilon(2) = \varepsilon_a, \quad z_1 < z < z_2 \\ \ldots \\ \varepsilon(l_1) = \varepsilon_d, \quad z_{l_1-1} < z < z_{l_1} \\ \ldots \\ \varepsilon(l_2) = \varepsilon_d, \quad z_{l_2-1} < z < z_{l_2} \\ \ldots \\ \varepsilon(N) = \varepsilon_b, \quad z_{N-1} < z < z_N \\ \varepsilon(s) = 1. \end{cases} \quad (15)$$

For the structure under consideration, the electric field distribution in layers has the form

$$E_x(n, t) = [A(n)\exp(\alpha(n)(z - z_n)) + B(n)\exp(-\alpha(n)(z - z_n))]\exp(-i\omega t), \quad (16)$$

where $A(n)$ and $B(n)$ are the incident and reflected wave amplitudes, respectively, in the nth layer,

$$\alpha(n) = \frac{\sqrt{\varepsilon(n)}\omega}{c}. \quad (17)$$

The magnetic field distribution in layers can be written as

$$H_y(n, t) = [A(n)\exp(i\alpha(n)(z - z_n)) - B(n)\exp(-i\alpha(n)(z - z_n))]\sqrt{\varepsilon(n)}\exp(-i\omega t). \quad (18)$$

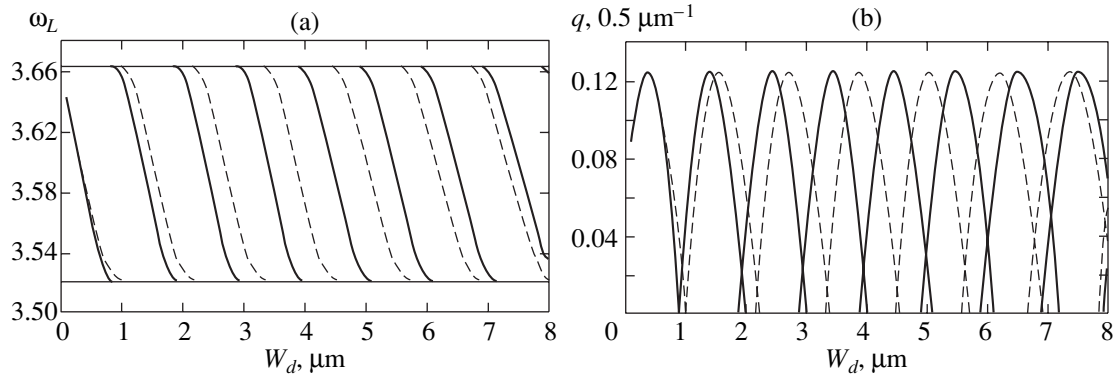


Fig. 2. Dependences of localized mode (a) frequency and (b) damping factor on defect layer thickness. The dashed line corresponds to $n_d = 1.5$ at $\theta = 0$, and the solid line, to $n_d = 1.7$ at $\theta = \pi/2$; ω_L is in c/W units.

The continuity condition for E_x and H_y at the interfaces, $z = z_{n-1}$, gives a system of equations which can be written as the matrix equation

$$\begin{pmatrix} A(n-1) \\ B(n-1) \end{pmatrix} = T_{n-1,n} \begin{pmatrix} A(n) \\ B(n) \end{pmatrix}. \quad (19)$$

Here, the transfer matrix has the form

$$T_{n-1,n} = D^{-1}(n-1)D(n)P(n), \quad (20)$$

where

$$D(n) = \begin{pmatrix} 1 & 1 \\ \sqrt{\varepsilon(n)} & -\sqrt{\varepsilon(n)} \end{pmatrix}, \quad (21)$$

$$P(n) = \begin{pmatrix} e^{-i\alpha(n)\gamma_n} & 0 \\ 0 & e^{i\alpha(n)\gamma_n} \end{pmatrix}, \quad (22)$$

and $\gamma_n = z_n - z_{n-1}$, $n = 1, 2, \ldots, N$. It follows from (19) that the $A(0)$ and $B(0)$ amplitudes are related to $A(s)$ and $B(s)$ as

$$\begin{pmatrix} A(0) \\ B(0) \end{pmatrix} = \hat{M} \begin{pmatrix} A(s) \\ B(s) \end{pmatrix}, \quad (23)$$

where

$$M = \hat{T}_{01}\hat{T}_{12}\ldots\hat{T}_{N-1,N}\hat{T}_{Ns}, \quad (24)$$

$s \equiv N + 1$, and $\gamma_{N+1} \equiv 0$. On the assumption that electromagnetic waves are not reflected on the right side of the photonic crystal sample, the $t(\omega)$ transmission coefficient is given by the equation

$$t(\omega) = \left| \left(\frac{A(s)}{A(0)} \right)_{B(s)=0} \right|^2. \quad (25)$$

Using (23) yields

$$t(\omega) = \frac{1}{|\hat{M}_{11}|^2}, \quad (26)$$

where $\hat{M}_{11}$ is the $\hat{M}$ matrix element.

4. CALCULATION RESULTS AND DISCUSSION

Equations (9) and (10), which determine the frequencies and damping factors of localized modes of an infinite photonic crystal, were solved numerically. The results obtained for a photonic crystal with layer thicknesses $W_a = W_b = 1$ μm and layer permittivities $\varepsilon_a = 4$ and $\varepsilon_b = 2.25$, respectively, are given below. The refractive indexes of the defect layer in the IR region

$$n_d\left(\frac{\pi}{2}\right) = \sqrt{\varepsilon_\parallel} = n_\parallel = 1.7$$

and

$$n_d(0) = \sqrt{\varepsilon_\perp} = n_\perp = 1.5$$

corresponded to a 5TsB nematic liquid crystal at 20°C [25].

The electromagnetic excitation spectrum of an ideal layered medium has a band character [26]. A defect layer in a photonic crystal can cause the appearance of discrete frequencies within the forbidden bands of the unperturbed layered medium and electromagnetic field localization in defect modes. The dependences of the frequency and damping factor of defect modes in the second forbidden band ($n = 2$) on the thickness of the defect layer are shown in Fig. 2 for the normal and tangential orientations of the optical axis of the nematic. The frequencies

$$\omega_1 = 3.521\frac{c}{W} = 5.281 \times 10^{14} \text{ Hz}$$

and

$$\omega_2 = 3.663\frac{c}{W} = 5.495 \times 10^{14}\ \text{Hz}$$

in this figure determine the boundaries of the forbidden band spectral range. The change in the orientation of the director of the nematic not only substantially changes the frequency and damping factor of localized modes but can cause the appearance of new localized modes at certain defect layer thicknesses. Note that restoring translational invariance in the photonic crystal with $W_d = 1$ μm and $n_d = 1.5$ causes the disappearance of discrete frequencies within forbidden bands. An increase in the defect layer thickness at a given liquid crystal optical axis orientation increases the number of discrete frequencies in the photonic crystal forbidden band. At $W_d = 100$ μm, the number of localized modes at $\theta = 0$ and $\theta = \pi/2$ increases to five. At a qualitative level, this result can be given a visual explanation. Localized modes in the vicinity of defects in photonic band gap structures have much in common with cavity resonators. Indeed, a defect photonic crystal mode can be described as a standing wave formed as a result of reflection from semibounded superlattices, or, in other words, cavity walls. And the number of modes in a cavity in a given frequency range is proportional to its length.

Figure 3 illustrates the possibility of controlling the spectrum of defect modes and the spatial distribution of the square of the electric field modulus in photonic crystal defect modes with a $W_d = 4.5$ μm defect layer thickness. The center of the defect layer coincides with the origin. The curves at negative z values are obtained by mirror reflection in the $z = 0$ plane. The curve shown in Fig. 3b was constructed for $\theta = \pi/2$ and $n_{||} = 1.7$. The localized mode frequency near the forbidden band center then equals $\omega_L = 5.380 \times 10^{14}$ Hz, and the corresponding damping factor is $q = 0.125$. The localization of the square of the electric field modulus in the vicinity of the defect layer is clearly seen. Changing the orientation of the optical axis of the nematic from tangential to normal ($\theta = 0$, $n_\perp = 1.5$) causes the appearance of two defect modes near the continuous spectrum boundaries (Figs. 3a, 3c). The frequencies and the damping factors of the modes are $\omega_L = 5.491 \times 10^{14}$ Hz, $q = 0.026$ (Fig. 3a) and $\omega_L = 5.282 \times 10^{14}$ Hz, $q = 0.019$ (Fig. 3c). It follows that changes in the director orientation induce the appearance of new defect modes much less localized. Indeed, the mode damping factors (Figs. 3a, 3c) decrease 4.8 and 6.6 times, respectively. It follows from Fig. 2 that there exist defect layer thicknesses at which director reorientation from tangential to normal does not change the number of defect modes and only shifts their frequencies and damping factors.

Consider the special features of the transmission spectrum of a finite photonic crystal by numerically solving Eq. (26) for the transmission coefficient at various layered medium parameters. As with an infinite photonic crystal, we assume that $W_a = W_b = 1$ μm, $\varepsilon_a = 4$, and $\varepsilon_b = 2.25$.

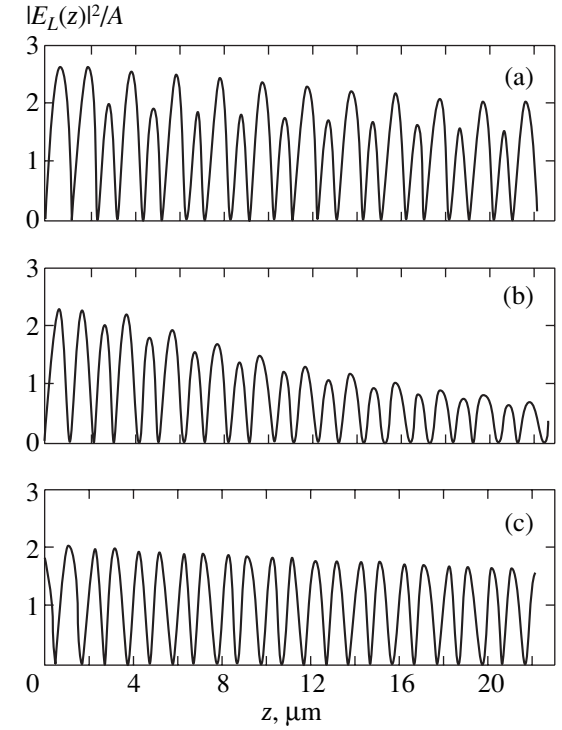


Fig. 3. The square of the electric field modulus of a localized mode. Defect layer thickness $W_d = 4.5$ μm, $n_d =$ (a, c) 1.5 and (b) 1.7, $A = \left|A^r_{La}\right|^2$. The center of the defect layer coincides with the origin ($z = 0$).

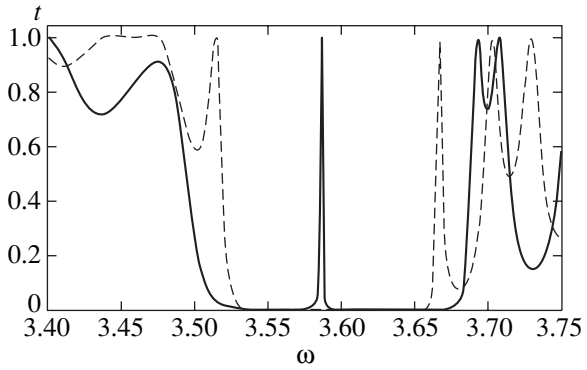


Fig. 4. Frequency dependence of the transmission coefficient of a photonic crystal with $N = 85$ layers. The thickness of the defect layer situated in the center of the layered medium is $W_d = 4.5$ μm. The dashed line corresponds to $n_d = 1.5$ at $\theta = 0$, and the solid line, to $n_d = 1.7$ at $\theta = \pi/2$; ω is in c/W units.

The frequency dependences of the photonic crystal transmission coefficient in the frequency range of the second forbidden band of the ideal photonic crystal at

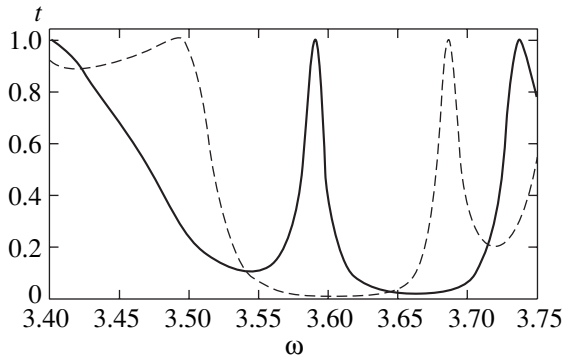


Fig. 5. Frequency dependences of the transmission coefficient of a sample with $N = 45$ layers. The defect is situated in the center of the photonic crystal. The other parameters have the same values as in Fig. 4; ω is in c/W units.

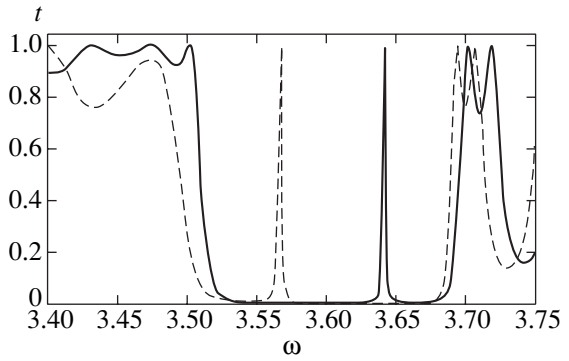


Fig. 6. Frequency dependences of the transmission coefficient of a photonic crystal with a defect layer of thickness $W_d = 5.2$ μm. The other parameters have the same values as in Fig. 4; ω is in c/W units.

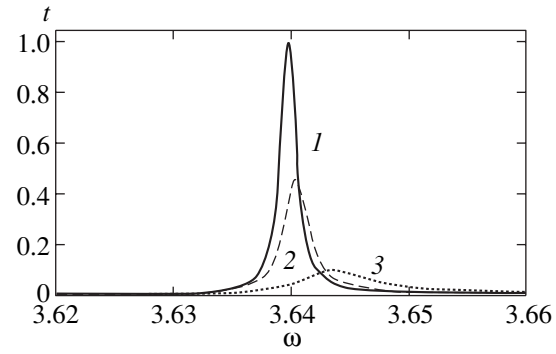


Fig. 7. Fragments of the transmission spectra of a photonic crystal at $\theta = \pi/2$ with various defect layer positions in the lattice. The defect is (*1*) in the center of the sample, $l = 43$, and at $l =$ (*2*) 33 and (*3*) 23. The other parameters are the same as in Fig. 6.

various nematic optical axis orientations are shown in Fig. 4. It follows from these dependences that the defect mode spectrum of a finite sample is close to the spectrum of localized modes in an infinite photonic crystal if the other layered structure parameters are equal (Figs. 2, 3). Indeed, at a $\theta = \pi/2$ orientation angle, a defect mode with an $\omega_L = 5.381 \times 10^{14}$ Hz frequency, which is close to the corresponding frequency of the defect layer in an infinite photonic crystal (Fig. 3b), appears in the transmission spectrum. The width of the transmission curve is about 10 Å. At $\theta = 0$, the transmission spectrum contains two defect modes with frequencies $\omega_L = 5.271 \times 10^{14}$ Hz and $\omega_L = 5.502 \times 10^{14}$ Hz. The frequency dependence of transmission for a sample with a smaller number of layers is shown in Fig. 5. A comparison of Figs. 5 and 4 shows that an almost twofold decrease in the number of layers in a photonic crystal with a lattice defect causes noticeable changes in the transmission spectrum. The transmission curves broaden, the position of the minimum changes, and the frequency corresponding to maximum transmission shifts.

It has been mentioned that, for an infinite photonic crystal, there exist defect layer thicknesses at which nematic optical axis reorientation shifts the frequency of the defect mode but does not cause the appearance of new defect levels in the forbidden band. An example is a photonic crystal with a $W_d = 5.2$ μm defect layer thickness (Fig. 2). For comparison, the frequency dependence of the transmission coefficient of a finite photonic crystal with a defect layer of the same thickness is shown in Fig. 6.

A characteristic feature of the curves shown in Figs. 4–6 is the high penetrating power of *H* waves. The transmission coefficient of *H* waves increases virtually to one when defect layers in the forbidden band appear. Note also that polarization of radiation that passes through photonic crystal samples with lattice defects can be controlled. Indeed, reorientation of the nematic liquid crystal optical axis from normal to tangential results in that *H*- and *E*-type modes acquire a phase difference after passage through the nematic layer.

Fragments of the transmission spectra of photonic crystals with different defect layer positions are shown in Fig. 7. The figure illustrates typical behavior of the transmission coefficient of *H* waves caused by the appearance of a defect level in the forbidden band when the defect layer is displaced from the symmetrical position to the boundary between the sample and the vacuum. We see that the transmission ability of the photonic crystal decreases, the transmission curve width increases, and the frequency of the defect mode slightly shifts. These effects can be given a simple physical interpretation. As mentioned, localized modes in the vicinity of the defect in a photonic crystal have much in common with a cavity resonator. The displacement of the defect to the boundary between the sample and the vacuum decreases the *Q* factor of the cavity, that is, modifies the transmission spectra of the photonic crystal.

Lastly, the transmission spectra of a photonic crystal with two differently spaced identical defect lattices are shown in Fig. 8. Increasing the distance between defect

layers causes qualitative changes in the transmission spectrum, namely, two transmission curves of defect modes coalesce; that is, localized electromagnetic modes become degenerate. Frequency degeneracy first appears for the normal orientation of the nematic liquid crystal director. Changes in the positions of transmission curve maxima can be qualitatively explained as follows.

The characteristic length determining the region of electromagnetic wave localization in the vicinity of a defect is $l = 1/q$. According to Fig. 2, $l(0) = 1/q \approx 17$ μm for $W_d = 5.2$ μm and $\theta = 0$ and $l(\pi/2) \approx 21$ μm for the same W_d thickness and $\theta = \pi/2$. When the distance between the defects approaches the characteristic length, $r \approx l$, two modes become strongly coupled, and their frequencies become split (Fig. 8a). Further, the distance between the defect mode frequencies increases as nematic liquid crystal layers approach each other. When the distance between the defects decreases to $r = 11.2$ μm, the high- and low-frequency defect mode transmission curves for $\theta = \pi/2$ and $\theta = 0$, respectively (Fig. 8a), coalesce with the continuous transmission spectrum of the photonic crystal. The modes are weakly coupled if the distance between the defects exceeds characteristic length l. The defect mode frequency is then doubly degenerate. At $\theta = 0$, the frequencies become degenerate at a smaller distance between the defects than when $\theta = \pi/2$ (Figs. 8b, 8c) because the corresponding characteristic localization lengths are related as $l(0) < l(\pi/2)$. Placing the defects $r = 71.2$ μm apart (the defects are then situated close to the boundaries between the photonic crystal and the vacuum) makes defect mode frequencies at $\theta = 0$ and $\theta = \pi/2$ doubly degenerate, the transmission curves of the localized modes are then strongly broadened, especially for $\theta = 0$, and maximum transmission coefficients are then much smaller than one. Note that if coupling between defect modes is close to critical (to coupling at which mode frequencies become degenerate), the transmission curve has a steep slope, and the transmission coefficient weakly depends on frequency (Fig. 8c, $\theta = \pi/2$). This circumstance can be used to create band filters.

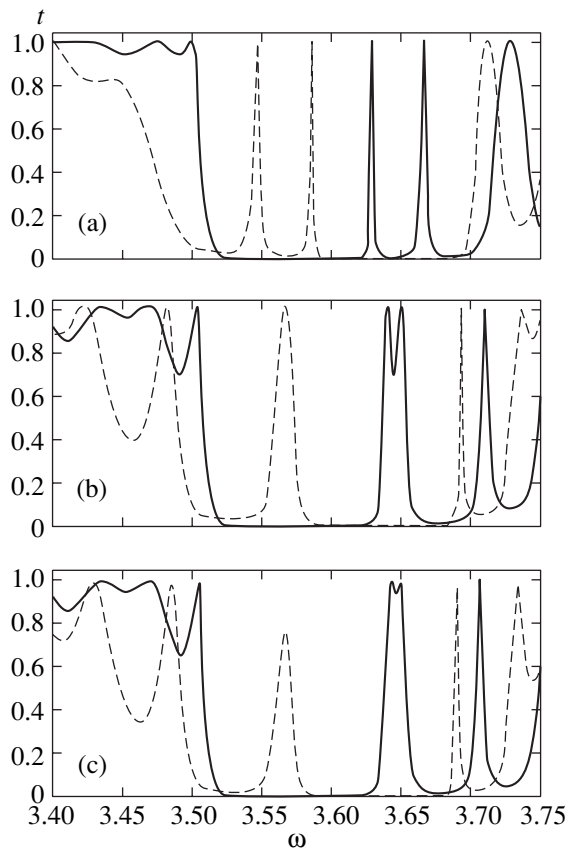


Fig. 8. Frequency dependences of the transmission coefficient of a photonic crystal with two defect layers. The number of layers in the sample $N = 87$, the other parameters are the same as in Fig. 4; (a) numbers of defect layers are $l_1 = 35$ and $l_2 = 53$, the distance between the layers is $r = 22.2$ μm, the defect modes are split; (b) $l_1 = 23$, $l_2 = 65$, and $r = 46.2$ μm, the frequencies for $\theta = 0$ are degenerate; and (c) $l_1 = 21$, $l_2 = 67$, and $r = 50.2$ μm, the frequencies are virtually degenerate also for $\theta = \pi/2$.

5. CONCLUSION

To summarize, we showed in this work that the spectrum of defect modes and the field distribution in defect modes of a one-dimensional photonic crystal had certain special features largely because of a strong permittivity anisotropy and a high sensitivity of the nematic, which played the role of a structural defect layer, to external fields.

Importantly, there exist liquid crystal layer thicknesses at which changes in the orientation of the nematic optical axis cause qualitative changes in the spectrum of defect modes, the appearance of new defect levels, and substantial changes in the degree of field localization in defect modes. We also showed that the transmission spectrum of a photonic crystal with one and two lattice defects experienced substantial rearrangement when the optical axis of the nematic liquid crystal was reoriented. In addition, the transmission spectrum of photonic crystals with two defects could be qualitatively modified by changing the distance between the defect layers in the lattice.

In practical applications, such photonic band gap structures offer promise for creating filters and polarizers with controllable characteristics. Lastly, note that the possibility of controlling the degree of localization of electromagnetic field along the direction of laser beam propagation appears to be promising for controlling the efficiency of nonlinear optical interactions.

REFERENCES

1. J. Joannopoulos, R. Meade, and J. Winn, *Photonic Crystals* (Princeton Univ. Press, Princeton, 1995).

2. E. Yablonovitch, Phys. Rev. Lett. **58**, 2059 (1987).
3. E. Yablonovitch, J. Opt. Soc. Am. B **10**, 283 (1993).
4. S. John and T. Quang, Phys. Rev. A **50**, 1764 (1994).
5. S. John, Phys. Rev. Lett. **53**, 2169 (1984).
6. S. John, Phys. Rev. Lett. **58**, 2486 (1987).
7. D. R. Smith, R. Dalichaouch, S. Schultz, *et al.*, J. Opt. Soc. Am. B **10**, 314 (1993).
8. S. Y. Lin, E. Chow, V. Hietala, *et al.*, Science **282**, 274 (1998).
9. A. M. Zheltikov, S. A. Magnitskiĭ, and A. V. Tarasishin, Pis'ma Zh. Éksp. Teor. Fiz. **70**, 323 (1999) [JETP Lett. **70**, 323 (1999)].
10. A. M. Zheltikov, S. A. Magnitskiĭ, and A. V. Tarasishin, Zh. Éksp. Teor. Fiz. **117**, 691 (2000) [JETP **90**, 600 (2000)].
11. M. Bayinder, B. Temelkuran, and E. Ozbay, Phys. Rev. B **61**, R11855 (2000).
12. B. Temelkuran and E. Ozbay, Appl. Phys. Lett. **74**, 486 (1999).
13. M. V. Alfimov, A. M. Zheltikov, A. A. Ivanov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 714 (2000) [JETP Lett. **71**, 489 (2000)].
14. M. Scalora, J. P. Dowling, C. M. Bowden, and M. J. Bloemer, Phys. Rev. Lett. **73**, 1368 (1994).
15. P. R. Villeneuve, D. S. Abrams, S. Van, and J. Joannopoulos, Opt. Lett. **21**, 2017 (1996).
16. A. Stingl, M. Jenzner, Ch. Spielmann, *et al.*, Opt. Lett. **20**, 602 (1995).
17. B. Temelkuran, E. Ozbay, J. P. Kavanaugh, *et al.*, Appl. Phys. Lett. **72**, 2376 (1998).
18. H. Kosaka, T. Kawashima, A. Tomita, *et al.*, Appl. Phys. Lett. **74**, 1370 (1999).
19. A. de Lustrac, F. Gadot, S. Cabaret, *et al.*, Appl. Phys. Lett. **75**, 1625 (1999).
20. K. Bush and S. John, Phys. Rev. Lett. **83**, 967 (1999).
21. S. Ya. Vetrov and A. V. Shabanov, Zh. Éksp. Teor. Fiz. **101**, 1340 (1992) [Sov. Phys. JETP **74**, 719 (1992)].
22. I. O. Bogul'skiĭ, S. Ya. Vetrov, and A. V. Shabanov, Opt. Spektrosk. **84**, 823 (1998) [Opt. Spectrosc. **84**, 742 (1998)].
23. K.-Q. Chen, X.-H. Wang, and B.-Y. Gu, Phys. Rev. B **61**, 12 075 (2000).
24. P. Yeh, J. Opt. Soc. Am. **69**, 742 (1979).
25. V. Ya. Zyryanov and V. Sh. Épshteĭn, Prib. Tekh. Éksp., No. 2, 164 (1987).
26. A. Yariv and P. Yeh, in *Optical Waves in Crystals: Propagation and Control of Laser Radiation* (Wiley, New York, 1984; Mir, Moscow, 1987), p. 169.

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 93, No. 5, 2001, pp. 985–997.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 5, 2001, pp. 1135–1149.

ATOMS, SPECTRA, RADIATION

Three-Dimensional Interference Effects in the Mechanical Action of Weak Biharmonic Fields upon Particles with the $J = 0 \longrightarrow J = 1$ Quantum Transition

S. A. Gavrilyuk, I. V. Krasnov*, and S. P. Polyutov**

Institute of Computational Modeling, Siberian Division, Russian Academy of Sciences, Krasnoyarsk, 660036 Russia
**e-mail: krasn@ksc.krasn.ru*
***e-mail: psp@beep.ru*
Received April 23, 2001

Abstract—Explicit expressions are derived for the rectified radiative forces (RRFs) related to the action of a weak interfering optical field of an arbitrary three-dimensional (3D) configuration upon resonance particles featuring the $J = 0 \longrightarrow J = 1$ quantum transition. It is shown that, in contrast to the case of a monochromatic field, there are simple 3D biharmonic field configurations for which the ratio of the vortex and potential RRF components can be controlled by adjusting frequencies and polarizations of the interfering light waves. This modification of the RRF structure gives rise to qualitatively different types of both vortex and potential light-induced particle motions that may lead to a 3D spatial localization (confinement) of these particles within the cells of an effective optical lattice with a period significantly greater than the light wavelength. In particular, the particles may perform a stable rotational motion along closed trajectories inside the elementary cells. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Effective optical control of the motion and spatial localization of resonance atoms [1–3] can be based on the use of the so-called rectified radiative forces (RRFs) [4] induced by interfering biharmonic light fields. This idea was originally formulated in [4–7] and then significantly developed in a number of subsequent theoretical and experimental investigations [8–20]. In a strong field, the RRF possesses a magnitude on the order of an induced light-pressure force, exhibits no saturation with increasing radiation intensity, and has a constant sign over a macroscopic spatial scale significantly exceeding the light wavelength. Another remarkable property of the RRFs, manifested in both strong and weak biharmonic fields, is the possibility of controlling the spatial structure of this force [5, 6, 14, 15]. In a strong biharmonic field, the RRF contains a vortex component besides the potential component, but the ratio of these components can be modified by adjusting the directions of propagation of the interfering waves. This ratio significantly affects the character of the light-induced motion of resonance particles [6, 21, 22]. In the case of a weak biharmonic field, the vortex RRF component can be suppressed, as demonstrated for two-dimensional (2D) field configurations [6], by properly selecting the radiation parameters. This would remove some fundamental limitations (of the type related to the Earnshaw theorem [23]) hindering stable localization of the particles by the forces of spontaneous light pressure in a monochromatic field.

However, these considerations concerning the attractive properties of RRFs were based on the results of calculations performed within the framework of a simple scalar model describing the interaction of an atom with a resonance electromagnetic field. Therefore, strictly speaking, the above conclusions cannot be rigorously transferred to the most interesting case (e.g., for solving the problem of purely optical 3D confinement of atomic species).[1] A correct problem solution requires taking into account degeneracy of the quantum states with respect to the magnetic quantum number M.

Recently, the possibility of ensuring the optical confinement of resonance particles with the aid of RRFs induced by strong fields of a certain 3D configuration was studied by Wasik and Grimm [18] for atoms featuring the quantum transition

$$J_g = 1/2 \longrightarrow J_e = 3/2,$$

where J_g and J_e are the moments of the ground and excited states, respectively.

Degeneracy of the ground state of a resonance particle is a very important feature of the physical situation studied in [18]. This factor predetermines the possibility of existence of a highly successful combination of the effects related to the 3D macroscopic confinement of atoms in a superlattice (induced by a potential RRF), the sub-Doppler (polarization-gradient) cooling, and the microscopic confinement in potential wells with

[1] Previous calculations [5, 6] showed only the possibility of an effective 2D localization of particles.

dimensions on the order of a light wavelength. However, a theoretical model proposed in [18] and the theoretical consequences are inapplicable to many atomic species (e.g., such as the alkaline-earth isotopes with even-even nuclei) possessing nondegenerate ground states with $J_g = 0$.

In this study, the theory of interference phenomena in the resonance light pressure is developed for the case of weak bichromatic fields with arbitrary polarizations and 3D spatial configuration and the particles featuring the quantum transition

$$J_g = 0 \longrightarrow J_e = 1.$$

General expressions obtained for the RRFs contain essentially new (in comparison to the scalar model) interference terms related to a nonlinear mixing of the contributions from waves with different frequencies and polarizations to the resonance light pressure. Nevertheless, it was found that the main conclusion made previously about the possibility of controlling the spatial structure of RRFs remains valid. We have established that there exist simple 3D symmetric configurations of the interfering waves (with a zero average total radiation flux density) for which the ratio of the vortex and potential RRF components can be controlled by adjusting frequencies. An additional control factor is provided by polarizations of the interfering light waves which allows, for example, the changing of the axis of rotation for the particles performing a vortex motion in a field with the 3D*Lin* $\perp$ *Lin* configuration. Rearrangement of the spatial structure of RRFs is accompanied by the appearance of qualitatively different types of vortex or potential light-induced particle motions that may lead, in particular, to a stable 3D spatial localization of these particles within the cells of an effective macroscopic optical lattice (superlattice) with a period significantly greater than the light wavelength λ.

There is an important circumstance following from the results of our investigation which can be related to the use of weak biharmonic fields for purely optical (nonmagnetic) 3D confinement of atomic species. Even for a relatively small level of saturation of the quantum transition, the RRFs can be still sufficiently large to hold cold particles (with a temperature of $T \sim 10^{-3}$ K corresponding to the Doppler cooling limit in this problem) provided that the field parameters are selected so as to construct a "correct" spatial structure of the RRFs. The advantages of using weak biharmonic fields are (i) a small magnitude of the light-induced Stark shift of the energy levels (not exceeding a natural width γ of the optical resonance), (ii) the possibility of using wide nondiverging laser beams for purely optical confinement of large-size bunches of resonance particles, and (iii) the simplicity of controlling the spatial RRF structure (and, hence, the character of the light-induced motion of particles) without modification of the base geometry of intersecting light rays.

2. A MODEL OF THE ATOM–ELECTROMAGNETIC FIELD INTERACTION

Let us consider an atom possessing the mass m, moving with the velocity $\mathbf{v}$ in a bichromatic field with the complex amplitude

$$\mathbf{E} = \mathbf{E}_0(\mathbf{r})e^{-i\Delta_0 t} + \mathbf{E}_1(\mathbf{r})e^{-i\Delta_1 t}, \quad (1)$$

where Δ_0 and Δ_1 are the frequency detunings of the fields $\mathbf{E}_0$ and $\mathbf{E}_1$, respectively, from the frequency ω_0 of the quantum transition between the ground state $|J_g = 0, M_g = 0\rangle$ and the excited states $|J_e = 1, M_e = 0, \pm 1\rangle$ of the atom.

The state of this atom in the field will be described in terms of the density matrix $\hat{\rho}$ in a Cartesian representation $\rho_{jj'}^{\beta\beta'}$ [24] using basis functions φ_j^β of the type

$$\varphi_0^g = |00\rangle, \quad \varphi_x^e = \frac{|1,-1\rangle - |11\rangle}{\sqrt{2}},$$

$$\varphi_y^e = -i\frac{|1,-1\rangle + |11\rangle}{\sqrt{2}}, \quad \varphi_z^e = |10\rangle.$$

The matrix elements of the operator $\hat{\mathbf{d}}$ of the dipole transition between the atomic states are directed along the axes of the Cartesian coordinate system $\mathbf{e}_j$ ($j = x, y, z$):

$$\left\langle \varphi_j^e \middle| \hat{\mathbf{d}} \middle| \varphi_0^g \right\rangle = \mathbf{e}_j d,$$

the amplitude $\mathbf{P}(t)$ of the field-induced atomic dipole moment $\mathrm{Sp}(\rho\hat{\mathbf{d}})$ determined using an expansion

$$\mathbf{p} = \sum_j d\rho_j \mathbf{e}_j,$$

where

$$d = \frac{\langle 1\|d\|0\rangle}{\sqrt{3}},$$

$\langle 1\|d\|0\rangle$ is the reduced matrix element, and

$$\rho_j = \rho_{j0}^{eg}\exp(i\omega_0 t).$$

Exposed to a field of type (1), an atom experiences the action of a force [1]

$$\mathbf{F} = \hbar\sum_j \rho_j \nabla V_j^* + \text{c.c.}, \quad (2)$$

where

$$V_j = \frac{d(\mathbf{e}_j \cdot \mathbf{E})}{\hbar} = \sum_{j\alpha} V_{j\alpha} e^{-i\Delta_\alpha t}, \quad j = x, y, z,$$

$$\alpha = 0, 1, \quad V_{j\alpha}(\mathbf{r}) = \frac{d(\mathbf{e}_j \cdot \mathbf{E}_\alpha(\mathbf{r}))}{\hbar}$$

are the local Rabi frequencies and ρ_j are the projections of the complex amplitude of the induced dipole moment (expressed in units of d) onto the Cartesian coordinate axes. The latter quantities are determined from the optical Bloch equations considered (in the approximation of a preset motion [1]) along the classical atomic trajectory $\mathbf{r} = \mathbf{r}(t)$:

$$i\left(\frac{d}{dt} + \gamma_\perp\right)\rho_i = \sum_j q_{ij}V_j, \quad j, i = x, y, z,$$

$$i\left(\frac{d}{dt} + \gamma\right)q_{ij} = -i\gamma\delta_{ij} + (\rho_i V_j^* - V_i\rho_j^*) \qquad (3)$$

$$-\delta_{ij}\sum_{l=x,y,z}(\rho_l^* V_l - \text{c.c.}).$$

Here, we introduced the rate of the spontaneous decay of the excited state $\gamma_\perp = \gamma/2$, and the combinations of elements of the density matrix

$$q_{ij} = \rho_{ij} - \rho\delta_{ij}, \quad \rho_{ij} = \rho_{ij}^{ee},$$
$$\rho = \rho_{00}^{gg} = 1 - (\rho_{xx} + \rho_{yy} + \rho_{zz}).$$

Apparently, ρ has a sense of the relative population of the bottom level, q_{ii} are the differential populations of the working levels, and the quantities q_{ij} $(i \neq j)$ describe the effects due to the coherency between states of the excited atom.

We will consider the case of weak fields:[2]

$$\left|\frac{V_{j\alpha}}{\nu_\alpha}\right|^2, \left|\frac{V_{j\alpha}^2}{\nu_1\nu_2}\right|, \left|\frac{V_{j\alpha}^2}{\nu_\alpha\gamma}\right| \leq g \ll 1.$$

In this case, the solutions to the Bloch equations and the radiative force can be determined using the perturbation theory. To this end, the unknown quantities are expanded into series in powers of the field strength (in fact, with respect to the small parameter $g \ll 1$):

$$\rho_j = \rho_j^{(1)} + \rho_j^{(3)} + \ldots,$$
$$q_{ij} = -\delta_{ij} + q_{ij}^{(2)} + q_{ij}^{(4)} + \ldots, \qquad (4)$$
$$\mathbf{F} = \mathbf{F}^{(2)} + \mathbf{F}^{(4)} + \ldots,$$

where the superscripts indicate the order of smallness.

Nontrivial interference effects in the light-induced pressure appear in the fourth order of smallness with respect to the weak field [5, 6]. For this reason, we restrict the expansion of the induced dipole moment ρ_j to terms of the third order. The resulting Bloch equations possess the following structure:

$$\rho_j^{(1)} = \sum_{\alpha=0}^{1} A_{j\alpha}e^{-i\Delta_\alpha t},$$
$$\rho_j^{(3)} = \sum_{\alpha=0}^{1} A_{j\alpha}^{(3)}e^{-i\Delta_\alpha t} + N_j e^{-i(\delta+\Delta_0)t} + M_j e^{i(\delta-\Delta_1)t}, \qquad (5)$$

where $\delta = \Delta_0 - \Delta_1$ $(j = x, y, z)$. The functions $A_{\alpha j}$ and $A_{\alpha j}^{(3)}$ are sequentially determined from the following system of linear inhomogeneous equations:

$$\left(\frac{d}{dt} - i\nu_\alpha\right)A_{j\alpha} = iV_{j0}(\mathbf{r}),$$

$$\nu_\alpha = \Delta_\alpha + i\gamma_\perp, \quad \mathbf{r} = \mathbf{r}(t),$$

$$\left(\frac{d}{dt} - i\nu_0\right)A_{i0}^{(3)} = -i\sum_j (C_{ij}V_{j0}(\mathbf{r}) + D_{ij}V_{j1}(\mathbf{r})),$$

$$\nu_{01} = \delta + i\gamma,$$

$$\left(\frac{d}{dt} - i\nu_1\right)A_{i1}^{(3)} = -i\sum_j (C_{ij}V_{j1}(\mathbf{r}) + D_{ji}^* V_{j0}(\mathbf{r})), \qquad (6)$$

$$\left(\frac{d}{dt} + \gamma\right)C_{ij} = i\sum_\alpha (A_{j\alpha}^* V_{i\alpha}(\mathbf{r}) - A_{i\alpha}V_{j\alpha}^*(\mathbf{r}))$$

$$-i\delta_{ij}\sum_{l,\alpha}(A_{l\alpha}V_{l\alpha}^*(\mathbf{r}) - \text{c.c.}),$$

$$\left(\frac{d}{dt} - i\nu_{01}\right)D_{ij} = i(V_{i0}(\mathbf{r})A_{j1}^* - V_{j1}(\mathbf{r})A_{i0})$$

$$+ i\delta_{ij}\sum_l (V_{l0}(\mathbf{r})A_{l1}^* + V_{l1}^*(\mathbf{r})A_{l0}).$$

In what follows, we will assume that the frequency detunings Δ_0 and Δ_1 are not very close to each other $(|\Delta_0 - \Delta_1| > g\gamma)$. This allows us to write an expression for the force with neglect of the terms oscillating at the frequencies representing the multiples of $\delta = \Delta_0 - \Delta_1$.[3] It is also not necessary to determine explicitly the components $\rho_j^{(3)} \propto M_j$ and N_j.

[2] These conditions provide for both a small occupancy of the excited atomic states and a small relative value of the light-induced Stark shift as compared to the resonance width: for $\Delta_\alpha \gg \gamma$, we always have $\left|V_\alpha^2/\gamma\Delta_\alpha\right| \ll 1$. The perturbation theory employed here is inapplicable to the RRF determination, for example, in the case of $|\gamma/\Delta_\alpha|$, $|V_{j\alpha}/\Delta_\alpha|^2 \ll 1$ if $\left|V_{j\alpha}^2/\gamma\Delta_\alpha\right| \gg 1$ (see [1, 6]).

[3] An allowance for these terms leads to small (in the quasi-classical limit of $mv \gg \hbar\omega_0/c$) oscillating corrections $\delta p < \hbar\omega_0/c$ to the particle momentum [5].

3. RECTIFIED RADIATION FORCES

As can be readily seen from the above Eqs. (2) and (4)–(6), the interfering light fields are of the plane wave superposition type with the wave vectors $\{\mathbf{k}_\alpha^\sigma\}$:

$$\mathbf{E}_\alpha(\mathbf{r}) = \sum_\sigma \mathbf{E}_\alpha^\sigma \exp(i\mathbf{k}_\alpha^\sigma \cdot \mathbf{r}), \tag{7}$$

in which the radiative force is a nonlinear function of both the wave amplitude $|\mathbf{E}_\alpha^\sigma|$ and the ratios of phases and polarizations; in other words, the field dependence of the force exhibits a pronounced interference character.

Let us assume that the sets of wavevectors $\{\mathbf{k}_0^\sigma\}$, $\{\mathbf{k}_1^\sigma\}$ contain the pairs $\mathbf{k}_\alpha^\sigma$, $\mathbf{k}_{\alpha'}^{\sigma'}$ such that

$$|\mathbf{k}_\alpha^\sigma - \mathbf{k}_{\alpha'}^{\sigma'}| \sim \Delta k \ll k.$$

In this case, a quasi-periodic spatial structure of the force is characterized by two sharply different scales: microscopic (on the order of the light wavelength $\lambda \sim 1/k$) and macroscopic ($\lambda_M \sim \Delta k^{-1} \gg \lambda$) related to beats of the spatial harmonics with close wavevectors $\mathbf{k}_\alpha^\sigma$.

The rectified radiative force [5, 6] is a smooth component of the radiative force $\langle \mathbf{F} \rangle$ averaged over the microscopic spatial oscillations. Note that the spatial variation of this force is determined by the macroscopic quantity λ_M, while the characteristic magnitude is determined on the microscopic (!) scale:[4]

$$|\langle \mathbf{F} \rangle| \propto k \sim 1/\lambda.$$

We will restrict the consideration to the case of slow atoms (frequently encountered in modern experiments),

$$kv \ll \gamma, \tag{8}$$

and take into account the nonlocal (retarding) part of the field-induced dipole moment in solving Eq. (6) in the linear approximation with respect to the velocity. Under these conditions, Eqs. (2) and (4)–(6) give the following expressions for the RRF (to within the terms on the order of $\sim g^2$):

$$\langle \mathbf{F} \rangle = \mathbf{F}_{0R} + \mathbf{F}_{1R} + \mathbf{F}_R. \tag{9}$$

Here, $\mathbf{F}_{0R}$ and $\mathbf{F}_{1R}$ are the RRF components of the second order in the field amplitude, representing the sums of independent contributions of the $\mathbf{E}_0$ and $\mathbf{E}_1$ fields,

$$\mathbf{F}_{0R} = \hbar\gamma \sum_{\alpha, j} \frac{\langle \mathbf{J}_{j\alpha} \rangle}{|\nu_\alpha|^2}, \tag{10}$$

$$\mathbf{J}_{j\alpha} = \frac{i}{2}(V_{j\alpha} \nabla V_{j\alpha}^* - \text{c.c.}), \quad j = x, y, z,$$

$$\mathbf{F}_{1R} = \hbar\gamma \sum_{\alpha, j} \frac{\Delta_\alpha}{|\nu_\alpha|^4} \langle (\mathbf{v} \cdot \nabla V_{j\alpha}) \nabla V_{j\alpha}^* + \text{c.c.} \rangle - \hbar i \sum_{\alpha, j} \frac{\Delta_\alpha^2 - \gamma_\perp^2}{|\nu_\alpha|^4} \langle [\mathbf{v} \times [\nabla V_{j\alpha} \times \nabla V_{j\alpha}^*]] \rangle, \tag{11}$$

and $\mathbf{F}_R$ is the RRF component of the fourth order in the field amplitude:

$$\begin{aligned} \mathbf{F}_R = -2\hbar\gamma \sum_\alpha \frac{\langle \mathbf{J}_\alpha I_\alpha \rangle}{|\nu_\alpha|^4} + \frac{\hbar}{|\nu_0|^2 |\nu_1|^2} \Bigg\{ \Lambda \Bigg[\langle I_0 \nabla I_1 \rangle \\ + \sum_j \langle I_{j0} \nabla I_{j1} \rangle + \frac{1}{2} \sum_{j \neq l} (\langle I_0^{lj} \nabla I_1^{jl} \rangle + \text{c.c.}) \Bigg] \\ - \Lambda_1 \Bigg[\langle I_0 \mathbf{J}_1 \rangle + \sum_j \langle \mathbf{J}_{j0} I_{j1} \rangle + \frac{1}{2} \sum_{j \neq l} (\langle \mathbf{J}_0^{jl} I_1^{jl} \rangle + \text{c.c.}) \Bigg] \\ - \Lambda_0 \Bigg[\langle I_1 \mathbf{J}_0 \rangle + \sum_j \langle \mathbf{J}_{j1} I_{j0} \rangle + \frac{1}{2} \sum_{j \neq l} (\langle \mathbf{J}_1^{jl} I_0^{lj} \rangle + \text{c.c.}) \Bigg] \Bigg\}, \end{aligned} \tag{12}$$

$$\mathbf{J}_\alpha = \sum_j \mathbf{J}_{j\alpha},$$

$$\mathbf{J}_\alpha^{jl} = \frac{i}{2}(V_{l\alpha} \nabla V_{j\alpha}^* - V_{j\alpha}^* \nabla V_{l\alpha}), \quad \mathbf{J}_{j\alpha} = \mathbf{J}_\alpha^{jj},$$

$$I_\alpha = \sum_j I_{j\alpha}, \quad I_{j\alpha} = |V_{j\alpha}|^2,$$

$$I_\alpha^{jl} = V_{j\alpha}^* V_{l\alpha}, \quad j, l = x, y, z,$$

$$\Lambda = [(\Delta_1 - \Delta_0)(1 - \cos\chi) - \gamma \sin\chi],$$

$$\Lambda_1 = [\gamma(1 + \cos\chi) - 2\Delta_1 \sin\chi],$$

$$\Lambda_0 = [\gamma(1 + \cos\chi) + 2\Delta_0 \sin\chi],$$

$$\nu_\alpha = |\nu_\alpha| \exp(i\chi_\alpha), \quad \chi = 2(\chi_1 - \chi_0), \tag{13}$$

where angular brackets denote averaging over the microscopic spatial oscillations.

It should be noted that the quantities $I_{j\alpha}(\mathbf{r})$ and $\mathbf{J}_{j\alpha}(\mathbf{r})$ are proportional, respectively, to the "intensity" (square modulus of the complex amplitude) and the energy flux density of the field components (polarized in the direction of the unit vector $\mathbf{e}_j$ and belonging to the same αth mode) in the superposition (7). The supplementary

[4] In other words, the RRF exists provided that the averaging procedure does not reduce the force magnitude $|\langle \mathbf{F} \rangle| \sim |\mathbf{F}| \propto k$ in the order of magnitude (with respect to the parameter $\lambda/\lambda_M \ll 1$) [5].

quantities I_α^{jl} and $\mathbf{J}_\alpha^{jl}$ with $j \neq l$ represent a measure of "mixing" of the waves with different polarizations belonging to the same mode (indeed, I_α^{jl}, $J_\alpha^{jl} = 0$ provided that at least one of the amplitudes $|V_{j\alpha}|$ or $|V_{l\alpha}|$ is zero). In a particular case of the 2D field configuration, when all waves in the superposition (7) are polarized along one of the Cartesian axes $\mathbf{e}_z$ (i.e., $\mathbf{E}_\alpha^\sigma \propto \mathbf{e}_z$), we obtain $J_\alpha^{jl} = I_\alpha^{jl} = 0$ for $j \neq l$ and $I_{j\alpha} \propto J_{j\alpha} \propto \delta_{zj}$, so that expressions (12) convert into relationships (25) derived in [5].

Essentially new interference terms in expression (12), which are due to the polarization effects, are related to correlators of the intensity–flux type ($\langle \mathbf{J}_{j\alpha}, I_{l\alpha} \rangle$, $j \neq l$) referring to the waves of different polarizations with the same frequencies (intramode interference) and correlators of the intensity–flux and intensity–intensity types ($\langle \mathbf{J}_{j\alpha'}, I_{l\alpha} \rangle$, $\langle I_{j\alpha} \nabla I_{l\alpha'} \rangle$, $j \neq l$, $\alpha \neq \alpha'$) referring to the waves of different polarizations and different frequencies (intermode interference). The intermode interference is also determined by the correlators involving mixed products of the projections of the complex field amplitudes and their derivatives of the types $\langle \mathbf{J}_\alpha^{jl} I_{\alpha'}^{lj} \rangle$ and $\langle I_\alpha^{jl} \nabla I_{\alpha'}^{lj} \rangle$, where $\alpha \neq \alpha'$ and $j \neq l$.

One of the most important factors determining the light-induced motion of resonance particles is the character of the spatial RRF structure. It should be kept in mind that, under the conditions studied, the principal part of the RRF expansion into perturbative series (i.e., the $\mathbf{F}_{0R}$ force component) always possesses a purely vortex structure. Indeed, calculation of the average radiation flux densities $\langle \mathbf{J}_\alpha \rangle$ in the case of field superpositions of the type (7) yields

$$\langle \mathbf{J}_{j\alpha} \rangle = \sum_\sigma \mathbf{k}_\alpha^\sigma |a_{j\alpha}^\sigma|^2 + \frac{1}{2} \times \left[\sum_{\eta > \gamma} (\mathbf{k}_\alpha^\eta + \mathbf{k}_\alpha^\gamma)(a_{j\alpha}^\eta)^* a_{j\alpha}^\gamma \exp[i(\phi_\alpha^\gamma - \phi_\alpha^\eta)] + \text{c.c.} \right], \quad (14)$$

where

$$\phi_\alpha^\gamma = \mathbf{k}_\alpha^\gamma \cdot \mathbf{r}, \quad a_{j\alpha}^\gamma = \mathbf{e}_j \cdot \mathbf{E}_\alpha^\gamma.$$

In the double sum, the indices γ, η refer to all possible pairs of wavevectors with close orientations:

$$|\mathbf{k}_\alpha^\gamma - \mathbf{k}_\alpha^\sigma| \ll k.$$

Expressions (14) and (10) are clearly indicative of the vortex character of the $\mathbf{F}_{0R}$ force component, since

$$\text{div}\mathbf{F}_{0R} = 0.$$

Note that this is essentially an expression of the Earnshaw theorem [23] for the RRF. Such a "defect" in the spatial RRF structure formed in a weak monochromatic field ($\mathbf{E}_1 = 0$) is basically inavoidable. Indeed, although a correction $\mathbf{F}_R$ to the rectified radiative force in the fourth order of smallness with respect to the field amplitude contains both vortex and potential components, the ratio of the former to the latter is fixed and cannot be changed arbitrarily by modifying the field parameters and configurations [6]. This circumstance significantly limits the possibility of using weak monochromatic fields for controlling the motion and spatial localization of resonance particles. An essentially different physical situation is observed in the case of bichromatic fields, provided that the total radiation flux densities for each frequency mode turn zero,

$$\sum_j \langle \mathbf{J}_{j\alpha} \rangle = 0, \quad j = x, y, z, \quad (15)$$

which implies that the principal vortex component of the RRF is suppressed ($\mathbf{F}_{0R} = 0$).

Now we will consider three examples of particular 3D field configurations satisfying conditions (15). In these examples, $\mathbf{F}_{1R}$ is the friction force and the $\mathbf{F}_R$ exhibits either a purely potential or potential–vortex character with a fully controllable ratio of the vortex and potential components.

3.1. Mutually Orthogonal Standing Waves

Let us consider mutually orthogonal standing waves (Fig. 1a):

$$V_{x\alpha} = V_\alpha \cos(k_\alpha z), \quad V_{y\alpha} = V_\alpha \cos(k_\alpha x), \quad V_{z\alpha} = V_\alpha \cos(k_\alpha y). \quad (16)$$

Upon substituting (16) into Eqs. (10)–(12), we obtain

$$\mathbf{F}_{1R} = -m\kappa\mathbf{v}, \quad \kappa = -\frac{\hbar k^2 \gamma}{m}\left(\frac{V_0^2 \Delta_0}{|\nu_0|^4} + \frac{V_1^2 \Delta_1}{|\nu_1|^4}\right), \quad (17)$$

$$\mathbf{F}_R = -\nabla U(\mathbf{r}),$$

$$U(\mathbf{r}) = -\frac{\hbar \Gamma_1 k V_1^2 V_0^2}{4\delta k |\nu_1|^2 |\nu_0|^2}\left\{\sum_j \cos(2\delta k \mathbf{e}_j \cdot \mathbf{r}) + \frac{1}{2}\sum_{i \neq j} (\cos[\delta k(\mathbf{e}_i - \mathbf{e}_j) \cdot \mathbf{r}] + \cos[\delta k(\mathbf{e}_i + \mathbf{e}_j) \cdot \mathbf{r}])\right\},$$

where

$$\delta k = k_1 - k_0 = \frac{\Delta_1 - \Delta_0}{c},$$

m is the particle mass, and Γ_1 is a function of the relaxation constants and frequency detunings:

$$\Gamma_1 = \gamma\gamma_\perp(\Delta_1 - \Delta_0)\left(\frac{1}{|\nu_1|^2} + \frac{1}{|\nu_0|^2}\right).$$

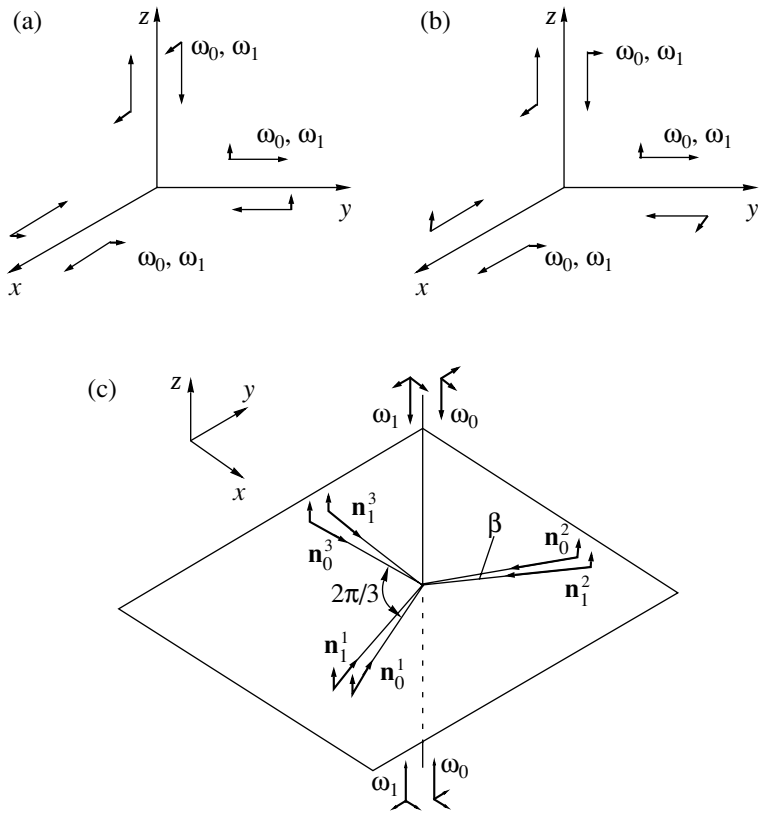


Fig. 1. Three-dimensional optical field configurations satisfying conditions (15): (a) a superposition of mutually orthogonal standing waves; (b) 3D*Lin* ⊥ *Lin* field configuration; (c) a superposition of standing (along the *z* axis) and linearly polarized (in the same direction) traveling waves. Long arrows show the directions of wave propagation; short arrows indicate polarizations of the interfering waves; ω_α is the wave frequency, $\mathbf{n}_\alpha^\sigma$ are the unit vectors in the propagation directions of waves with frequencies ω_α; $\beta \ll 1$ is the angular deviation.

Thus, $\mathbf{F}_{1R}$ is a friction force (for $\kappa(\Delta_1, \Delta_0) > 0$), while the rectified radiative force $\mathbf{F}_R$ exhibits a purely potential character. In the vicinity of the RRF nodes $\mathbf{r}_0$ corresponding to the point of minima for the $U(\mathbf{r})$ function,

$$\mathbf{r}_0 = \frac{\pi}{\delta k}(m\mathbf{e}_x + [m+2n]\mathbf{e}_y + [m+2p]\mathbf{e}_z), \qquad (18)$$

the potential has a spherically symmetric character. For $2\delta kR < 1$, this potential can be presented in the following form:

$$U(\mathbf{r}) \approx \frac{1}{2}U_0\delta k^2R^2,$$
$$U_0 = \hbar\omega_0\left(\frac{\gamma^2}{|\nu_1|^2} + \frac{\gamma^2}{|\nu_0|^2}\right)\frac{V_1^2V_0^2}{|\nu_1|^2|\nu_0|^2}, \qquad (19)$$

where $m, n, p \in Z$ (Z is the set of integers) and $R = |\mathbf{r} - \mathbf{r}_0|$ is the displacement from the RRF node; in writing (19), an insignificant constant additive was omitted.

It should be noted that the RRF nodes of the $\mathbf{r}_0$ type form a body-centered-cubic (bcc) lattice, the period of which ($\pi/\delta k$) can be controlled by varying the field detunings.

3.2. A Lin ⊥ Lin Wave Superposition (3D Lin ⊥ Lin Configuration)

Let us consider a superposition of waves with *Lin* ⊥ *Lin* configuration:

$$V_{x\alpha} = V_\alpha(e^{ik_\alpha z} + e^{-ik_\alpha y}),$$
$$V_{y\alpha} = V_\alpha(e^{ik_\alpha x} + e^{-ik_\alpha z}), \qquad (20)$$
$$V_{z\alpha} = V_\alpha(e^{ik_\alpha y} + e^{-ik_\alpha x}).$$

In this superposition, each wave propagating along one of the three Cartesian coordinate axes is supplemented with an opposite wave (counterwave) of the same frequency, polarized in the perpendicular direction (Fig. 1b). In comparison with the configuration depicted in Fig. 1a, only the polarization direction of one wave in each pair is changed. However, a spatial structure of the RRF exhibits a significant qualitative variation, which can be considered as a manifestation

of the polarization effects in the resonance light-induced pressure.

Indeed, the friction coefficient κ_1 and the RRF component $\mathbf{F}_R$ in this case are described by the following expressions:

$$\kappa_1 = 4\kappa, \quad \mathbf{F}_R = -\nabla U + \mathrm{rot}\mathbf{A}, \tag{21}$$

where κ is a coefficient determined by formula (17). The scalar (U) and vector ($\mathbf{A}$) RRF potentials introduced in (21) can be expressed as follows:

$$\mathbf{A} = A_0\Psi(\mathbf{r})\mathbf{e}, \quad \mathbf{e} = \frac{1}{\sqrt{3}}(\mathbf{e}_x + \mathbf{e}_y + \mathbf{e}_z),$$

$$A_0 = \frac{2\sqrt{3}\hbar k V_0^2 V_1^2 \Gamma}{|\nu_0|^2|\nu_1|^2\delta k},$$

$$\Psi(\mathbf{r}) = \sin[\delta k(y+z)] + \sin[\delta k(x+z)] + \sin[\delta k(x+y)],$$

$$\Gamma = \gamma(1+\cos\chi) + (\Delta_0 - \Delta_1)\sin\chi = (\Delta_1\Delta_0 + \gamma_\perp^2)\left(\frac{\gamma}{|\nu_0|^2} + \frac{\gamma}{|\nu_1|^2}\right), \tag{22}$$

$$U = -U_0\left(\sum_j \cos(2\delta k\mathbf{e}_j\cdot\mathbf{r}) + \frac{1}{2}\sum_{i\neq j}(3\cos[\delta k(\mathbf{e}_i+\mathbf{e}_j)\cdot\mathbf{r}] + 2\cos[\delta k(\mathbf{e}_i-\mathbf{e}_j)\cdot\mathbf{r}])\right),$$

$$U_0 = -\tan\left(\frac{\chi}{2}\right)A_0 = \hbar\omega_0\left(\frac{\gamma^2}{|\nu_1|^2} + \frac{\gamma^2}{|\nu_0|^2}\right)\frac{V_1^2V_0^2}{|\nu_0|^2|\nu_1|^2}.$$

It is seen that the RRF represents a combination of the potential force and the vortex component

$$\mathbf{F}_{\mathrm{vort}} = \mathrm{rot}\mathbf{A}.$$

The ratio of the two components is proportional to

$$\tan\left(\frac{\chi}{2}\right) = \gamma_\perp\frac{\Delta_0 - \Delta_1}{\Delta_1\Delta_0 + \gamma_\perp^2}$$

and can be controlled (virtually arbitrarily) by adjusting the field frequencies. Indeed, for

$$\Delta_1\Delta_0 = -\gamma_\perp^2$$

the RRF is purely potential, while for

$$\Delta_1\Delta_0 \sim -\gamma_\perp^2 \quad \left(\left|\tan\left(\frac{\chi}{2}\right)\right| \gg 1\right),$$

a small vortex "admixture" appears in the still dominating potential component, and for

$$\left|\tan\left(\frac{\chi}{2}\right)\right| \ll 1 \quad \left(\gamma g < |\Delta_0 - \Delta_1| \ll \left|\gamma_\perp + \frac{\Delta_0\Delta_1}{\gamma_\perp}\right|\right)$$

the vortex component begins to prevail and the potential component becomes a small admixture.

The positions of the RRF nodes $\mathbf{r}_0$ in which the potential function exhibits absolute minima is determined (as well as in the preceding example) by formula (18). However, the potential in a small vicinity of these nodes is no longer spherically symmetric. In a small region near $\mathbf{r}_0$, such that $2\delta kr < 1$, the potential can be expressed as

$$U(r) \approx \frac{U_0}{2}(16\eta + 13[\xi^2 + \zeta^2]) + \mathrm{const},$$

$$\eta = \frac{\delta k(x+y+z)}{\sqrt{3}}, \quad \xi = \frac{\delta k(2z-x-y)}{\sqrt{6}}, \quad \zeta = \frac{\delta k(x-y)}{\sqrt{2}}.$$

According to this, the potential level surfaces at the nodes $\mathbf{r}_0$ possess the shape of an ellipsoid of revolution with an axis parallel to the bisector of the first octant of the Cartesian coordinate system—the straight line C determined by the equation

$$\mathbf{r} = \mathbf{r}(\sigma) = \sigma\mathbf{e}, \quad -\infty < \sigma < \infty.$$

This circumstance is directly related to the symmetry of the optical field, since the line C is a third-order symmetry axis for the 3D$Lin \perp Lin$ field configuration: rotation of all wave vectors and polarization vectors about this axis by an angle of $2\pi/3$ leaves the initial configuration unchanged. The presence of the symmetry axis C also determines to a considerable extent the structure of the vortex field $\mathbf{F}_{\mathrm{vort}}$. Indeed, according to expressions (21) and (22),

$$\mathbf{F}_{\mathrm{vort}}\cdot\mathbf{e} = 0,$$

which implies that the vector field lines of the vortex RRF component are lying in the Π_c planes (determined by the equations $\mathbf{r}\cdot\mathbf{e} = c$) perpendicular to the symmetry axis $\mathbf{e}$. The set of periodically arranged lines

$$\mathbf{r} = \mathbf{r}(\sigma, \mathbf{r}_0) = \mathbf{r}_0 + \sigma\mathbf{e},$$

parallel to the axis C represents the nodal lines for the vortex RRF force component: $\mathbf{F}_{\mathrm{vort}}(\mathbf{r}(\sigma, \mathbf{r}_0)) = 0$. In addition, taking into account that

$$\nabla\Psi\cdot\mathbf{F}_{\mathrm{vort}} = 0,$$

we infer that the vector lines of the vortex force field represent the curves of intersection of the level planes Π_c with the level surfaces of the function $\Psi(\mathbf{r})$. In a small vicinity of the nodal lines $\mathbf{r} = \mathbf{r}(\sigma, \mathbf{r}_0)$, these curves appear as circumferences in the Π_c plane with

the centers occurring at the points of intersection of the planes Π_c and the nodal lines $\mathbf{r} = \mathbf{r}(\sigma, \mathbf{r}_0)$. As is demonstrated below, the resonance particles can perform the light-induced rotational motions about these centers (see the next Section).

The symmetry axis C in the 3D*Lin* ⊥ *Lin* configuration (and, hence, the "particle rotation axes") can be readily changed by consistently changing the wave polarizations. For example, the direction of this axis in a field of the 3D*Lin* ⊥ *Lin* configuration is determined by the formulas

$$V_{x\alpha} = V(\exp(-ik_\alpha z) + \exp(-ik_\alpha y)),$$
$$V_{y\alpha} = V(\exp(ik_\alpha x) + \exp(ik_\alpha z)),$$
$$V_{z\alpha} = -V(\exp(-ik_\alpha x) + \exp(ik_\alpha y)),$$

and represented by the vector

$$\mathbf{e}_1 = \frac{\mathbf{e}_x + \mathbf{e}_y - \mathbf{e}_z}{\sqrt{3}}.$$

3.3. A Superposition of Standing and Linearly Polarized Traveling Waves

Let us consider a superposition of standing (along the z axis) and linearly polarized (in the same direction) traveling waves with a symmetric triangular configuration, intersecting in the xy plane (Fig. 1c):

$$V_{x1} = \hat{V}_1 \cos[(k + \delta k)z],$$
$$V_{y1} = -\hat{V}_1 \cos[(k + \delta k)z],$$
$$V_{j0} = \hat{V}_0 \cos kz, \quad j = x, y, \tag{23}$$
$$V_{z\alpha}(\mathbf{r}) = V_\alpha \sum_{\sigma=1}^{3} \exp[i\hat{\varphi}_\alpha^\sigma(\mathbf{r})],$$

where

$$\hat{\varphi}_0^\sigma(\mathbf{r}) = k\mathbf{n}_0^\sigma \cdot \mathbf{r} + \varphi_0^\sigma,$$
$$\hat{\varphi}_1^\sigma(\mathbf{r}) = (k + \delta k)\mathbf{n}_1^\sigma \cdot \mathbf{r} + \varphi_1^\sigma, \quad k = k_0,$$

the unit vectors determining the directions of wavevectors $\mathbf{n}_\alpha^\sigma$ are lying in the xy plane,

$$\mathbf{n}_0^\sigma = (\cos(2\pi\sigma/3), \sin(2\pi\sigma/3), 0), \quad \sigma = 1, 2, 3,$$

and the system of vectors $\mathbf{n}_1^\sigma$ is "rigidly" rotated about the z axis relative to $\mathbf{n}_0^\sigma$ vectors by a small angle $\beta \ll 1$. Assuming that

$$\delta k/k \ll \beta \ll 1,$$

restricting the consideration to a region

$$\rho = \sqrt{x^2 + y^2} < 1/\delta k,$$

and using Eqs. (10)–(12) and (23), we obtain the following expressions:

$$\mathbf{F}_{1R} = -m(\kappa_\perp \mathbf{v}_\perp + \kappa_z \mathbf{e}_z v_z),$$
$$\kappa_\perp = -\frac{3\hbar k^2}{m}\gamma\left(\frac{V_1^2\Delta_1}{|\nu_1|^4} + \frac{V_0^2\Delta_0}{|\nu_0|^2}\right),$$
$$\kappa_z = -\frac{2\hbar k^2}{m}\gamma\left(\frac{\hat{V}_1^2\Delta_1}{|\nu_1|^4} + \frac{\hat{V}_0^2\Delta_0}{|\nu_0|^2}\right),$$
$$\mathbf{F}_R = -\frac{2\hbar\Gamma}{|\nu_0|^2|\nu_1|^2}\left\{\frac{V_1^2 V_0^2}{\beta}\right.$$
$$\times\left[\nabla U(\mathbf{r}) + \tan\left(\frac{\chi}{2}\right)\mathrm{rot}\,\mathbf{a}(\mathbf{r})\right]$$
$$\left. + \tan\left(\frac{\chi}{2}\right)\frac{\hat{V}_1^2\hat{V}_0^2}{4\delta k}\mathbf{e}_z\frac{\partial U_1(z)}{\partial z}\right\}, \tag{24}$$

where

$$\mathbf{v}_\perp = \mathbf{e}_x v_x + \mathbf{e}_y v_y, \quad U_1(z) = 2\cos(2\delta k z),$$
$$\mathbf{a}(\mathbf{r}) = 2\Psi_1(\mathbf{r})\mathbf{e}_z,$$
$$\Psi_1(\mathbf{r}) = \sum_{\sigma=1}^{3} \cos(\beta\sqrt{3}k\mathbf{n}_0^\sigma \cdot \mathbf{r} + \xi_\sigma),$$
$$U(\mathbf{r}) = -\frac{2}{\sqrt{3}}\sum_{\sigma=1}^{3} \sin(\beta\sqrt{3}k\mathbf{n}_0^\sigma \cdot \mathbf{r} + \xi_\sigma),$$
$$\xi_1 = [\varphi_0^2 - \varphi_1^2] - [\varphi_0^3 - \varphi_1^3],$$

and the constant phases ξ_2 and ξ_3 are obtained from ξ_1 by the cyclic permutation of indices. In this force field, the motions in the xy plane and along the z axis are completely separated and independent. A force acting along the z axis is always potential and possesses a periodic spatial structure with the period $\pi/\delta k$. A ratio of the vortex and potential force components acting in the directions parallel to the xy plane is proportional to $\tan(\chi/2)$ and, hence, can be fully controlled by selecting appropriate field frequency detunings Δ_1 and Δ_0 Here, both potential and vortex RRF components possess a periodic spatial structure.

Figure 2 shows the pattern of level lines for the potential function $U(x, y)$. The point of intersection of the separatrix lines represent the saddle points forming a planar hexagonal lattice with the period

$$\lambda_M = 4\pi/3k\beta,$$

determined by the angular deviation β. Located at the centers of the triangular separatrix cells are the points of minima (denoted by dots in Fig. 2) and maxima of

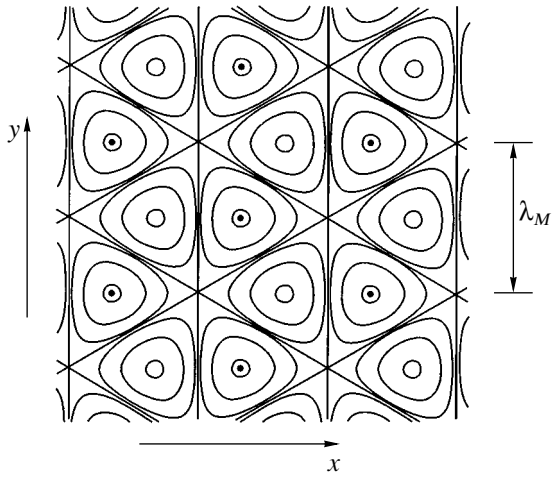


Fig. 2. A schematic diagram showing the level lines of the potential function $U(r)$ described by Eq. (24) determining the potential RRF component in a field configuration of the third type, acting in the directions parallel to the xy plane. Dots at the centers of some cells indicate the points of minima of the function $U(r)$; $\lambda_M = 4\pi/3k\beta$ is the macroscopic spatial scale.

the function $U(x, y)$. The extremal points of both types also form mutually shifted planar hexagonal lattices.

Figure 3 shows the vector field lines of the vortex RRF component. Here, the network of the separatrix lines (described by the equation $\Psi_1(\mathbf{r}) = -1$) forms the so-called kagome lattice. Vector lines inside the triangular and hexagonal cells of this lattice represent closed curves surrounding special points (centers) coinciding with the positions of saddle point (in the hexagonal cells) and extremal points (in the triangular cells) of the potential function $U(x, y)$.

4. FEATURES OF THE LIGHT-INDUCED MOTION OF PARTICLES

Let b denote the characteristic size of a region featuring intersection of real laser beams and containing resonance particles. The interference effects in the light-induced motion of the resonance particles can be fully manifested provided that the macroscopic spatial scale λ_M (in the field model under consideration, this value corresponds to the period of an optical superlattice formed as a result of the superposition of plane waves (7)) does not exceed b:

$$\lambda_M < b.$$

Taking into account that

$$\lambda_M \sim \frac{1}{|\delta k|} = \frac{c}{|\Delta_1 - \Delta_0|},$$

and a typical value of the spontaneous relaxation rate is $\gamma \sim 10^8\ \mathrm{s}^{-1}$, this condition can be fulfilled even for wide beams ($b \sim 10$–20 cm) if the field frequency difference is significantly greater than the resonance width:

$$|\Delta_1 - \Delta_0| \gg \gamma.$$

With an allowance for this circumstance, an optimum set of the field frequencies and amplitudes (selected based on the criterion of maximum RRF at a fixed value of the parameter g) must obey the following relationships:

$$|\nu_0| \sim \gamma, \quad |\Delta_1| \gg \gamma, \quad \left|V_{j1}^2/\nu_1\nu_0\right| \sim \left|V_{j0}/\nu_0\right|^2 \sim g \ll 1,$$
$$\left|V_{j1}/\nu_1\right|^2 \ll g.$$

For $\Delta_0 < 0$, the field $\mathbf{E}_0$ is responsible for the cooling process and the field $\mathbf{E}_1$ (together with $\mathbf{E}_0$), for the man-

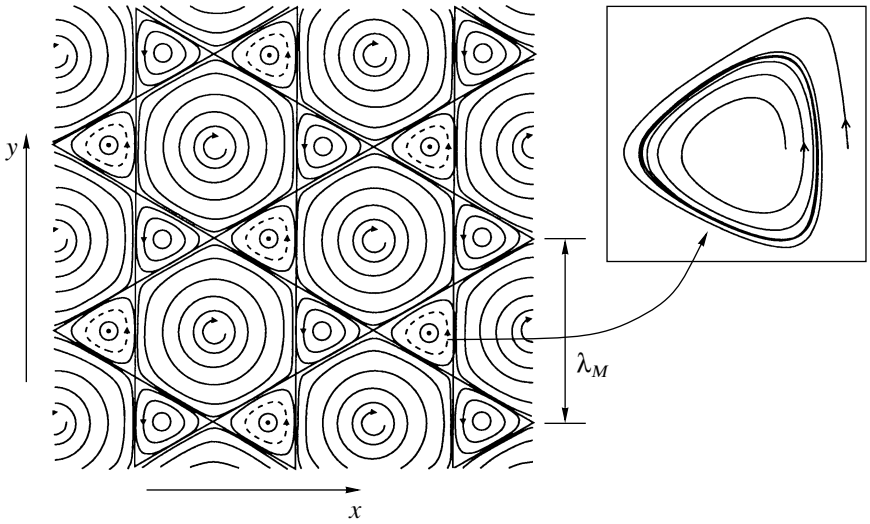


Fig. 3. A schematic diagram of the kagome lattice formed by vector lines of a vortex RRF component induced by interfering optical fields with a configuration of the third type. Dashed lines show stable closed particle trajectories (limiting cycles), which may appear only in the triangular cells containing the points of minima of the potential function $U(\mathbf{r})$. The inset shows a typical trajectory of a particle falling within the separatrix lattice cell.

ifestation of interference effects. The characteristic values of the friction coefficient and the RRF can be estimated from the following simple relationships:

$$\kappa = \omega_r g, \quad \omega_R = \frac{\hbar k^2}{m}, \quad |\mathbf{F}_R| = F \sim \hbar k \gamma g^2.$$

The RRF component $\mathbf{F}_R$ will be the main factor determining (together with the friction force) the light-induced motion of an ensemble of resonance particles with the temperature T (expressed in the energy units) under the following conditions:

$$U_g \ll T \ll \Delta U, \tag{25}$$

where U_g is the depth of microscopic potential wells formed under the action of rapidly oscillating (with a period $\sim\lambda$) gradient forces [1, 2], $\Delta U \sim F\lambda_M$ is the characteristic work performed by the RRF for the transfer of particles over a macroscopic distance λ_M (for the potential RRFs, ΔU is the depth of macroscopic potential wells). The temperature T corresponding to the so-called Doppler cooling limit is established within a characteristic time $\sim\kappa^{-1}$ as a result of the competition between the Doppler cooling and the diffusion processes in the velocity space [1, 2] determined by the quantum fluctuations of radiative forces.

For a selected ratio of the main problem parameters, the rate of the velocity diffusion is [1, 2]

$$D \sim \left(\frac{\hbar k}{m}\right)^2 \gamma g, \quad T \sim \frac{mD}{\kappa} \sim \hbar\gamma, \quad U_g \sim \hbar\gamma g.$$

For $g \ll 1$, the left-hand inequality in (25) is always fulfilled, which implies impossibility of confining particles at the small-scale potential wells. The right-hand inequality (25) indicates that the field must not be very weak:

$$1 \gg g^2 > 1/k\lambda_M.$$

An important feature of the light-induced motion of particles in a weak bichromatic field for $\kappa > 0$ is the overdamped character of this motion which is caused by a large friction force:

$$\frac{\Omega^2}{\kappa^2} = \varepsilon \sim \frac{\gamma}{\omega_R}\frac{\Delta k}{k} \ll 1, \tag{26}$$

where $\Omega = \sqrt{F\Delta k/m}$ is the characteristic frequency of motion in the absence of friction; $\Delta k = |\delta k|$ for a field configuration of the first or second type and $\Delta k = k|\beta|$, for the third type (see the preceding section). The smallness of the parameter ε is related to smallness of the ratio of the microscopic and macroscopic scales. Indeed, for the typical values of $\gamma/\omega_R \sim 10^2$–$10^3$, $k = 10^5$ cm^{-1}, and $\Delta k \sim 1$ cm^{-1}, relationship (26) yields an estimate $\varepsilon \sim 10^{-2}$–10^{-3}.

When it is necessary to provide for the condition $|\tan(\chi/2)| \ll 1$, an interesting situation takes place when both detunings are large:

$$\gamma \ll |\Delta_0| < |\Delta_1|.$$

In this case,

$$\left|\tan\left(\frac{\chi}{2}\right)\right| = \left|\frac{(\Delta_1 - \Delta_0)\gamma_\perp}{\Delta_1\Delta_0}\right| \ll 1,$$

and the overdamping condition takes the form of inequality

$$\varepsilon = \frac{\Delta_0^2}{\omega_R\gamma}\frac{\Delta k}{k} \ll 1$$

which is well fulfilled in a broad range of parameters provided that $\Delta k/k \ll 1$.

Mathematically, the condition (26) is manifested by the fact that, upon the passage to dimensionless variables

$$\mathbf{r} \longrightarrow \Delta k\mathbf{r}, \quad \mathbf{v} \longrightarrow \mathbf{v}/u_0,$$

$$t \longrightarrow \tau = \Omega^2 t/\kappa, \quad \mathbf{F}_R \longrightarrow \mathbf{u}(\mathbf{r}) = \mathbf{F}_R/m\kappa u_0,$$

(where $\mathbf{v}$ is the particle velocity and $u_0 = F/m\kappa$), the equations of particle motion under the RRF action transform into a system of singular perturbed differential equations

$$\frac{d\mathbf{r}}{d\tau} = \mathbf{v}, \quad \varepsilon\frac{d\mathbf{v}}{d\tau} + \mathbf{v} = \mathbf{u}(\mathbf{r}), \tag{27}$$

where the notations $\mathbf{v}$ and $\mathbf{r}$ are retained for the dimensionless quantities. The methods of investigation of the systems of this type and their reduction to the equations of lower dimensionality are well developed in the theory of differential equations [25, 26]. A solution to the system (27) with arbitrary initial conditions $\{\mathbf{r}_0, \mathbf{v}_0\}$ can be represented in the form of a combination of a rapid transient process described by the boundary functions of the type [25]

$$\Pi\mathbf{v}(\tau) \sim \exp(-\tau/\varepsilon), \quad \Pi\mathbf{r}(\tau) \sim \varepsilon\exp(-\tau/\varepsilon),$$

exponentially decaying within a characteristic time $\tau \sim \varepsilon(t \sim \kappa^{-1})$ and a slow motion over a surface (integral manifold [26]) of the type

$$\mathbf{v} = \mathbf{G}(\mathbf{r}, \varepsilon) \tag{28}$$

in the phase space. In our case, the function $\mathbf{G}$ can be determined using a regular expansion into series with respect to the parameter ε

$$\mathbf{G} = \mathbf{G}_0 + \varepsilon\mathbf{G}_1 + \ldots. \tag{29}$$

Substituting this expansion into Eqs. (27), we obtain a sequence of $\mathbf{G}_n$ values

$$\mathbf{G}_0 = \mathbf{u}(\mathbf{r}), \quad \mathbf{G}_1 = -\varepsilon(\mathbf{u}(\mathbf{r})\nabla)\mathbf{u}(\mathbf{r}) \tag{30}$$

and an equation describing the particle trajectory in a light wave field (for $\tau < 1/\varepsilon^2$, expansion (29) can be restricted to the first two terms):

$$\frac{d\mathbf{r}}{d\tau} = \mathbf{u}_1(\mathbf{r}) = \mathbf{u}(\mathbf{r}) - \varepsilon(\mathbf{u}(\mathbf{r})\nabla)\mathbf{u}(\mathbf{r}),$$
$$\mathbf{r}(0) = \mathbf{r}_0 + O(\varepsilon). \tag{31}$$

According to Eqs. (28) and (31), the stationary velocity of a resonance particle (established by the time $t > \kappa^{-1}$) adiabatically "follows" its spatial position. A relationship between the particle velocity $\mathbf{u}_1(\mathbf{r})$and the RRF component $\mathbf{F}_R(\mathbf{r})$ has a nonlocal character. The velocity $\mathbf{u}_1$ at each point $\mathbf{r}$ depends both on the force $\mathbf{F}_R(\mathbf{r})$ at this point and on the derivative of $\mathbf{F}_R$ with respect to the spatial coordinates. Therefore, the RRF vector lines in the general case do not coincide with the particle trajectory: $\mathbf{u}_1(\mathbf{r}) \neq \mathbf{u}(\mathbf{r})$. An allowance for the terms on the order of ε in the right-hand part of (31) is important when the RRF vortex component induced by a bichromatic field is dominating, since this very contribution may account for the instability of a rotational particle motion in this case.

The results of numerical calculations of the particle trajectories based on Eqs. (31) showed that the character of motion is highly sensitive with respect to both the spatial configuration and parameters of the light field (frequency detunings). For all configurations, there exists a broad range of these parameters for which the particles may perform finite motions in the cells of effective superlattices.

For the field configuration of the first type described by Eq. (16) (Fig. 1a), the motion is always potential ($\operatorname{curl}\mathbf{u}_1(\mathbf{r}) = 0$) and leads for $t \sim t_0 = \kappa/\Omega^2$ to the localization of particles at the sites of a cubic lattice corresponding to the local minima of a potential $U(\mathbf{r})$ determined by Eq. (17).

For the field configuration of the second type (3D*Lin* $\perp$ *Lin*, Fig. 1b), an analogous potential motion takes place only for specially selected detunings of the field frequencies: $\Delta_1\Delta_0 = -\gamma_\perp^2$ with $|\tan(\chi/2)| \longrightarrow \infty$. In the general case, when this condition is not fulfilled, $\operatorname{curl}\mathbf{u}_1(\mathbf{r}) \neq 0$ and the motion exhibits a vortex character. Figure 4 shows a typical particle trajectory for $|\tan(\chi/2)| \sim \varepsilon \ll 1$, representing a helix with nonmonotonically (!) varying radius wound on a nodal line of the RRF vortex component. As was noted in Section 3, the axes of particle rotation (parallel to the field symmetry axis) can be readily controlled by consistently changing the field polarizations.

For the field configuration of the third type described by Eq. (23) (Fig. 1c), the motions along the *z* axis and in the directions parallel to the *xy* plane are completely separated and independent. The motion

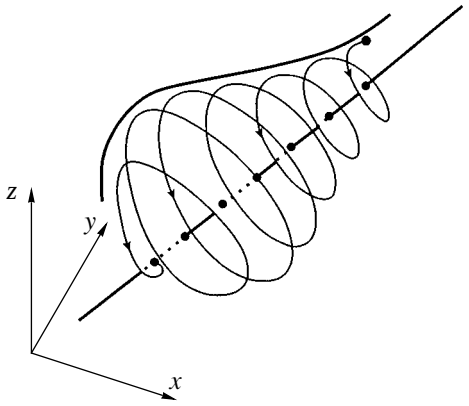


Fig. 4. A trajectory of the light-induced vortex motion ($\operatorname{curl}\mathbf{u}_1 \neq 0$) of resonance particles in a field with the 3D*Lin* $\perp$ *Lin* configuration.

along the *z* axis leads eventually to a stable particle grouping in the planes

$$z = (\pi/\delta k)n,$$

where *n* are the arbitrary integer numbers. The character of the vortex motion of particles in these planes is determined by the ratio of the $\tan(\chi/2)$ and ε. For

$$|\tan(\chi/2)| \lesssim 1/\sqrt{\varepsilon}, \quad \Gamma > 0,$$

the points $\mathbf{r}_m$ (indicated by dots in Figs. 2 and 3) corresponding to minima of the $U(\mathbf{r})$ function (see Eqs. (24)) are stable focuses of the system of differential equations (31) and, hence, the points of localization of the particles forming a regular hexagonal lattice. For

$$|\tan(\chi/2)| \gtrsim 1/\sqrt{\varepsilon}$$

these focuses become unstable, while stable limiting cycles (depicted by dashed lines in Fig. 3) appear inside the triangular regions of the kagome lattice. An explicit form of the parameter ξ for this bifurcation (Hopf bifurcation) can be derived from an analysis of stability of system (31):

$$\xi = |\tan(\chi/2)| - s, \tag{32}$$

where

$$s = \frac{\kappa}{G}, \quad \kappa \approx -\frac{3\omega_R\Delta_0 g_0\gamma}{|\nu_0|^2},$$
$$G^2 = \omega_R g_0 g_1 3\sqrt{3}|\Gamma\beta|, \quad g_\alpha = \left|\frac{V_\alpha}{\nu_\alpha}\right|^2,$$
$$\tan\left(\frac{\chi}{2}\right) \approx -\frac{\gamma_\perp\Delta_1}{\Delta_1\Delta_0 + \gamma_\perp^2}, \quad \Delta_0 < 0,$$

and the consideration was restricted for simplicity to the most interesting case of

$$|\Delta_0| < \gamma \ll |\Delta_1|, \quad g_0 \sim \left|\frac{V_1^2}{\nu_1 \nu_0}\right| \sim g.$$

The limiting cycles appear for $\xi > 0$; in the case of $0 < \xi/s \ll 1$, these cycles acquire the form of circular trajectories with the radius proportional to $\sqrt{\xi/s}$. The direction of rotation depends on the sign of the angular deviation β. As the sign of Γ changes to opposite, the limiting cycles pass to the neighboring triangular cells. Note that the adjustment to the regime of stable particle rotation depends in a sharp and complicated manner on the field frequency detunings and in a smooth manner, on the geometric factor (angular deviation).

For $|\tan(\chi/2)| \longrightarrow \infty$, when the RRF vortex component fully dominates, the orbits of rotating particles are "pressed" arbitrarily close to the boundaries of separatrix cells. Here, we may introduce small additive fluctuating terms (with a broad frequency spectrum) into the right-hand part of Eqs. (31) so as to model the real quantum fluctuations of the radiative forces [2]. In this case, the particles on the trajectories pressed to the cell boundaries will cross this boundary in the region of saddle points (i.e., the points of intersection of the separatrix lines in Fig. 3) and pass to the boundaries of the adjacent cells. As a result, the particles perform the infinite motion appearing as a random (brownian) walk over edges of the kagome lattice.[5] A similar phenomenon of the light-induced random walk of particles over the edges of effective square lattices in the case of strong fields with a 2D configuration was originally reported in [6].

Finally, let us present some numerical estimates illustrating the possibilities of a mechanical action of light upon atoms in the system studied. For certainty, we will consider resonance particles with $m = 40$ amu exposed to a field of mutually orthogonal standing weaves with $\hbar\omega_0 \approx 3$ eV, $\gamma \approx 3 \times 10^8$ s^{-1}, $\Delta_0 = -\gamma/5$, and $|\Delta_1| = 10^{12}$ s^{-1}. The electromagnetic wave intensities I_α for each frequency component ($\alpha = 0, 1$) were selected so as to satisfy the weak field criterion:

$$g = |V_{j0}/\nu_0|^2 = \left|V_{j1}^2/\nu_0\nu_1\right| = 2 \times 10^{-2},$$

which was achieved for $I_0 \approx 1.5$ mW/cm^2 and $I_1 \approx 10$ W/cm^2 (in this case, $|\rho^{(3)}/\rho^{(1)}| \sim 10^{-1}$). Then, using Eq. (17) and formulas derived in this section, we obtain the potential

$$\Delta U \approx U_0 \approx 7 \times 10^{-3} \text{ K},$$

the macroscopic lattice period

$$\lambda_M = 2\pi/|\delta k| = 1.8 \text{ mm},$$

and the other quantities:

$$T \approx 10^{-3} \text{ K } ((\Delta U/T) \approx 7), \quad \kappa \sim 10^4 \text{ s}^{-1},$$

$$t_0 \sim (\Omega^2/\kappa)^{-1} \sim 10^{-3} \text{ s} \quad (\varepsilon \sim 0.1),$$

$$F \sim 10^{-5} \text{ eV/cm}.$$

By significantly decreasing detuning of the field $\mathbf{E}_1$ (thus, increasing the macroscopic spatial scale λ_M), it is possible to create a superdeep potential well—a purely optical 3D trap for atoms—using bichromatic laser beams of a large diameter ($b \gtrsim 10$ cm). For example, in the case of $|\Delta_1| = 1.5 \times 10^{10}$ s^{-1} ($\lambda_M \approx 10$ cm), $I_0 \approx 1.5$ mW/cm^2, and $I_1 \approx 150$ mW/cm^2, the potential well depth is $\Delta U = 0.5$ K and $\Delta U/T \sim 500$(!) for the characteristic time of particle localization $t_0 \approx 0.03$ s. Thus, the right-hand inequality (25) characterizing the efficacy of the RRF action upon the resonance particles, can be satisfied with a large margin under quite realistic conditions.

For the comparison, it is interesting to note that the passage to a "strong field" regime [4]

$$\left|V_1^2/\gamma\Delta_1\right| \sim \left|V_0^2/\gamma\Delta_0\right| = g_1 \gg 1, \quad |\Delta_0| \gg \gamma,$$

for the same level of saturation ($\left|V_0^2/\Delta_0^2\right| = g$) and detuning $|\Delta_1|$ as in the example given above, the field intensities E_0 and E_1 must be increased by a factor of g_1^2/g and g_1/g, respectively. As a result, the RRF magnitude and the potential ΔU will increase only by a factor of g_1 ($g_1 \ll g_1^2/g$, g_1/g) and the $\Delta U/T$ ratio will remain unchanged (because $T \sim \hbar\gamma g_1 \gg \hbar\gamma$ [14]).

5. CONCLUSION

Under the physical conditions studied, the system is always characterized by a small degree of occupation of the atomic states ($\rho_{ii} \ll 1$) and by small values of the light-induced Stark shift (the latter circumstance is important for the spectroscopic applications). Nevertheless, the RRFs can be sufficiently large to effectively act upon an atomic ensemble with a temperature corresponding to the Doppler cooling limit.

Based on an analysis of the general relationships derived for the RRFs, we found symmetric configurations of the interfering waves for which the spatial structure of the light-induced force field can be effectively controlled by consistently changing the frequencies of the optical fields, which allows the ratio of the

[5] It is interesting to note that, according to the results of numerical calculations, the walk over boundaries of the separatrix cells is also observed in the absence of small fluctuating forces, provided that the calculation time is sufficiently large, which is probably related to the "noise" introduced by the so-called discretization errors. The time of a particle escape from the cell may depend on the numerical method selected. Similar discretization effects, arising during a dynamic chaos simulation in the Hamiltonian systems, are considered, for example, in monograph [27].

vortex and potential RRF components to be varied almost arbitrarily.

It was found for the field configurations studied that it is possible to remove, in a broad range of the control parameters, some fundamental limitations of the type of the Earnshaw optical theorem (proved for weak monochromatic fields [23], see also [5, 6]) prohibiting the 3D localization (confinement) of the resonance particles by means of the spontaneous light-induced pressure. The light-induced motion of a confined resonance particle proceeds inside an elementary cell of an effective optical superlattice (with a cubic or hexagonal structure for the field configurations studied). This motion is finite and exhibits a vortex or potential character, depending on the frequency detunings selected, and leads eventually to the localization (confinement) of particles at the RRF nodes or to their stable rotation along closed orbits inside the elementary lattice cells. The transition from a light-induced potential motion to the vortex motion and a change of the axes of particle rotation during the vortex motion can be also provided by consistently varying polarizations of the interfering waves without altering their propagation directions. A polarization effect of this type is manifested in a field configuration of the 3D*Lin* ⊥ *Lin* type.

In a situation of the absolutely dominating vortex RRF component, a very interesting regime of the light-induced infinite motion of particles is possible in the form of their random walk over edges of a planar superlattice of the kagome type (for the field configuration depicted in Fig. 1c).

The rectified radiative forces can be used for controlling the motion of resonance particles, creating stable periodic 3D structures in a cold atomic gas, and constructing purely optical (nonmagnetic) macroscopic traps (using laser beams of large diameter) capable of trapping large-size bunches of resonance particles. An example of interesting application is offered by the purely optical confinement of an ultracold rarefied plasma bunch with resonance ions [19, 28, 29].

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research, project nos. 99-02-16873, 01-02-06478-mas, and 01-02-06479-mas.

REFERENCES

1. A. P. Kazantsev, G. I. Surdutovich, and V. P. Yakovlev, *The Mechanical Action of Light on Atoms* (Nauka, Moscow, 1991).
2. V. G. Minogin and V. S. Letokhov, *The Pressure of Laser Radiation on Atoms* (Nauka, Moscow, 1986).
3. V. S. Letokhov, M. A. Ol'shanii, and Yu. B. Ovchinnikov, Quantum Semiclassic. Opt. **7**, 5 (1995).
4. A. P. Kazantsev and I. V. Krasnov, Pis'ma Zh. Éksp. Teor. Fiz. **46**, 264 (1987) [JETP Lett. **46**, 332 (1987)].
5. A. P. Kazantsev and I. V. Krasnov, Zh. Éksp. Teor. Fiz. **95**, 103 (1989) [Sov. Phys. JETP **68**, 59 (1989)].
6. A. P. Kazantsev and I. V. Krasnov, J. Opt. Soc. Am. B **6**, 2140 (1989).
7. A. P. Kazantsev and I. V. Krasnov, Phys. Lett. A **127**, 33 (1988).
8. R. Grimm, Yu. B. Ovchinnikov, A. I. Sidorov, and V. S. Letokhov, Phys. Rev. Lett. **65**, 1415 (1990).
9. V. S. Voĭtsekhovich, M. V. Danileĭko, A. M. Negriĭko, *et al.*, Zh. Éksp. Teor. Fiz. **99**, 393 (1991) [Sov. Phys. JETP **72**, 219 (1991)].
10. J. Juvanainen, Phys. Rev. Lett. **4**, 519 (1990).
11. R. Grimm, J. Soding, and Yu. B. Ovchinnikov, Pis'ma Zh. Éksp. Teor. Fiz. **61**, 362 (1995) [JETP Lett. **61**, 367 (1995)].
12. J. Soding, R. Grimm, Yu. B. Ovchinnikov, *et al.*, Phys. Rev. Lett. **78**, 1420 (1997).
13. T. T. Grove, B. C. Duncan, V. Sanchez-Villicana, *et al.*, Phys. Rev. A **51**, R4325 (1995).
14. I. V. Krasnov, Laser Phys. **4**, 906 (1994).
15. I. V. Krasnov, Zh. Éksp. Teor. Fiz. **107**, 1135 (1995) [JETP **80**, 632 (1995)].
16. H. Pu, T. Cai, N. P. Bigelow, *et al.*, Opt. Commun. **118**, 261 (1995).
17. E. A. Korsunsky and D. V. Kosachiov, J. Phys. B **30**, 5701 (1997).
18. G. Wasik and R. Grimm, Opt. Commun. **137**, 406 (1997).
19. A. P. Gavrilyuk, I. V. Krasnov, and N. Ya. Shaparev, Pis'ma Zh. Éksp. Teor. Fiz. **63**, 316 (1996) [JETP Lett. **63**, 324 (1996)].
20. A. P. Gavrilyuk, I. V. Krasnov, S. P. Polyutov, and N. Ya. Shaparev, Vychisl. Tekhnol. **4**, 43 (1999).
21. A. Hemmerich and T. W. Hansh, Phys. Rev. Lett. **68**, 1492 (1992).
22. T. Walker, P. Feng, D. Hoffmann, and R. S. Williamson, Phys. Rev. Lett. **69**, 2168 (1992).
23. A. Ashkin and J. P. Gordon, Opt. Lett. **8**, 511 (1983).
24. A. P. Kazantsev, V. S. Smirnov, G. I. Surdutovich, *et al.*, J. Opt. Soc. Am. **2**, 1731 (1985).
25. A. B. Vasil'eva and V. F. Butuzov, *Asymptotic Expansions of Singular-Perturbed Equations* (Nauka, Moscow, 1973).
26. Yu. A. Mitropol'skiĭ and O. B. Lykova, *Integral Manifolds in Nonlinear Mechanics* (Nauka, Moscow, 1973).
27. G. M. Zaslavskiĭ, V. Z. Sagdeev, D. A. Usikov, and A. A. Chernikov, *Weak Chaos and Quasi-Regular Structures* (Nauka, Moscow, 1991).
28. A. P. Gavrilyuk, I. V. Krasnov, and N. Ya. Shaparev, Laser Phys. **8**, 653 (1998).
29. S. A. Gavrilyuk, I. V. Krasnov, S. P. Polyutov, and N. Ya. Shaparev, in *Proceedings of 5th Russian-Chinese Symposium on Laser Physics and Technologies* (Tomsk State Univ., Tomsk, 2000).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 93, No. 5, 2001, pp. 998–1003.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 5, 2001, pp. 1150–1156.

ATOMS, SPECTRA, RADIATION

Anomalous Behavior of the Fluorescence Spectra of Perylene in *n*-Octane under Pressure and Temperature Variations

S. N. Balashov*, V. P. Karpov, O. N. Korotaev, and M. F. Shchanov
Moscow State Pedagogical University, ul. Malaya Pirogovskaya 1, Moscow, 119882 Russia
**e-mail: bserg@enigmaenter.ru*
Received May 23, 2001

Abstract—Pressure-induced transformations of the vibronic spectra of perylene in *n*-octane were studied. An increase in pressure from normal to 10 kbar caused a sharp change in the multiplet structure of the spectra. An anomalous, almost linear temperature dependence of spectral line widths was observed in the temperature range 4.2–15 K. Both temperature and baric effects were explained in terms of the same model of a two-well adiabatic potential of impurity centers, which caused the appearance of low-frequency splittings in the ground and excited electronic level regions. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In recent years, much attention has been paid to studying various aspects of manifestations of so-called two-level systems in glasses. Anderson *et al.* [1] and Phillips [2] were the first to suggest the two-level system model for explaining anomalous temperature dependences of the heat capacity and heat conductivity of organic glasses at low temperatures ($T < 10$ K). Later, it was found that two-level systems were present in various glasses and polymers and were capable of influencing the optical properties of glass-like solutions containing dye molecules as admixtures. In particular, the interaction of an impurity molecule with a two-level system caused the appearance of such effects as spectral diffusion [3], nonphotochemical stable spectral hole burning [4], etc.

Two-level systems are usually described in terms of a two-well adiabatic potential; currently, we know almost nothing about their structure. This raises the question of whether two-level systems only occur in glasses or can exist in other media. The results obtained in our studies suggest that two-level systems may occur in crystallites containing dye molecules as admixtures. In particular, the spectra of the perylene–*n*-octane admixture system exhibit an almost linear dependence of phononless line widths in the temperature range 4.2–15 K [5]; such dependences are anomalous for crystals but characteristic of glasses. The spectra of the same admixture system experienced transformations caused by increasing external pressure, which we were able to explain by a complex form of the adiabatic potential [6].

In this work, we undertook further studies of the fluorescence spectra of the perylene–*n*-octane admixture system. These studies were performed under laser irradiation with the use of a continuous He–Cd laser and a tunable pulsed dye laser; the external pressure was varied from normal to 10 kbar. The results allow both effects described in [5, 6] to be explained in terms of one model of admixture interactions with a two-level system.

At low temperature and normal pressure, the solutions were a snowlike mass consisting of *n*-paraffin microcrystals. Increasing pressure noticeably improved the optical quality of the samples, and, starting with approximately 0.5 kbar, the samples turned optically transparent. The samples contained perylene admixtures in concentrations of 10^{-6}–10^{-5} mol/l. They were loaded into a special high-pressure chamber that ensured quasi-hydrostatic compression of samples; the chamber was contained in a helium cryostat. After freezing, the samples were subjected to the action of external pressure. Pressure was controlled by a calibrated dynamometer and was also determined from shifts of the *R* line of rubidium, whose microcrystal was specially introduced into the sample. Fluorescence was excited by a He–Cd laser (in what follows, continuous laser) line of wavelength 441.6 nm or by a tunable pulsed ($\tau \approx 10$ ns, $f = 10$ Hz) dye laser (DL-midi ESTLA, in what follows, pulsed laser) with a generation band about 2 cm^{-1} wide. The mean power of excitation was 4–10 mW. The spectra were recorded by a DFS-24 spectrometer with a 4.5 Å/mm linear dispersion.

2. RESULTS AND DISCUSSION

When perylene fluorescence is excited by a laser, the result is a superposition of two spectra shifted along the wavelength scale and differing in the character of optical excitation (Fig. 1a). One of these is the so-called "quasi-line spectrum" [7]; such spectra are characteristic of *n*-paraffin solutions of complex organic molecules at low temperatures. The line character of these

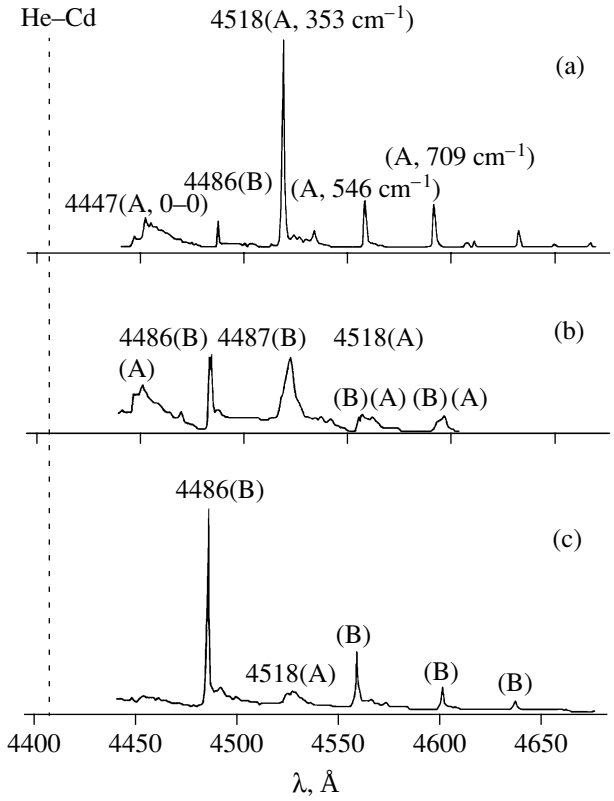


Fig. 1. Fluorescence spectra of perylene in *n*-octane at various pressures (T = 4.2 K): (a) atmospheric pressure, (b) P = 3.2 kbar, and (c) P = 4 kbar; A and B are the quasi-line and LFLN spectrum lines, respectively.

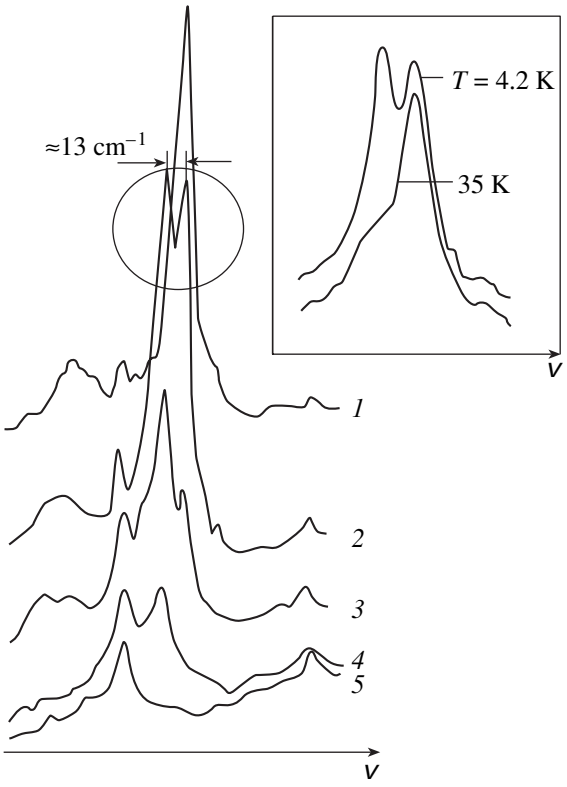


Fig. 2. Transformation of perylene fluorescence spectrum in the region of the 353 cm^{-1} vibronic transition (λ = 451.8 nm) caused by increasing pressure P = 0.6 (*1*), 1.22 (*2*), 1.83 (*3*), 2.44 (*4*), and 3.05 (*5*) kbar. Shown in the inset is an enlarged spectrum fragment circled in the main figure.

spectra is caused by the presence of narrow (1–10 cm^{-1}) peaks of a inhomogeneous distribution function of impurity centers with respect to the $S_1 \longrightarrow S_0$ electronic transition frequency, which is an envelope of a set of purely electronic phononless lines of separate impurity centers.

Apart from quasi-line spectra, we observe the so-called "laser-induced fluorescence line narrowing" (LFLN) spectra [8] of perylene caused by a selective character of laser excitation. The LFLN spectra are only formed by those impurity centers whose $S_0 \longrightarrow S_1$ vibronic transition is in resonance with laser radiation. These spectra are observed upon excitation to the region of both the main maximum and the background of the inhomogeneous distribution function. Unlike quasi-line spectra, they trace the laser radiation frequency. We studied the LFLN spectra of perylene in *n*-octane under excitation in a wide spectral range, 437–448 nm, and the influence of external pressure and temperature on the fluorescence spectra of perylene. Consider the action of these factors on the quasi-line and LFLN spectra separately.

2.1. Quasi-Line Spectra of Perylene

We found that increasing pressure caused several important changes in the spectra:

(1) the intensity of the spectra rapidly decreased as pressure increased;

(2) the spectrum experienced dramatic changes in the region of the 0–0 transition and its vibronic recurrences; these changes were well noticeable in the pressure range from atmospheric to 3 kbar;

(3) the rate of temperature broadening of phononless lines changed, and the temperature dependence itself in the region of helium temperatures was quasi-linear, which was anomalous for crystals. Consider these changes separately.

A decrease in the intensity of the quasi-line spectrum. The quasi-line spectrum of perylene in *n*-octane at atmospheric pressure (Fig. 1a) consists of an inhomogeneously broadened electronic phononless line corresponding to the 0–0 transition with wavelength 444.7 nm and vibronic recurrences of this line accompanied by weak phonon wings. The most intense line at 451.8 nm corresponds to the most active vibration with a 353 cm^{-1} frequency. The laser-induced line narrowing spectrum of perylene excited by the He–Cd laser contains a single line at 448.6 nm shifted by the fundamental vibration frequency (353 cm^{-1}) from the exciting laser line. The intensity of the quasi-line spectrum gradually decreases as pressure increases (Fig. 1b), whereas

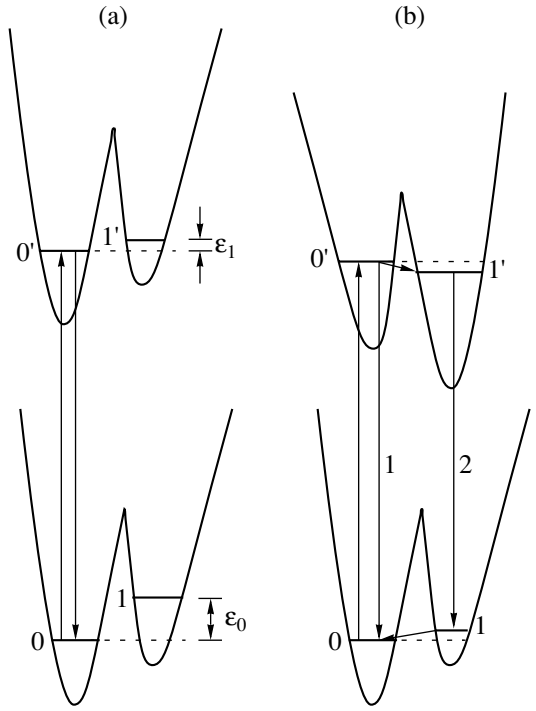


Fig. 3. Pressure-induced changes in the shape of a two-well adiabatic potential: (a) atmospheric pressure and (b) high pressure.

the laser-induced line narrowing spectrum becomes more intense under pressure actions, in particular, the 448.6 nm line becomes stronger. Increasing pressure to 4 kbar causes the quasi-line spectrum to virtually disappear, and all spectral lines except weak remains of a broadened 451.8 nm line are those of the LFLN spectrum (Fig. 1c).

The reason for pressure effects on the intensity of the spectra is easy to determine considering the difference in character between excitations of the quasi-line and LFLN spectra by laser radiation. Unlike the LFLN spectra, which are excited into phononless lines of impurity centers, quasi-line spectra are excited through wide phonon wings. According to the data of special studies, phonon wings are uniformly present in the whole spectral range of excitation, whereas the inhomogeneous distribution function has no noticeable singularities in this range. The observed effect cannot therefore be explained by a baric shift of the spectrum. At the same time, this effect is not related to a decrease in the concentration of impurity centers that contribute to the quasi-line spectrum. This is proved by the observation of intense fluorescence under excitation into the main distribution function peak at all pressures. The same argument proves that the observed effect is not related to fluorescence quenching processes.

Considering the aforesaid, we believe that a decrease in the intensity of the quasi-line spectrum (and also "flaring up" of the LFLN spectra) caused by increasing pressure should be related to a decrease in the probability of transitions with creating phonons with respect to phononless transitions and, accordingly, to intensity transfer from the phonon wings to the phononless line. The intensity of the phonon wing is determined by an electron–phonon interaction operator term which is linear with respect to the coordinates of nuclei [9]. It follows that our results are evidence of a sharp decrease in linear electron–phonon interaction caused by increasing pressure.

Quasi-line spectrum transformation. The fine structure of the quasi-line fluorescence spectrum of perylene excited by the continuous laser experiences synchronous changes in the region of the 0–0 electronic transition and all its vibronic recurrences in the pressure range from atmospheric to 2 kbar (Fig. 2). These changes include a rapid decrease in the intensity of the main maximum and the appearance of side maxima on its long-wave side. The side maxima alternately grow and then decrease as pressure increases. At the same time, the specular absorption spectrum does not experience such changes. As a result, pressure increase causes the loss of resonance between the phononless 0–0 lines of the fluorescence and absorption spectra. The character of fluorescence spectrum transformations depends on the spectral region of excitation. In particular, when the spectrum is excited by mercury light at 365 nm, we observe broadening and smooth shift of phononless lines to the longer waves, which can be described as behavior of the envelope of the spectrum obtained under laser radiation.

Pressure-induced spectrum transformations can be described by the model of a two-well (multiwell) adiabatic potential of impurity centers. According to this model, pressure distorts the adiabatic potential and causes transitions from one well to another. For instance, an impurity molecule that, after photon absorption, occurs in the left well (Fig. 3b) can tunnel to the right well and, after getting rid of excess energy, emit a photon with a larger wavelength.

If the picture described above is correct, an increase in temperature should in part restore the occupancy of the left well, and a change in the backward intensity distribution should be observed. A special experiment performed by us (its results are shown in the inset in Fig. 2) proved the validity of our conclusions.

The two-well potential model is, of course, approximate. The form of the adiabatic potential may be much more complex, and this potential may have many minima. Precisely for this reason, the effect depends of the wavelength of exciting radiation. At a small excitation energy excess with respect to the energy of the purely electronic 0–0 transition, as when fluorescence is excited by the continuous laser (λ = 441.6 nm), only one or several side potential wells are activated, and several peaks are observed in the region of the 0–0 transition. However, if the energy of exciting quanta is much larger than the energy of the 0–0 transition (excitation by the 365 nm mercury line), all potential wells

are activated, which results in a continuous spectral shift distribution.

The experimental results described above are insufficient for solving the problem of the nature of two-well potentials in the admixture system under consideration. One of possible explanations of the existence of several adiabatic potential minima may, in our view, be the presence of local crystal structure defects near an impurity molecule; these defects may be sensitive to both external pressure and the electronic state of the impurity molecule.

Anomalous temperature broadening of phononless lines. The low-frequency electronic level splittings considered above, which are caused by a complex shape of the adiabatic potential of the impurity molecule, can be treated as an analogue of two-level systems in glasses. The interaction with many such two-level systems in glasses results in a quasi-linear temperature dependence of phononless line widths. As mentioned, we observed a quasi-linear dependence uncharacteristic of crystals also for *n*-octane solutions of perylene.

Pressure in the sample strongly influences the temperature dependence of the phononless line. The experimental temperature dependences of the width of the quasi-line spectrum phononless line at 451.8 nm, from which the inhomogeneous component was subtracted, are shown in Fig. 4. These temperature dependences contain two distinct temperature regions. At high temperatures, line width γ follows the law $\gamma \propto T^2$, as is characteristic of the phonon mechanism of broadening. In the low-temperature region, a $\gamma \propto T^{1.1}$ quasi-linear dependence is observed; such dependences are characteristic of glasses and correspond to interactions with two-level systems. It is natural to suggest that the temperature dependence is determined by two different mechanisms, largely by interactions with a two-level system at low temperatures and largely by electron–phonon interactions at high temperatures. The $\gamma(T)$ dependence shown in Fig. 4 has one more remarkable feature, namely, its low-temperature portion tends to saturation at a temperature of about 15 K.

In conformity with the aforesaid, the temperature broadening of the phononless line was described by the equation

$$\gamma(T) = \gamma_{\text{ph}}(T) + \gamma_{\text{TLS}}(T), \qquad (1)$$

where the first and second terms corresponded to the contributions of electron-phonon interactions and interactions with a two-level system, respectively. Interactions of the latter kind were theoretically considered in several works in terms of both stochastic and dynamic approaches (see review [10]). We selected the tunnel theory advanced by Osad'ko [10] because this theory predicted the effect of $\gamma(T)$ dependence saturation mentioned above. In terms of this theory, phononless line broadenings are caused by interaction with two-level system excitation quanta, tunnelons, which, unlike phonons, are Fermi-type particles. In the simplest

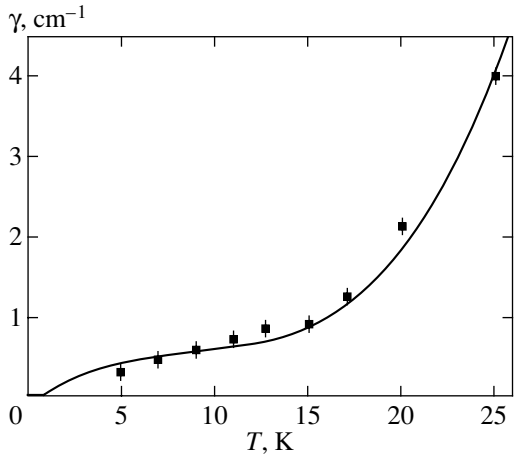


Fig. 4. Temperature dependence of the homogeneous width of the 451.8 nm line in the spectrum of perylene in *n*-octane: solid line corresponds to calculations by (1)–(3) with the parameters (cm^{-1}) $\varepsilon_0 = 16$, $b = 57$, $\omega_g = 74$, and $\gamma_0 = 0.45$.

model of interaction with a single tunnelon, the formula describing temperature broadening of the phononless line takes the form

$$\gamma_{\text{TLS}}(T) = \frac{\gamma_0}{\cosh^2(\varepsilon_0/2kT)} + \frac{\gamma_1}{\cosh^2(\varepsilon_1/2kT)}, \qquad (2)$$

where $\varepsilon_{0,1}$ and $2\gamma_{0,1}$ are the energy and reciprocal lifetime of the tunnelon in the ground (0) and excited (1) impurity molecule electronic states. The ε_0 and ε_1 values are related as $\varepsilon_1 = \varepsilon_0 + \Delta$, where Δ is the electron-tunnel interaction coupling constant. This constant was not selected arbitrarily in our calculations but was determined from the experimental data described in the preceding section. As can be seen from Fig. 3, it is the Δ value that determines the line splitting observed on increasing the pressure. In agreement with these results, $\Delta = -13$ cm^{-1}. We determined this value from the experimental temperature dependence under saturation conditions on the assumption that $\gamma_0 = \gamma_1$. The only remaining indefinite calculation parameter, ε_0, was found by adjustment. The phonon contribution to the line width was calculated by the formula obtained in the weak coupling limit for the interaction with a quasi-local phonon of frequency ω_g [11],

$$\gamma_{\text{ph}}(T) = \frac{b}{\sinh^2(\hbar\omega_g/2kT)}, \qquad (3)$$

where the b and ω_g parameters were selected empirically. Figure 4 shows that the theoretical temperature broadening curve calculated by (1)–(3) closely agrees with the experimental dependence, which substantiates the correctness of the selected interaction model.

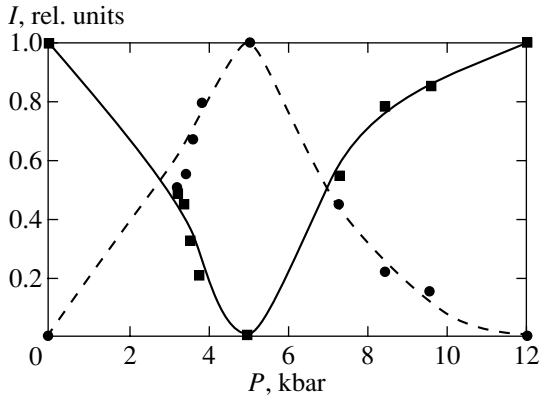


Fig. 5. Pressure-induced changes in the relative intensities of (■) short-wave and (●) long-wave doublet components.

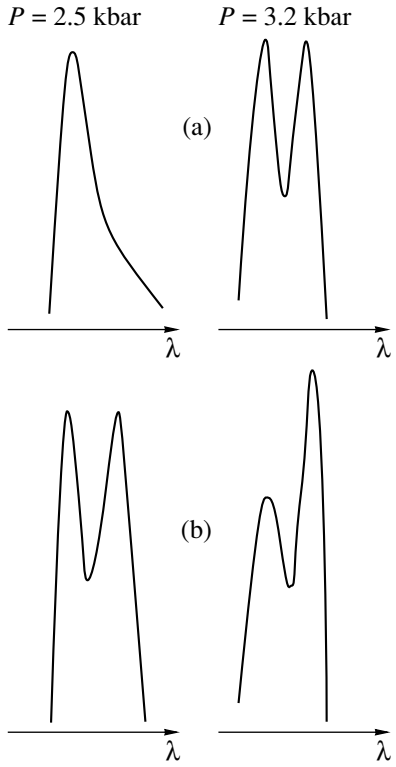


Fig. 6. Doublet component intensity distributions for different excitation sources at two different pressures: (a) continuous He–Cd laser and (b) pulsed laser (excitation light wavelength was 441.6 nm in all spectra).

2.2. Laser-Induced Line Narrowing Spectra of Perylene

The laser-induced line narrowing spectra of perylene at normal pressure comprise isolated lines whose positions depend on the wavelength of laser radiation used to excite fluorescence. During pressure increase, starting with approximately 2 kbar, the doublet structure of the laser-induced line narrowing spectra becomes manifest. This structure is best defined in the region of vibrations at 353 cm^{-1}, which are most active in the spectrum (line 448.6 nm in Fig. 1b). Doublet splittings vary over the spectrum. In particular, for the 353 cm^{-1} vibration, for which up to four vibrational quanta are observed, the splitting is approximately proportional to the number of vibrational recurrence. This is evidence that the doublet structure of the spectrum may be related to the presence in solution of two types of impurity centers with slightly different normal vibration frequencies. Centers with the fundamental vibration frequency equal to 353 cm^{-1} will be called centers of the first type, and centers with a shifted frequency (approximately 357 cm^{-1}) will be called centers of the second type. A further increase in pressure causes an unusual behavior of the spectra: different doublet components alternatively increase in intensity as pressure grows (Fig. 5), and the spectrum alternates between singlet and doublet. Further studies showed that, apart from pressure, the multiplet structure of the spectra depended on two other factors, namely, (a) excitation wavelength and (b) continuous or pulsed excitation character. At normal pressure, centers of the first type were present in the whole spectral range of our measurements, whereas centers of the second type were only observed in separate spectral regions. Increasing pressure might cause doublets to disappear in the regions where they were observed earlier or appear in the regions where they were absent.

This doublet structure behavior cannot be explained by baric shifts of the spectra. The observed pressure-induced intensity transfer between doublet components (Fig. 5) is evidence that centers of the first and second types can transform into each other. This transformation can be described by the two-well adiabatic potential model for impurity molecules, when the left and right wells alternatively become deeper depending on pressure. The question arises what is the difference between centers of the two types. In our view, the presence of centers of two types can be related to deformation of perylene molecules (D_{2h} point symmetry group) along one of two twofold symmetry axes situated in the symmetry plane of the molecule.

In conclusion, consider the dependence of the multiplet structure on the character of laser excitation mentioned above in more detail. The ratio between the intensities of doublet lines at a constant pressure depends on whether the laser used to excite fluorescence is continuous or pulsed (Fig. 6). Indeed, at a 2.5 kbar pressure, only the short-wave component of the doublet is clearly observed when the spectrum is excited by the continuous laser, whereas the second component only appears as a weak wing (Fig. 6a). At the same pressure and under pulsed laser irradiation, the two components have equal intensities (Fig. 6b). Increasing pressure to 3.2 kbar causes an increase in the intensity of the long-wave doublet component in both spectra, but the increase is approximately eight times larger when the continuous laser is used.

These results are evidence of the existence of an additional exchange channel between centers of the first and second types. This channel may, in our view, be light-induced transitions of centers of different types into each other. Our explanation of the effect under consideration is based on the assumption that the intensity ratio between doublet components in each of the spectra shown in Fig. 6 is largely determined by the ratio between their concentrations at the laser frequency, because the centers responsible for different doublet components have the same nature. The observed dramatic differences in the intensity distribution in the spectra excited by the continuous and pulsed lasers is evidence that optical excitation can cause substantial changes in the ratio between these concentrations. This can be explained by an electronic transition-induced deformation of the two-well adiabatic potential of impurity centers, which affects the probability of transitions between the wells. The kinetics of the process and, accordingly, the intensity ratio between doublet components can also depend on the character of optical excitation (pulsed or continuous). Precisely this dependence was observed in our experiments. Interestingly, a similar phenomenon is known in spectroscopy of isolated molecules characterized by "jumps of spectral lines" between two positions. According to [12], this effect can cause the appearance of either single or doublet lines in the spectra of single molecules depending on the sweep rate.

3. CONCLUSION

To summarize, we obtained evidence that low-frequency splittings similar to those characteristic of two-level glass systems can be observed in the region of impurity electronic excitation in crystallites such as formed in the admixture system under consideration. These splittings can be caused by the presence of several equilibrium configurations either of the impurity molecule itself or of its nearest environment. We believe that, in the latter case, instability is impurity-induced, and the presence of related effects can depend on the impurity–matrix pair.

REFERENCES

1. P. W. Anderson, B. J. Halperin, and C. M. Varma, Philos. Mag. **25**, 1 (1972).
2. W. A. Phillips, J. Low Temp. Phys. **7**, 351 (1972).
3. W. Brein, J. Friedrich, and D. Haarer, J. Chem. Phys. **81**, 3915 (1984).
4. J. M. Hayes and G. Small, J. Chem. Phys. **27**, 151 (1978).
5. O. N. Korotaev, I. P. Kolmakov, M. F. Shchanov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **55**, 417 (1992) [JETP Lett. **55**, 424 (1992)].
6. O. N. Korotaev, M. F. Shchanov, V. P. Karpov, and I. P. Kolmakov, Pis'ma Zh. Éksp. Teor. Fiz. **58**, 715 (1993) [JETP Lett. **58**, 687 (1993)].
7. É. V. Shpol'skiĭ, Usp. Fiz. Nauk **77**, 321 (1962) [Sov. Phys. Usp. **5**, 522 (1962)].
8. R. I. Personov, E. I. Al'shits, and L. A. Bykovskaya, Pis'ma Zh. Éksp. Teor. Fiz. **15**, 609 (1972) [JETP Lett. **15**, 431 (1972)].
9. I. S. Osad'ko, Zh. Éksp. Teor. Fiz. **72**, 1575 (1977) [Sov. Phys. JETP **45**, 827 (1977)].
10. I. S. Osad'ko, Phys. Rep. **206**, 37 (1991).
11. M. A. Krivoglaz, Fiz. Tverd. Tela (Leningrad) **6**, 1707 (1964) [Sov. Phys. Solid State **6**, 1340 (1964)].
12. I. S. Osad'ko, in *Selective Spectroscopy of Single Molecules* (Fizmatlit, Moscow, 2000), p. 288.

Translated by V. Sipachev