Journal of Experimental and Theoretical Physics, Vol. 93, No. 6, 2001, pp. 1137–1154.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 6, 2001, pp. 1315–1335.
Original Russian Text Copyright © 2001 by V. Polyachenko, E. Polyachenko.

GRAVITATION, ASTROPHYSICS

A Linear Theory of Resonance Structures in Spiral Galaxies

V. L. Polyachenko and E. V. Polyachenko*
Institute of Astronomy, Russian Academy of Sciences, Pyatnitskaya ul. 48, Moscow, 109017 Russia
**e-mail: evgenii@orc.ru*

Received May 25, 2001

Abstract—The resonance mechanism for the formation of galactic spirals is considered. Expressions are derived for the resonance responses of disks with circular and nearly circular stellar orbits. The spiral responses produced by the central oval-shaped structures (bars) available in many galaxies are shown to have the characteristic properties of the spirals observed in these galaxies. In the most interesting case of a quasi-steady state, the spiral responses possess a similarity property: the spiral thickness and inclination are proportional to the mean size of an epicycle (an analog of the Larmor circle in plasma). © *2001 MAIK “Nauka/Interperiodica”.*

1. INTRODUCTION

Astronomical objects such as galaxies are investigated by using models in the form of simple figures of equilibrium: a flat layer, a cylinder, a sphere, a disk, and an ellipsoid. The first two figures are unbounded at least in one direction. These models share many common features with homogeneous plasma. In particular, appropriate plasma analogies can generally be found for various kinetic effects in such systems [1, 2]. On the other hand, the particle motion in the models of disks and spherical or ellipsoidal systems is finite in all directions. The kinetic effects in such systems are attributable to resonances between waves and orbital particle motion. In plasma physics, such effects are studied, for example, in connection with plasma instability in closed magnetic traps: tokamaks, the Earth's magnetosphere, and the like (see [3]).

The most important parameter of the wave whose resonant interaction with galactic stars we will be interested in below is its frequency ω. If the gravitational potential of the system under consideration admits a complete separation of variables in the Hamilton–Jacobi method [4], then the motion of stars is also characterized by the frequencies

$$\Omega_i = \frac{\partial H}{\partial J_i}, \tag{1}$$

where $i = 1, 2, 3$; and $H(J_1, J_2, J_3)$ is the Hamiltonian of the star expressed in terms of the actions J_1, J_2, and J_3. The condition of resonance between the wave and particles (stars) can then be written as

$$\omega = l_1\Omega_1 + l_2\Omega_2 + l_3\Omega_3 \tag{2}$$

(l_1, l_2, and l_3 are integers). Equation (2) is a special form of the general resonance condition

$$\omega_{\text{outer}} = \omega_{\text{inner}}, \tag{3}$$

where ω_{outer} is the frequency of the outer action on particles of the system under consideration (in the special case of the greatest interest, ω_{outer} is the frequency ω of the wave excited in the system in one way or another), and

$$\omega_{\text{inner}} = l_1\Omega_1 + l_2\Omega_2 + l_3\Omega_3$$

is one of the possible inner frequencies of the orbital particle motion.

If the outer force is produced by a satellite, then, similar to Eq. (2),

$$\omega_{\text{outer}} = n_1\Omega_1^s + n_2\Omega_2^s + n_3\Omega_3^s,$$

where Ω_1^s, Ω_2^s, and Ω_3^s are the satellite oscillation frequencies, and n_1, n_2, and n_3 are its own set of integers. The structures associated with this kind of resonances are directly observed, for example, in Saturn's rings as so-called wave trains. These trains are the density waves that decay as they propagate away from the resonances (see [5]).[1] As another example with the general resonance (3), we mention a Landau-type collision term [7] for systems with finite particle motion [8]. This collision term, which corresponds to pair interactions of particles (a and b), contains the frequency δ functions

$$\delta[(l_1^a\Omega_1^a + l_2^a\Omega_2^a + l_3^a\Omega_3^a) - (l_1^b\Omega_1^b + l_2^b\Omega_2^b + l_3^b\Omega_3^b)].$$

These δ functions mean that particle a interacts only with those particles b for which the resonance condition is satisfied. The orbital motion of particle b may be said

---

[1] This interpretation of the observed wave trains is supported, in particular, by the proper correspondence of their localization (farther from the planet than the resonance position) to the direction of the density-wave group velocity (away from the planet) that follows from their characteristic dispersion relation (see Eq. (1) from [6]). The bending waves in resonance regions, which are more difficult to observe, were also detected (see [5] for more details); they are closer to the planet (relative to the resonance positions), which is also in agreement with the direction of the bending wave group velocity.

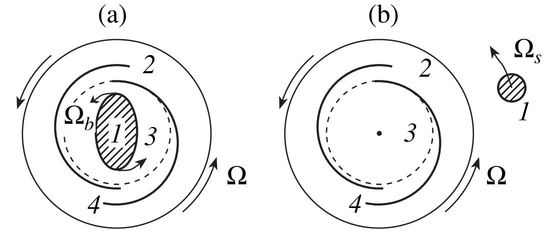


Fig. 1. Possible ways of generating resonance spirals in galaxies: (a) by a central bar (*1* is the bar rotating at angular velocity Ω_b, *2* is the galactic disk, *3* is the resonance circumference, and *4* are the spirals); and (b) by a galactic satellite (*1* is the satellite rotating about the galaxy in its plane at angular velocity Ω_s, *2* is the galactic disk, *3* is the resonance circumference, and *4* are the spirals). The disks rotate at angular velocity $\Omega(r)$.

to produce a gravitational field fluctuating with frequencies

$$\omega = l_1^b\Omega_1^b + l_2^b\Omega_2^b + l_3^b\Omega_3^b$$

with which the orbital motion of particle *a* must be in resonance:

$$l_1^a\Omega_1^a + l_2^a\Omega_2^a + l_3^a\Omega_3^a = \omega;$$

by contrast, the average effect of nonresonant interaction proves to be zero.

Resonances show up most clearly and easily in disk systems with nearly circular particle motion (spiral galaxies, planetary rings). The reason is that, in this case, the resonances have a specific spatial localization (near resonance circumferences), while, in general, the resonance conditions define some sets in phase space, which is naturally more difficult to study, in particular, observationally.

Below, we restrict our analysis to spiral galaxies. Many observable structures in galaxies, notably, rings and spirals, are clearly associated with resonances. They can be induced, for example, by satellites or central bars (Fig. 1). In principle, galactic satellites act just like planetary satellites. However, satellites appear to be much less important for galaxies than for planetary rings. Planetary rings are much older systems if the age is measured in units of dynamical time, the revolution time of a typical particle in a circular orbit. For planetary rings, this age is about 10^{12} years, whereas, for spiral galaxies, it is often only on the order of several tens of galactic years. In particular, for this reason, even resonances of very high orders (e.g., those that correspond to a 34 : 35 ratio of the revolution frequencies of a satellite and dust particles of a ring) have managed to clearly manifest themselves in planetary rings; the galaxy lifetime is evidently not enough for this manifestation (see [5] for more details on a comparison of the properties for various astrophysical disks).

Central bars (Fig. 1a) are encountered approximately in half of all spiral galaxies (this type of barred galaxies is commonly designated as SB, to distinguish them from normal SA spirals). Being within the galactic disk, the nonaxisymmetric gravitational potential of the bar strongly affects the motion of stars (especially in the resonance regions that are closest to the bars), producing spirals, rings, and other structures.

Here, our goal is to construct a linear theory of resonance galactic structures. Most of them are associated with the density perturbations that arise from stellar motions in the disk plane. Therefore, for the appropriate resonances, we may set $l_3 = 0$ in the general expression (2) for frequency ω_{inner} (if Ω_3 denotes the frequency of the vertical *z* oscillations of the star; the disk is assumed to be located in the *x*, *y* plane). The resonance condition (2) can then be rewritten as

$$\omega = m\Omega_2 + l\Omega_1, \qquad (4')$$

where the following notation is used: *m* is the azimuthal wave number [we assume the perturbation to be expanded in a Fourier series and consider one of the harmonics proportional to $\exp(im\varphi)$]; *l* is the radial index; Ω_1 and Ω_2 are the radial and azimuthal frequencies, respectively. At a fixed *m*, the resonances are numbered by one *l* (positive, negative, or zero), and it is convenient to rewrite Eq. (4') [by dividing both sides of Eq. (4') by *m*] as

$$\Omega_p = \Omega_2 + l\Omega_1/m, \qquad (4'')$$

where $\Omega_p = \omega/m$ is the angular velocity of the wave in an inertial frame of reference. If, for example, the perturbations are produced by a bar that rotates at angular velocity Ω_b, then $\Omega_p = \Omega_b$.

Here, we use the approximation of an infinitely thin disk. Allowance for the finite thickness or, in other words, for the particle motion along the third (*z*) axis is a separate problem, which can be analyzed only numerically. Since there are no systematic studies of this kind as yet, we are unable to accurately predict the extent to which the corresponding effects can alter the results of a two-dimensional analysis. Note, however, that in the available specially carried-out numerical experiments (in the *N*-body approach), the evolutions of perturbations in three- and two-dimensional disks usually turned out to be almost indistinguishable. We may allude, for example, to [9], in which a similar problem of the formation of bars and their accompanying spiral pattern is considered.

Note also that, in general, allowance for the finite thickness is more important for a gaseous disk than it is for a stellar one. The point is that the latter (in contrast to the former) is collisionless; hence, it can, in principle, have a highly anisotropic pressure. In this case, as suggested by observations, the mean thermal stellar velocity along the *z* axis (c_z) is always lower than that in the galactic disk plane (c_r). For our Galaxy,

$$c_z^2/c_r^2 \approx 1/4,$$

but it can be easily imagined that there are also specimens with much larger anisotropies among the great variety of spiral galaxies. In essence, the only fundamental limiter here is fire-hose instability, but it grows only under a very large anisotropy. Theory shows (see [10] for a review) that a single stellar disk becomes unstable only at[2]

$$c_z^2/c_r^2 \approx 1/10.$$

Given the stabilizing role of the spheroidal galactic components, the anisotropy required for instability becomes even larger.

In the next section, we show (Subsection 2.2) how the resonance condition (4") simplifies for disks with nearly circular orbits. We also give the equilibrium distribution functions of stars in such disks (Subsection 2.3) and derive (Subsection 2.1) the laws of stellar motion in the approximation needed for their subsequent use (Subsection 3.1) in deriving expressions for the disk resonance responses. These responses are then analyzed in detail (Subsection 3.2).

In conclusion, we discuss some of the fundamental points related to our results.

2. EQUILIBRIUM OF AXISYMMETRIC STELLAR DISKS

2.1. Nearly Circular Stellar Orbits in an Axisymmetric Gravitational Field

In this section, our goal is to derive expressions for the coordinates of a star, $r(t)$ and $\varphi(t)$, in a nearly circular orbit, to within quadratic terms in small deviation $\delta r = r - r_0$ of the current stellar radius r from the radius r_0 of the circular orbit that corresponds to angular momentum L. We need these expressions below (in Section 3) to analyze the responses of galactic disks near the resonance circumferences.

Writing the equations of two-dimensional motion for a star in an axisymmetric potential $\Phi_0(r)$ in cylindrical coordinates r and φ and eliminating φ using the angular momentum conservation law

$$L = r^2\dot{\varphi} = \text{const},$$

we obtain

$$\ddot{r} = -\frac{d\Phi_0}{dr} + \frac{L^2}{r^3} = -\frac{dW}{dr}, \tag{5}$$

where

$$W = \Phi_0 + L^2/2r^2$$

is the effective potential energy.

We will seek a solution to Eq. (5) in the form of an expansion in powers of $\delta r = r - r_0$, the small deviation of the current stellar radius r from the circular-orbit radius r_0:

$$r(t) = r_0 + r_1(t) + r_2(t) + \ldots, \tag{6}$$

where

$$\frac{L^2}{r_0^3} = \frac{d\Phi_0(r_0)}{dr_0}; \tag{7}$$

the latter equation expresses an equality between the gravitational and centrifugal forces. Once the radius $r(t)$ of the star has been calculated, its azimuth $\varphi(t)$ can be determined by solving the equation

$$\dot{\varphi} = L/r^2.$$

Instead of the azimuthal velocity $v_\varphi(t)$ in an inertial frame of reference, it would be more natural to work with the residual velocity

$$\tilde{v}_\varphi = v_\varphi(t) - r(t)\Omega[r(t)], \tag{8}$$

which is the deviation of $v_\varphi(t)$ from the local [at the current point $r(t)$] circular velocity: the velocities $\tilde{v}_\varphi$ and v_r are then the same, in order of magnitude; the function

$$\Omega(r) = L/r^2$$

denotes the angular velocity of the star in a circular orbit of radius r. To derive the required equations for the orbits, all the constants involved in solving the equations of motion must be expressed in terms of the initial coordinates and velocities. Retaining the designations

$$r = r(t = 0), \quad \varphi = \varphi(t = 0),$$
$$v_r = v_r(t = 0), \quad \tilde{v}_\varphi = \tilde{v}_\varphi(t = 0),$$

for the latter, we write the orbit (so far in the first order in δr) as

$$r(t) = r + \sin(\kappa_0 t)\frac{v_r}{\kappa_0} + (1 - \cos(\kappa_0 t))\left(\frac{2\Omega_0}{\kappa_0}\right)\frac{\tilde{v}_\varphi}{\kappa_0}, \tag{9}$$

$$\begin{aligned}\varphi(t) = \varphi + \Omega_0 t - \frac{2\Omega_0}{r_0\kappa_0}(1 - \cos(\kappa_0 t))\frac{v_r}{\kappa_0} \\ + \sin(\kappa_0 t)\frac{4\Omega_0^2}{\kappa_0^2 r_0}\frac{\tilde{v}_\varphi}{\kappa_0},\end{aligned} \tag{10}$$

where

$$\kappa_0^2 = \kappa^2(r_0), \quad \kappa^2(r) = \Phi_0''(r) + 3L^2/r^4, \quad \Omega_0 = \Omega(r_0).$$

The linear (in δr) approximation described above is commonly called (for historical reasons that date back to Ptolemaeus!) epicyclic. The point is that, as can be

---

[2] It is for this reason that fire-hose instability has long been considered to be of little interest in gravitating systems. This instability has again attracted attention after it was shown in [11] (see also [1, 2]) that it could account for the observed maximum oblateness of elliptical galaxies. Elliptical galaxies are slowly rotating systems whose equilibrium is provided by an anisotropic pressure. They are hotter in the rotation plane (compared to spiral galaxies) and, hence, are more easily subjected to fire-hose instability.

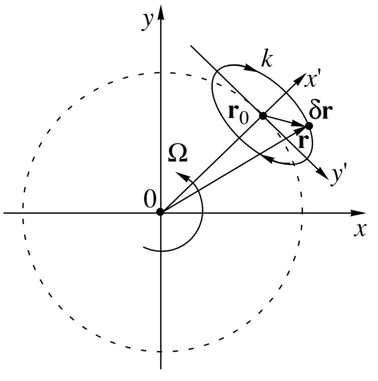


Fig. 2. A schematic representation of epicyclic motion in a disk with nearly circular particle motions. The disk angular velocity is $\Omega(r)$. The current particle radius vector is $\mathbf{r} = \mathbf{r}_0 + \delta\mathbf{r}$, where vector $\mathbf{r}_0$ uniformly rotates with angular velocity $\Omega_0 = \Omega(r_0)$. The tip of vector $\delta\mathbf{r}$ in a coordinate system that rotates with velocity Ω_0 describes a small ellipse (epicycle) with an epicyclic frequency κ in the sense opposite to the disk rotation.

shown from Eqs. (9) and (10), the stars in a frame of reference rotating with angular velocity Ω_0 move along azimuthally elongated small ellipses (Fig. 2); the same ellipses that appear when describing nearly circular planetary motions in the Solar System were called epicycles. The vector

$$\delta\mathbf{r} = \mathbf{r} - \mathbf{r}_0$$

has the coordinates

$$\delta\mathbf{r}(r_1\cos\phi - r_0(\delta\varphi)\sin\phi,\ r_1\sin\phi + r_0(\delta\varphi)\cos\phi),$$

where $\phi \equiv \varphi_0 + \Omega t$, so the current stellar coordinates in local Cartesian coordinates (x', y') (see Fig. 2) are

$$x' = \delta\mathbf{r}\cdot\mathbf{k}, \quad y' = \delta\mathbf{r}\cdot\mathbf{n}, \tag{11}$$

where $\mathbf{k} = (\sin\phi, -\cos\phi)$ and $\mathbf{n} = (\cos\phi, \sin\phi)$ are the corresponding unit vectors. Calculations yield (α = const is a constant phase)

$$\begin{aligned} x' &= -r_0\delta\varphi = \frac{2\Omega_0}{\kappa_0}\sin(\kappa_0 t + \alpha), \\ y' &= r_1 = a\cos(\kappa_0 t + \alpha), \end{aligned} \tag{12}$$

i.e., the rotation in an ellipse ($x'^2/b^2 + y'^2/a^2 = 1$) with the axial ratio

$$b/a = 2\Omega_0/\kappa_0$$

in the sense opposite to the main rotation of stars in the disk (with an epicyclic frequency κ_0). The original expression for the square of the epicyclic frequency can be easily transformed to

$$\kappa^2 = 4\Omega^2 + r(\Omega^2)', \tag{13}$$

where

$$(\Omega^2)' \equiv \frac{d\Omega^2}{dr}.$$

It follows from (13) that, at $\Omega' < 0$ (this condition is always satisfied), the ratio

$$b/a = 2\Omega_0/\kappa_0 > 1;$$

i.e., the epicycle is actually elongated azimuthally.

As we already noted above, to calculate the responses (in Section 3) requires equations for the stellar orbits not in the epicyclic (linear in δr) approximation but in the next, post-epicyclic one, which also include terms quadratic in δr. Omitting calculations [which are naturally more cumbersome than those used to derive the epicyclic formulas (9) and (10)], we give the required equations for $r(t)$ and $\varphi(t)$:

$$r(t) = r + \delta r, \tag{14}$$

$$\begin{aligned} \delta r = {} & \sin\kappa_0 t\frac{v_r}{\kappa_0} + (1 - \cos(\kappa_0 t))\frac{1}{\kappa_0}\left(\frac{2\Omega_0\tilde{v}_\varphi}{\kappa_0}\right) \\ & + (1 - \cos(\kappa t))\frac{2\Omega\alpha}{\kappa^4}\left(\frac{2\Omega\tilde{v}_\varphi}{\kappa}\right)^2 + (1 - \cos(\kappa t))\frac{2\Omega\beta}{\kappa^4}v_r^2, \end{aligned} \tag{15}$$

$$\varphi(t) - \varphi = (\Omega_0 + \delta\Omega)t + \delta\varphi, \tag{16}$$

$$\begin{aligned} \delta\varphi = {} & -\frac{2\Omega_0}{r_0\kappa_0^2}(1 - \cos(\kappa_0 t))v_r \\ & + \frac{2\Omega_0}{r_0\kappa_0^2}\sin(\kappa_0 t)\frac{2\Omega_0\tilde{v}_\varphi}{\kappa_0} \\ & + \sin(\kappa t)\frac{4\Omega^2\alpha}{\kappa^5 r}\left(\frac{2\Omega\tilde{v}_\varphi}{\kappa}\right)^2 - \sin(\kappa t)\frac{4\Omega^2\beta}{\kappa^5 r}v_r^2, \end{aligned} \tag{17}$$

where we slightly changed the meaning of δr and designated

$$\alpha = \frac{\mu}{6\Omega} + \frac{1}{2}(r\Omega'' + 2\Omega') - \frac{\Omega}{r}, \tag{18}$$

$$\beta = \frac{\mu}{3\Omega}, \quad \mu = -\frac{1}{2}\left[(\kappa^2)' - \frac{3\kappa^2}{r}\right], \tag{19}$$

$$\begin{aligned} \delta\Omega &= \left(\frac{3}{2r^2} - \frac{\mu}{r\kappa^2}\right)\frac{\Omega}{\kappa^2}\left[v_r^2 + \left(\frac{2\Omega\tilde{v}_\varphi}{\kappa}\right)^2\right] \\ &= \frac{\kappa'}{\kappa}\frac{\Omega}{r\kappa^2}\left[v_r^2 + \left(\frac{2\Omega\tilde{v}_\varphi}{\kappa}\right)^2\right]. \end{aligned} \tag{20}$$

Note that in the expression for $\varphi(t) - \varphi$ we set off the linearly increasing (with time) term $(\delta\Omega)t$ that appears here, where $\delta\Omega$ denotes a nonlinear correction to the azimuthal frequency of the star in a nearly circular orbit:

$$\Omega_2 = \Omega + \delta\Omega.$$

As must be the case, $\delta\Omega = 0$ for a rigidly rotating disk in which $\Omega' = \kappa' = 0$. Rigid rotation is provided by the quadratic potential

$$\Phi_0 = \Omega^2 r^2/2 + \text{const},$$

in which the general orbit is known (see, e.g., [4]) to be an ellipse centered on $r = 0$. Clearly, the star executes two complete radial oscillations (irrespective of its amplitude) in the time of a complete turn in such an ellipse; i.e., in this case, the radial frequency Ω_1 is twice the azimuthal frequency Ω_2:

$$\Omega_1 = 2\Omega_2.$$

For nearly circular orbits, this equality can be derived from (13) at $\Omega' = 0$ in the form $\kappa = 2\Omega$. In the opposite limiting case of a Coulomb potential Φ_0, we obtain the frequency correction $\delta\Omega$ from (20), which can also be derived from Kepler's third law:

$$\delta\Omega = -3a^2\Omega_0/2r_0^2.$$

2.2. Characteristic Oscillation Frequencies of Stars in Nearly Circular Orbits

Let us derive expressions for the oscillation frequencies of stars in nearly circular orbits in a circular disk, Ω_1 and Ω_2, from their definition (1). According to [4], the actions J_1 and J_2 are given by

$$J_1 = \frac{1}{2\pi}\oint p_r dr = \frac{1}{\pi}\int_{r_{\min}}^{r_{\max}} dr\sqrt{2E - 2\Phi_0(r) - L^2/r^2}, \qquad (21)$$

$$J_2 = \frac{1}{2\pi}\oint p_\varphi d\varphi = L, \qquad (22)$$

where $p_r = \sqrt{2E - 2\Phi_0(r) - L^2/r^2}$ and $p_\varphi = L$ are the corresponding generalized momenta; E and L are the energy and angular momentum of the star (its mass is assumed to be unity); and the minimum ($r_{\min}$) and maximum ($r_{\max}$) orbital radii are the roots of the equation

$$p_r^2 = 2E - 2\Phi_0(r) - L^2/r^2 = 0.$$

Equations (21) and (22) implicitly define the function $E(J_1, J_2)$. Differentiating (21) with respect to J_1 and J_2 yields the required frequencies in the form

$$\Omega_1 = \frac{\partial E}{\partial I_1} = \frac{\pi}{I_1}, \quad \Omega_2 = \frac{\partial E}{\partial I_2} = \frac{I_2}{I_1}, \qquad (23)$$

where we denoted the integrals by

$$I_1 = \int_{r_{\min}}^{r_{\max}} \frac{dr}{\sqrt{2[E - W(r)]}}, \qquad (24)$$

$$I_2 = \int_{r_{\min}}^{r_{\max}} \frac{Ldr}{r^2\sqrt{2[E - W(r)]}}; \qquad (25)$$

$$W(r) = \Phi_0(r) + L^2/2r^2$$

is the effective potential energy.

Clearly, the function $W(r)$ has a minimum at the circular-orbit radius r_0: $W'(r_0) = 0$, i.e., $\Phi_0' = L^2/r_0^3$ [this is the condition (7) of equality between the centrifugal and gravitational forces]. Representing $W(r)$ near $r = r_0$ by a parabola and calculating integrals (24) and (25), we obtain

$$I_1 \approx \pi/\sqrt{\Phi_0''(r_0) + 3L^2/r_0^4},$$
$$I_2 \approx L\pi/\left(r_0^2\sqrt{\Phi_0''(r_0) + 3L^2/r_0^4}\right); \qquad (26)$$

hence,

$$\Omega_1 \approx \sqrt{\Phi_0''(r_0) + 3L^2/r_0^4} = \kappa_0, \qquad (27)$$

$$\Omega_2 \approx L/r_0^2 = \Omega_0, \qquad (28)$$

where we use the previously introduced notation for the epicyclic frequency κ and angular velocity Ω of the star in a circular orbit. Note that the limiting values (27) and (28) for the characteristic frequencies of a star in a disk could also be obtained from the epicyclic formulas (9) and (10) of the preceding section, given that, by the meaning of (23),

$$\Omega_1 = 2\pi/T_r, \quad \Omega_2 = 2\pi/T_\varphi,$$

where T_r is the time in which r changes from $r_{\max}$ to $r_{\min}$ and then back to $r_{\max}$, and T_φ is the time in which the radius vector turns through 2π (the integral I_2 is equal to the angle $\Delta\varphi$ through which the radius vector turns in time $T_r/2 = I_1$).

Accordingly, the resonance condition (4") for nearly circular orbits can be rewritten as

$$\Omega(r) - \Omega_p = -l\kappa(r)/m. \qquad (29)$$

The following are the main three resonances in spiral galaxies with special names: the corotation resonance (CR) at which $\Omega = \Omega_p$ corresponds to $l = 0$; the resonances that correspond to $|l| = 1$ are called Lindblad resonances. If we move from the corotation resonance toward the galactic center, then the local angular velocity will increase and we can eventually encounter (but not always) a circle on which Ω exceeds Ω_p by $\kappa/|m|$. This resonance is called an inner Lindblad resonance (ILR): $l = -1$ for it at $m > 0$. The other resonance with $|l| = 1$ is located in the galaxy outside the corotation radius (an outer Lindblad resonance, OLR). The above resonances take place where the frequency with which a star crosses the crests and troughs of the wave potential, $|\omega - m\Omega|$, is either zero (i.e., the star is always in phase with the wave) or equal to the oscillation frequency of the star (κ) about a circular orbit). The reso-

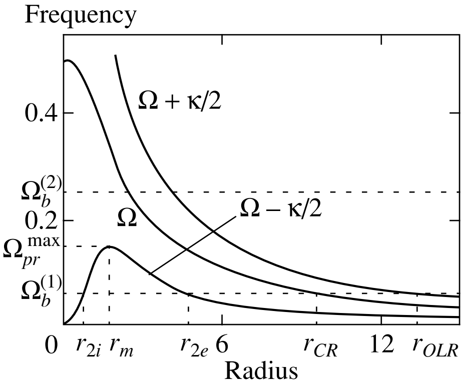


Fig. 3. The main characteristic frequencies of stellar orbits: $\Omega(r)$ is the angular velocity; $\Omega - \kappa/2 = \Omega_{pr}$ is the precession velocity for nearly circular orbits; r_{CR} is the corotation; r_{2i} and r_{2e} are the inner Lindblad resonances; r_{OLR} is the outer Lindblad resonance; $\Omega_b^{(1)} < \Omega_{pr}^{\max}$ is the velocity of a slow bar; $\Omega_b^{(2)} > \Omega_{pr}^{\max}$ is the velocity of a fast bar.

nances of higher orders (than those listed above) are dynamically less important. In addition, the outer resonances often lie outside galaxies, while the inner resonances lie close to the galactic nucleus; therefore, both are of no importance in forming the main spiral pattern.

For $|m| = 2$, the inner and outer Lindblad resonances in the Galaxy are roughly separated by the disk size; for $|m| \geq 3$, they all approach corotation.

For the observationally most interesting cases of two-arm structures ($m = 2$) and for realistic rotation curves $\Omega = \Omega(r)$ (Fig. 3), there is always only one corotation and one outer Lindblad resonance at which

$$\omega_*(r_{OLR}) \equiv \omega - m\Omega(r_{OLR}) = \kappa(r_{OLR}),$$

i.e.,

$$\Omega_p = \Omega(r_{OLR}) + \kappa(r_{OLR})/2.$$

In Fig. 3, the inner Lindblad resonances where $\omega_*(r_{ILR}) = -\kappa(r_{ILR})$ correspond to the intersections of the $\Omega = \Omega_p$ straight line with the $\Omega = \Omega_{pr} \equiv \Omega(r) - \kappa(r)/2$ curve. As we see from the figure, there are several possibilities: (i) if $\Omega_p > \Omega_{pr}^{\max}$, then there are no inner Lindblad resonances at all; (ii) if $\Omega_p < \Omega_{pr}^{\max}$, then two resonances emerge: internal ILR(r_{2i}) and external ILR(r_{2e}).

In conclusion, note that the condition for a Lindblad resonance can also be interpreted as the condition of equality between the wave frequency Ω_p and the angular velocity of orbital precession; at $m = 2$, $\Omega_{pr} = \Omega_2 - \Omega_2/2$ or $\Omega_{pr}^{+} = \Omega_2 + \Omega_2/2$ ($= \Omega \pm \kappa/2$ in the epicyclic approximation); the precession velocity is equal to the angular velocity of the rotating frame of reference in which the orbit is closed (an ellipsoidal oval). Recall that, for example, the stability theory for low-frequency perturbations in gravitating systems is most naturally formulated [12] in the language of orbital precession (rather than particle) velocities.

2.3. Equilibrium Distribution Functions of Stars in a Circular Disk in the Epicyclic and Post-Epicyclic Approximations

The distribution function of a disk with circular particle orbits (which is also called a cold one) is

$$f_0 \propto \delta(v_r)\delta(\tilde{v}_\varphi).$$

Accordingly, the distribution function of a stellar system with nearly circular orbits can be represented as a formal series of the δ functions of v_r and v_φ and their derivatives [1, 2, 13]:

$$\begin{aligned} f_0 &= a_1\delta(v_r)\delta(\tilde{v}_\varphi) + b_1\delta'(v_r)\delta(\tilde{v}_\varphi) \\ &+ b_2\delta(v_r)\delta'(\tilde{v}_\varphi) + c_1\delta''(v_r)\delta(\tilde{v}_\varphi) \\ &+ c_2\delta'(v_r)\delta'(\tilde{v}_\varphi) + c_3\delta(v_r)\delta''(\tilde{v}_\varphi). \end{aligned} \qquad (30)$$

Substituting expansion (30) into the equilibrium kinetic equation

$$v_r\frac{\partial f_0}{\partial r} - \frac{\partial\Phi_0}{\partial r}\frac{\partial f_0}{\partial v_r} + \frac{v_\varphi^2}{r}\frac{\partial f_0}{\partial v_r} - \frac{v_r v_\varphi}{r}\frac{\partial f_0}{\partial v_\varphi} = 0 \qquad (31)$$

and setting the coefficients in the various combinations of the derivatives of δ functions equal to zero, we obtain the equalities

$$c_2 = 0 \quad (\text{i.e., } \overline{v_r\tilde{v}_\varphi} = 0);$$

$$b_1 = 0 \text{ (i.e., there is no radial flux, } \bar{v}_r = 0);$$

$$c_3 = c_1\kappa^2/4\Omega^2,$$

which corresponds to a certain relationship between the radial (c_r) and azimuthal (c_φ) stellar velocity dispersions, $c_\varphi = c_r\kappa/2\Omega$, and the following general expression for the azimuthal stellar flux:

$$b_2 = \Pi_\varphi = \frac{1}{2\Omega r}\left[(r\sigma_0 c_r^2)' - \frac{\kappa^2\sigma_0}{4\Omega^2}c_r^2\right]. \qquad (32)$$

Accordingly, instead of (30), we have a general expression for the equilibrium distribution function of stars with nearly circular orbits in the form

$$\begin{aligned} f_0 = \sigma_0(r)\Bigg\{&\delta(v_r)\delta(\tilde{v}_\varphi) - \frac{1}{2\Omega r\sigma_0} \\ &\times\left[(r\sigma_0 c_r^2)' - \frac{\kappa^2\sigma_0}{4\Omega^2}c_r^2\right]\delta(v_r)\delta'(\tilde{v}_\varphi) + \frac{1}{2}c_r^2(r) \\ &\times\delta''(v_r)\delta(\tilde{v}_r) + \frac{1}{2}\frac{\kappa^2}{4\Omega^2}c_r^2(r)\delta(v_r)\delta''(\tilde{v}_\varphi)\Bigg\}. \end{aligned} \qquad (33)$$

In general, we may arbitrarily specify the system surface density $\sigma_0(r)$ and radial velocity dispersion $c_r(r)$ (which must be small). The function $\Omega(r)$ can be calculated from the equilibrium condition for a star in a circular orbit:

$$\Omega^2 r = d\Phi_0/dr,$$

where the potential $\Phi_0(r)$ is determined by the distribution of all masses in the galaxy, one of the components of which (the flattest, disk one) we consider.

Let us derive the "spread" distribution function that corresponds to (33),

$$f_0(E, L) = F_0(r_0, \mathscr{E}), \tag{34}$$

where E and L are, respectively, the energy and angular momentum of the star (per unit mass):

$$E = \frac{1}{2}(v_r^2 + v_\varphi^2) + \Phi_0(r),$$

$L = rv_\varphi$; and r_0 and $\mathscr{E}$ are the epycyclic integrals, which can be determined as functions of E and L from the equations [14]

$$r_0^2\Omega(r_0) = L, \quad \mathscr{E} = E - E_c(r_0), \tag{35}$$

$$\begin{gathered} E_c(r_0) = \frac{1}{2}r_0^2\Omega^2(r_0) + \Phi_0(r_0), \\ r_0\Omega^2(r_0) = \frac{d\Phi_0}{dr_0}. \end{gathered} \tag{36}$$

We assume the deviations from circular orbits to be small, i.e.,

$$|\mathscr{E}| \ll |E_c|, \quad |r - r_0| \ll r_0.$$

An example of the distribution function (34) is the generalized Schwarzschild distribution function

$$F_0 = P(r_0)\exp\left[-\frac{\mathscr{E}}{c_0^2(r_0)}\right], \tag{37}$$

which was considered by Shu [14], where the functions $P(r_0)$ and $c_0(r_0)$ should be expressed in terms of the surface density and radial velocity dispersion. The distribution function (37) can be written in variables r, v_r, and $\tilde{v}_\varphi$ with the accuracy required for subsequent use as follows:

$$\begin{gathered} F_0(r, v_r, \tilde{v}_\varphi) = \frac{2\Omega}{\kappa}\frac{\sigma_0}{2\pi c_0^2}\exp\left[-\frac{v_r^2 + \left(\frac{2\Omega\tilde{v}_\varphi}{\kappa}\right)^2}{2c_0^2}\right] \\ \times\left\{1 + \left(\frac{\sigma_0'}{\sigma_0} + \frac{\Omega'}{\Omega} - \frac{\kappa'}{\kappa}\right)\frac{1}{\kappa}\left(\frac{2\Omega\tilde{v}_\varphi}{\kappa}\right)\right. \\ \left. - \left(\frac{\Omega'}{\Omega} - \frac{2\kappa'}{3\kappa}\right)\frac{1}{2\kappa c_0^2}\left(\frac{2\Omega\tilde{v}_\varphi}{\kappa}\right)^3\right\}, \end{gathered} \tag{38}$$

where

$$\tilde{v}_\varphi = \frac{r_0^2\Omega(r_0)}{r} - r\Omega(r) \tag{39}$$

denotes the residual azimuthal velocity. In (38), we assumed that $c_r = c_0 \neq c_0(r)$; this slightly simplifies the calculations but is of no fundamental importance.

It is easy to verify that the distribution function (38) actually belongs to type (37) with

$$P(r_0) = \frac{2\Omega(r_0)}{\kappa(r_0)}\frac{\sigma_0(r_0)}{2\pi c_0^2}.$$

This primarily requires finding a relation between $r_0 - r$ and $\tilde{v}_\varphi$ by expanding the right-hand side of Eq. (39) in powers of the small difference $r_0 - r$:

$$r_0 - r \approx \frac{1}{\kappa}\frac{2\Omega\tilde{v}_\varphi}{\kappa} - \frac{2\Omega}{\kappa^4}\left(\frac{2\Omega\tilde{v}_\varphi}{\kappa}\right)^2. \tag{40}$$

Expanding the right-hand side of the equation

$$\begin{gathered} \mathscr{E} = \frac{1}{2}v_r^2 + \frac{r_0^4\Omega^2(r_0)}{2r^2} \\ - \frac{1}{2}r_0^2\Omega^2(r_0) + \Phi_0(r) - \Phi_0(r_0) \end{gathered} \tag{41}$$

in powers of $r_0 - r$ and then substituting $r_0 - r$ from (40) into the derived expression, we obtain after some transformations

$$\mathscr{E} \approx \frac{v_r^2}{2} + \frac{1}{2}\left(\frac{2\Omega\tilde{v}_\varphi}{\kappa}\right)^2 + \frac{1}{2}\left(\frac{2\Omega\tilde{v}_\varphi}{\kappa}\right)^3\frac{1}{\kappa}\left(\frac{\Omega'}{\Omega} - \frac{2\kappa'}{3\kappa}\right). \tag{42}$$

Finally, writing

$$\begin{gathered} P(r_0) \approx P(r)[1 + (r_0 - r)P'(r)] \\ \approx \frac{2\Omega}{\kappa}\frac{\sigma_0}{2\pi c_0^2}\left\{1 + \left(\frac{\sigma_0'}{\sigma_0} + \frac{\Omega'}{\Omega} - \frac{\kappa'}{\kappa}\right)\frac{1}{\kappa}\left(\frac{2\Omega\tilde{v}_\varphi}{\kappa}\right)\right\}, \end{gathered}$$

by using (40) and representing the exponent in (37) as

$$\begin{gathered} \exp\left(-\frac{\mathscr{E}}{c_0^2}\right) \approx \exp\left[-\frac{v_r^2 + \left(\frac{2\Omega\tilde{v}_\varphi}{\kappa}\right)^2}{2c_0^2}\right] \\ \times\left\{1 - \left(\frac{\Omega'}{\Omega} - \frac{2\kappa'}{3\kappa}\right)\frac{1}{2\kappa c_0^2}\left(\frac{2\Omega\tilde{v}_\varphi}{\kappa}\right)^3\right\}, \end{gathered}$$

we arrive at (38).

Calculating the azimuthal flux

$$\Pi_\varphi = \int\tilde{v}_\varphi F_0 dv_r d\tilde{v}_\varphi,$$

we derive an expression for it that agrees with (33) (at $c_0 = \text{const}$).

The distribution function in the epicyclic approximation can be inferred from (38) if we retain only unity in the entire expression in curly braces:

$$f_0 = \frac{2\Omega(r)}{\kappa(r)}\frac{\sigma_0(r)}{2\pi c_0^2}\exp\left(-\frac{v_r^2}{2c_r^2} - \frac{\tilde{v}_\varphi^2}{2c_\varphi^2}\right), \tag{43}$$

where

$$c_\varphi^2 = c_r\kappa^2/4\Omega^2.$$

This (Schwarzschild) distribution function is most commonly used both in theoretical studies (e.g., in the stability theory for gravitating disks [15–17]) and in reducing observational data and comparing them with the theory. Clearly, the Schwarzschild distribution function is an anisotropic Maxwell distribution function, but with a certain temperature anisotropy:

$$\frac{T_r}{T_\varphi} = \frac{c_r^2}{c_\varphi^2} = \left(\frac{2\Omega}{\kappa}\right)^2.$$

For our purposes, however, the purely Schwarzschild distribution is not enough, and we will have to use a distribution function of the more general form (38).

3. RESPONSES OF GRAVITATING DISKS WITH A FINITE STELLAR VELOCITY DISPERSION

3.1. Deriving Expressions for the Resonance Responses

The linearized kinetic equation can be written as (see, e.g., [1, 2, 18])

$$\frac{df_1}{dt} = \frac{\partial\Phi_1}{\partial\mathbf{r}}\cdot\frac{\partial f_0}{\partial\mathbf{v}}, \tag{44}$$

where d/dt is the time derivative with respect to an unperturbed stellar orbit in the phase space, and f_1 and Φ_1 are the distribution-function and potential perturbations, respectively.

For the distribution function (34), the right-hand side of (44) can be transformed to

$$\frac{\partial\Phi_1}{\partial\mathbf{r}}\cdot\frac{\partial f_0}{\partial\mathbf{v}} = \frac{\partial F_0}{\partial\mathcal{E}}\frac{d\Phi_1}{dt} + i\omega_*(r_0)\frac{\partial F_0}{\partial\mathcal{E}}\Phi_1 + \frac{2im\Omega(r_0)}{r_0\kappa^2(r_0)}\frac{\partial F_0}{\partial r_0}\Phi_1. \tag{45}$$

Writing the solution $f_1(r, v_r, v_\varphi)\exp(-i\omega t + im\varphi)$ (ω is the frequency) of Eq. (44) with the right-hand side (45) as the path integral (see, e.g., [1, 2]) and integrating f_1 over velocities yields the perturbation of the surface density $\sigma_1(r)\exp(-i\omega t + im\varphi)$:

$$\begin{aligned}\sigma_1(r) &= \int dv_r d\tilde{v}_\varphi f_1 = -\frac{1}{c_0^2}\int dv_r d\tilde{v}_\varphi \\ &\times F_0\left\{\Phi_1(r) + i\omega_*(r_0)\int_{-\infty}^{0}\Phi_1[r(t)]\right. \\ &\left.\times\exp[-i\omega t + im[\varphi(t)-\varphi]]dt\right\} \\ &+ \int dv_r d\tilde{v}_\varphi\frac{2im\Omega(r_0)}{r_0\kappa^2(r_0)}\frac{\partial F_0}{\partial r_0} \\ &\times\int_{-\infty}^{0}\Phi_1[r(t)]\exp[-i\omega t + im[\varphi(t)-\varphi]]dt,\end{aligned} \tag{46}$$

where we designated

$$\omega_*(r_0) = \omega - m\Omega(r_0);$$

we assume that the functions $r(t)$ and $\varphi(t)$ represent the unperturbed stellar orbit that has coordinates (r, φ) and velocities (v_r, v_φ) at time $t = 0$ and that the equilibrium distribution function has the generalized Schwarzschild form (37) at $c_0 = \text{const}$. Thus, it is clear that the proper passage to the limiting case of a cold disk ($c_0 \longrightarrow 0$) requires the expressions for $r(t)$ and $(\varphi(t) - \varphi)$ calculated in the post-epicyclic approximation. Below, we also make this passage but without expanding the resonance denominators.

Note that the post-epicyclic approximation for the radius $r(t) = r + \delta r$ was previously used by Shu [14] in his theory of tightly wound spiral density waves, which generalizes the theory of Lin and Shu [16, 17]; however, there are errors in Shu's formulas for δr and the relation between r_0 and r.

Let us expand the function $\Phi_1[r(t)]$ appearing in the path integral in a Taylor series of powers of $\delta r = r(t) - r$ to within terms $\sim(\delta r)^2$:

$$\Phi_1[r(t)] \approx \Phi_1(r) + \delta r\Phi_1'(r) + \frac{1}{2}(\delta r)^2\Phi_1''(r); \tag{47}$$

accordingly, we also expand in powers of

$$\delta\varphi = \varphi(t) - \varphi - (\Omega_0 + \delta\Omega)t$$

the exponent:

$$\begin{aligned}&\exp[-i\omega t + im[\varphi(t) - \varphi]] \\ &= \exp[-i[\omega_*(r_0) - m\delta\Omega]t]\exp[im\delta\varphi] \\ &\approx \exp[-i[\omega_*(r_0) - m\delta\Omega]t]\left\{1 + im\delta\varphi - \frac{m^2}{2}(\delta\varphi)^2\right\}.\end{aligned} \tag{48}$$

Clearly, these expansions may be used if the perturbing potential $\Phi_1(r, \varphi)$ changes only slightly on the epicycle length. The two mentioned ways of generating resonance responses (by a bar or a satellite) satisfy this condition.

It is convenient to represent the expression derived from calculations for the surface density perturbation as

$$\sigma_1 = \Sigma_1\Phi_1''(r) + \Sigma_2\Phi_1'(r) + \Sigma_3\Phi_1(r), \tag{49}$$

where

$$\begin{aligned}\Sigma_1 = {} & -\frac{1}{c_0^2}\int dv_r d\tilde{v}_\varphi F_0 \times \frac{1}{2}\omega_*(r)\int\limits_{-\infty}^{0} dt\exp(-i\omega_*(r)t) \\ & \times\left[\frac{1}{2}(1-\cos(2\kappa t))\frac{v_r^2}{\kappa^2}\right. \\ & \left. + \frac{1}{2}(3-4\cos(\kappa t)+\cos(2\kappa t))\left(\frac{2\Omega\tilde{v}_\varphi}{\kappa}\right)^2\right],\end{aligned} \tag{50}$$

$$\begin{aligned}\Sigma_2 = {} & -\frac{1}{c_0^2}\int dv_r d\tilde{v}_\varphi F_0 \times i\omega_*(r_0)\int\limits_{-\infty}^{0} dt\exp(-i\omega_*(r_0)t) \\ & \times\left[(1-\cos(\kappa_0 t))\frac{1}{\kappa_0}\left(\frac{2\Omega\tilde{v}_\varphi}{\kappa_0}\right) + \sin(\kappa_0 t)\frac{v_r}{\kappa_0} + \frac{2\Omega\alpha}{\kappa^4}\right. \\ & \left.\times\left(\frac{2\Omega\tilde{v}_\varphi}{\kappa}\right)^2(1-\cos(\kappa t)) + \frac{2\Omega\beta}{\kappa^4}v_r^2(1-\cos(\kappa t))\right] \\ & \times\left\{1 + im\left[\sin(\kappa_0 t)\frac{2\Omega_0}{\kappa_0^2 r_0}\frac{2\Omega_0\tilde{v}_\varphi}{\kappa_0}\right.\right. \\ & \left.\left. -(1-\cos(\kappa_0 t))\frac{2\Omega_0}{\kappa_0^2 r_0}v_r\right]\right\},\end{aligned} \tag{51}$$

$$\begin{aligned}\Sigma_3 = {} & -\frac{1}{c_0^2}\int dv_r d\tilde{v}_\varphi F_0 \times i\omega_*(r_0)\int\limits_{-\infty}^{0} dt\exp(-i\omega_*(r_0)t) \\ & \times\left\{im\left[\sin(\kappa_0 t)\frac{2\Omega_0}{\kappa_0^2 r_0}\left(\frac{2\Omega_0\tilde{v}_\varphi}{\kappa_0}\right) - \frac{2\Omega_0 v_r}{\kappa_0^2 r_0}(1-\cos(\kappa_0 t))\right.\right. \\ & \left. + \sin(\kappa t)\left(\frac{4\Omega^2}{\kappa^2}\right)\frac{\alpha}{\kappa^3 r}\left(\frac{2\Omega\tilde{v}_\varphi}{\kappa}\right)^2 - \sin(\kappa t)\left(\frac{4\Omega^2}{\kappa^2}\right)\frac{\beta}{\kappa^3 r}v_r^2\right] \\ & -\frac{m^2\Omega^2}{\kappa^4 r^2}\left[(1-\cos(2\kappa t))\left(\frac{2\Omega\tilde{v}_\varphi}{\kappa}\right)^2\right. \\ & \left.\left. + (3-4\cos\kappa t + \cos(2\kappa t))v_r^2\right]\right\}\end{aligned} \tag{52}$$

$$\begin{aligned} & + \frac{1}{c_0^2}\int dv_r d\tilde{v}_\varphi F_0\frac{m\delta\Omega}{\omega_*(r_0)} \\ & + \frac{2im\Omega}{r\kappa^2}\int dv_r d\tilde{v}_\varphi \left.\frac{\partial F_0}{\partial r_0}\right|_{r_0 = r}\int\limits_{-\infty}^{0} dt\exp(-i\omega_*(r)t).\end{aligned}$$

The integrals over t in (50)–(52) can be calculated from the formulas

$$\int\limits_{-\infty}^{0} dt\exp(-i\omega_*(r_0)t) = \frac{1}{-i\omega_*(r_0)},$$

$$\int\limits_{-\infty}^{0}\exp(-i\omega_*(r_0)t)\sin(\kappa_0 t)dt = \frac{\kappa_0}{\omega_*^2(r_0)-\kappa_0^2},$$

$$\int\limits_{-\infty}^{0}\exp(-i\omega_*(r_0)t)\cos(\kappa_0 t)dt = -\frac{\omega_*(r_0)}{i(\omega_*^2(r_0)-\kappa_0^2)}.$$

When integrating over velocities, one should first keep in mind that the equilibrium distribution function of a disk with nearly circular orbits satisfies the conditions

$$\int v_r F_0 dv_r d\tilde{v}_\varphi = 0, \quad \int v_r\tilde{v}_\varphi F_0 dv_r d\tilde{v}_\varphi = 0, \tag{53}$$

$$\int v_r^2 F_0 dv_r d\tilde{v}_\varphi = \int\left(\frac{2\Omega\tilde{v}_\varphi}{\kappa}\right)^2 F_0 dv_r d\tilde{v}_\varphi. \tag{54}$$

It is also important to take into the fact that the azimuthal flux is

$$\Pi_\varphi = \int\tilde{v}_\varphi F_0 dv_r d\tilde{v}_\varphi \approx \sigma_0\frac{c_0^2}{\Omega r} \neq 0. \tag{55}$$

Properties (53)–(55), along with a specific expression for the flux Π_φ, were derived in the preceding section.

Making the symmetrizing change

$$v_r \longrightarrow x = \frac{v_r}{c_0}, \quad \tilde{v}_r \longrightarrow y = \frac{2\Omega\tilde{v}_\varphi}{\kappa c_0}, \tag{56}$$

we obtain

$$F_0(r, v_r, \tilde{v}_\varphi)dv_r d\tilde{v}_\varphi = f_0(r, x, y)dxdy, \tag{57}$$

$$\begin{aligned} f_0(r, x, y) = {} & \frac{\sigma_0}{2\pi}\exp\left(-\frac{x^2+y^2}{2}\right) \\ & \times\left\{1 + \rho y\left(\frac{\sigma_0'}{\sigma_0} + \frac{\Omega'}{\Omega} - \frac{\kappa'}{\kappa}\right) - \frac{1}{2}\rho y^3\left(\frac{\Omega'}{\Omega} - \frac{2\kappa'}{3\kappa}\right)\right\},\end{aligned} \tag{58}$$

where $\rho = c_0/\kappa$ is the epicycle scale size. Simple calculations allow expressions (50)–(52) for Σ_1, Σ_2, and Σ_3 to be reduced to

$$\Sigma_1 = -\frac{\sigma_0}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dy \frac{y^2 e^{-y^2/2}}{\omega_*^2(r_0) - \kappa^2(r_0)}, \tag{59}$$

$$\Sigma_2 = C^{(1)} + C^{(2)}, \tag{60}$$

$$C^{(1)} = -\frac{\sigma_0}{\sqrt{2\pi}\rho}\int_{-\infty}^{\infty} dy \frac{e^{-y^2/2}}{\omega_*^2(r_0) - \kappa^2(r_0)} \times\left[y + y^2\rho\left(\frac{\sigma_0'}{\sigma_0} + \frac{2\Omega'}{\Omega} - \frac{\kappa'}{\kappa}\right) - \frac{1}{2}y^4\rho\left(\frac{\Omega'}{\Omega} - \frac{2\kappa'}{3\kappa}\right)\right], \tag{61}$$

$$C^{(2)} = -\frac{2\Omega}{\kappa^2}\frac{\sigma_0}{\sqrt{2\pi}} \times \int_{-\infty}^{\infty} dy \frac{e^{-y^2/2}}{\omega_*^2(r_0) - \kappa^2(r_0)}(\alpha y^2 + \beta), \tag{62}$$

$$\Sigma_3 = A + D^{(1)} + D^{(2)}, \tag{63}$$

$$A = -\frac{2m\Omega}{r\kappa^2}\left(\frac{\sigma_0'}{\sigma_0} + \frac{\Omega'}{\Omega}\right)\frac{\sigma_0}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dy \frac{e^{-y^2/2}}{\omega_*(r_0)} + \frac{m\Omega}{r\kappa^2}\frac{\kappa'}{\kappa}\frac{\sigma_0}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dy \frac{e^{-y^2/2}(3 + y^2)}{\omega_*(r_0)}, \tag{64}$$

$$D^{(1)} = \frac{2m\Omega\omega_*}{\kappa^2 r}\frac{\sigma_0}{\sqrt{2\pi}\rho}\int_{-\infty}^{\infty} dy \frac{e^{-y^2/2}}{\omega_*^2(r_0) - \kappa^2(r_0)} \times\left[y + y^2\rho\left(\frac{\sigma_0'}{\sigma_0} + \frac{3\Omega'}{\Omega} - \frac{m\Omega'}{\omega_*} - \frac{3\kappa'}{\kappa} - \frac{1}{r}\right) - \frac{1}{2}y^4\rho\left(\frac{\Omega'}{\Omega} - \frac{2\kappa'}{3\kappa}\right)\right], \tag{65}$$

$$D^{(2)} = \frac{4\Omega^2}{\kappa^2}\frac{\sigma_0}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dy \frac{e^{-y^2/2}}{\omega_*^2(r_0) - \kappa^2(r_0)} \times\left[(\alpha y^2 + \beta)\frac{m\omega_*}{r\kappa^2} + \frac{m^2}{r^2}\right]. \tag{66}$$

All the functions that appear in (59)–(66) as the coefficients of the integrals are assumed to be calculated at the resonance position $r = r_c$. In all these formulas, we must assume that

$$r_0 \approx r + \rho y, \tag{67}$$

and that, for example, near corotation,

$$\omega_*(r_0) \approx m|\Omega'(r_c)|[(r - r_c) + \rho y + i\delta], \tag{68}$$

$\delta = \gamma/m|\Omega'(r_c)|$, and γ is the imaginary part of the frequency: $\omega = m\Omega_p + i\gamma$, and Ω_p is the angular velocity of the wave.

To pass to the limit of a cold disk, we must expand the resonance denominators in (59)–(66):

$$\frac{1}{\omega_*(r_0)} \longrightarrow \frac{1}{\omega_*(r)} + \left(\frac{1}{\omega_*(r)}\right)'\rho y,$$

$$\frac{1}{\omega_*^2(r_0) - \kappa_0^2} \longrightarrow \frac{1}{\omega_*^2(r) - \kappa^2} + \left(\frac{1}{\omega_*^2(r) - \kappa^2}\right)'\rho y;$$

take into account the fact that

$$\alpha + \beta = \frac{\kappa^2(1 - r\Omega'/2\Omega)}{2\Omega r};$$

and calculate the emerging integrals over y. It is easy to verify that passing to the limit yields the following expression for $\sigma_1(r)$:

$$\sigma_1(r) = -\frac{1}{r}\frac{d}{dr}\left(r\varepsilon\frac{d\Phi_1}{dr}\right) + \frac{4}{r^2}\varepsilon\Phi_1 + \frac{4}{r\omega_*}\Phi_1\frac{d}{dr}(\varepsilon\Omega). \tag{69}$$

Here,

$$\varepsilon(r) = \frac{\sigma_0(r)}{\omega_*^2 - \kappa^2}$$

is the gravitational analog of the dielectric constant.

For a convenience of comparison, it is useful to transform expression (69) to

$$\begin{aligned}\sigma_1 = {} & -\Phi_1''\frac{\sigma_0}{\omega_*^2 - \kappa^2} - \frac{\Phi_1'}{r}\times\left\{\frac{\sigma_0}{\omega_*^2 - \kappa^2} + \frac{r\sigma_0'}{\omega_*^2 - \kappa^2} + \sigma_0 r\left(\frac{1}{\omega_*^2 - \kappa^2}\right)'\right\} \\ & + \Phi_1\left\{\frac{2m\Omega\sigma_0'}{r\kappa^2}\left(\frac{\omega_*}{\omega_*^2 - \kappa^2} - \frac{1}{\omega_*}\right) + \frac{m^2}{r^2}\frac{\sigma_0}{\omega_*^2 - \kappa^2}\right. \\ & + \frac{2m\Omega'\sigma_0}{r\kappa^2}\left(\frac{\omega_*}{\omega_*^2 - \kappa^2} - \frac{1}{\omega_*}\right) - \frac{2m\Omega\sigma_0}{r\kappa^2}\frac{\omega_*(\omega_*^2 - \kappa^2)'}{(\omega_*^2 - \kappa^2)^2} \\ & \left. - \frac{4m^2\Omega\sigma_0}{r^2\kappa^2}\frac{r\Omega'}{\omega_*^2 - \kappa^2} - \frac{4m\Omega\sigma_0}{r(\omega_*^2 - \kappa^2)\omega_*}\frac{\kappa'}{\kappa}\right\}.\end{aligned} \tag{70}$$

As we see from (69), only the principal galactic resonances appear in the cold-disk approximation: corotation ($\omega_* = 0$) and Lindblad ($\omega_*^2 = \kappa^2$) ones.

Note that expression (69) can be directly derived from the hydrodynamic equations with a zero pressure (which corresponds to a purely circular motion of disk particles). Indeed, linearizing these hydrodynamic equations (written in cylindrical coordinates) yields

$$\frac{\partial\sigma_1}{\partial t} + \Omega\frac{\partial\sigma_1}{\partial\varphi} + \frac{1}{r}\frac{\partial}{\partial r}(r\sigma_0 v_{r1}) + \frac{\sigma_0}{r}\frac{\partial v_{\varphi 1}}{\partial\varphi} = 0, \qquad (71)$$

$$\frac{\partial v_{r1}}{\partial t} + \Omega\frac{\partial v_{r1}}{\partial\varphi} - 2\Omega v_{\varphi 1} = -\frac{\partial\Phi_1}{\partial r}, \qquad (72)$$

$$\frac{\partial v_{\varphi 1}}{\partial t} + \Omega\frac{\partial v_{\varphi 1}}{\partial\varphi} + \frac{\kappa^2}{2\Omega}2\Omega v_{r1} = -\frac{1}{r}\frac{\partial\Phi_1}{\partial\varphi}, \qquad (73)$$

where v_{r1} and $v_{\varphi 1} = v_\varphi - r\Omega(r)$ are the perturbed velocities. We find from the last two equations that

$$v_{r1} = \frac{1}{\omega_*^2 - \kappa^2}\left(-i\omega_*\frac{\partial\Phi_1}{\partial r} + \frac{2im\Omega}{r}\Phi_1\right), \qquad (74)$$

$$v_{\varphi 1} = \frac{1}{\omega_*^2 - \kappa^2}\left(\frac{m\omega_*}{r}\Phi_1 - \frac{\kappa^2}{2\Omega}\frac{\partial\Phi_1}{\partial r}\right). \qquad (75)$$

Substituting (74) and (75) into (71), we obtain relation (69) between Φ_1 and σ_1 after some transformations. Note that this relation in the compact form (69) was first given in [19].

Surprising as it may seem, nobody has ever performed the passage to the limiting case of a cold disk from a kinetic treatment (even the proper expressions for orbits in the post-epicyclic approximation required for this passage to the limit have not been derived). However, we are interested in the effect of a small velocity dispersion, which modifies the resonance denominators. Being interested in the disk regions near one resonance or another, we must separate the dominant terms from (59)–(66).

A particularly simple expression for the surface-density response is derived near the corotation resonance:

$$\sigma_1 = A\Phi_1. \qquad (76)$$

The expression near the Lindblad resonances is much more cumbersome:

$$\sigma_1 \approx B\Phi_1'' + C\Phi_1' + D\Phi_1, \qquad (77)$$

where we designated

$$B = \Sigma_1, \quad C = \Sigma_2, \quad D = \Sigma_3 - A.$$

3.2. Disk Responses near the Principal Resonances

There is a variety of possible resonance disk responses. However, we focus only on spiral responses [although the other forms of resonance responses are described by the same general expressions (76) and (77) and can be easily obtained from them]. Spiral responses are most commonly encountered in galaxies. At the same time, for example, rings often consist of pairs of nearly circular half-turn spirals revealed in high-resolution observations.

As was already noted above, a galactic disk is capable of responding to the action from a central bar or a satellite in a resonant way (see Fig. 1). Our derived expressions for resonance responses are suitable for use in both cases, but below, for definiteness, we restrict ourselves to a more detailed analysis of bar-generated responses.

Let us first determine the disk responses obtained in the limit $\gamma \longrightarrow 0$; these are of particular interest, because they correspond to quasi-steady states (see also the Conclusion). For this passage to the limit, the resonance denominators simplify as follows:

$$\frac{1}{\omega_*(r_0)} \propto \frac{1}{y + \delta r' + i\delta'} = P\frac{1}{y + \delta r'} - i\pi\delta(y + \delta r'), \qquad (78)$$

$$\frac{1}{\omega_*(r_0) + \kappa(r_0)} \propto \frac{1}{-\mathrm{sgn}(\Omega_{pr}')(y + \delta r') + i\delta''}$$

$$= \begin{cases} \dfrac{1}{y + \delta r' + i\delta''} = P\dfrac{1}{y + \delta r'} - i\pi\delta(y + \delta r'), & (79) \\ \Omega_{pr}' < 0, & \\ -\dfrac{1}{y + \delta r' - i\delta''} = -P\dfrac{1}{y + \delta r'} - i\pi\delta(y + \delta r'), & (80) \\ \Omega_{pr}' > 0, & \end{cases}$$

$$\begin{aligned} \frac{1}{\omega_*(r_0) - \kappa(r_0)} &\propto \frac{1}{y + \delta r' + i\delta'''} \\ &= P\frac{1}{y + \delta r'} - i\pi\delta(y + \delta r'), \end{aligned} \qquad (81)$$

where

$$\delta r' = \frac{r - r_c}{\rho}, \quad \delta' = \frac{\gamma}{2\rho|\Omega'|},$$

$$\delta'' = \frac{\gamma}{2\rho|\Omega_{pr}'|}, \quad \delta''' = \frac{\gamma}{2\rho|\Omega_{pr}^{+\prime}|};$$

P means the principal value; recall also the previously introduced (at the end of Subsection 2.2) notation for the orbital precession velocities in the epicyclic approximation:

$$\Omega_{pr} = \Omega - \kappa/2, \quad \Omega_{pr}^{+} = \Omega + \kappa/2.$$

For the response at corotation, we therefore have

$$\sigma_1 \propto A_c(\delta r')\cos\{2[\varphi - \varphi_0^{(c)}(\delta r')]\}, \qquad (82)$$

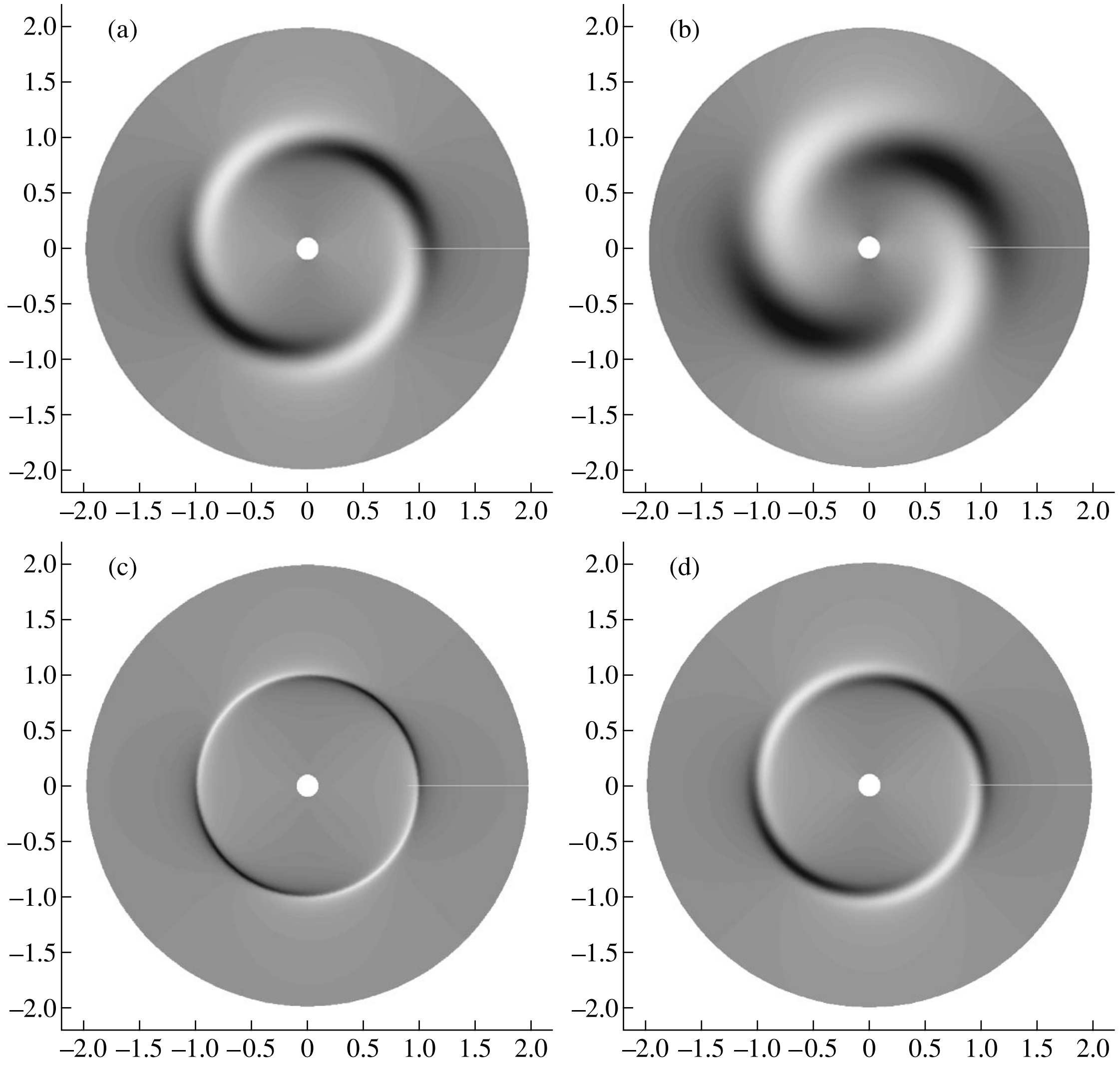


Fig. 4. Resonance disk responses at corotation ($r_c = 1$). The disk rotates counterclockwise. The bar is oriented vertically. The degree of blackening increases with perturbed surface density. (a) A universal spiral at $\rho = 0.1$; (b) a universal spiral at $\rho = 0.2$; (c) a cold-disk response at $\gamma = 0.05$; and (d) a disk response at $\gamma = 0.05$ and $\rho = 0.05$.

where

$$A_c(\delta r') = \sqrt{R_c^2(\delta r') + I_c^2(\delta r')}, \tag{83}$$

$$\begin{aligned} R_c(\delta r') &= P\int_{-\infty}^{\infty} dy \frac{e^{-y^2/2}}{y + \delta r'} \\ &\times \left[\frac{\kappa'}{\kappa}(3 + y^2) - 2\left(\frac{\sigma_0'}{\sigma_0} + \frac{\Omega'}{\Omega}\right)\right], \end{aligned} \tag{84}$$

$$\begin{aligned} I_c(\delta r') &= \pi \exp(-(\delta r')^2/2) \\ &\times \left\{\frac{\kappa'}{\kappa}[3 + (\delta r')^2] - 2\left(\frac{\sigma_0'}{\sigma_0} + \frac{\Omega'}{\Omega}\right)\right\}, \end{aligned} \tag{85}$$

$$\tan[2\varphi_0^{(c)}(\delta r')] = \frac{I_c(\delta r')}{R_c(\delta r')}. \tag{86}$$

According to (82), the response σ_1 expressed in variables $\delta r'$ and φ is a universal function (for a given equilibrium model, i.e., in our case, at fixed σ_0'/σ_0, Ω'/Ω, and κ'/κ on the corotation circle). In particular, the equation for the spiral that corresponds to maximum density (at each radius) is universal:

$$\varphi = \varphi_0^{(c)}(\delta r'). \tag{87}$$

Accordingly, the changes in azimuth $\Delta\varphi$ along this spiral are also the same when the amplitude A decreases by a certain factor (because the amplitude is also a given function of $\delta r'$). On the other hand, it follows from the above analysis that the thickness of the spiral response and the spiral inclination in real coordinates (r, φ) are proportional to the epicycle size ρ. These properties of the responses at corotation are illustrated in Figs. 4a and 4b, where we assume that $r_c = 1$ and take a model with the following values typical of spiral galaxies:

$$\frac{\sigma_0'}{\sigma_0} = -\frac{1}{r_c/2}, \quad \frac{\Omega'}{\Omega} = -\frac{1}{r_c}, \quad \frac{\kappa'}{\kappa} = -\frac{1}{r_c}. \tag{88}$$

Similarly, at the external ILR, we obtain (retaining the principal terms in the expression for the response)

$$\sigma_1 \propto -(n - n_{cr})A_L(\delta r')\cos\{2[\varphi - \varphi_0^{(L)}(\delta r')]\}, \tag{89}$$

where n is the exponent in the dependence

$$\Phi_1 \propto r^{-n}\cos 2\varphi$$

for the bar potential ($n > 0$, the bar is assumed to be vertical), $n_{cr} = 4\Omega/\kappa|_{r_c}$,

$$A_L(\delta r') = \sqrt{R_L^2(\delta r') + I_L^2(\delta r')}, \tag{90}$$

$$\tan[2\varphi_0^{(L)}(\delta r')] = \frac{I_L(\delta r')}{R_L(\delta r')}, \tag{91}$$

and we give expressions for the functions $R_L(\delta r')$ and $I_L(\delta r')$ here only for the specific model (88):

$$R_L(\delta r') = P\int_{-\infty}^{\infty} dy \frac{ye^{-y^2/2}}{y + \delta r'}, \tag{92}$$

$$I_L(\delta r') = -\pi\delta r' e^{-(\delta r')^2/2}. \tag{93}$$

The responses calculated from (90)–(93) for $\rho = 0.1$ and 0.2 are shown in Figs. 5a and 5b. In this case, the rotation curve $V_{0\varphi} = r\Omega(r)$ is assumed to be flat: $V_{0\varphi} =$ const (which roughly holds for the most of the disk in most spiral galaxies); for this rotation curve, $n_{cr} = 2\sqrt{2} \approx 2.82$.

Here, we give no expressions for the responses at the two remaining principal resonances (internal ILR and OLR); the responses at the internal ILR calculated from the corresponding formulas are shown in Figs. 6a and 6b; as for the responses at the OLR, they are similar to the responses at the external ILR for $n < n_{cr}$ (Figs. 5a and 5b). Note the following peculiarity of the responses at the internal ILR: they have the form of leading spirals; i.e., they unwind in the sense of galactic-disk rotation (counterclockwise). In all the remaining cases (see Figs. 4 and 5), the spirals are trailing ones: they seem to follow the galactic rotation (see the Conclusion for more details).

Let us now consider the responses of cold disks ($\rho = 0$). Since this case has previously been studied [20, 21], we restrict ourselves to several remarks pertaining to spiral responses (to compare them with the responses in disks with nearly circular, but not exactly circular, orbits).

(1) Near corotation [$\omega_* \approx 0$, i.e., $\Omega_p = \Omega(r_c)$ and $\Omega_p = \omega/2$], we obtain from Eq. (69)

$$\sigma_1(r, \varphi) \approx \frac{2r^{-n-1}}{\sqrt{\Omega_c'^2(r - r_c)^2 + \gamma^2/4}} \times \frac{d}{dr}\left(\frac{\Omega\sigma_0}{\kappa^2}\right)\cos 2(\varphi - \varphi_c(r)), \tag{94}$$

$$\tan 2\varphi_c(r) = -\frac{\gamma}{2\Omega_c'(r - r_c)}, \tag{95}$$

where $\Omega_c' = d\Omega/dr|_{r_c}$. These formulas describe a two-arm trailing (because $\Omega_c' < 0$; see Fig. 3) spiral. A typical response is shown in Fig. 4c; it refers to a simple exponential-disk model,

$$\sigma_0(r) = \sigma_0(0)e^{-r/r_d}$$

[at $\sigma_0(0) = 1$, $r_d = 1/2$], and a flat rotation curve, $\Omega = V_0/r$ ($V_0 = 1$); the bar rotates with an angular velocity Ω_p such that $r_c = 1$. Note that to provide this rotation curve, the disk must be within a sufficiently massive spherical component (which is actually the case); the latter is essentially not involved in perturbations in the disk plane, because the orbits of its constituent stars are highly eccentric.

Clearly, the maximum angular length of each arm is $\pi/2$; the longest arms are obtained at small γ. Figure 4c corresponds to the condition $|\sigma_0'/\sigma_0| \gg |\Omega'/\Omega|$ (which appears to be always satisfied). For the inverse inequality, we would have an unobservable arrangement of spirals relative to the bar, which is derived from the arrangement shown in the figure by counterclockwise rotation through $\pi/2$. The intermediate cases $|\sigma_0'/\sigma_0| \geqslant |\Omega'/\Omega|$ describe a realistic pattern of spirals that go slightly behind the bar and that branch off from it at an angle smaller than $\pi/2$.

(2) Near one of the ILRs ($r \approx r_{2i}$ or $r \approx r_{2e}$), Eq. (69) gives

$$\frac{\sigma_1(r)}{\sigma_0(r)} \approx (n_{cr} - n)r\Omega_{pr}' \frac{r^{-n-2}}{\kappa[4(\Omega_b - \Omega_{pr}(r))^2 + \gamma^2]} \times \cos 2(\varphi - \varphi_2(r)), \tag{96}$$

$$\tan\varphi_2(r) = -\frac{\gamma}{2\Omega_{pr}'(r - r_c)}, \tag{97}$$

where

$$\Omega_{pr}' = \left.\frac{d\Omega_{pr}}{dr}\right|_{r_c}, \quad n_{cr} = \left.\frac{4\Omega}{\kappa}\right|_{r_c}$$

(for the assumed flat rotation curve, $n_{cr} = 2\sqrt{2} \approx 2.82$). It follows from these formulas that, at $r \approx r_{2e}$, the reso-

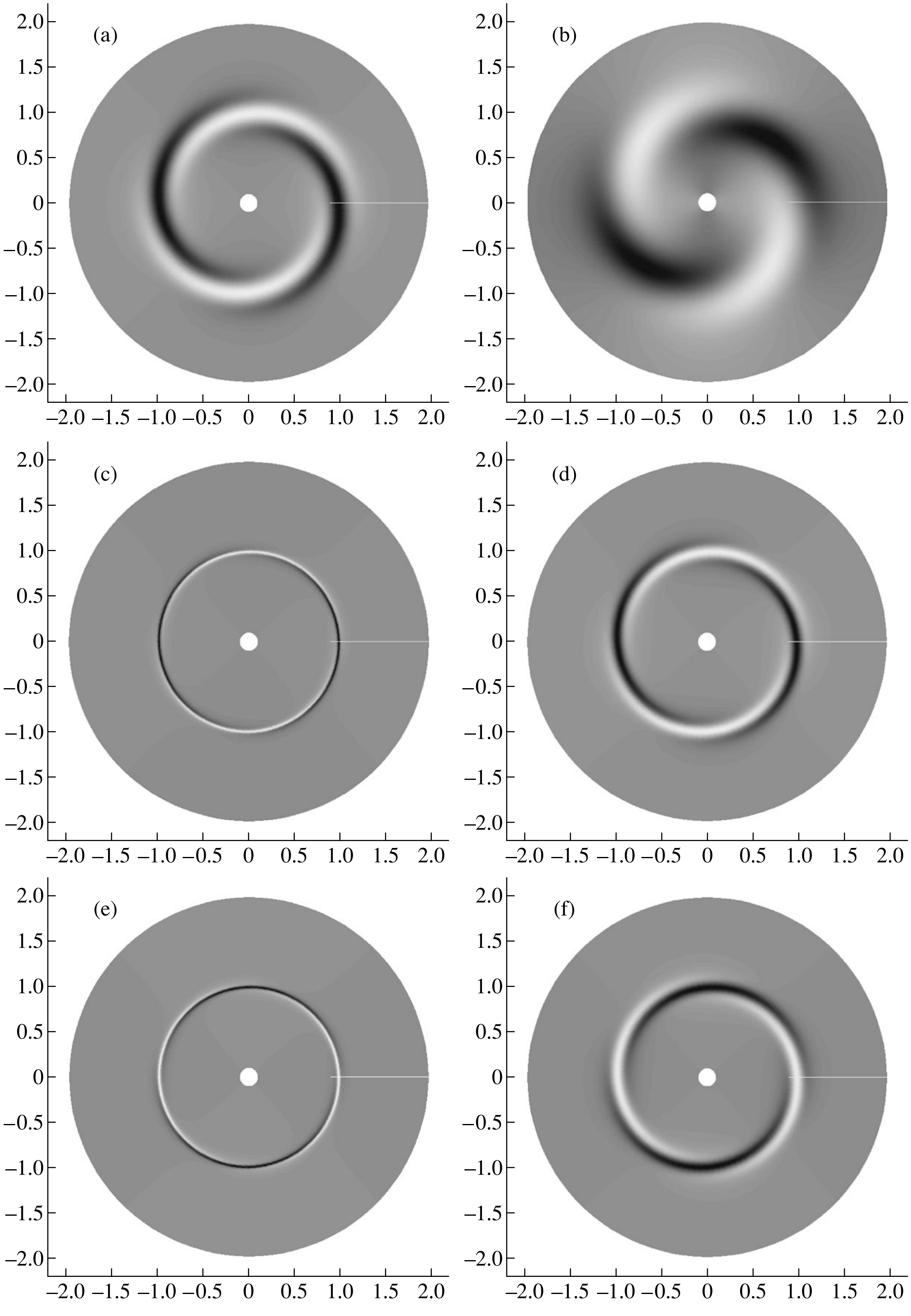


Fig. 5. Resonance responses of a disk at the external ILR ($r_c = 1$). The disk rotates counterclockwise. The bar is oriented vertically. The degree of blackening increases with perturbed surface density. (a) A universal spiral at $\rho = 0.1$, $n = 2$, $n < n_{cr}$, $n_{cr} = 2.82$; (b) a universal spiral at $\rho = 0.2$, $n = 2$, $n < n_{cr}$; (c) a cold-disk response at $\gamma = 0.05$, $n = 2$, $n < n_{cr}$; (d) a disk response at $\gamma = 0.05$, $\rho = 0.05$, $n = 2$, $n < n_{cr}$; (e) a cold-disk response at $\gamma = 0.05$, $n = 3$, $n > n_{cr}$; (f) a disk response at $\gamma = 0.05$, $\rho = 0.05$, $n = 3$, $n > n_{cr}$.

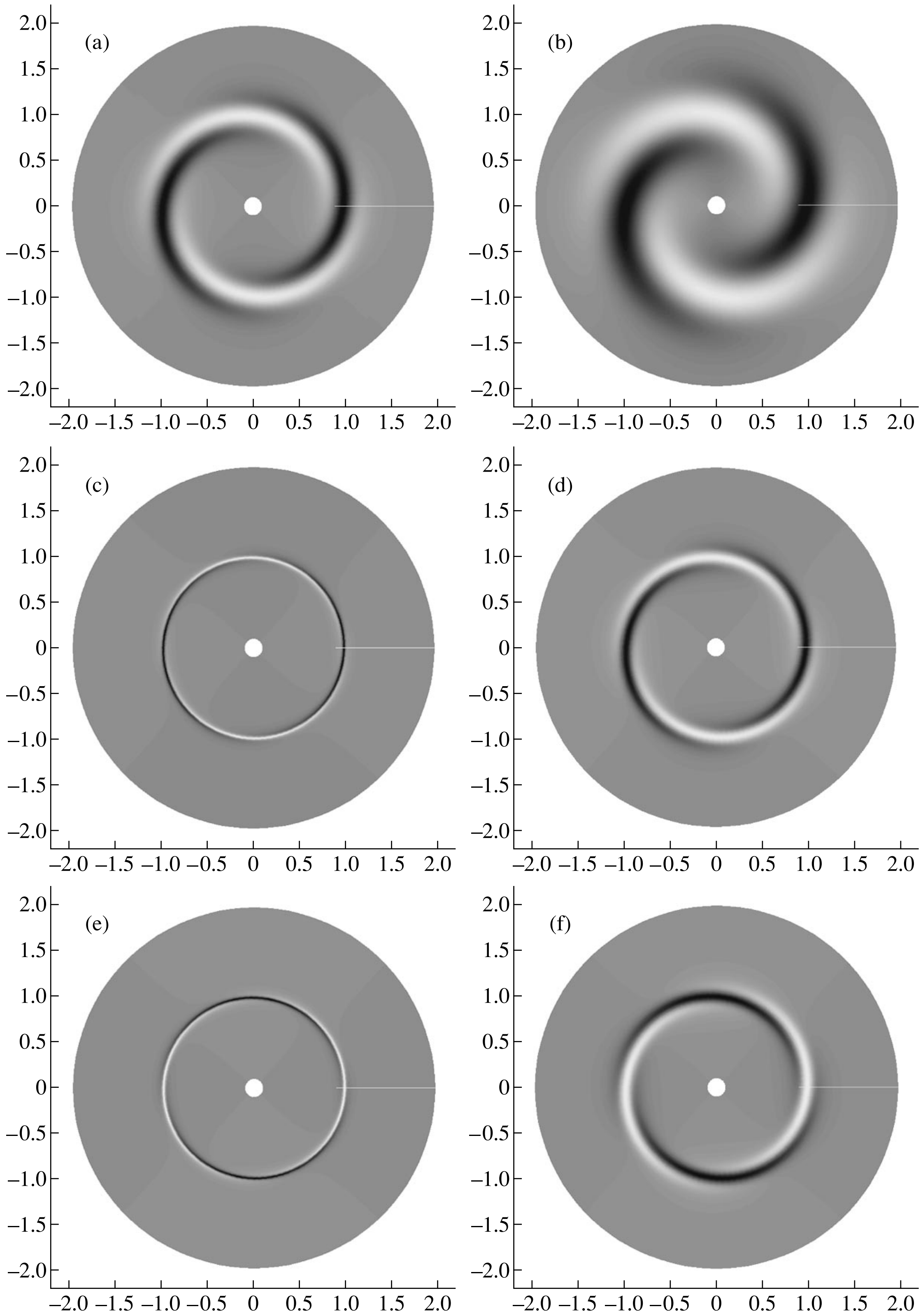


Fig. 6. Resonance responses of a disk at the internal ILR ($r_c = 1$). The disk rotates counterclockwise. The bar is oriented vertically. The degree of blackening increases with perturbed surface density. (a) A universal spiral at $\rho = 0.1$, $n = 3$, $n > n_{cr}$, $n_{cr} = 2.82$; (b) a universal spiral at $\rho = 0.2$, $n = 3$, $n > n_{cr}$; (c) a cold-disk response at $\gamma = 0.05$, $n = 3$, $n > n_{cr}$; (d) a disk response at $\gamma = 0.05$, $\rho = 0.05$, $n = 3$, $n > n_{cr}$; (e) a cold-disk response at $\gamma = 0.05$, $n = 2$, $n < n_{cr}$; (f) a disk response at $\gamma = 0.05$, $\rho = 0.05$, $n = 2$, $n < n_{cr}$.

nance responses have the form of trailing two-arm spirals shown in Figs. 5c and 5e (respectively, for positive and negative $n_{cr} - n$). The maximum length of each of the spiral arms is π. It is reached at small γ when the spirals themselves are nearly circular and the pair of spirals forms a ringlike configuration. At $r \approx r_{2i}$, the resonance responses (Figs. 6c and 6e) are similar to those just considered but have two important distinctions. The main distinction is that the spiral arms are now leading (because $\Omega'_{pr} > 0$ at $r = r_{2i}$; see Fig. 3). The second distinction (the reason is the same) is that, for $r \approx r_{2i}$, the spirals branch off from the bar ends at $n > n_{cr}$ (and lie on the bar ends at $n < n_{cr}$); recall that the bar is assumed to be oriented vertically. Note that the universal spirals for $n > n_{cr}$ and $n < n_{cr}$ also differ from each other [in Figs. 5a and 5b (6a and 6b), these universal responses are shown only for $n < n_{cr}$ ($n > n_{cr}$)].

The disk responses at various resonances for $\gamma = 0.05$ and $\rho = 0.05$ are shown in Figs. 4d, 5d, 5f, 6d, and 6f; they were calculated from the general formulas (89) and (90) for the equilibrium model described above. We see from a comparison of Figs. 4c and 4a, 4b, 4d for corotation (and similar figures for the other resonances) that the principal effect of a stellar velocity dispersion is an increase in the thickness of the spiral response and in the degree of its opening; the latter takes place at $\rho > \delta = \gamma/2|\Omega'|$ for CR, $\rho > \delta_1 = \gamma/2\left|\Omega'_{pr}\right|$ for ILR, and $\rho > \delta_2 = \gamma/2\left|\Omega^{+'}_{pr}\right|$ for OLR.

4. CONCLUSION

Below, we note some of the fundamental points related to our analysis of the resonance generation of galactic spiral arms.

(1) In general, the effect of a finite stellar velocity dispersion (even if it is assumed to be small compared to the circular velocity) on the galactic disk response is difficult to take into account. So far, this problem has been solved (in the analytical theory of Lin and Shu [16, 17] only for tightly wound spirals. This theory assumes the potential Φ_1 as well as all other perturbations (surface density σ_1, distribution function f_1, etc.) to be proportional to e^{ikr}, where k is the wave number, with

$$kr \gg 1, \quad d\ln k/d\ln r \ll 1.$$

Thus, strictly speaking, the theory properly describes the galactic disk responses only under the effect of multiturn, tightly wound spirals.

Recall that the relation between the potential and surface density perturbations that follows from the Poisson equation,

$$\Phi_1 = -\frac{2\pi G\sigma_1}{|k|}$$

(which is also local in our approximation), can be used to derive the standard Lin–Shu dispersion relation. It is one of the basic relations in the modern theory of galactic spiral patterns (see [6] for more details).

The Lin–Shu theory is definitely not applicable to our problem of the disk response to the bar potential: the bar potential can be described as the limiting case that is opposite to the tightly wound spirals considered by Lin and Shu. A general solution of the kinetic equation in our problem is difficult to find (if at all possible). However, as we saw above, we can easily determine how the resonance denominators $1/\omega_*$, $1/(\omega_*^2 - \kappa^2)$, and $1/(\omega_*^2 - \kappa^2)^2$ in the cold-disk response are modified under the effect of a finite (though small) velocity dispersion (separately for each of the principal resonances).

We would specially note the quasi-steady-state spiral responses derived in Subsection 3.2. If we measure the width Δr of the ring in which a spiral is concentrated in units of mean epicyclic size $\rho = c_0/\kappa$, then we obtain a certain number for such a width (a different number for each of the resonances: $\Delta r/\rho \approx 2.5$ for CR and $\Delta r/\rho \approx 3.7$ for ILR). This provides a simple method for estimating the stellar velocity dispersion near the bar ends. If, on the other hand, there is an independent estimate of the velocity dispersion for a galaxy, then we obtain a test to check the nature (resonant or nonresonant) of the spirals adjacent to the bar.

(2) Noteworthy is the fact that, as follows from our results, the amplitude of the disk response to the action from the bar is virtually independent of the stellar velocity dispersion. This is in conflict with the popular (and, at first glance, natural) belief that the response amplitude must always decrease with increasing velocity dispersion.

The apparent paradox can be explained as follows. As was shown in the Lin–Shu theory [16, 17], to obtain the response of a disk with a velocity dispersion c_r to the action from tightly wound spiral arms, the cold-disk response must be multiplied by the reduction factor $\mathcal{F}_\nu(k^2\rho^2)$, where $k = 2\pi/\lambda$ is the radial wave number (λ is the wavelength) and $\rho = c_r/\kappa$ is the epicycle size; $\mathcal{F}_\nu(0) = 1$, $\mathcal{F}_\nu(x) < 1$ for $x > 0$ and monotonically decreases with increasing x; these properties of the function $\mathcal{F}_\nu(x)$ justify its name by showing that the response amplitude must actually decrease with increasing velocity dispersion. The function $\mathcal{F}_\nu$ is appreciably smaller than 1 at $x \sim 1$, i.e., at $\rho \sim \lambda$. Since the wavelength λ in most normal spiral galaxies is much smaller than the disk radius R, the inequality $\rho > \lambda$ is easily achieved by stars with quite moderate velocity dispersions $c_r \ll v_0$ (where v_0 is the circular velocity). Therefore, the effect of response reduction with increasing velocity dispersion is easy to observe in this case, and this is the source of the above popular belief. However, for a bar, $k \longrightarrow 0$ formally and $k \sim 1/R$ actually, so the effect of a velocity dispersion must become noticeable only at $\rho \sim R$, i.e., $c_r \sim v_0$. Consequently, in our approximation of nearly circular orbits ($c_r \ll v_0$), the response cannot significantly depend on the velocity dispersion.

Moreover, since the width of the bar-generated spirals increases with velocity dispersion, we may say that, in this sense, the larger the velocity dispersion, the more pronounced the effect.

(3) Let us consider the disk responses near the internal ILR in more detail: they have the form of leading spirals (see Fig. 6). Meanwhile, the trailing spirals are natural and expected. For example, any (randomly generated) density perturbations in a differentially rotating disk with a radially decreasing angular velocity $\Omega(r)$ become such spirals: they are stretched by this differential rotation into segments of trailing spirals (see [6] for more details). In addition, all the hitherto observed spirals were trailing ones, at least in spiral galaxies with a regular spiral patterns. Therefore, the recent discovery [22] (see also [23]) of leading spirals in the central region of the galaxy NGC 157 (they were revealed by an analysis of its photometric measurements) came as a great surprise. These spirals may be the disk response at the internal ILR to a perturbation from the short bar of NGC 157. The authors of the discovery [22] offered a similar interpretation in which the observed leading spiral was considered as the response of a cold ($\rho = 0$) disk to a growing barlike perturbation with a small increment γ; this possibility was previously explored in [20, 21]. Thus, if the bar is inside the internal ILR, then the resonance spiral pattern begins with a leading spiral.

(4) Our results are also of interest in connection with the long-discussed problem of the so-called antispiral theorem that was first formulated by Lynden-Bell and Ostriker [24] for a gaseous disk; later, it was also proven for a stellar disk [25]. The theorem states that a spiral shape cannot exist as a neutral oscillation mode of a nondissipative gaseous or stellar disk.

This theorem is a corollary of certain symmetry of the equations of motion (for gas or stars) in a gravitational field $\Phi(r, \varphi)$. If we reverse the direction of time and simultaneously reflect all motions in the meridional plane, $\varphi \longrightarrow -\varphi$, then we arrive at a state where a perturbation with the opposite sense of spiral winding is superimposed on the same axisymmetric background (including the sense of rotation). Since the equations of motion are symmetric relative to this transformation, the oscillation frequencies in the original and transformed states must be equal, and we have no grounds to prefer one of them. In general, the corresponding solutions to the equations for normal modes must be nonspiral. However, if there are resonance stars in the system, the arguments for the antispiral theorem are inapplicable. This comes as no surprise, because the antispiral theorem is largely a reflection of the time reversibility of the equations of motion that is violated by resonance stars.

The structures we considered in SB galaxies, which include a bar and spiral arms, are the components of a single mode; under certain conditions, one of these components (spirals) may be assumed to be induced by the other (bar). This approach seems most natural for SB galaxies. In particular, one could hardly suspect the existence of running spiral waves in these galaxies. The corotation region can act as a generator, because there are anticyclonic (or, occasionally, cyclonic) vortices here (see [26] for the latter).

Because of the clear presence of spiral arms with a certain sense of winding (most frequently trailing ones, as in Fig. 1a), the question arises as to what causes the inconsistency with the antispiral theorem. The simplest cause could be instability of the mode under consideration: at a sufficiently large instability growth rate γ, the opening of the spiral arms can be significant even in a cold disk (while, at small γ, the spirals are close to arcs of circumferences, similar to the spirals in Figs. 4c–6c). However, the assumption of strong instability for galactic systems is not very reasonable: such instability must have long been saturated (the available time margin is more than enough). It would be more plausible to consider quasi-steady states when γ is small or even $\gamma \longrightarrow 0$. In this case, however, as follows from the antispiral theorem, spirals can appear only in the presence of resonances. This seems to have been the case in reality: the bar ends lie near one of the principal galactic (corotation or Lindblad) resonances. We saw that thermal stellar motion “spreads” these resonances over rings with widths of the order of several epicyclic radii. Spiral arms, usually trailing ones (Figs. 4 and 5) but, occasionally, leading ones (Fig. 6), are localized in these rings.

(5) In principle, resonance responses can completely cover the entire variety of the spiral and ring shapes observed in SB galaxies. Thus, a survey of photographs for SB galaxies from available atlases (for example, in the well-known Hubble atlas) shows that, in general, the trailing spiral arms that branch off from the bar are clearly traceable over approximately half of the turn around the center. This is similar to the resonance responses in Figs. 4 and 5. On the other hand, short nuclear bars, as we see from Fig. 6, can give rise to leading spirals (provided that the bar is within the internal ILR).

Note that the possible generation of spirals by a bar is not a new problem. However, so far, almost exclusively numerical (mostly N-body) methods have been used to solve it, because it seemed that the problem was too complex and could not be solved analytically. Unfortunately, our analytic results are difficult to compare with numerous N-body studies, because the latter are not systematic.

(6) In conclusion, we note several ways of improving the theory proposed above.

Its most evident drawback is that it disregards the self-gravitation of a spiral response, although it can occasionally be significant (for sufficiently massive galactic disks). Now, we see no possibility of introducing the self-gravitation of a response in terms of the analytical theory. The obvious idea of using the WKB approximation for this purpose (in an iterative scheme) proves to be untenable. The point is that, as may be shown, the potentials of a sufficiently short, near circular spiral segment (which the resonance response is) calculated from the formula of the WKB theory (i.e., $\Phi_1 = -2\pi G\sigma_1/|k|$) and (numerically) from an exact formula for the simple-layer potential differ severalfold (by a factor of 2 to 4). For this reason, this improvement apparently has to be made by numerical methods.

The finite galactic-disk thickness seems to be even more difficult to rigorously take into account if, of course, the use of N-body theories is not considered. As a first step, we could try various formfactors on the thickness, which are widely used in the stability theory for a gravitating disk (see [1, 2] for details).

ACKNOWLEDGMENTS

We thank the referee for important remarks, which improved the quality of this paper. This study was supported in part by the Russian Foundation for Basic Research (project nos. 99-02-18432 and 00-15-96528).

REFERENCES

1. V. L. Polyachenko and A. M. Fridman, *Equilibrium and Stability of Gravitating Systems* (Nauka, Moscow, 1976).
2. A. M. Fridman and V. L. Polyachenko, *Physics of Gravitating Systems* (Springer-Verlag, New York, 1984), Vols. 1, 2.
3. A. B. Mikhaĭlovskiĭ, *Plasma Instability in Magnetic Traps* (Atomizdat, Moscow, 1978).
4. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 1: *Mechanics* (Nauka, Moscow, 1982; Pergamon, New York, 1988).
5. N. M. Gor'kavyĭ and A. M. Fridman, *Physics of Planetary Rings* (Nauka, Moscow, 1994).
6. V. L. Polyachenko and E. V. Polyachenko, Zh. Éksp. Teor. Fiz. **112**, 771 (1997) [JETP **85**, 417 (1997)].
7. L. D. Landau, Zh. Éksp. Teor. Fiz. **7**, 203 (1937).
8. V. L. Polyachenko and I. G. Shukhman, Astron. Zh. **59**, 228 (1982) [Sov. Astron. **26**, 140 (1982)].
9. F. Combes and B. G. Elmegreen, Astron. Astrophys. **271**, 391 (1993).
10. V. L. Polyachenko, in *Barred Galaxies*, Ed. by R. Buta, D. A. Crocker, and B. G. Elmegreen, Astron. Soc. Pac. Conf. Ser. **91**, 405 (1996).
11. V. L. Polyachenko and I. G. Shukhman, Astron. Zh. **56**, 957 (1979) [Sov. Astron. **23**, 539 (1979)].
12. V. L. Polyachenko, Zh. Éksp. Teor. Fiz. **101**, 1409 (1992) [Sov. Phys. JETP **74**, 755 (1992)].
13. A. B. Mikhaĭlovskiĭ, A. M. Fridman, and Ya. G. Épel'baum, Zh. Éksp. Teor. Fiz. **59**, 1608 (1970) [Sov. Phys. JETP **32**, 878 (1971)].
14. F. H. Shu, Astrophys. J. **160**, 99 (1970).
15. A. Toomre, Astrophys. J. **139**, 1217 (1964).
16. C. C. Lin and F. H. Shu, Proc. Natl. Acad. Sci. USA **55**, 229 (1966).
17. C. C. Lin, C. Yuan, and F. H. Shu, Astrophys. J. **155**, 721 (1969).
18. A. B. Mikhailovskii, *Theory of Plasma Instabilities* (Atomizdat, Moscow, 1970; Consultants Bureau, New York, 1974), Vol. 1.
19. G. S. Bisnovatyĭ-Kogan and A. B. Mikhaĭlovskiĭ, Astron. Zh. **50**, 312 (1973) [Sov. Astron. **17**, 205 (1973)].
20. I. I. Pasha and V. L. Polyachenko, Mon. Not. R. Astron. Soc. **266**, 92 (1994).
21. V. L. Polyachenko, in *Physics of Gaseous and Stellar Disks of the Galaxy*, Ed. by I. R. King; Astron. Soc. Pac. Conf. Ser. **66**, 103 (1994).
22. A. M. Fridman and O. V. Khoruzhii, Phys. Lett. A **276**, 199 (2000).
23. A. M. Fridman, O. V. Khoruzhii, V. A. Minin, *et al.*, Astron. Soc. Pac. Conf. Ser. **230**, 187 (2001).
24. D. Lynden-Bell and J. P. Ostriker, Mon. Not. R. Astron. Soc. **136**, 293 (1967).
25. F. H. Shu, Astrophys. J. **160**, 89 (1970).
26. A. M. Fridman, O. V. Khoruzhii, E. V. Polyachenko, *et al.*, Phys. Lett. A **264**, 85 (1999).

Translated by V. Astakhov

Journal of Experimental and Theoretical Physics, Vol. 93, No. 6, 2001, pp. 1155–1162.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 6, 2001, pp. 1336–1345.
Original Russian Text Copyright © 2001 by Volkova, Popov, Tikhonova.

Ionization and Stabilization of a Three-Dimensional System with a Short-Range Potential in a Strong Laser Field

E. A. Volkova, A. M. Popov*, and O. V. Tikhonova

Skobel'tsyn Institute of Nuclear Physics, Moscow State University, Vorob'evy gory, Moscow, 119899 Russia

**e-mail: popov@mics.msu.su*

Received May 14, 2001

Abstract—The interaction of a three-dimensional atomic system in a short-range potential with intense laser radiation is investigated by the direct numerical integration of the nonstationary Schrödinger equation. The calculations helped to discover a stabilization regime, which is interpreted as a result of forming a Kramers–Henneberger atom "dressed" in a field. Dynamics of the energy spectrum of photoelectrons depending on the increase of the laser field intensity is investigated, and conditions of a photodetachment of an electron from a bound state of the Kramers–Henneberger potential are analyzed. These results reveal specific features of the stabilization process of the three-dimensional system with a short-range potential compared to the similar process in a system with a long-range (Coulomb) potential. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Substantial restructuring of the spectrum of the initial atomic system and the appearance of a qualitatively new object—atom "dressed" in a field—is one of the most interesting phenomena in the theory of interaction of atomic and molecular systems with a strong laser field. A variety of analytical and numerical approaches to the investigation of this object properties have been proposed; the greater part of these approaches is based on the analysis of the phenomenon in terms of quasi-energy and quasi-energy wave functions [1–4] and on the concept of "dressed atom" states [5, 6]. One property inherent in quantum systems interacting with intense laser radiation is stabilization with respect to ionization. The stabilization regime manifests itself in that the ionization rate of the system, or the probability of ionization per pulse, does not increase or even decreases (beginning with a certain threshold value) as the laser radiation intensity grows. At present, two stabilization models (mechanisms) are most often discussed in the literature. These are the interference stabilization of Rydberg atoms [7] and adiabatic stabilization (by Kramers–Henneberger) [8, 9]. The latter is based on the transformation of an atom in a strong field into a Kramers–Henneberger (KH) atom whose stability with respect to the ionization increases with laser intensity. Stabilization of the initial atomic system is directly related to the stabilization of the KH atom and is caused by the adiabatic "flow" of a part of the population of unperturbed atomic states into the corresponding KH states and backwards on the fronts of the laser pulse.

For one-dimensional systems, the existence of the KH potential was proved in computer experiments, and specific features of its energy spectrum are well studied in a wide range of intensities and frequencies of radiation. In particular, it was shown in [10] that, under single photon ionization in the case when the quantum energy $\hbar\omega$ exceeds the binding energy of the unperturbed atomic level E_b, it is physically justified to investigate the system dynamics in a strong field in terms of the KH atom and its eigenstates, especially in the case of a large difference in the ionization potentials of the KH and unperturbed atoms.

However, in the case of three-dimensional systems with a short-range potential, issues of choosing one or another basis and the possibility of observing the KH stabilization regime remain open, since recent studies cast doubt on the very existence of stabilization. For example, it is demonstrated in [11] that in the case of strong fields the number of bound KH states not only does not increase with laser intensity but can even decrease, which leads to the absence of bound states in the KH potential; as a result no KH stabilization can occur. However, in our opinion, the rate of the appearance (or disappearance) of bound states in the KH potential depends on the atomic potential parameters. Hence, the assumption that the number of bound KH states in the three-dimensional case increases much slower than in the one-dimensional case seems more general.

On the other hand, the ionization of a three-dimensional system with a δ potential was investigated analytically in [12]. It was shown that nonmonotonic dependence of the ionization rate on laser intensity is explained by the "channel closing " due to the Stark shift in the energy of the initial level and the continuum boundary. No KH stabilization regime was discovered. However, the applicability limits of this method in a wide range of frequencies and laser radiation intensities remain not completely understood. In addition, we note that the δ potential is characterized by a zero radius of action; thus, it presents a special case among all short-range potentials, since the threshold intensity for the above-barrier ionization is infinitely large:

$$P_{BSI} \longrightarrow \infty.$$

The purpose of this paper is to investigate the interaction of a three-dimensional atomic system characterized by a short-range potential with intense laser radiation by direct numerical integration of the nonstationary Schrödinger equation. On the basis of our computations, we derive a dependence of the ionization probability on laser intensity for various quantum energies and analyze the appearance of the stabilization regime. In addition, we analyze the dependence of photoelectron energy on laser radiation intensity. The resulting data make it possible to observe the change of the energy spectrum of the dressed atom with increasing field and determine conditions under which the ionization occurs from bound states of the KH potential. Formation of the Kramers–Henneberger atom results in violation of the initial spherical symmetry of the original potential, which manifests itself in that, in a strong field, photoelectrons escape chiefly in the direction perpendicular to that of the intensity of the electric field.

2. NUMERICAL MODEL

We numerically investigate the dynamics of ionization of the model three-dimensional system with a short-range centrally symmetric potential

$$V(r) = -\frac{1}{\sqrt{a^2 + r^2}}\exp(-\lambda r), \tag{1}$$

where $a = 0.5a_0$ and $\lambda = 0.6$ Å^{-1}. The characteristic width of this potential is $a_{at} \approx 1.5$ Å ≈ 3 a.u.; it has a unique bound stationary s state with the energy $E_b \approx -3.73$ eV and the wave function $\varphi_b(r)$. Since the magnetic quantum number is preserved in a linearly polarized field, the system's wave function in our case depends only on two space coordinates. In the cylindrical reference frame (the z axis is directed along the electric field vector), it satisfies the equation

$$i\frac{\partial\psi}{\partial t} = -\frac{1}{2}\left(\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial\psi}{\partial\rho}\right) + \frac{\partial^2\psi}{\partial z^2}\right) + V(\rho, z)\psi(\rho, z, t) + z\varepsilon(t)\cos(\omega t)\psi(\rho, z, t), \tag{2}$$

where $\varepsilon(t)$ is the envelope of the laser pulse and ω is the radiation frequency.

Calculations of the ionization dynamics were performed for radiation with the quantum energy $\hbar\omega = 4$–10 eV in a wide range of intensities for every value of frequency. The shape of the pulse envelope was assumed to be trapezoidal with smoothed fronts of duration $\tau_f = 5T_\omega$ and the plateau $\tau_{pl} = 10T_\omega$ (T_ω is the duration of the optical cycle):

$$\varepsilon(t) = \begin{cases} \varepsilon_0 \sin^2\dfrac{\pi t}{2\tau_f}, & t \le \tau_f, \\ \varepsilon_0, & \tau_f \le t \le \tau_f + \tau_{pl}, \\ \varepsilon_0 \sin^2\dfrac{\pi[t - (2\tau_f + \tau_{pl})]}{2\tau_f}, & \tau_f + \tau_{pl} \le t \le 2\tau_f + \tau_{pl}. \end{cases} \tag{3}$$

The nonstationary Schrödinger equation was integrated on a nonuniform two-dimensional mesh in the reference frame (ρ, z) with 960 × 480 nodes.

The wave function determined from Eq. (2) makes it possible to calculate the population of the initial state W_b at any moment in time

$$W_b = |C_1(t)|^2, \quad C_1(t) = \int\psi(\rho, z, t)\varphi_b(\rho, z)2\pi\rho d\rho dz, \tag{4}$$

the probability of the system's ionization

$$W_i = 1 - W_b,$$

and the wave function

$$\psi_c(\rho, z, t) = \psi(\rho, z, t) - C_1(t)\varphi_b(\rho, z)\exp(-iE_b t) \tag{5}$$

that describes the packet in continuum. The function $\psi_c(\rho, z, t)$ was used to obtain the pulse distribution $W(\mathbf{k})$ for the electron in continuum:

$$W(\mathbf{k}) = |\tilde{\psi}_c(\mathbf{k})|^2, \tag{6}$$

where

$$\tilde{\psi}_c(\mathbf{k}) = \int\psi_c(\mathbf{r})\exp(-i\mathbf{k}\cdot\mathbf{r})d^3r. \tag{7}$$

For a concrete value of the electron's ejection angle with respect to the vector of the electric field polarization $\theta_{\varepsilon k}$, we obtain from (7)

$$\tilde{\psi}_c(k, \theta_{\varepsilon k}) = 2\pi\int\psi_c(\rho, z)J_0(k\rho\sin\theta_{\varepsilon k}) \times \exp(-ikz\cos\theta_{\varepsilon k})\rho d\rho dz. \tag{8}$$

In particular, for the angles $\theta_{\varepsilon k} = 0$ and $\theta_{\varepsilon k} = \pi/2$, we have

$$\tilde{\psi}_c(k, \theta_{\varepsilon k} = 0) = 2\pi\int\psi_c(\rho, z)\exp(-ikz)\rho d\rho dz, \tag{9}$$

$$\tilde{\psi}_c(k, \theta_{\varepsilon k} = \pi/2) = 2\pi\int\psi_c(\rho, z)J_0(k\rho)\rho d\rho dz. \tag{10}$$

Since, after the termination of the laser pulse, the wave packet in continuum is localized outside the potential's range, the pulse distribution of electrons can be easily converted to the energy distribution

$$W(E, \theta_{\varepsilon k})dE = W(k, \theta_{\varepsilon k})k^2 dk. \tag{11}$$

3. RESULTS AND DISCUSSION

3.1. Large Energy of the Field Quantum

Since the KH stabilization was traditionally investigated for high-frequency laser radiation, the case when the quantum energy substantially exceeds the initial state binding energy is most important. Figure 1 shows the remaining probability for the system to be in a bound state as a function of the laser radiation for $\hbar\omega$ = 10 eV; it was calculated by formula (4) at the moment of the laser pulse termination. These data confirm the appearance of the stabilization regime in the range of intensities $(2–30) \times 10^{15}$ W/cm^2; moreover, the maximal value of the residual population in the stabilization regime is about 10%. An analysis of the first peak location in the photoelectron spectrum testifies that it shifts monotonically to the domain of large energies as the field intensity increases (Fig. 2). This means that in the presence of the field the initial level shifts upwards in terms of energy, and this shift is greater than the shift of the continuum boundary. Thus, as the field amplitude increases, the electron escapes from a weaker bound state, and this state is one of the KH atom states, which is characteristic of the atom in a field. (Note that, for all intensities under consideration, there exists a unique bound state in the KH potential characterized by a positive parity that can be populated in the process of pulse switching.)

To prove the fact that a KH atom appears in the field, we calculated the photoelectron's energy, assuming that it is ionized from the KH potential, for various values of the laser intensity. The results are also presented in Fig. 2 (the solid curve); they are in good agreement (within the accuracy of determining the peak locations) with the electron energies found from the energy spectra. Thus, we conclude that it is preferable to analyze the dynamics of atomic systems in a wave field in terms of eigenstates of the KH potential, the atom in a field is a KH atom, and the stabilization discovered is a KH stabilization.

Figure 3 presents the time dependence of the squared absolute value of the projection of the desired wave function on the bound KH state corresponding to the laser pulse intensity on the "plateau" for various intensity values. This magnitude characterizes the part of population that is bound in the KH potential. In weak fields, the population decays on the plateau. As the intensity increases up to 2×10^{15} W/cm^2, the rate of decay increases and then decreases or the KH state is even additionally populated at the pulse plateau. The stabilization region on the curve $W_b(P)$ corresponds to the slowdown of the population decay rate or to additional population of the KH state. This is because a considerable part of the population of KH states on the trailing front of the pulse is "redistributed" to the population of the initial atomic state. Note that the stabilization threshold $P_{th} \approx 2 \times 10^{15}$ W/cm^2 corresponds to the situation when the characteristic value of the KH

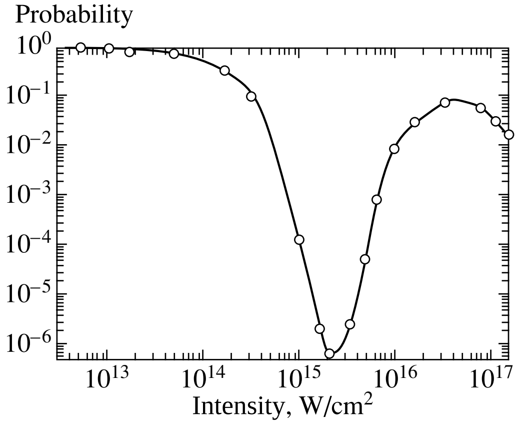


Fig. 1. Probability of the nonionization of the system upon the laser pulse termination as a function of laser intensity for the quantum energy $\hbar\omega$ = 10 eV.

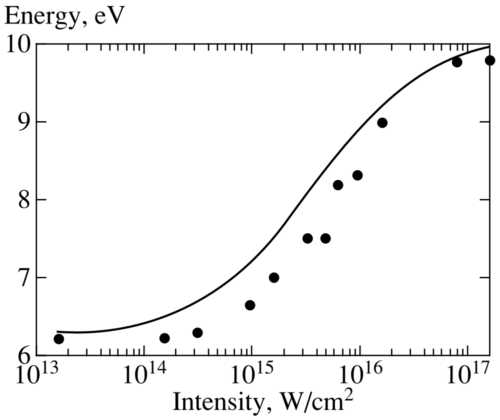


Fig. 2. Dependence of the first peak energy in the photoelectron spectrum on laser intensity obtained by calculation (dots) and under the assumption that the electron photoescapes from the KH potential (solid curve).

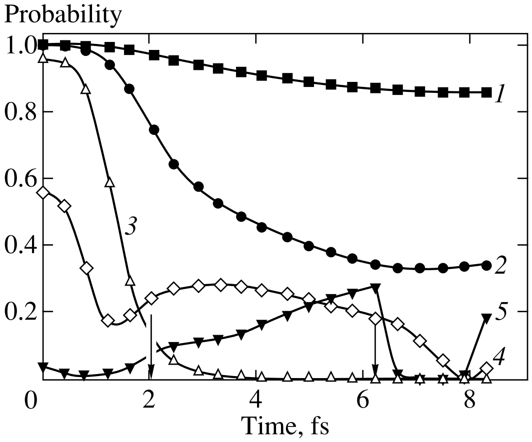


Fig. 3. Time dependence of the squared absolute value of the projection of the wave function $\psi(\rho, z, t)$ on the bound KH state corresponding to the laser pulse intensity on the "plateau" for the following laser intensity values (in W/cm^2): (*1*) 1.6×10^{13}; (*2*) 1.6×10^{14}; (*3*) 1.6×10^{15}; (*4*) 1.6×10^{16}; (*5*) 9.6×10^{16}. Arrows point to the beginning and end of the generation "plateau."

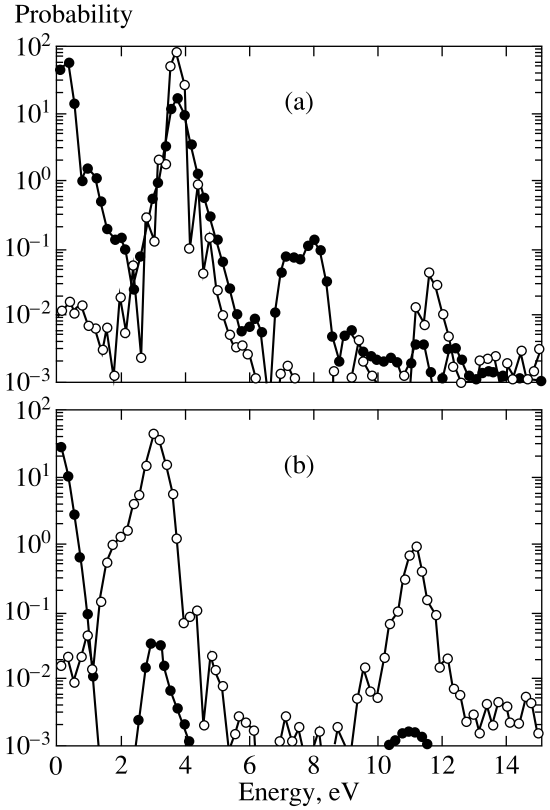


Fig. 4. The energy spectrum of photoelectrons ejected along (filled circles) and perpendicular (empty circles) to the polarization direction of the electric field of the wave at $\hbar\omega$ = 4 eV and for different values of $P = 3 \times 10^{13}$ (a) and 10^{14} (b) W/cm^2.

potential $a_e = \varepsilon_0/\omega^2$ is about the characteristic value of the atomic potential $a_{at} \approx 3$ a.u.

The maximal value of the residual probability in the stabilization regime is about 10%, and it corresponds to $a_e \approx 7$ a.u. Such a small probability for the system to remain in the bound state is explained by the existence of a single bound KH state (up to $a_e \approx 12$ a.u.) in which the population on the pulse plateau can be confined. This situation is drastically different from the case of one-dimensional systems for which the number of bound KH states is considerably greater than unity at $a_e \gg a_{at}$.

On the basis of all available data, one can conclude that, for high radiation frequency, the atomic system interacting with laser radiation can be considered as a KH atom, and the observed stabilization is easily explained in terms of the KH potential and its stationary states.

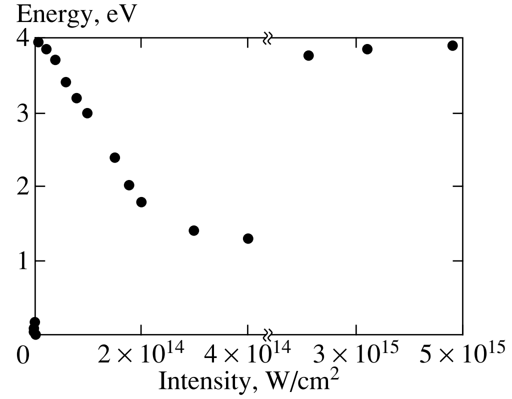


Fig. 5. Dependence of the first peak energy in the photoelectron spectrum on laser intensity for $\hbar\omega$ = 4 eV.

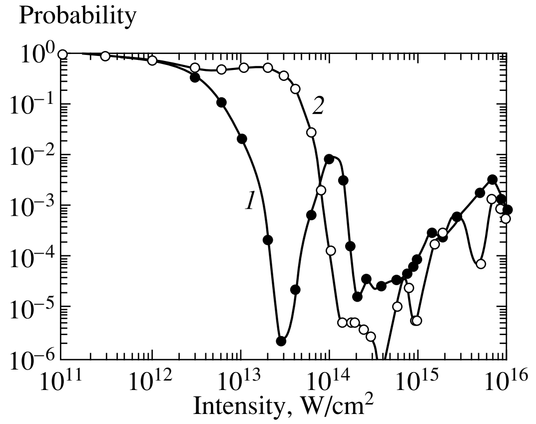


Fig. 6. Probability of the system nonionization after the laser pulse termination as a function of laser intensity for the quantum energy $\hbar\omega$ = 5 (*1*) and 4 eV (*2*).

3.2. Quantum Energies Comparable with the Initial State Binding Energy

The case of frequencies that cause ionization in the vicinity of the threshold is most difficult for both analytical and numerical analysis. Specific features of the ionization of the system under such quantum energy values first of all manifest themselves in the dynamics of photoelectron spectra depending on the intensity of laser radiation; this dynamics is characterized by the following properties. Firstly, in the domain of weak and moderate fields, the first peak in the photoelectron spectrum moves in the direction of smaller energies with increasing laser intensity; thus, at a certain magnitude of the intensity P^* ($P^* \approx 10^{14}$ W/cm^2 for $\hbar\omega$ = 5 eV and $P^* \approx 3 \times 10^{13}$ W/cm^2 for $\hbar\omega$ = 4 eV), the “channel closes” and the minimal number of quanta required for ionization becomes equal to two. A typical photoelectron spectrum is shown in Fig. 4a. The dependence of the first peak location in the photoelectron spectrum on laser intensity is shown in Fig. 5. The “channel closing” manifests itself in a sharp peak on the dependence of the nonionization probability on intensity in the vicinity of P^* (Fig. 6). This qualitatively agrees with the data obtained in [12] about the decrease in the ionization rate with the change of the multiphoton order of ionization. Note that the lower the frequency, the more often

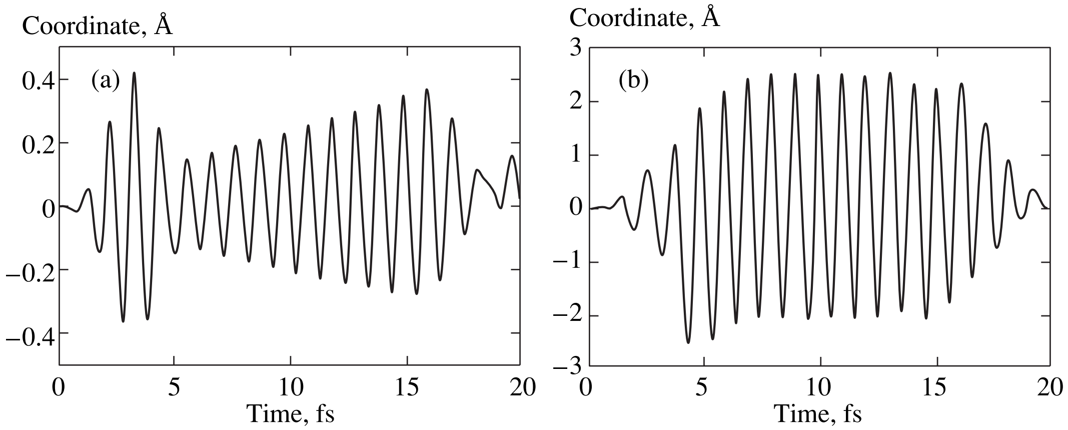


Fig. 7. Dependence of the mean electron coordinate (over the quantum state) on time under laser action for $\hbar\omega = 4$ eV. The intensity of radiation is $P = 3 \times 10^{13}$ (a) and 3×10^{14} (b) W/cm^2.

the multiphoton order can change; however, the first channel closing is primarily caused by the Stark shift of the initial atomic level, while the shift of the continuum boundary can be neglected. Another specific feature of the ionization under the condition of channel closing is the change in the angular distribution of the probability of the photoelectron ejection. Since the ionization is a two-photon one due to Stark's shift of the initial level, the final states in continuum can have the values of the orbital moment $L = 0$ and $L = 2$. As a result, the probability for an electron to escape perpendicularly to the field increases (Fig. 4b). The occurrence of the perpendicular ejection of electrons was also demonstrated in [13] for the case when the number of absorbed photons is greater than two. Moreover, for the initial 1*s* state, the transverse ejection will manifest itself for even above-threshold peaks.

Note that chiefly transverse ejection of electrons in sufficiently strong fields was also observed for the frequency $\hbar\omega = 10$ eV; this appears to be caused by the increase in the probability of two-photon ionization compared to the single-photon one as the radiation intensity increases.

In strong fields, in addition to peaks characterizing photodetachment of electrons from the original atom (with regard for Stark's shift of the atomic level and the continuum boundary), a new structure of peaks appears in the photoelectron spectrum. The energy of the first peak in this case is near $\hbar\omega$, which can be interpreted as the photodetachment of an electron from the KH potential, which contains a single weakly bound level, for almost all intensities in the range under consideration (see the right-hand part of the plot in Fig. 5). In addition, the new system of peaks is shifted upwards in terms of energy as the laser intensity increases.

Thus, the energy spectra confirm the existence of two fundamentally different systems describing an atom in the presence of an external field. This is a weakly perturbed atom characterized, in weak fields, by Stark's shift and channel closing, and an atom restructured by a strong field, which is in essence a KH atom. In addition, the transformation of a free atom in a field into a KH atom occurs in fields that correspond to the above-barrier ionization of the system ε_{BSI}. Similar conclusions were obtained as a result of the analysis of one-dimensional system ionization with a short-range potential by a low-frequency field [14]. More precisely, a KH atom characterizing a negative ion in a strong field appears only in fields with $\varepsilon > \varepsilon_{BSI}$.

It must be noted that in determining ε_{BSI} one must take into account the shift of the original atomic level with respect to the continuum boundary in the presence of a laser field. For this reason, the intensity of the above-barrier ionization of the system can be different for different laser radiation frequencies. Since the standard method for determining ε_{BSI} (see [15]) can be complicated in the three-dimensional case, one can use in numerical calculations the coincidence criterion of the mean quantum coordinate of the electron with the classical coordinate of the free electron in the electromagnetic wave field. We estimated ε_{BSI} with regard for the shift of the original level in the field on the basis of an analysis of the time dynamics of the mean quantum coordinate for various intensities. In the high-frequency limit, the magnitude P_{BSI} is approximately equal to 3×10^{13} W/cm^2, while in the case of the near-threshold ionization ($\hbar\omega = 4$ eV) the intensity of the field that suppresses the barrier corresponds to much greater intensities and turned out to be $P_{BSI} \approx 3 \times 10^{14}$ W/cm^2. Hence, for $\hbar\omega = 10$ eV, the ionization is an above-barrier one almost in the entire range of intensities, and the dressed atom is essentially a KH atom. A substantial increase in P_{BSI} in the case of $\hbar\omega \geq I_i$ (here $I_i = -E_b$) is caused by a considerable shift of the original level downwards with respect to the continuum boundary, in contrast to the high-frequency regime.

Figure 7 illustrates the time dynamics of the mean electron coordinate (over the quantum state) for two values of laser intensity. It is seen that, for weak fields

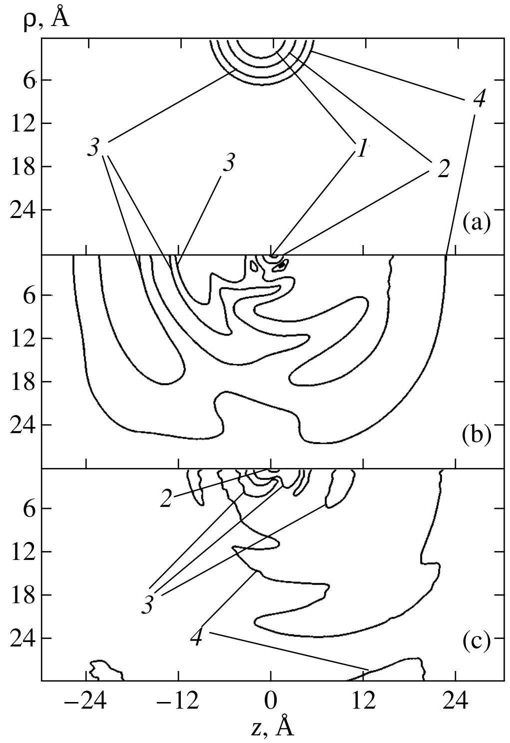


Fig. 8. Distribution of the electron density at different moments in time measured from the beginning of a rectangular laser pulse in the case $\hbar\omega = 5$ eV and $P = 2 \times 10^{14}$ W/cm^2. At the initial moment in time, the system is in the stationary state of the KH potential. The time moments are $t = 0$ (a), $3T_\omega$ (b), and $5T_\omega$ (c), where T_ω is the duration of the optical cycle. The level lines correspond to the values 0.01 (*1*), 10^{-3} (*2*), 10^{-4} (*3*), and 10^{-5} (*4*).

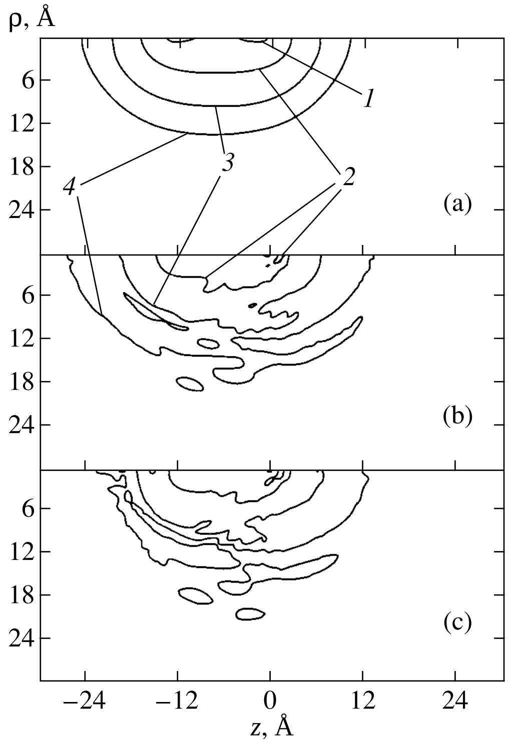


Fig. 9. The same as in Fig. 8. The intensity of radiation is $P = 7 \times 10^{15}$ W/cm^2. The time moments are $t = 0$ (a), $5T_\omega$ (b), and $10T_\omega$ (c).

(Fig. 7a), the electron continues to be influenced by the potential, and its motion is drastically different from that of the classical free electron in the wave field. The electron behaves almost freely in the wave field when the intensity corresponds to the above-barrier ionization of the system (Fig. 7b).

Thus, it was discovered that the KH ionization regime occurs only in fields for which $P > P_{BSI}$, and the field magnitude that "suppresses" the barrier is significantly different in the ranges $\hbar\omega \gg 1$ and $\hbar\omega \geq 1$ due to different direction of the original atomic level shifts in the field wave in these limiting cases. For this reason, for $\hbar\omega \geq I_i$, the atom behaves like a KH atom only in sufficiently strong fields.

We also note that the appearance of a new structure in the spectrum, which corresponds to the ionization from the KH potential, is observed only for intensities for which the dependence W_b has a stabilization regime (Fig. 6). Although the observed stabilization is characterized by a small value of the residual probability of nonionization, it is caused by the formation of a KH atom and by the increase in stability of KH states with the field intensity. In order to prove this fact, we made special calculations with the initial state in the form of a bound KH state.

Figures 8 and 9 illustrate the evolution of the system state from a stationary state in the KH potential in the case of a rectangular laser pulse for various laser intensities. If the field is not very strong ($P = 2 \times 10^{14}$ W/cm^2), the state rapidly disintegrates, and only the part of the wave packet that corresponds to the population of the initial atomic state remains localized. These data show that the evolution of the "atom + field" system must be characterized in terms of the states of the unperturbed Hamiltonian, and the KH potential and KH states are unrelated to physical reality.

The situation in strong fields is quite different (see Fig. 9). The data show that the system's state remains close to its initial state. This fact testifies that the dressed atom is a KH atom, and KH states are preferable for describing the system dynamics.

Similar to the high-frequency case, the stabilization threshold P_{th} corresponds to the values $a_e \sim a_{at}$ ($a_e \approx 4$ a.u. for $\hbar\omega = 4$ eV and $a_e \approx 2.5$ a.u. for $\hbar\omega = 5$ eV), and the stabilization itself is associated with improved stability of the bound KH states for $P \geq P_{th}$.

A similar conclusion was made in the analysis of a one-dimensional system ionization [10, 14]; however, in the three-dimensional case, the residual probability for the system to be in a bound state is rather small in the stabilization regime and does not exceed 1% for $\hbar\omega = 4$ and 5 eV. This fact is caused by a very slow increase in the number of bound states in the KH potential (with increasing intensity) in the three-dimensional case compared with a similar one-dimensional system. For example, at $a_e = 18$ a.u., only the second even KH state manifests itself in our case, whereas the number of even bound KH states for one-dimensional systems in a similar case is about 10.[1] This results in a substantially smaller value of the residual population in three-dimensional systems in the stabilization regime.

Breakdown of stabilization caused by the disappearance of bound states in the three-dimensional KH potential was studied in [11]. In our opinion, the rate of increase (or disappearance) in the number of bound KH states depends on parameters of the short-range potential.

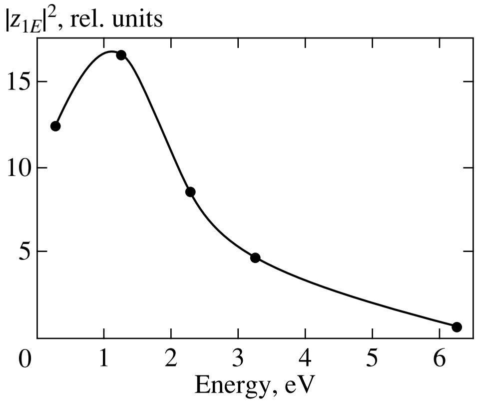


Fig. 10. Squared absolute value of the dipole matrix element $|z_{1E}|^2$ as a function of the final state energy in continuum.

3.3. System Ionization Rate and the Dipole Matrix Element of Transition to Continuum Depending on the Quantum Energy

The results presented in the previous sections show that the system dynamics is different in the two limiting cases $\hbar\omega \gg I_i$ and $\hbar\omega \geq I_i$. Although in both cases we have one-photon ionization, characteristic features of the KH atom for near-threshold radiation frequencies manifest themselves only in sufficiently strong fields due to high threshold intensity of the above-barrier ionization, which is caused by a considerable downward shift of the initial level with respect to the continuum boundary.

The difference of the shift directions of the peak energy in the spectrum of photoelectrons in not very strong fields for low and high frequencies can be qualitatively explained within the framework of perturbation theory. The energy shift of the original level in the wave field is determined as

$$\Delta E = -\frac{1}{4}\alpha\varepsilon_0^2, \tag{12}$$

and the polarizability α of the system with a single bound state accurate to the second order of perturbation theory is written as

$$\alpha = \int_0^\infty |z_{1E}|^2 \frac{2(E-E_b)}{(E-E_b)^2-(\hbar\omega)^2} dE, \tag{13}$$

where z_{1E} is the dipole matrix element of transition to continuum with account for the angular part and E_b is the energy of the ground state. The sign of polarizability determines the direction of the energy shift in the wave field; this sign can be different only due to a nonmonotonic dependence of the matrix element z_{1E} on the electron energy in continuum. Indeed, if $d\alpha/dE$ is positive in the neighborhood of the pole $E = E_b + \hbar\omega$, then the domain for which $\alpha > 0$ makes a greater contribution to integral (13); hence, the Stark's level shift is negative. In the converse case $d\alpha/dE < 0$, we have $\alpha < 0$, and the level energy will increase. In order to determine $|z_{1E}|^2$, we calculated the dependence of the probability of the initial level population on time in the range of intensities corresponding to the domain where perturbation theory is applicable. We found that this dependence is exponential on the generation plateau:

$$W_b \propto \exp(-\Gamma_i t). \tag{14}$$

Hence, the parameter Γ_i determines the ionization width of the state.[2] In relatively weak fields, the dependence of the ionization width on radiation intensity $\Gamma_i(P)$ proved to be linear. Using the relationship of the ionization width with the magnitude of the matrix element $|z_{1E}|^2$

$$\Gamma_i = \frac{\pi}{2}|z_{1E}|^2 P, \tag{15}$$

obtained by perturbation theory, we determined $|z_{1E}|^2$ for various radiation frequencies. The resulting dependence $|z_{1E}|^2$ on electron energy in continuum is plotted in Fig. 10; indeed, it shows an increase in $|z_{1E}|^2$ with energy in the near-threshold domain. Such a dependence of the matrix element is characteristic of negative ions whose one-electron model is described by a short-range potential [16].

Thus, the nonmonotonic dependence of the dipole matrix element on electron energy in continuum explains the specific features of the dynamics of the initial atomic state in the wave field in the near-threshold

---

[1] Note that odd KH states are of no importance for the stabilization regime, since they remain almost unpopulated when the laser pulse is on.

[2] Note that in the range of stronger fields, when perturbation theory is inapplicable, the function $W_b(t)$ is not exponential, and the width Γ_i cannot be introduced.

domain. These specific features are the formation of a KH atom and the appearance of the stabilization regime only in the domain of high radiation intensity.

4. CONCLUSION

Our calculations based on the direct numerical integration of the nonstationary Schrödinger equation for a system with a short-range potential in a field of strong laser radiation allow us to draw some general conclusions.

For three-dimensional systems, as well as for one-dimensional ones, there exists a range of laser radiation parameters for which the stabilization regime is realized. The cause of stabilization is the formation of a KH atom, which characterizes a dressed atom, and in the capture of the atomic population into bound KH states, which become more stable with increasing field. A substantial difference in the dynamics of three-dimensional and one-dimensional systems is in a considerably slower increase in the number of bound KH states with increasing radiation intensity in the three-dimensional case, which results in a very small value of the residual probability for the system to be in a bound state in the stabilization regime.

Note that the stabilization threshold corresponds to such intensities at which the characteristic width of the KH potential is of order of magnitude of the unperturbed atomic potential characteristic width, and the wave function of the stationary KH state is close to the wave function of the unperturbed atom. This fact leads to the explanation of the cause of stabilization in terms of the unperturbed atom states. Stabilization is caused by an increase in the probability of free–free transitions and by the interference of the direct ionization amplitudes and processes with a large multiphoton order involving intermediate states in continuum [17].

Our results also make it possible to reveal specific features of the ionization dynamics for systems with a short-range potential compared to systems with a long-range (Coulomb) potential. The nonmonotonic dependence of the dipole matrix element z_{1E} on electron energy, even for a one-photon bond with continuum, leads, in the case of a short-range potential, to the existence of a frequency range for which P_{BSI} lies in the domain of strong fields. Hence, the transition from the free atom to the KH atom and the realization of the KH ionization regime occurs only in sufficiently strong fields.

ACKNOWLEDGMENTS

We are grateful to M.V. Fedorov for the discussion of the problem statement and helpful remarks.

The work was supported by the Russian Foundation for Basic Research (project nos. 00-02-16046 and 00-15-96554) and by the U.S. Civilian Research & Development Foundation (project no. RP1-2259).

REFERENCES

1. Ya. B. Zel’dovich, Usp. Fiz. Nauk **110**, 139 (1973) [Sov. Phys. Usp. **16**, 427 (1973)].
2. N. L. Manakov *et al.*, Phys. Rep. **141**, 319 (1986).
3. M. V. Fedorov and A. E. Kazakov, Prog. Quantum Electron. **13**, 1 (1989).
4. R. Bhatt, B. Piraux, and K. Burnett, Phys. Rev. A **37**, 98 (1988).
5. R. M. A. Vivirito and P. L. Knight, J. Phys. B **28**, 4357 (1995).
6. A. S. Fearnside, R. M. Potvliege, and R. Shakeshaft, Phys. Rev. A **51**, 1471 (1995).
7. M. V. Fedorov and A. M. Movsesian, J. Phys. B **21**, L155 (1988).
8. M. Gavrila and J. Kaminski, Phys. Rev. Lett. **52**, 613 (1984).
9. M. Pont and M. Gavrila, Phys. Rev. Lett. **65**, 2362 (1990).
10. A. M. Popov, O. V. Tikhonova, and E. A. Volkova, Laser Phys. **9**, 201 (1999).
11. R. M. Potvliege, Phys. Rev. A **62**, 013403 (2000).
12. N. L. Manakov, M. V. Frolov, B. Borca, and A. F. Starace, in *Super-Intense Laser-Atom Physics*, Ed. by B. Piraux and K. Rzazewski (Kluwer, Dordrecht, 2001), p. 295.
13. G. F. Gribakin and M. Yu. Kuchiev, Phys. Rev. A **55**, 3760 (1997).
14. A. M. Popov, O. V. Tikhonova, and E. A. Volkova, J. Phys. B **32**, 3331 (1999).
15. N. B. Delone and V. P. Krainov, *Multiphoton Processes in Atoms* (Springer-Verlag, Berlin, 1994).
16. H. S. W. Massey, *Negative Ions* (Cambridge Univ. Press, Cambridge, 1976; Mir, Moscow, 1979).
17. A. M. Popov, O. V. Tikhonova, and E. A. Volkova, Laser Phys. **10**, 188 (2000).

Translated by A. Klimontovich

Journal of Experimental and Theoretical Physics, Vol. 93, No. 6, 2001, pp. 1163–1167.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 6, 2001, pp. 1346–1350.

Special Features of the Coherent Transient Radiation of an Electron Bunch Accelerated in a Microtron

A. V. Serov*, S. V. Levonyan, and B. M. Bolotovskiĭ

Lebedev Institute of Physics, Russian Academy of Sciences, Moscow, 117924 Russia

**e-mail: serov@sgi.lpi.msk.ru*

Received May 23, 2001

Abstract—The angular distribution of radiation generated by a relativistic electron bunch penetrating through a metal foil was experimentally measured at angles much greater than the characteristic angle γ^{-1}. The experiments showed that the emission at large angles exhibits an asymmetric angular distribution with a pronounced maximum. The results of numerical calculations of the transient radiation of an electron bunch accelerated in a microtron agree with the experimental data. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The simplest type of a transient radiation generated by a relativistic particle escaping from a conductor perpendicularly to the surface was studied in sufficient detail both theoretically and experimentally [1, 2]. As is well known, the energy radiated by a particle is zero in the direction of the particle motion, reaches maximum at an angle of $\theta_m \approx \gamma^{-1}$ measured from this direction (γ is the relative energy of the particle), and gradually decreases as the angle θ grows further. Power radiated along the surface (i.e., for $\theta \approx 90°$) is lower by a factor of γ^2 as compared to the maximum value (radiated at $\theta_m \approx \gamma^{-1}$).

When an extended bunch of charged particles crosses the conductor surface, the transient radiation of this object, representing interference of the fields generated by a large number of particles, may significantly differ from the radiation of a single charge. The difference is especially pronounced if the emission is measured at a wavelength comparable to the bunch size. In connection with this, it should be noted that the transient radiation of bunches possessing finite dimensions was theoretically studied by many researchers (see, e.g., monograph [2]) who were interested predominantly in establishing the conditions under which the radiation of a bunch containing N particles would be the same as that of an effective point particle with a charge of eN. This implied certain restrictions to the size of bunches. In practice, an alternative problem formulation can also be of interest: how the bunch size and the particle distribution affect the angular distribution and spectrum of the transient radiation. In this case, the maximum coherency of radiation cannot be attained, in a sense that the radiation intensity will not be N^2 times that of a separate particle.

Below, we report on the results of experimental measurements of the spatial distribution of the intensity of radiation generated when a bunch of particles accelerated in a microtron penetrates through a metal foil. It was found that the radiation of such bunches exhibits the properties of both transient radiation and Vavilov–Cherenkov radiation.

2. EXPERIMENTAL MEASUREMENTS OF THE TRANSIENT RADIATION

Figure 1 shows a schematic diagram of the experimental setup. The source of electrons was a microtron operating in the first acceleration mode. The particles acquired an energy of 7.4 MeV, with a total current of 40 mA in a 4-μs pulse. The electron bunch was extracted from microtron via a magnetic channel with an internal diameter of 8 mm. The electron beam passed a 1.5-m-long transient space and penetrated through a 100-μm-thick foil *1* on the microtron flange to escape into atmosphere. Then, the extracted electron bunch crossed foil *2* to produce radiation measured by detector D. The detector was based on a silicon point diode of the D404 type possessing sensitivity in several bands within a 6–12 mm wavelength range. Situated in the plane of the microtron orbit (xz plane in Fig. 1), detector D could be positioned at various distances from the beam axis z and moved parallel to this axis.

The second foil could be changed in size and moved relative to the beam. These experiments were performed with a 100-μm-thick copper foil with a length of 300 mm and a height of 200 mm. In order to protect detector D from the radiation generated in foil *1*, the second foil was bent at a right angle and positioned in front of the microtron flange as depicted in Fig. 1. The distance d from the beam axis to the bending line was 85 mm.

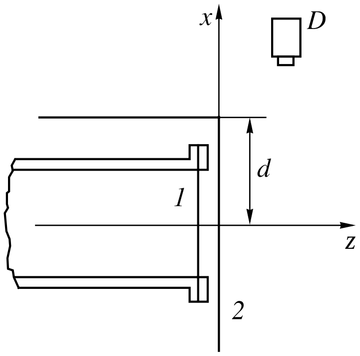


Fig. 1. Schematic diagram of the experimental geometry.

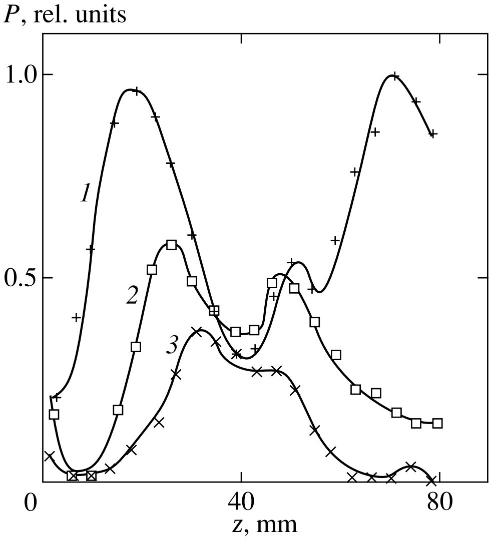


Fig. 2. Experimental plots of the radiation intensity P versus detector coordinate z for $\gamma = 15$, $\lambda \approx 8$ mm, and various fixed transverse coordinates $x = 95$ (*1*), 135 (*2*), and 165 mm (*3*).

In this experimental configuration, the relative energy was $\gamma \approx 15$. Therefore, the critical angle corresponding to a maximum intensity of the transient radiation was

$$\theta_m = \gamma^{-1} \approx 3.5°.$$

The angles θ at which the radiation was detected varied from 45° to 90°, that is, were significantly greater than θ_m. According to the theory [2], the transient radiation intensity is proportional to

$$\frac{\sin^2\theta}{(1-\beta^2\cos^2\theta)^2},$$

where $\beta = v/c$ is the ratio of the charge velocity v to the speed of light c. Therefore, the intensity of radiation of a single relativistic particle must gradually decrease by half when the angle increases from 45° to 90°.

Figure 2 shows the experimental curves of the radiation intensity versus the detector coordinate z for various fixed transverse coordinates x. The measurements were performed at a wavelength $\lambda \approx 8$ mm, comparable with the bunch size. As is seen from Fig. 2, the transient radiation intensity distribution significantly differs from that theoretically predicted for a single charge: the experimental curve exhibits pronounced maxima at large angles. For example, curve *1* shows a peak at $x = 95$ mm and $z \approx 20$ mm, which corresponds to $\theta \approx 70°$.

Our experiments revealed another significant distinction of the transient radiation of a bunch from that of a single electron: the angular distribution of the radiation intensity at large angles in the former case is asymmetric relative to the z axis. In other words, the radiation intensities are different at two points with the same z and opposite x, being significantly (5–8 times) greater for positive x than for the equal but negative coordinate.

The discrepancy between the observed angular distribution and that predicted by the theory is related to the coherent character of the radiation. In the experiments, the detector measures the radiation resulting from interference of the waves emitted by all electrons in the bunch, rather than radiation of a single electron. It is this interference that accounts for such a big difference between angular distributions of the transient radiation of an electron bunch and single electron. The main factor determining the angular distribution of the bunch is the mutual arrangement of emitting particles.

In order to explain the experimental results, we performed numerical calculations of the charge distribution in the bunch accelerated in a microtron and the angular distribution of the transient radiation generated by this bunch.

3. NUMERICAL CALCULATION OF THE TRANSIENT RADIATION

The motion of charged particles in a microtron is traditionally described using specially designed numerical methods [3]. Experimental investigations of the particle acceleration modes confirmed that these methods describe the process of electron capture and bunch formation with sufficient precision. As is known [4], the operation of a microtron depends on a number of parameters, including the size and shape of the accelerating cavity, position of the electron emitter, and amplitudes of the accelerating high-frequency field and the guiding magnetic field. For this reason, numerical calculations must take into account the particular cavity geometry and the microtron operation mode.

The cavity of the microtron in which the experiments were performed possessed certain features influencing the distribution of accelerated particles. The cavity size was such that the accelerated beam current excited a spurious high-frequency E_{011} mode. This led to an increase in the electron capture coefficient and to a change in the electron bunch size (due to the focusing effect of the induced mode). The results of calculation of the dynamics of particles accelerated in such a microtron were reported in [5].

Figure 3 shows the spatial distribution of particles in the bunch prior to penetrating through a foil. This distribution was obtained by numerically calculating the dynamics of electrons in the acceleration mode employed in our experiments. As can be seen from this figure, the electron bunch exhibits rather sharp boundaries. In the transverse direction, the boundaries are determined by the output channel diameter and the transit path length, while, in the longitudinal direction, the bunch size is determined by features of the phase motion of particles in the microtron. In the microtron operation mode employed, the longitudinal bunch size in front of the foil was ≈9 mm, while the horizontal and vertical dimensions were ≈28 and ≈4 mm, respectively.

The calculated spatial distribution of particles in a bunch was used to determine the transient radiation of the bunch penetrating through a metal foil. According to the theory, the radiation spectrum of a charged particle crossing a plane metal–vacuum interface is described by the relationship

$$E_\omega = \frac{q}{\pi c R}\left(\frac{\beta\sin\theta}{1-\beta^2\cos^2\theta}\right)\exp\left(i\frac{\omega}{c}R - i\omega t_i\right), \qquad (1)$$

where ω is the frequency, q is the particle charge, R is the distance from the point of crossing the interface to the point of observation, and t_i is the time of particle escape from the metal. The radiation field of the bunch is a sum of the fields generated by individual particles.

We have calculated the transient radiation intensity $P_\omega \sim E_\omega^2$ as a function of the longitudinal coordinate z at a fixed coordinate x. The calculations were performed for various radiation wavelengths λ. The results of these calculations are presented in Fig. 4. Figure 4a refers to the case of $\lambda = 8$ mm, at which the experimental measurements of the radiation intensity P depending on the coordinate z were carried out (see Fig. 2, curve *1*). As can be seen, the results of calculations qualitatively agree with experiment: the $P_\omega(z)$ curve for $x = 95$ mm exhibits a pronounced maximum at which the amplitude is significantly higher than that at the same z for $x = -95$ mm. However, there is some quantitative difference between theoretical and experimental data: according to the calculation, the radiation intensity maximum takes place at $\theta \approx 60°$, whereas the measurements showed the peak at $\theta \approx 70°$. This discrepancy can probably be explained by the fact that a model used for calculating the electron motion did not take into account some factors influencing the spatial distribution of electrons in the bunch. The point is that even small changes in this distribution may significantly affect the radiation characteristics.

Figure 4b shows a series of the $P_\omega(z)$ curves calculated for the same x and various radiation wavelengths

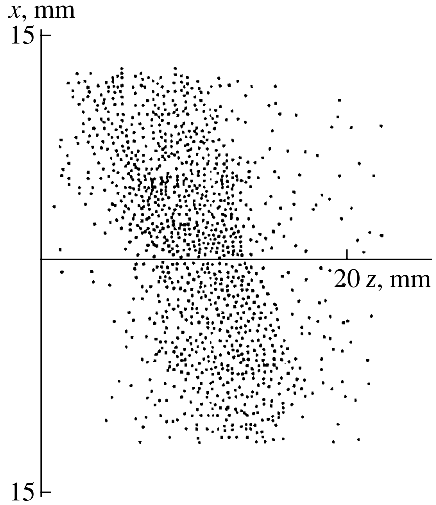


Fig. 3. A typical spatial distribution of particles in the bunch prior to penetrating through a foil.

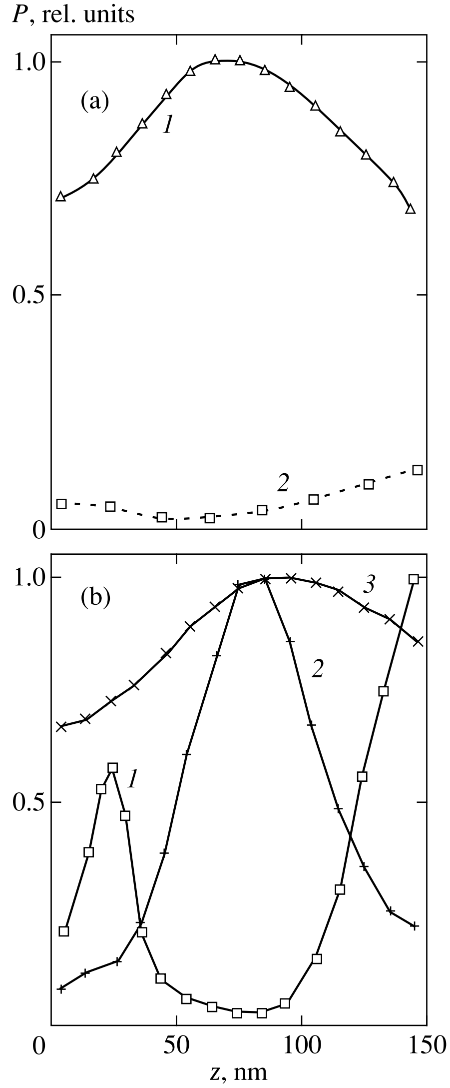


Fig. 4. Numerically calculated plots of the radiation intensity P versus detector coordinate z: (a) $\lambda = 8$ mm, $x = 95$ (*1*) and −95 mm (*2*); (b) $x = 95$ mm, $\lambda = 1$ (*1*), 4 (*2*), and 12 mm (*3*).

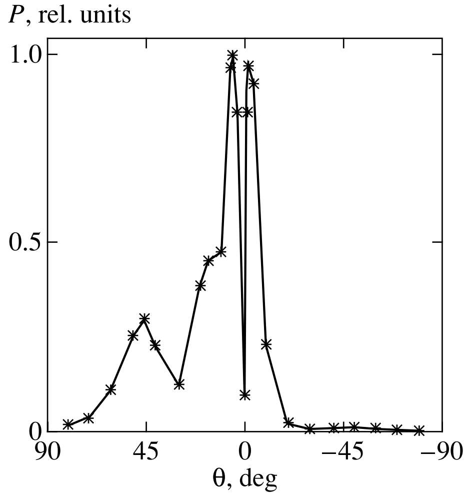


Fig. 5. The angular profile of the transient radiation intensity $P_\omega(\theta)$ in the xz plane calculated for λ = 8 mm and R = 100 mm.

λ. All curves are normalized to the maximum intensity of radiation at a given wavelength for the coordinate z varying from zero to 150 mm. These results indicate that the intensity of radiation at large angles exhibits a maximum in a broad range of wavelengths.

Figure 5 presents the angular dependence of the radiation intensity $P_\omega(\theta)$ in the xz plane. In this calculation, the distance from the point of interface crossing to the detector was taken equal to R = 100 mm. As can be seen from this figure, the transient radiation of a bunch detected at large angles relative to the direction of particle motion is characterized by a sharply asymmetric profile. In addition to the peaks at $\theta = \pm 1/\gamma$, there is a pronounced maximum at $\theta \approx 60°$.

In interpreting the results, an important circumstance is that the bunch front is sloped with respect to the velocity vector. In other words, the normal to the front is not parallel to the velocity vector, the angle between these directions being about 10°. Therefore, various points of the front cross the foil surface at various time instants. Note that the front–surface intercept moves along the foil at a velocity exceeding the speed of light. The size of the foil region from which the transient radiation is emitted varies with time. As can be readily shown, the boundary of the emitting region moves at a superluminous velocity. As a result, the process exhibits the properties of both transient radiation and Vavilov–Cherenkov radiation. Effectively, the former radiation source moves along the normal to the foil surface, while the Vavilov–Cherenkov radiation source moves in the foil plane. Note that each individual particle in the bunch crossing the surface produces only transient radiation, but the source of this radiation moves at a superluminous velocity along the interface. Interference of the waves emitted by all individual particles imparts the process directionality characteristic of Vavilov–Cherenkov radiation.

In comparing the experimental results to classical Vavilov–Cherenkov radiation, one must take into account the experimental conditions. If the path length L of the superluminous source along the surface were sufficiently large, the radiation would be characterized by a narrow directivity pattern typical of traditional Vavilov–Cherenkov radiation sources. In our experiments, the path length L of the superluminous source along the surface was on the order of a few (3–4) wavelengths. In this case, the width $\Delta\theta$ of the angular distribution of the Vavilov–Cherenkov radiation is approximately given by the formula

$$\Delta\theta = \frac{\lambda}{2L}\sin\theta_0\cos\theta_0, \tag{2}$$

where θ_0 is the Cherenkov angle.

In addition, it must be taken into account that the velocity of the superluminous source appearing in these experiments is not constant because the bunch boundary is not flat and the bunch density is not homogeneous. These factors additionally increase the angular distribution width. As can be seen from the plots presented in Fig. 4, the angular distribution width decreases with decreasing wavelength.

It should also be noted that the above features of coherent radiation can be observed in the experiments with relativistic beams on most of the linear accelerators. The maxima of the radiation intensity at large angles will be most pronounced in cases when the transverse size of a bunch is greater than the longitudinal size. Such bunches are obtained, for example, in the accelerator of the Nuclear Physics Laboratory of Tokhoku University (Sendai, Japan).

4. CONCLUSION

The angular distribution of intensity of the transient radiation generated by a relativistic electron bunch accelerated in a microtron was experimentally measured at a wavelength λ = 8 mm (comparable to the bunch size) at angles much greater than the characteristic angle $\theta_m = 1/\gamma$. The experiments showed that the coherent emission at large angles exhibits an asymmetric angular distribution with a pronounced maximum at $\theta \approx 70° \gg \theta_m$. Using the results of numerical calculations of the motion of electrons in a microtron, the spatial distribution of particles in a bunch penetrating through a metal foil was determined. Based on these data, the angular distribution of the intensity of a coherent transient radiation of the electron bunch was calculated. The results of these calculations qualitatively agree with the experimental data.

ACKNOWLEDGMENTS

The authors are grateful to B.S. Dumesh (Institute of Spectroscopy, Russian Academy of Sciences, Moscow) for kindly providing the D404 diodes.

This study was supported by the Russian Foundation for Basic Research, project nos. 99-02-18183 and 00-02-17381.

REFERENCES

1. V. L. Ginzburg and I. M. Frank, Zh. Éksp. Teor. Fiz. **16**, 15 (1946).
2. V. L. Ginzburg and V. N. Tsytovich, *Transient Radiation and Transient Scattering* (Nauka, Moscow, 1984).
3. S. P. Kapitsa, V. N. Melekhin, I. G. Krutikova, and G. P. Prudkovskiĭ, Zh. Éksp. Teor. Fiz. **41**, 376 (1961) [Sov. Phys. JETP **14**, 266 (1962)].
4. S. P. Kapitsa and V. N. Melekhin, *Microtron* (Nauka, Moscow, 1969).
5. K. A. Belovintsev, S. V. Levonyan, and A. V. Serov, Zh. Tekh. Fiz. **51**, 752 (1981) [Sov. Phys. Tech. Phys. **26**, 447 (1981)].
6. T. Takahashi, Y. Shibata, R. Ishi, *et al.*, Phys. Rev. E **50**, 4041 (1994).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 93, No. 6, 2001, pp. 1168–1177.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 6, 2001, pp. 1351–1361.
Original Russian Text Copyright © 2001 by Lebedev, Misochko, Ainbund.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Manifestations of the Two-Quantum Photoeffect and Photon Statistics in the Photoelectron Multiplier Pulse Amplitude Distribution

M. V. Lebedev[a], *, O. V. Misochko[a], and M. R. Ainbund[b]
[a]*Institute of Solid State Physics, Russian Academy of Sciences, p/o Chernogolovka, Moscow oblast, 142432 Russia*
**e-mail: lebedev@issp.ac.ru*
[b]*Central Research Institute, Electron Corporation, St. Petersburg, Russia*

Received July 17, 2001

Abstract—The contribution of the two-quantum photoeffect to the pulse amplitude distribution at the output of a photomultiplier is considered. An expression generalizing the Mandel formula is derived that takes into account the second-order photoeffect. The influence of a temporal and spatial coherency of the radiation field on the photomultiplier count statistics is studied. Possible applications of the obtained results to determining statistical characteristics of the optical fields are discussed. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The two-quantum photoeffect was previously studied both experimentally and theoretically in connection with the problem of determining the count statistics for so-called two-photon photoelectron multipliers tubes (PMTs) [1–4]. In these devices, the photocathode electron work function is too high for detecting single photons but allows an electron to be emitted from the cathode when two energy quanta of the incident radiation are simultaneously absorbed [1–4]. On the other hand, Artem'ev [5, 6] demonstrated that the two-quantum photoeffect may significantly contribute to the pulse amplitude distribution at the output of a usual "one-photon" PMT. It was established that pulses produced upon absorption of the correlated photons at the PMT cathode are characterized, on the average, by a twice greater amplitude as compared to that of the one-photon response pulses; the two-photon response is extremely sensitive to the degree of coherency of the incident light.

In recent years, the interest of researchers in the quantum-statistical analysis of optical radiation has markedly increased in connection with the study of nonclassical states of light fields (including squeezed light [7]), with the development of quantum photometry [8], and with the commonly recognized importance of taking the coherency into account in descriptions of the collective states of quantum ensembles (e.g., superconductivity and Bose–Einstein condensation). In this context, development of new highly sensitive methods for measuring the photon statistics is of urgent importance.

We aimed at investigating the two-quantum photoeffect in detail and assessing the possibility of using this phenomenon in the study of coherence of the optical radiation. The material is arranged as follows. Section 2 describes the experimental setup and conditions. The results of the measurements are presented in Section 3. In Section 4, we analyze the pulse amplitude distribution at the output of a PMT (FEU-64) and describe a theoretical model providing interpretation of the main experimental data. Section 5 is devoted to application of the model to an analysis of the whole body of experimental results presented in Section 3. In Section 6, we derive a generalized Mandel formula which takes into account both one- and two-quantum photoeffect contributions. In the Conclusion, we summarize the main results and discuss possibilities of the further experimental investigation of the photon statistics.

2. EXPERIMENTAL SETUP AND PROCEDURE

We have studied the photoelectric effect on an antimony–cesium photocathode by measuring the distribution of pulse amplitudes at the output of a FEU-64 photoelectron multiplier tube detecting light from sources possessing various statistical properties. Figure 1 shows a general schematic diagram of the experimental setup. The light from a radiation source passed through a grating monochromator (MDR-6U) with a PMT (FEU-64 tube) at the output placed into a cooled housing to reduce the dark noise level. The experiments were performed using different light sources, including an incandescent lamp with a ribbon emitting element (a SI6-100 lamp power supplied from a stabilized power unit) and a He–Ne laser (ILGN-104 type) generating a set of He and Ne spectral lines (lasing at 6328 Å was suppressed by detuning one of the resonator mirrors).

The above light sources are characterized by significantly different radiation coherence times. As is well known, the spectral lines of neon in a He–Ne laser are broadened as a result of the Doppler effect, with a typical linewidth amounting to 1700 MHz, corresponding to a coherence time on the order of 0.2 ns. For an incandescent lamp, the coherence time can be estimated using the spectral width of the monochromator exit slit cutting a quasi-monochromatic spectral range from the incident radiation. In our experiments, the exit slit width was varied from 0.1 to 2 mm, which corresponded to the coherence time changing from 1 to 0.06 ps (for an MDR-6U monochromator with 1200-mm^{-1} gratings and a 13-Å/mm dispersion).

For the convenient comparison of different sources, the light intensities were leveled by a neutral filter with variable optical density. In addition, the short-wavelength radiation passing through the monochromator in the second diffraction order of the gratings was eliminated by using various spectral filters. The monochromator entrance slit was illuminated in both single-mode and multimode regimes. In the former case, the light source was spaced 70 cm from the 100-μm-wide entrance slit. The laser tube had an internal diameter close to 1 mm, whereas the extended light source (incandescent lamp) was diaphragmed with a 1-mm aperture. As a result, the size of a coherently illuminated spot in the entrance slit plane [9] was approximately equal to the entrance slit size. In the transverse direction, the radiation was confined by the second slit oriented at a right angle to and positioned immediately in front of the entrance slit (to provide for the maximum proximity of the slit planes). The size of the second slit was also 100 μm. In the multimode regime, the entrance slit was illuminated with the aid of a condenser.

The PMT (FEU-64) output pulses were amplified by a preamplifier possessing a bandwidth of 300 MHz and a gain coefficient of 20. The preamplifier was mounted inside the cooled PMT housing. The preamplified signal was fed to a matched splitter with a bandwidth of 18 GHz which divided the signal between two identical channels. In the first channel, the signal was amplified by Ortec 9302 and 474 amplifiers, inverted, and fed to a multichannel analyzer. In the second channel, the signal was analyzed using an Ortec 584 CF-discriminator, the output pulses from which were detected by an Ortec 9315 photon counter.

The pulse amplitude distribution at the PMT output is highly sensitive both to the character of illumination of the photocathode surface and to instabilities in the operation of electronic equipment. It is well known that too intense an illumination of the photocathode and sharp fluctuations in the level of high voltage supplied to the PMT may lead to long-term (on the order of hours and above) variations in the pulse amplitude distribution. In the course of long-time data accumulation exposures, a certain distortion can be introduced by slow drift in the parameters of the electronics. In order to minimize the effect of such a drift, all the experimental setup was switched on and trained for 1.5–2 h prior to measurements. The influence of photocathode inhomogeneities and a scatter of the takeoff coefficient for photoelectrons ejected from various cathode areas were eliminated by gluing a diaphragm with an aperture of 2.5 mm onto the PMT entrance window, which restricted illumination to a small central region of the photocathode.

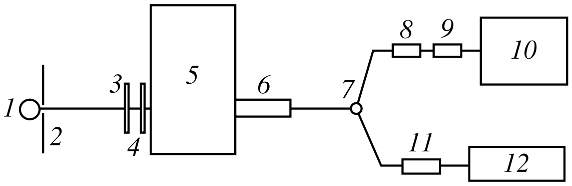


Fig. 1. Schematic diagram of the experimental setup: (*1*) light source; (*2*) aperture; (*3, 4*) neutral and spectral filters; (*5*) monochromator; (*6*) photoelectron multiplier; (*7*) matched splitter; (*8, 9*) amplifiers; (*10*) multichannel analyzer; (*11*) discriminator; (*12*) photon counter.

Using the two-channel measuring scheme described above, it was possible to perform the photon count experiments with a fast-response discriminator under conditions of a permanent monitoring of the pulse amplitude distribution. If the amplitude distribution deviated in shape from the normal (which sometimes took place, e.g., as a result of electric breakthroughs), the data were rejected. The dark noise level of the cooled PMT was as low as a few tens of counts per second (cps) and, hence, did not significantly influence the results of measurements.

3. EXPERIMENTAL RESULTS

Figure 2 shows a typical pulse amplitude distribution at a photoelectron multiplier tube output. In this curve, we may separate the regions of dynode noise (I), one-electron peak (II), and large-amplitude-pulse contribution (III). The one-electron peak is separated from the dynode noise by a dip, the level and position of which will be characterized by a “contrast” parameter K representing the ratio of the one-electron peak height to the left-hand minimum level.

Figure 3 shows the pulse amplitude distributions (plotted on the usual and logarithmic scales) measured with a FEU-64 photomultiplier operated at various intensities of the photocathode illumination with an incandescent lamp. In the logarithmic plot (Fig. 3b), the distributions were normalized to the one-electron peak height, which provides for a convenient comparison of the curve shapes. At small illumination intensities, the distribution shows a clearly pronounced one-electron peak. As the intensity grows, an additional “shoulder” appears and rapidly increases on the right-hand wing, while the one-electron peak position remains unchanged. However, on further increasing the illumi-

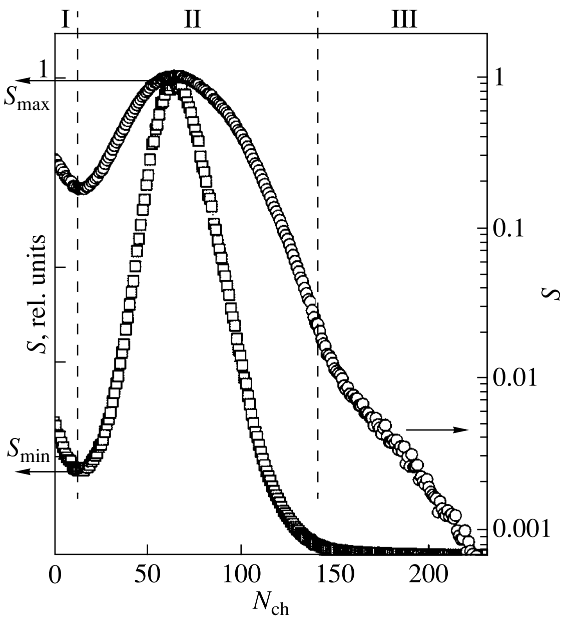


Fig. 2. A plot of the pulse count rate S versus channel number N_{ch} showing the typical pulse amplitude distribution at a PMT output (on the usual and logarithmic scales) and illustrating the one-electron peak "contrast" ($K = S_{max}/S_{min}$) determination: (I) dynode noise; (II) one-electron peak; (III) large-amplitude contribution (the curves are normalized to maximum).

nation intensity, the one-electron peak begins to shift toward small amplitudes.

The additional shoulder in the amplitude characteristic of the FEU-64 was originally observed by Artem'ev [5, 6] and interpreted as a "two-electron" peak arising due to the events whereby two electrons are ejected from the photocathode upon simultaneous absorption of two photons. This interpretation was based on the fact that the shoulder is observed in the region where the pulse amplitudes are approximately twice as large as the one-electron peak amplitude. The two-electron peak due to the simultaneous absorption of photons follows simply from energy considerations.

A simple approach to verification of this concept consists in establishing the dependence of the two-electron peak height on the illumination intensity. The results of such measurements are presented in Fig. 4, showing a plot of the photon count rate S_{2e2ph} in the region of the two-electron peak (Fig 2, region III) versus the photon count rate S_{1e1ph} corresponding to the one-electron peak. Since the latter value is proportional to the illumination intensity, we may conclude that the two-electron peak height is proportional, with a good precision, to the square of the incident light intensity. Thus, the quadratic relationship is indicative of a two-photon nature of the two-electron peak. For convenience, the one-electron one-photon and two-electron two-photon maxima in the pulse amplitude distributions are referred to below as the 1*e*1*ph* and 2*e*2*ph* peaks, respectively.

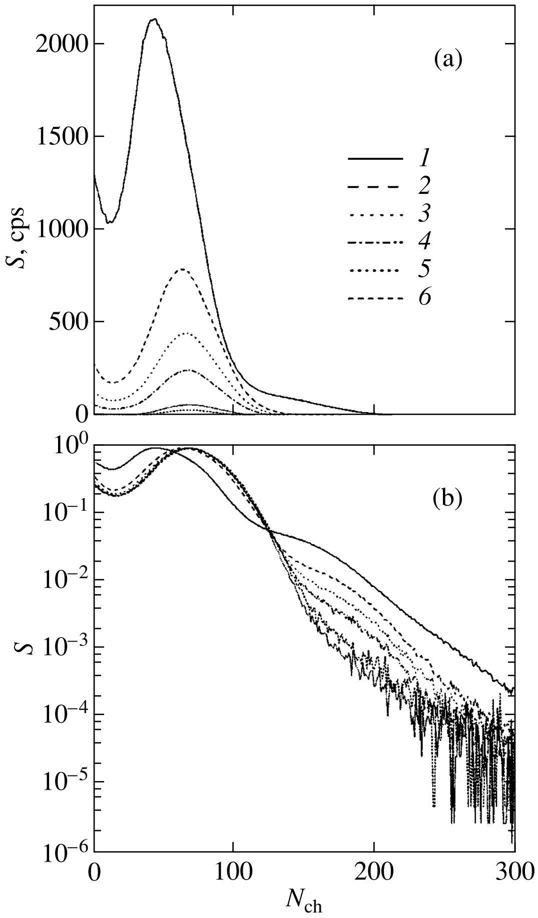


Fig. 3. The pulse amplitude distributions profiles on the (a) usual and (b) logarithmic scales measured using a FEU-64 photomultiplier at various intensities of the photocathode illumination with the light of an incandescent lamp ($\lambda = 700$ nm).

The response of a one-photon photodetector (i.e., the number of counts in a one-electron peak) is proportional to the average illumination intensity $\overline{I}(T)$. The $\overline{I}(T)$ value as such does not bear information about statistical properties of the measured light, since these properties depend on the second- and high-order moments. Therefore, the 2*e*2*ph* peak is of considerable interest from the standpoint of the photon statistics. The number of counts of a two-photon photodetector over a certain time interval T is proportional to $\overline{I^2}(T)$, which is the square of the light source intensity averaged over the time T. The $\overline{I^2}(T)$ value significantly depends on the nature of the light source. Importance of the 2*e*2*ph* peak for the photon statistics was correctly understood by Artem'ev, but his statement that this peak arises only in response to coherent light sources [6] and, hence, allows correlated photons to be directly detected is (see below) misleading.

Once the amplitude characteristics measured at various illumination levels have been considered, let us turn to the spectral dependence of the pulse amplitude distribution at the FEU-64 output. In order to study the spectral dependence, the pulse amplitude distributions were measured by illuminating a PMT with a quasimonochromatic light beam obtained by passing the incandescent lamp radiation through the monochromator. Figure 5 shows the pulse amplitude distributions measured under these conditions for various wavelengths. As is well known, the sensitivity of the antimony–cesium photocathode is maximal in the wavelength range from 200 to 400 nm and drops rapidly when the light wavelength increases above 400 nm.

In order to maintain the measurement accuracy on an approximately constant level, we controlled the light intensity with the aid of a neutral light filter so as to obtain the same signal intensity (~10^4 cps) at the PMT output for all wavelengths. Independent estimates obtained using a photomultiplier of the XPH277 type (with the spectral sensitivity range extended toward red spectral interval) showed that the FEU-64 output signal on a level of 10^4 cps at a wavelength of 800 nm could be obtained using a light flux on the order of 7×10^7 photons/s. The main feature revealed by the data presented in Fig. 5 is that the PMT response in the region of small amplitudes exhibits a significant growth with increasing light wavelength, which is manifested by a sharp drop in "contrast" of the 1*e*1*ph* peak. A similar decrease in the contrast is observed with increasing illumination intensity (Fig. 6).

Finally, let us consider dependence of the PMT pulse amplitude distribution on the illuminated cathode area. This is an important relationship because the 2*e*2*ph* peak intensity is a nonlinear (quadratic) function of the light intensity. The 1*e*1*ph* peak height depends linearly both on the light intensity and on the illuminated spot area. Therefore, the same 1*e*1*ph* peak can be obtained with a large spot at low intensity and with a small spot at proportionally high intensity. In other words, the 1*e*1*ph* peak intensity depends only on the total number of photons striking the cathode. The 2*e*2*ph* peak may behave differently: if we maintained a constant number of photons in the light beam and varied the cathode spot area, the intensity would increase with decreasing spot size and, hence, the 2*e*2*ph* peak height would grow relative to that of the 1*e*1*ph* peak height. In that case, it would be necessary to perform measurements with permanent control of the spot size on the PMT cathode.

In order to study how the 2*e*2*ph* peak height depends on the illuminated spot area, we measured the PMT pulse amplitude distributions for various values of the exit slit width of a monochromator illuminated with the light of an incandescent lamp. Using a condenser lens mounted in the cooled PMT housing, the exit slit image was projected onto the photocathode. It was found that the 2*e*2*ph* peak height grows by a quadratic law with increasing spot area. In other words, the peak height is proportional to the square of the number of photons in the beam. As a result, the ratio of the 1*e*1*ph* and 2*e*2*ph* peak intensities is independent of the illuminated spot area on the photocathode and is determined only by the number of photons in the beam.

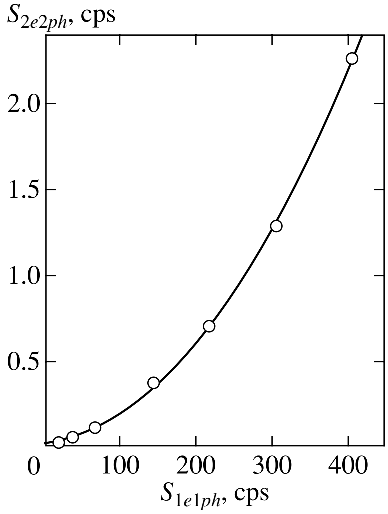


Fig. 4. Experimental plot of the 2*e*2*ph* peak intensity versus the illumination intensity (incandescent lamp); solid curve shows the result of approximation by a quadratic function.

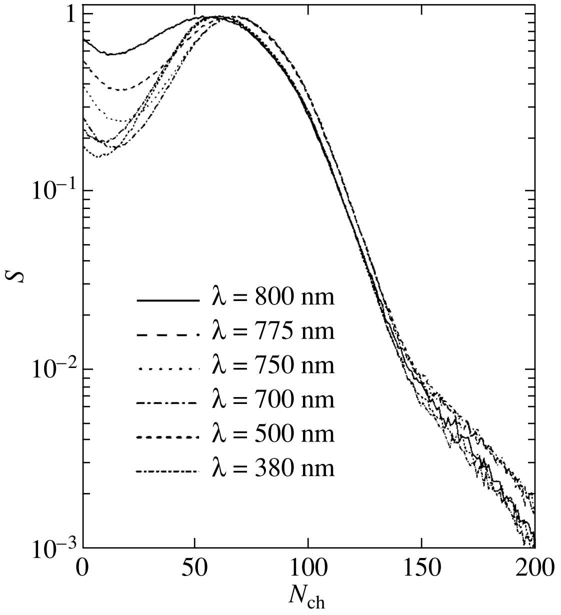


Fig. 5. Pulse amplitude intensity distributions for a PMT illuminated with a quasimonochromatic light of various wavelengths from an incandescent lamp. The one-electron peak contrast decreases significantly with increasing wavelength (the curves are normalized to maximum).

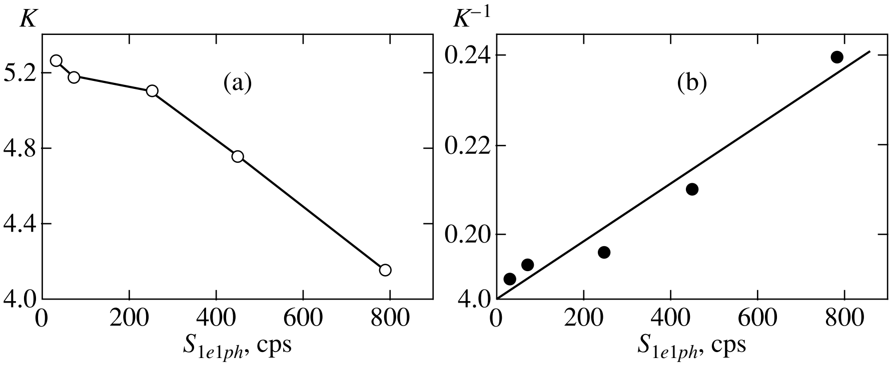


Fig. 6. Plots of (a) the one-electron peak contrast and (b) the inverse contrast (approximated by a linear function) versus the illumination intensity (proportional to the count rate at the 1*e*1*ph* peak).

4. A MODEL DESCRIBING THE PHOTOMULTIPLIER PULSE AMPLITUDE DISTRIBUTION

A real photoelectron multiplier is a sufficiently complicated device that may feature various processes directly and indirectly influencing the output pulse distribution. In order to rationalize the obtained experimental data, it is necessary to develop a theoretical model correctly reflecting the main experimental trends. Such a model can be based on the approach proposed and developed by Prescott [10]. According to this approach, the process of electron multiplication on the PMT dynodes is considered as a Poisson process characterized by an average secondary electron emission coefficient μ. This allows a recurrent formula to be derived for the generating function of the electron number distribution past M dynodes. Assuming (as an initial condition) that a single electron arrives at the first dynode from the photocathode, we can determine a distribution of the probability for m electrons to be ejected after the Mth dynode.

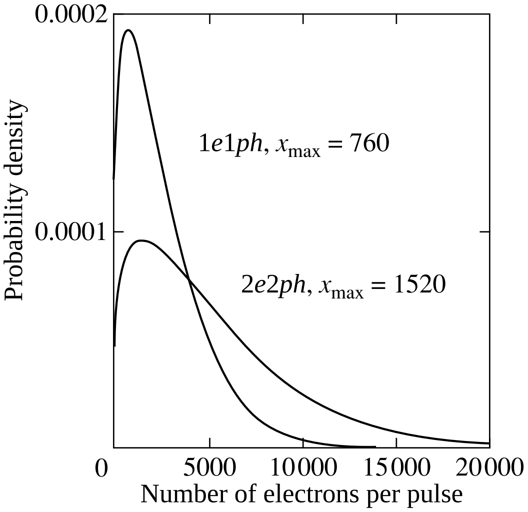


Fig. 7. Pulse amplitude distributions for the 1*e*1*ph* and 2*e*2*ph* peaks calculated by the Prescott formulas with $\mu = 2$, $b = 0.1$, and $M = 11$.

In order to take into account the PMT nonideality, Prescott introduced a coefficient b that is much smaller than unity for a good photomultiplier [10]. If the initial condition is modified so that two electrons (rather than one) are ejected from the photocathode, the Prescott formulas must describe the pulse amplitude distribution for the 2*e*2*ph* peak as well. The corresponding distribution profiles are presented in Fig. 7. In confirmation of intuitive expectations, (i) the 2*e*2*ph* peak turns out to be two times as wide as the 1*e*1*ph* peak, and (ii) the pulse amplitude at which the distribution exhibits a maximum for the 2*e*2*ph* peak is two times that for the 1*e*1*ph* peak. These properties of the distributions calculated by the Prescott formulas agree well with the properties of 1*e*1*ph* and 2*e*2*ph* peaks observed in experiment. However, the experimental distribution obtained for the FEU-64 is less asymmetric than that theoretically predicted. Therefore, the real peak amplitude distribution can hardly be fitted to the Prescott formulas with reasonable values of the model parameters.

Let us consider in more detail the region of small pulse amplitudes to the left of the 1*e*1*ph* peak observed in experiment. This region is traditionally related to the PMT dynode noise. Investigations of the dynode noise showed evidence of rather unusual properties [11], one of which is the sensitivity toward PMT illumination. This effect is conventionally explained by the probability of electron escape from the dynode increasing as a result of the dynode bombardment by electrons of the avalanche, that is, by a “memory” of the previously detected photon retained for a certain time. The memory trace can be sufficiently long, since an increase in the dynode noise in some experiments was observed for more than one day after intense illumination [11].

For constructing the model, let us assume that all dynode noises are related only to the first dynode. This simplifying assumption is in fact quite realistic. Indeed, the average secondary electron emission coefficient of the dynode amounts to approximately 3.5 (as can be readily estimated from the amplitude of the one-electron peak, taking into account that the FEU-64 contains 11 dynodes with equal voltages between these units). Thus, one "noisy" electron ejected from the first dynode would give rise to an average output pulse amplitude 3.5 times smaller than that of the 1*e*1*ph* peak. This estimate provides for a correct scale of the dynode noise amplitude. The noise in the second and subsequent dynodes leads to a noise pulse amplitude at least 3.5 times lower compared to that of the first dynode noise, thus contributing to the distribution in the immediate vicinity of the zero-energy channel. We will also assume that the escape of more than one "noisy" electron from the first dynode has extremely low probability after this dynode is bombarded with either one or even two photoelectrons.

Under the above assumptions, the probability of detecting N "noisy" dynode pulses with an amplitude falling within the interval from x to $x + dx$ over a time T can be expressed as

$$P_{ND}(x, T)dx = P_N(T)p_D(x)dx, \tag{1}$$

where $p_D(x)$ is the probability density of the dynode noise pulse amplitude distribution and $P_N(T)$ is the probability of detecting N counts (pulses) over the time T. The distribution of counts with an allowance for both one- and two-quantum photoeffects is considered in detail in Section 6. Here, we only note that the noise pulses arise only during the detection of photons, whereas the dark noise is ignored. Therefore, the probability of detecting a dynode noise pulse of a given amplitude is a conditional probability, which is reflected by formula (1).

Averaged over the time T, the number of dynode noise pulses with a given amplitude $\overline{N_D}(x, T)dx$ can be expressed as

$$\begin{aligned}\overline{N_D}(x, T)dx &= \sum_{N=1}^{\infty} NP_{ND}(x, T)dx \\ &= \sum_{N=1}^{\infty} NP_N(T)p_D(x)dx = \bar{N}p_D(x)dx.\end{aligned} \tag{2}$$

As demonstrated below (see Eq. (18)),

$$\overline{N_D}(x, T)dx = [\alpha_1 \bar{I}(T)T + \alpha_2 \overline{I^2}(T)T]p_D(x)dx, \tag{2*}$$

where α_1 and α_2 are the efficiencies of the one- and two-quantum photoeffects, and $\bar{I}(T)$ and $\overline{I^2}(T)$ are the average intensity and average squared intensity.

5. DISCUSSION OF EXPERIMENTAL RESULTS

Now, we will attempt at explaining, within the framework of the model described in the preceding section, the whole body of the experimental data. As indicated above, both 1*e*1*ph* and 2*e*2*ph* peaks observed in the experimental distributions agree quite well in position and width with the one- and two-electron peaks in the model based on the Prescott formulas. For small light intensities, when the 2*e*2*ph* peak contribution to the distribution is insignificant, the dynode noise level must linearly depend (see formula (2*)) on the illumination intensity—and this is actually observed in experiment. As the light intensity increases, the contribution of the 2*e*2*ph* peak becomes significant (see curve *4* in Fig. 3b) and the dynode noise grows faster than does the 1*e*1*ph* peak, which accounts for deteriorated contrast of the latter peak. Note also that an analogous decrease in the 1*e*1*ph* peak contrast is observed when the photocathode is illuminated with light possessing a 750–800 nm wavelength. This fact can be seen in Fig. 5 and is clearly illustrated in Fig. 8, showing the contrast as a function of the wavelength. This drop in the contrast can be readily explained using formula (2*), since both $\alpha_1(\lambda)$ and $\alpha_2(\lambda)$—the efficiencies of one- and two-quantum photoeffects, respectively—depend on the light wavelength λ. An increase in the relative contribution of the two-quantum photoeffect is evidence that $\alpha_1(\lambda)$ decreases more rapidly than $\alpha_2(\lambda)$ with increasing λ. Note that a difference in the efficiency of the one- and two-quantum photoeffects was pointed out in [5, 6].

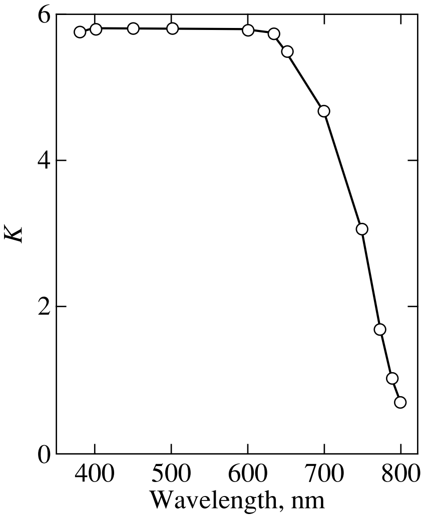


Fig. 8. A plot of the one-electron peak contrast versus quasi-monochromatic illumination wavelength for a constant total (integrated over all channels) count rate of ~10^4 cps.

Based on the notions underlying formula (2*) and on the analogy in variation of the pulse amplitude distribution with increasing illumination intensity and light wavelength, we may suggest a method for evaluation of the relative quantum efficiency of the two-quantum photoeffect. Indeed, it can be shown (see Section 6) that the average number of one- and two-electron pulses with a given amplitude x is

$$N_{1e1ph} = \alpha_1 \bar{I}(T)Tp_{1e1ph}(x)dx, \tag{3}$$

$$N_{2e2ph} = \alpha_2 \overline{I^2}(T)Tp_{2e2ph}(x)dx. \tag{4}$$

Then, the contrast K of the $1e1ph$ peak is determined as

$$K^{-1} = \frac{N_D(x_D) + N_{1e1ph}(x_D)}{N_D(x_{\max 1e1ph}) + N_{1e1ph}(x_{\max 1e1ph}) + N_{2e2ph}(x_{\max 1e1ph})}, \tag{5}$$

where

$$x_D = \frac{1}{\mu} x_{\max 1e1ph}.$$

It would be expedient to ignore the contributions of dynode noise and $2e2ph$ pulses in the denominator in comparison to the $1e1ph$ peak contribution, since usually

$$N_D(x_{\max 1e1ph}),\ N_{2e2ph}(x_{\max 1e1ph}) \ll N_{1e1ph}(x_{\max 1e1ph}).$$

Then, we obtain

$$\begin{aligned} K^{-1} &= \frac{[\alpha_1 \overline{I}(T) + \alpha_2 \overline{I^2}(T)] T p_D(x_D) + \alpha_1 \overline{I}(T) T p_{1e1ph}(x_D)}{\alpha_1 \overline{I}(T) T p_{1e1ph}(x_{\max 1e1ph})} \\ &= \frac{p_D(x_D) + p_{1e1ph}(x_D)}{p_{1e1ph}(x_{\max 1e1ph})} + \frac{\alpha_2 \overline{I^2}(T)}{\alpha_1 \overline{I}(T)} \frac{p_D(x_D)}{p_{1e1ph}(x_{\max 1e1ph})}. \end{aligned} \tag{6}$$

For a stabilized light source, $\overline{I^2}(T) = \overline{I}^2(T)$ and eventually we arrive at the formula

$$K^{-1} = A + \frac{\alpha_2}{\alpha_1} B \overline{I}(T), \tag{7}$$

where the coefficients A and $(\alpha_2/\alpha_1)B$ can be readily determined in experiment. Note that the ratio α_2/α_1 is wavelength dependent. Taking the dependence of K^{-1} on $\overline{I}(T)$ for various λ, we obtain an expression for the change in the relative quantum efficiency:

$$\frac{\alpha_2}{\alpha_2}(\lambda) / \frac{\alpha_2}{\alpha_1}(\lambda_{\max}),$$

where $\lambda_{\max}$ is the wavelength corresponding to a maximum spectral sensitivity of a given PMT. This result is confirmed by the contrast versus wavelength curve presented in Fig. 8, which shows that the contrast drops for $\lambda > 650$ nm and is constant ($K = \text{const}$) at lower wavelengths.

6. MANDEL FORMULA GENERALIZED TO TAKE INTO ACCOUNT BOTH ONE- AND TWO-QUANTUM PHOTOEFFECTS

Using the well-known Mandel formula (see, e.g., [12]), it is possible to calculate the probability of detecting a preset number of photons over the time interval T, provided that a one-quantum mechanism of the photoeffect is operative in the detector. Let us consider how the original Mandel formula for the probability $P_n(t, t')$ of detecting n counts at the PMT output should be modified if the two-quantum photoeffect plays a significant role in addition to the one-quantum mechanism.

Let us consider a PMT provided with the gate opened at a time instant t and closed at t', so that the light strikes the photocathode only within the time interval $T = t' - t$ referred to as the count time. We assume that the probability of exciting one photoelectron during a negligibly small time interval dt' upon the absorption of a single photon is proportional to the instantaneous light intensity I, $p_1 dt' = \alpha_1 I dt'$, while the absorption of two photons leads to the emission of two photoelectrons with a probability of $p_2 dt' = \alpha_2 I^2 dt'$. As will be shown below, it is the latter two-electron case that frequently takes place in practice. Artem’ev [5, 6] attributed the appearance of pulses with a double amplitude (related to the two-quantum photoeffect) in the amplitude distribution to the events in which two electrons are simultaneously ejected from the cathode.

The fact that one-quantum and two-quantum pulses differ in amplitude requires a certain specification of the experimental procedure used for detecting these pulses. Indeed, in the traditional count statistics, each photoelectron emitted from the cathode yields a pulse at the PMT output (provided the effective collection of photoelectrons with a takeoff probability equal to unity). However, this is no longer the case in our situation. The registration system may (i) measure the total charge delivered to the anode during the count time T, or (ii) count the total number of pulses, or (iii) measure the amplitude distribution of pulses. Apparently, the most detailed information will be gained in the latter case.

Taking into account all the above considerations, we will separately consider two cases: first, when the total number of photoelectrons emitted from the cathode is detected and, second, when the total number of pulses at the PMT output is counted without analysis of the amplitude distribution. As will be demonstrated below, this approach will allow us to study all other possible experimental situations as well (for example, by taking into account the two-quantum one-electron photoeffect, whereby the absorption of two photons leads to the emission of a single photoelectron).

Let us denote by p_0dt' the probability that not a single photoelectron is emitted from the cathode during the time interval dt'. Then, by virtue of the completeness of the space of elementary events taking place within a negligibly small time interval dt', we may write for the first case

$$p_0dt' + p_1dt' + p_2dt' = 1. \tag{8}$$

Here, it is implicitly assumed, in accordance with experimental practice, that

$$p_2dt' \gg (p_1dt')^2.$$

Taking into account that the detection process is continuous in time, the integral probability $P_n(t, t')$ of detecting n electrons over the time period $T = t' - t$ must obey the following equation:

$$\begin{aligned} P_n(t, t') &= P_n(t, t' - dt')(1 - p_1dt' - p_2dt') \\ &+ P_{n-1}(t, t' - dt')(1 - p_2dt')p_1dt' \\ &+ P_{n-2}(t, t' - dt')(1 - p_1dt')p_2dt', \end{aligned} \tag{9}$$

which yields

$$\begin{aligned} \frac{dP_n(t, t')}{dt'} &= -(p_1 + p_2)P_n(t, t') + P_{n-1}(t, t')p_1 \\ &+ P_{n-2}(t, t')p_2. \end{aligned} \tag{10}$$

Obviously, the right-hand part of the equation for $P_1(t, t')$ must contain only the first two terms. In the simplest differential equation describing $P_0(t, t')$, we should retain only the first term in the right-hand part:

$$\frac{dP_0(t, t')}{dt'} = -(p_1 + p_2)P_0(t, t'). \tag{11}$$

The initial condition is that $P_0(t, t) = 1$. A solution to this equation is expressed as

$$P_0(t, T) = \exp[-\alpha_1\overline{I}(t, T) - \alpha_2\overline{I^2}(t, T)], \tag{12}$$

where $\overline{I}(t, T)$ is the average intensity and $\overline{I^2}(t, T)$ is the average squared intensity over the count time, which are determined as

$$\overline{I}(t, T) = \frac{1}{T}\int_t^{t+T} I(t')dt', \tag{13}$$

$$\overline{I^2}(t, T) = \frac{1}{T}\int_t^{t+T} I^2(t')dt'. \tag{14}$$

Here, by an average over the count time, we imply the time-average for the optical field taken at the initial time instant t and freely evolving over the count time T. Since neither the field state at the time t nor the field evolution with time is known, the quantities $\overline{I}(t, T)$ and $\overline{I^2}(t, T)$ are essentially random values. Therefore, the final distribution of the probability of detecting a preset number of counts must be determined by averaging over an ensemble of the optical fields with all possible initial states. Assuming the process to be stationary and ergodic, we omit the argument t in what follows.

As can be readily shown, a solution to Eq. (10) for an arbitrary n can be written in the following form:

$$P_n(T) = F_n(T)P_0(T), \tag{15}$$

so that

$$\begin{aligned} F_0(T) &= 1, \\ F_1(T) &= \alpha_1\overline{I}(T)T, \\ F_2(T) &= \frac{[\alpha_1\overline{I}(T)T]^2}{2} + \alpha_2\overline{I^2}(T)T, \\ F_3(T) &= \frac{[\alpha_1\overline{I}(T)T]^3}{3!} + [\alpha_1\overline{I}(T)T][\alpha_2\overline{I^2}(T)T], \end{aligned} \tag{16}$$

and so on. A sum of all $F_i(T)$ represents an expansion into series for the exponent

$$\exp[\alpha_1\overline{I}(T)T + \alpha_2\overline{I^2}(T)T],$$

which ensures that the sum of all probabilities $P_i(T)$ is unity. Note that the obtained probability distribution is not of the Poisson type.

Now let us consider the second case, whereby the detection system counts the total number of pulses over the time period T not distinguishing between one- and two-electron pulses. Denoting this total number of counts by N and conducting considerations analogous to those described above, we eventually arrive at the following equation:

$$\begin{aligned} \frac{dP_N(t, t')}{dt'} &= -[p_1(t') + p_2(t')] \\ &\times P_N(t, t') + [p_1(t') + p_2(t')]P_{N-1}(t, t'). \end{aligned} \tag{17}$$

As can be readily seen, a solution to this equation is provided by the Poisson distribution

$$P_N(T) = \frac{\overline{N}^N}{N!}\exp(-\overline{N}), \tag{18}$$

where $\overline{N} = \alpha_1\overline{I}(T)T + \alpha_2\overline{I^2}(T)T$. Note that this case also describes the one-electron two-photon process, which may contribute to the one-electron peak in the pulse amplitude distribution. Note that the coefficients α_2 for the one- and two-electron photoeffects may differ significantly.

The Poisson distributions also describe the one- and two-electron pulses detected over the time period T in the important case when the PMT output pulses are well separated with respect to amplitude and the space of elementary events can be determined independently for the pulses of each type. The average number of pulses is

$$\bar{N}_{1e1ph} = \alpha_1 \bar{I}(T)T$$

in the first case and

$$\bar{N}_{2e2ph} = \alpha_2 \overline{I^2}(T)T$$

in the second case. The final expression for the photoresponse distribution function is obtained by averaging over an ensemble of the optical fields with all possible initial instantaneous intensities:

$$\tilde{P}(N, T) = \langle P_N(T) \rangle, \tag{19}$$

where the angle brackets denote averaging over the ensemble.

7. CONCLUSION

Having considered the various detection schemes, let us analyze the possibility of using the 2*e*2*ph* peak in the investigations of photon statistics. Unfortunately, the intuitive notion that the radiation of a light source possessing a large coherence time must contain a greater amount of photon pairs (and, hence, a larger 2*e*2*ph* to 1*e*1*ph* peak ratio) as compared to a source with a shorter coherence time is misleading. The above analysis indicates that the number of 2*e*2*ph* pulses counted over a time interval T is determined by the average square intensity of light over the same period. The difference in the number of 2*e*2*ph* counts for two stabilized light sources possessing equal average intensities but different coherence times $T_{c1} \neq T_{c2}$ vanishes in the limit

$$\frac{T_{c1, c2}}{T} \longrightarrow 0$$

(see [13]). Since the coherence times of the great majority of light sources do not exceed 1 ns, while the data accumulation times in the experiments analogous to those described above amount to hundreds of seconds and above, no differences between the sources of two types must be observed in the experiments.

This conclusion was thoroughly verified by comparing the pulse amplitude distributions for the light sources of two types possessing significantly different coherence times. The measurements were performed for a green neon line (5400 Å) and a quasimonochromatic radiation with the same wavelength obtained by passing the light of an incandescent lamp through a grating monochromator. It was found that the two distributions coincide to within the experimental accuracy. Any comparison of two different light sources always involves a risk of possible variation of some poorly controlled parameters (such as the character of the photocathode illumination). An alternative approach to checking for the sensitivity of the PMT pulse amplitude distribution with respect to the light coherence is to compare the distributions obtained using the same source operating in the single-mode (coherent) and multimode (incoherent) regimes. Such experiments also confirmed coincidence of the pulse amplitude distributions in the two cases.

In order to reveal a nontrivial dependence of the experimental results on the coherence time of a light source, it is necessary to count the number of pulses over a short period of time T (small time window), the duration of which is comparable to the coherence time. If the light intensity is not too high, so that the probability for more than one PMT pulse to fall within the time window is negligibly small (since we are speaking of a single spatial mode, this condition is valid with large margins for the great majority of light sources except for lasers), the average number of pulses detected in a multiply repeated experiment with the same time window is

$$P_1(T) = \bar{N}(T) = (\alpha_1 \bar{I}(T) + \alpha_2 \overline{I^2}(T))T.$$

As the time window decreases, $T \longrightarrow T_c$, the average squared intensity $\overline{I^2}(T)$ would significantly change:

$$\frac{\overline{I^2}(T) - \bar{I}^2(T)}{\bar{I}^2(T)} \longrightarrow 1 \text{ for } T \longrightarrow 0$$

(see [13]). Thus, using a series of experiments with the same light source and the time window decreased from $T \gg T_c$ to $T \approx T_c$, it is possible to directly measure the coherence time T_c value. Another possibility is offered by comparison of different sources with close $\bar{I}$ and $\overline{I^2}$ values but significantly different coherence times. Modern electronic equipment can provide for a time window as small as 100 ps and below, while the use of a microchannel plate photodetector allows the pulse front position to be determined to within 20 ps. This makes it possible to employ light sources with coherence times on the picosecond scale. Advantages of the proposed method are (i) use of the same photomultiplier, (ii) the possibility of simultaneously determining both $\overline{I^2}$ (2*e*2*ph* peak) and $\bar{I}$ (1*e*1*ph* peak), and (iii) the fact that the results are independent of the light intensity distribution in the illuminated cathode area.

Thus, we have studied the two-quantum photoeffect and proposed a new method for measuring the quantum-statistics properties of the light field. The experiments were performed with an antimony–cesium photocathode, which allows the obtained data to be compared to the results reported previously [5, 6]. Generalizing the well-known Mandel formula, the theoretical model developed in this study allows the whole body of experimental data to be rationalized from a common standpoint, which is very important for practical application of the obtained results.

ACKNOWLEDGMENTS.

This study was supported by the Russian Foundation for Basic Research, project no. 01-02-16480.

REFERENCES

1. M. C. Teich, J. M. Schroeer, and G. J. Wolga, Phys. Rev. Lett. **13**, 611 (1964).
2. P. Bloch, J. Appl. Phys. **35**, 2052 (1964).
3. M. C. Teich and G. J. Wolga, Phys. Rev. **171**, 809 (1968).
4. M. C. Teich and P. Diament, J. Appl. Phys. **40**, 625 (1969).
5. V. V. Artem'ev, Radiotekh. Élektron. (Moscow) **9**, 756 (1964).
6. V. V. Artem'ev, Radiotekh. Élektron. (Moscow) **13**, 316 (1968).
7. D. N. Klyshko, Usp. Fiz. Nauk **166**, 613 (1996) [Sov. Phys. Usp. **39**, 573 (1996)].
8. D. N. Klyshko and A. N. Penin, Usp. Fiz. Nauk **152**, 653 (1987) [Sov. Phys. Usp. **30**, 716 (1987)].
9. M. Born and E. Wolf, *Principles of Optics* (Pergamon, Oxford, 1969; Nauka, Moscow, 1970).
10. R. Prescott, Nucl. Instrum. Methods **39**, 173 (1966).
11. A. N. Pertsev and A. N. Pisarevskiĭ, *Photomultipliers: One-Electron Characteristics and Applications* (Atomizdat, Moscow, 1971).
12. R. Loudon, *The Quantum Theory of Light* (Clarendon, Oxford, 1973; Mir, Moscow, 1976).
13. G. Bedard, J. C. Chang, and L. Mandel, Phys. Rev. **160**, 1496 (1967).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 93, No. 6, 2001, pp. 1178–1183.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 6, 2001, pp. 1362–1368.
Original Russian Text Copyright © 2001 by Basiev, Doroshenko, Osiko, Prokhorov.

NUCLEI, PARTICLES,
AND THEIR INTERACTION

Highly Efficient Cooperative Energy Transfer from Ho^{3+} and Tm^{3+} Ions to Ce^{3+} Ions in Crystals

T. T. Basiev*, M. E. Doroshenko, V. V. Osiko, and A. M. Prokhorov
Laser Materials and Technology Research Center of the General Physics Institute,
Russian Academy of Sciences, Moscow, 119991 Russia
**e-mail: basiev@lst.gpi.ru*

Received August 20, 2001

Abstract—A phenomenon of highly efficient cooperative energy transfer from Ho^{3+} and Tm^{3+} ions to two-particle ($2Ce^{3+}$) cooperative acceptors in crystals of solid solutions of $La_{1-x}Ce_xF_3$ is revealed. The rates of cooperative energy transfer in $Ho^{3+} \longrightarrow 2Ce^{3+}$, $Tm^{3+} \longrightarrow 2Ce^{3+}$, and $Tb^{3+} \longrightarrow 2Yb^{3+}$ systems are measured, as well as their dependence on the magnitude of the matrix elements of donor transition. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The conventional mechanism of nonradiative energy transfer takes place between two particles, i.e., a donor and an acceptor, only in the case of electronic or electronic–vibrational transitions resonance, which corresponds to nonzero value of the overlapping integral between donor fluorescence and acceptor absorption [1, 2]. This mechanism leads to a linear (with respect to the acceptor concentration) dependence of the quenching rate at the initial and kinetic stages of the process [3–5]. At the same time, the possibility of cooperative transfer from one atom (ion) with a higher energy to two atoms (ions) with a lower energy and vice versa has been discussed by researchers in the field of physics of electron [6] and spin [7–9] excitations of activated crystals since the 1950s. In spite of numerous attempts to detect and investigate such processes in optics, only a few attempts may be regarded as successful and reliable.

In the 1970s, due to the wide search for materials and schemes for visible lasers, much effort was concentrated on the investigation of the up-conversion processes [10–13], which resulted in detecting [14, 15] in highly concentrated and strongly excited crystals the nonradiative energy transfer from a pair of Yb^{3+} ions with an energy of about 10 000 cm^{-1} each, acting as a cooperative donor to a simple Tb^{3+} ion with an energy of about 20 000 cm^{-1}, serving as an acceptor of energy. The probability of such a cooperative process proved to be very low [13], only 2.4 s^{-1}, which is two–three orders of magnitude less than the probability of radiative decay of the levels involved. The quantum yield of this elegant physical process turned out to be less than 1%, and the process failed to find any practical application.

In recent years, in the physics of activated laser materials, one feels the need for moving from the near infrared to medium infrared range [16]. In view of this, the processes of splitting or multiplication of excitations accompanied by a decrease in energy (with an increase in the number of excitations) may be of special interest; these processes were investigated theoretically more than forty years ago.

The process of nonradiative cooperative quenching or cooperative down-conversion was first described in [17, 18], where the concentration dependence of the neodymium Nd^{3+} impurity ions quenching in single crystals of $La_{1-x}Ce_xF_3 : Nd$ solid solutions was investigated. The optical excitation energy is transferred from a single Nd^{3+} ion with the energy of (${}^4F_{3/2}$–${}^4I_{15/2}$) donor transition of 5000 cm^{-1} simultaneously to a pair of Ce^{3+} ions (cooperative acceptor) with twice lower ${}^2F_{5/2}$–${}^2F_{7/2}$ transition energy of approximately 2500 cm^{-1}. The rate of such a process in a self-activated crystal of $CeF_3 : Nd^{3+}$,

$$W(\mathrm{Nd} \longrightarrow 2\mathrm{Ce}) = 1500\ \mathrm{s}^{-1},$$

is much lower than the rates of conventional resonant energy transfer (10^4 to 10^8 s^{-1}) [19–21]; however, it exceeds the up-conversion rate ($2Yb \longrightarrow Tb$) by almost three orders of magnitude and competes with the radiative decay rates for neodymium $A(\mathrm{Nd}) = 1300\ s^{-1}$. Therefore, the quantum yield of cooperative transfer and down-conversion exceeds 50%, this being of considerable interest from the standpoint of practical applications.

The relatively low value of the rate of the processes of cooperative up and down-conversion ($2D \longrightarrow A$, $D \longrightarrow 2A$) may be associated with the small magnitude of the electronic square elements of reduced matrix $U^{(2)}$,

$U^{(4)}$, and $U^{(6)}$ of the respective donor transitions of terbium and neodymium (10^{-3} to 10^{-2}) treated in [16–18], which are much smaller than the transition matrix elements of Yb^{3+} and Ce^{3+} (0.3 to 1). To realize higher cooperative energy transfer rates, the process should involve donor transitions with larger values of electronic reduced matrix elements. Having analyzed several ions and transitions, we have selected the ${}^3F_4$–${}^3H_6$ transition of thulium ion (Tm^{3+}) and the ${}^5I_7$–${}^5I_8$ transition of holmium (Ho^{3+}), which are in good resonance with the absorption of a cooperative pair acceptor ($2Ce^{3+}$), because the energy gap for the above-mentioned transitions is approximately twice as large as the energy of ${}^5F_{7/2}$–${}^2F_{5/2}$ transition of Ce^{3+} (see Fig. 1). The respective values of reduced matrix square elements for the above-identified transitions for Tm^{3+} ion,

$$U^{(2)} = 0.249, \quad U^{(4)} = 0.118, \quad U^{(6)} = 0.608,$$

and for Ho^{3+} ion,

$$U^{(2)} = 0.0249, \quad U^{(4)} = 0.1344, \quad U^{(6)} = 1.5210,$$

have the same order of magnitude as the values of the reduced matrix square elements for Yb^{3+} and Ce^{3+} ions,

$$U^{(2)} = 0.35, \quad U^{(4)} = 0.69, \quad U^{(6)} = 0.93.$$

The investigated dependence of luminescence decay time of Tm^{3+} and Ho^{3+} ions on the concentration of Ce^{3+} ions revealed the efficient cooperative energy transfer processes Ho ⟶ 2Ce and Tm ⟶ 2Ce with the efficiency exceeding 90% and the process rates of $(0.8–2) \times 10^4$ s^{-1}, which are two orders of magnitude higher than the radiative decay probabilities of respective transitions.

2. EXPERIMENTAL RESULTS AND THEIR DISCUSSION

We have investigated crystals of solid solutions of $La_{1-x}Ce_xF_3$: Tm (0.5%) and $La_{1-x}Ce_xF_3$: Ho (0.5%) with different contents of Ce^{3+} ions (x = 0, 0.2, 0.4, 0.6, 0.8, 1.0) which substitute for the optically inactive La^{3+} ions synthesized by the Bridgeman–Stockbarger technique in a fluorinating atmosphere. As was pointed out previously [17], the proximity of the ionic radii of La^{3+} and Ce^{3+} to one another provided for the invariability of the lattice constant, the structure, and the optical properties of solid solution upon variation of x from zero to unity.

The kinetics of luminescence decay of ${}^3F_4$–${}^3H_6$ transitions of Tm^{3+} ion and ${}^5I_7$–${}^5I_8$ transitions of Ho^{3+} ion were measured upon excitation by broadband radiation of a LiF crystal laser with F_2^- color centers with a generation maximum in the range of 1.12–1.16 μm, and with a GGG : Nd^{3+} laser used as the pumping source. The excitation was performed to a level located above the test level (${}^3H_5$ for Tm^{3+} and ${}^5I_6$ for Ho^{3+}) and related to the luminescent level by fast multiphonon relaxation. The luminescence was registered by an MDR-2 monochromator and a liquid-nitrogen-cooled Ge : Au photoresistor. For kinetics with a long decay time, the luminescence signal was recorded also by a liquid-nitrogen-cooled PbS photoresistor having a higher sensitivity for wavelengths in the region of 2 μm and the response time an order of magnitude longer. The scheme of the measuring system including a high-speed (400 MHz) Tektronix TDS-380 digital oscilloscope connected to a personal computer, which allowed us to accumulate and store the measured fluorescent signal with an increase in the signal-to-noise ratio, is presented in Fig. 2.

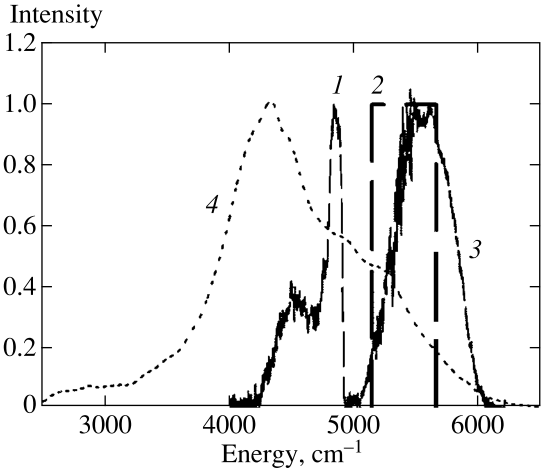


Fig. 1. The overlapping of luminescence spectra of transitions of Nd^{3+}, Ho^{3+}, and Tm^{3+} donor ions with virtual absorption spectrum of two-particle acceptors obtained by the convolution of spectra of one-particle acceptors: (*1*) luminescence spectrum of ${}^5I_7$–${}^5I_8$ transition of Ho^{3+} donor ion; (*2*) spectral region of luminescence of ${}^4F_{3/2}$–${}^4I_{15/2}$ transition of Nd^{3+} donor ion; (*3*) fluorescence spectra of ${}^3F_4$–${}^3H_6$ transition of Tm^{3+} donor ion; (*4*) absorption spectrum of $2Ce^{3+}$ two-particle cooperative acceptor, obtained from the absorption spectrum of Ce^{3+} one-particle acceptor as a convolution of the form

$$I(\nu_2) = \int_0^\infty I(\nu_1)I(\nu_2 - \nu_1)d\nu_1.$$

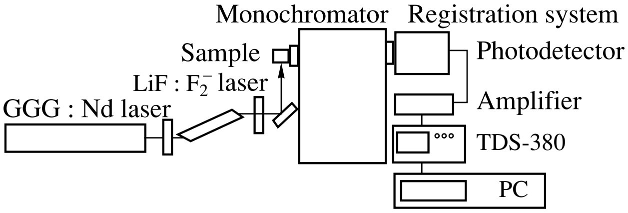


Fig. 2. The scheme of registration of luminescence decay kinetics.

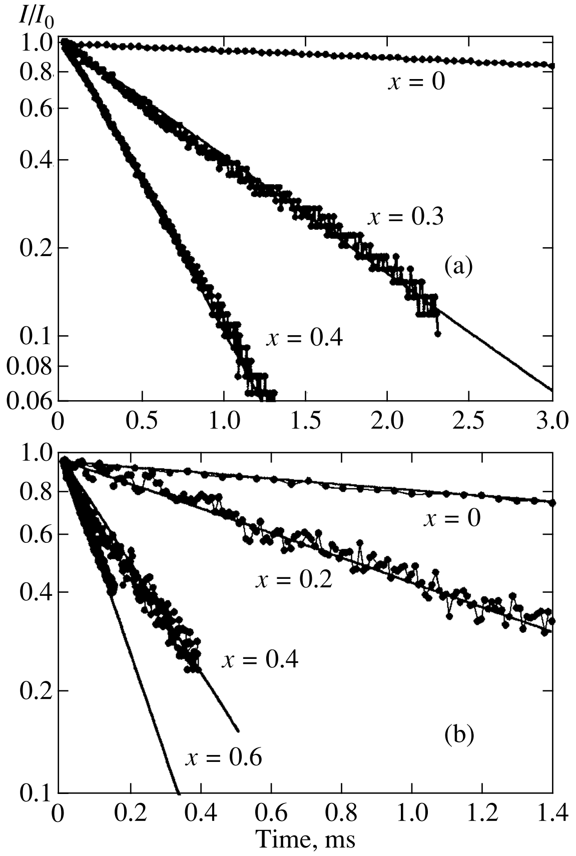


Fig. 3. Measured kinetics of luminescence decay of (a) levels $^5I_7$ of Ho^{3+} ion and (b) levels $^3F_4$ of Tm^{3+} ion in $La_{1-x}Ce_xF_3$ crystals for different concentrations x of Ce^{3+} ions.

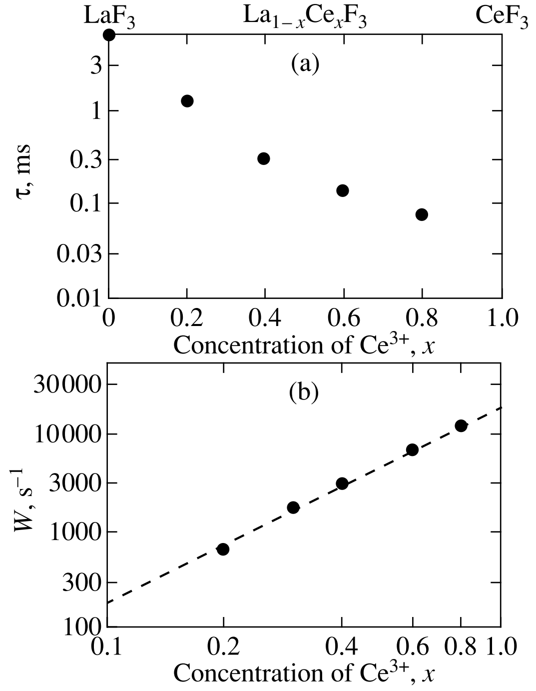


Fig. 4. (a) The initial decay time τ of the level $^3F_4$ of Tm^{3+} ion and (b) the cooperative quenching rate $W_{Tm \to 2Ce}$ as functions of the partial concentration x of Ce^{3+} ions in a $La_{1-x}Ce_xF_3$ crystal. The dots indicate the experimental results, and the dashes indicate the quadratic approximation of the experimental data.

Figure 3 gives examples of kinetics of luminescence decay of Ho^{3+} and Tm^{3+} ions at different concentrations x of Ce^{3+} ions. One can see that the decay time is reduced drastically with an increase in x and, at the initial stage, are well described by an exponential law, which corresponds to predictions and computer simulation of the process of cooperative quenching [22].

Figure 4a demonstrates the measured dependence of the initial time of decay $\tau(^5I_7)$ of the $^5I_7$ level of Ho^{3+} ion on the concentration x of Ce^{3+} ions. The corresponding rates of cooperative energy transfer, calculated from the measured decay times,

$$W_x(\mathrm{Ho} \longrightarrow 2\mathrm{Ce}) = \frac{1}{\tau(x)} - \frac{1}{\tau(0)}$$

are indicated in Fig. 4b by filled dots. For convenience, the graph of Fig. 4b is given on a log-log scale. The dashed line in Fig. 4b corresponds to quadratic approximation of the experimental data.

Figure 5a gives the measured decay time for the level $^3F_4$ of Tm^{3+} ion as a function of the concentration x of Ce^{3+} ions. As in the previous case, the filled dots in Fig. 5b correspond to the values of the rate of cooperative transfer,

$$W_x(\mathrm{Tm} \longrightarrow 2\mathrm{Ce}) = \frac{1}{\tau(x)} - \frac{1}{\tau(0)},$$

calculated from the measured values of decay time of the level $^3F_4$ of Tm^{3+} ions, and the dashed line indicates a quadratic approximation of the experimental data. The experimentally obtained values of the quenching rate fit well the quadratic dependence, $W \propto x^2$, whence it follows that the process of quenching of Tm^{3+} and Ho^{3+} ions by Ce^{3+} ions has a two-particle, cooperative nature. Here, it must be emphasized that $\tau(x)$ and $W(x)$ were measured at the initial stage of the kinetics of donor luminescence decay, where the conventional Forster energy transfer exhibits a linear dependence, $W \propto x$ [2–5, 16–18]. The data on the quenching rates, of different rare-earth ions (donors) by Ce^{3+} ions are summarized in the table for two values of the concentration of Ce^{3+} ions ($x = 0.4$ and 1.0). One can see in the table that the value of the rate of cooperative quenching in the case of Tm ⟶ 2Ce and Ho ⟶ 2Ce processes exceeds considerably the respective value for the case of Nd ⟶ 2Ce transition. This correlates with much higher values of the reduced matrix square elements $U^{(2)}$, $U^{(4)}$, and $U^{(6)}$, which define the line strength of the investigated transitions of Tm^{3+} and Ho^{3+} ions. It follows from the table that the energy efficiency of the processes of quenching of Tm^{3+} and Ho^{3+} ions and of sensitization of Ce^{3+} ions,

$$\eta = \frac{W(x)}{W(x)+A} \times 100\% = W(x)\tau(x) \times 100\%,$$

is close to 100% in both cases of the treated cooperative transfer, Tm $\longrightarrow$ 2Ce and Ho $\longrightarrow$ 2Ce, and is even somewhat higher than in the case of direct resonant transfer Er $\longrightarrow$ Ce [16]. The quantum yield of sensitization and down-conversion for the cooperative processes of Tm $\longrightarrow$ 2Ce and Ho $\longrightarrow$ 2Ce is higher than 100% (~190%) as a result of sharing the excitation. An analysis of the data given in the table enables one to compare the rates of overall energy transfer $W_{x=1}$ from the donor to Ce^{3+} ions in the CeF_3 lattice for three cases, namely, when the Ce^{3+} ions act as one-particle acceptors, two-particle cooperative acceptors, and three-particle cooperative acceptors. Indeed, the matrix elements of the $^4S_{3/2}$–$^4F_{9/2}$ transition of Er^{3+} ion are close in magnitude to the matrix elements of the $^4F_{3/2}$–$^4I_{15/2}$ transition of Nd^{3+} ion. This comparison leads one to conclude that the probability of transfer to two-particle cooperative acceptors is two to three orders of magnitude lower compared to conventional transfer to one-particle acceptors,

$$W_1(\mathrm{Nd} \longrightarrow 2\mathrm{Ce}) = 5 \times 10^{-3} W_1(\mathrm{Er} \longrightarrow \mathrm{Ce}).$$

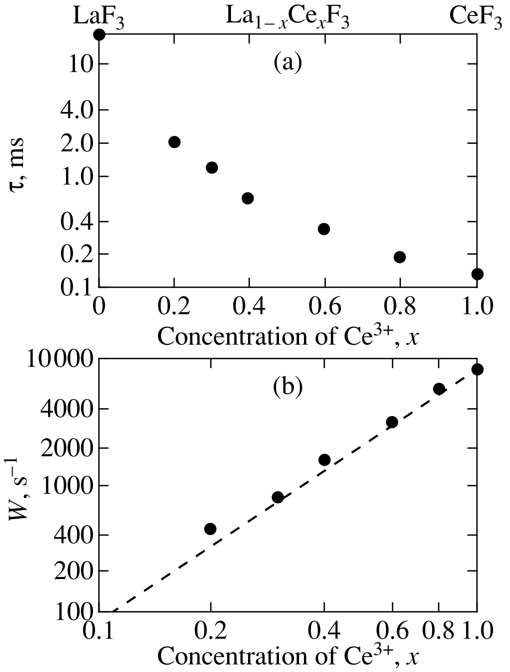


Fig. 5. (a) The initial decay time τ of the level $^5I_7$ of Ho^{3+} ion and (b) the cooperative quenching rate $W_{Ho \to 2Ce}$ as functions of the partial concentration x of Ce^{3+} ions in a $La_{1-x}Ce_xF_3$ crystal. The dots indicate the experimental results, and the dashes indicate the quadratic approximation of the experimental data.

The similar equality between the matrix elements of the transitions $^4I_{13/2}$–$^4I_{15/2}$ of Er^{3+} ions and transitions $^5I_7$–$^5I_8$ of Ho^{3+} ions enables one to compare the probabilities of

Table

Donor ion (transition)	Reduced matrix square elements of donor electronic transition	Acceptor	Reduced matrix square elements of acceptor electronic transition	Rate of transfer W [s^{-1}] and energy efficiency of sensitization $\eta = W(x)\tau(x) \times 100\%$ for CeF_3	
				$x = 0.4$	$x = 1$
Er^{3+} ($^4I_{13/2}$–$^4I_{15/2}$)	$U^{(2)} = 0.0195$ $U^{(4)} = 0.11$ $U^{(6)} = 1.43$	3Ce	$U^{(2)} = 0.12$ $U^{(4)} = 0.41$ $U^{(6)} = 0.86$	9.2 (9.2%)	1.2×10^2 (54%)
Tb^{3+} ($^5D_4$–$^7F_6$)	$U^{(2)} = 0.0009$ $U^{(4)} = 0.0008$ $U^{(6)} = 0.0013$	2Yb	$U^{(2)} = 0.12$ $U^{(4)} = 0.41$ $U^{(6)} = 0.86$		1.4×10^2 (14%)
Nd^{3+} ($^4F_{3/2}$–$^4I_{15/2}$)	$U^{(2)} = 0$ $U^{(4)} = 0$ $U^{(6)} = 0.0275$	2Ce	$U^{(2)} = 0.12$ $U^{(4)} = 0.41$ $U^{(6)} = 0.86$	3.5×10^2 (19.9%)	1.5×10^3 (52%)
Tm^{3+} ($^3F_4$–$^3H_6$)	$U^{(2)} = 0.249$ $U^{(4)} = 0.118$ $U^{(6)} = 0.608$	2Ce	$U^{(2)} = 0.12$ $U^{(4)} = 0.41$ $U^{(6)} = 0.86$	3.1×10^3 (94.3%)	2×10^4 (99.96%)
Ho^{3+} ($^5I_7$–$^5I_8$)	$U^{(2)} = 0.0249$ $U^{(4)} = 0.1344$ $U^{(6)} = 1.5210$	2Ce	$U^{(2)} = 0.12$ $U^{(4)} = 0.41$ $U^{(6)} = 0.86$	2.3×10^3 (96.8%)	8×10^3 (99.98%)
Er^{3+} ($^4S_{3/2}$–$^4F_{9/2}$)	$U^{(2)} = 0$ $U^{(4)} = 8 \times 10^{-5}$ $U^{(6)} = 0.0228$	1Ce	$U^{(2)} = 0.12$ $U^{(4)} = 0.41$ $U^{(6)} = 0.86$	1.24×10^5 (89.3%)	2.7×10^5 (97.2%)

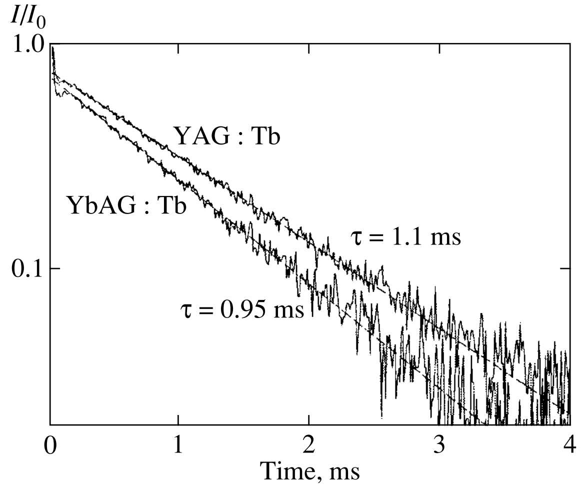


Fig. 6. Measured kinetics of luminescence decay of the level $^5D_4$ of Tb^{3+} ions in $Y_3Al_5O_{12}$ and $Yb_3Al_5O_{12}$ garnet crystals.

cooperative quenching energy transfer to three-particle acceptors with transfer to two-particle cooperative acceptors,

$$W_1(\mathrm{Er} \longrightarrow 3\mathrm{Ce}) = 1.5 \times 10^{-2} W_1(\mathrm{Ho} \longrightarrow 2\mathrm{Ce}),$$

whose probability likewise proved to be two orders of magnitude lower. Note that, so far, we have not taken into consideration the differences between the overlapping integrals and between the lattice sums of the processes being analyzed, which calls for additional investigations.

As a result of our detection of the heretofore unknown anomalously high rates of cooperative process ($10^3 - 2 \times 10^4$ s^{-1}), which are two orders of magnitude higher than the rates of radiative relaxation and three orders of magnitude higher than the rates of cooperative up-conversion, the question arose of the need to verify, using a more direct technique, the value of the rate of cooperative process of 2Yb ⟶ Tb equal to W = 2.4 s^{-1} [13] (estimated from indirect data by Ostermayer and Van Vitert [15]).

We have experimentally measured the inverse process of cooperative energy transfer ($Tb^{3+} \longrightarrow 2Yb^{3+}$) from one Tb^{3+} ion to a two-particle cooperative acceptor, namely, two Yb^{3+} ions in garnet crystals permitting 100% replacement of optically inactive Y^{3+} ions by active Yb^{3+} ions without distortion of the crystal lattice. The kinetics of luminescence decay of the level $^5D_4$ of Tb^{3+} ion were investigated in $Y_3Al_5O_{12}$: Tb (2%) and $Yb_3Al_5O_{12}$: Tb (2%) crystals, which corresponds to zero (x = 0) and 100% (x = 1) concentration of Yb^{3+} ions acting as cooperative acceptors. Yb^{3+} ions represent a peculiar kind of analog of Ce^{3+} ions in the series of rare-earth elements. They also possess one optical transition $^2F_{5/2}$–$^2F_{7/2}$ on the 4f subshell with the same magnitudes of electronic square elements of the reduced matrix as in the case of the $^2F_{5/2}$–$^2F_{7/2}$ transition of Ce^{3+} ion,

$$U^{(2)} = 0.12, \quad U^{(4)} = 0.41, \quad U^{(6)} = 0.86.$$

The measuring scheme was similar to that given in Fig. 2 and differed only in that a pulsed nitrogen laser with a lasing wavelength of 337 nm was used to excite the level $^5D_4$, and the luminescence was recorded by a FEU-79 photomultiplier. The measured fluorescence decay curves are given in Fig. 6. One can see in Fig. 6 that the kinetics of luminescence decay are single-exponential, with the respective lifetimes of $\tau(x = 0)$ = 1.1 ms and $\tau(x = 1)$ = 950 μs. The maximal rate of overall cooperative quenching, calculated from the measured lifetimes,

$$W_1(\mathrm{Tb} \longrightarrow 2\mathrm{Yb}) = \frac{1}{\tau(x=1)} - \frac{1}{\tau(x=0)}$$

was approximately 143.5 s^{-1}, which is much lower than the rate of cooperative quenching W_1(Ho ⟶ 2Ce) and W_1(Tm ⟶ 2Ce) but almost two orders of magnitude higher than the theoretical estimate of the rate of cooperative up-conversion, W_0(2Yb ⟶ Tb) = 2.4 s^{-1}, obtained in [13].

The physical meaning of the difference between the rates of cooperative up and down processes with summation and splitting of excitation consists apparently in the difference between the numbers of final states of the system to which it changes as a result of cooperative process. For a low concentration of impurity with a high-frequency transition (Tb) and for a high concentration of particles with low-frequency transitions (Yb), there is only one final state in the case of up-conversion of 2Yb ⟶ Tb, and the experiment in fact involves the measurement of an elementary act of transfer which has a fairly low probability (2.4 s^{-1}) due to the smallness of the matrix elements of the transition of Tb^{3+} ion (see table). In the mode of down-conversion, the number of final states of cooperative acceptors at high values of x may considerably exceed unity and, for a crystal of $Yb_3Al_5O_{12}$ or CeF_3 (x = 1), may reach several hundred (the number of pair combinations of the acceptor ions in the coordination spheres nearest to the donor).

3. CONCLUSIONS

Phenomena have been revealed of highly efficient cooperative energy transfer from thulium and holmium ions (donors) to two-particle cooperative acceptors (cerium ions) in crystals of lanthanum–cerium fluorides at rates of the order of 10^4 s^{-1}, exceeding those of radiative rates by several orders of magnitude. Nonlinear (quadratic) laws of cooperative concentration quenching Ho ⟶ 2Ce and Tm ⟶ 2Ce have been found experimentally. The rate of cooperative quenching energy transfer from a single Tb^{3+} ion to two-particle cooperative acceptor, namely Yb^{3+} ions, in $Yb_3Al_5O_{12}$ garnet crystals was measured. Rates of cooperative energy transfer with multiplication of excitations have been found which are two–three orders of magnitude higher compared to the known rates of cooperative sensitization and up-conversion.

REFERENCES

1. Th. von Förster, Ann. Phys. **2**, 53 (1948).
2. V. M. Agranovich and M. D. Galanin, *Electronic Excitation Energy Transfer in Condensed Matter* (Nauka, Moscow, 1978; North-Holland, Amsterdam, 1982).
3. S. I. Golubov and D. V. Konobeev, Fiz. Tverd. Tela (Leningrad) **13**, 3185 (1971) [Sov. Phys. Solid State **13**, 2679 (1972)].
4. V. P. Sakun, Fiz. Tverd. Tela (Leningrad) **14**, 2199 (1972) [Sov. Phys. Solid State **14**, 1906 (1973)].
5. A. I. Burshtein, J. Lumin. **34**, 167 (1985).
6. D. L. Dexter, Phys. Rev. **108**, 630 (1957).
7. W. B. Mims and J. D. Mc Gee, Proc. IRE **47**, 2120 (1959).
8. A. A. Manenkov and A. M. Prokhorov, Zh. Éksp. Teor. Fiz. **42**, 75 (1962) [Sov. Phys. JETP **15**, 54 (1962)].
9. A. M. Prokhorov and V. B. Fedorov, Zh. Éksp. Teor. Fiz. **46**, 1937 (1964) [Sov. Phys. JETP **19**, 1305 (1964)].
10. P. P. Feofilov and V. V. Ovsyankin, Appl. Opt. **6**, 1828 (1967).
11. F. Auzel, C. R. Seances Acad. Sci. **263**, 819 (1966).
12. V. V. Ovsyankin and P. P. Feofilov, Izv. Akad. Nauk SSSR, Ser. Fiz. **37**, 262 (1973).
13. T. Kushida, J. Phys. Soc. Jpn. **34**, 1318 (1973); **34**, 1327 (1973); **34**, 1334 (1973).
14. L. D. Livanova, I. G. Saitkulov, and A. L. Stolov, Fiz. Tverd. Tela (Leningrad) **11**, 918 (1969) [Sov. Phys. Solid State **11**, 750 (1969)].
15. F. W. Ostermayer and L. G. van Uitert, Phys. Rev. B **1**, 4208 (1970).
16. T. T. Basiev, M. E. Doroshenko, V. V. Osiko, *et al.*, in *Technical Digest of Advanced Solid-State Lasers Conference, 2001*, p. 93.
17. T. T. Basiev, M. E. Doroshenko, and V. V. Osiko, OSA TOPS **34**, 485 (2000).
18. T. T. Basiev, M. E. Doroshenko, and V. V. Osiko, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 14 (2000) [JETP Lett. **71**, 8 (2000)].
19. O. K. Alimov, M. Kh. Ashurov, T. T. Basiev, E. O. Kirpichenkova, and V. B. Murav'ev, in *Selective Laser Spectroscopy of Activated Crystals*, Ed. by V. V. Osiko (Nova Science, New York, 1990).
20. Yu. V. Orlovskii, T. T. Basiev, I. N. Vorob'ev, *et al.*, Laser Phys. Int. J. **6**, 448 (1996).
21. T. T. Basiev, Yu. V. Orlovskii, and Yu. S. Privis, J. Lumin. **69**, 187 (1996).
22. T. T. Basiev, I. T. Basieva, M. E. Doroshenko, *et al.*, J. Lumin. (2002) (in press).

Translated by H. Bronstein

Journal of Experimental and Theoretical Physics, Vol. 93, No. 6, 2001, pp. 1184–1189.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 6, 2001, pp. 1369–1374.
Original Russian Text Copyright © 2001 by Vaulina, Nefedov, Petrov, Samaryan, Fortov.

PLASMA,
GASES

Self-Oscillations of Macroparticles in the Dust Plasma of Glow Discharge

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

Institute of High Energy Densities, Institute of High-Temperatures Scientific Association (IVTAN),
Russian Academy of Sciences, Izhorskaya ul. 13/19, Moscow, 127412 Russia
**e-mail: ipdustpl@redline.ru*

Received April 11, 2001

Abstract—Two types of instabilities emerging in dust-plasma systems with a spatial gradient of the macroparticle charge are considered. It is shown that the change in the macroparticle charge is an effective mechanism for exciting dust self-oscillations in a laboratory plasma. The results of experimental observations and an analysis of the conditions for the development of various self-oscillations of macroparticles in the strata of a dc glow discharge are presented. *© 2001 MAIK “Nauka/Interperiodica”.*

The phenomena associated with the development of instabilities in dissipative systems of interacting particles are important for various branches of science such as plasma physics, biophysics, and astronomy. A laboratory dust plasma is a wonderful experimental model for studying instabilities in such systems. The dust plasma consists of electrons, ions, neutral gas molecules, and charged macroparticles of a micrometer size. Owing to the high mobility of electrons, nonemitting dust particles acquire an equilibrium negative charge matching with the parameters of the surrounding plasma. This charge may be a function of time and the position of the particle in a plasma with varying parameters.

A laboratory dust plasma is an open dissipative system. The combined effect of dissipation and other processes in such a plasma may lead to the emergence of stable stationary structures of macroparticles (similar to a liquid or a solid) as well as complex vibrational or chaotic modes [1–6]. The formation of self-oscillations in dust-plasma systems differs significantly from the formation of oscillations in conservative systems since such systems display dissipative energy losses, and energy “pumping” with the help of some mechanisms may also take place in them. One of the possible mechanisms that can convert the potential energy of an external electric field into the kinetic energy of macroparticles is associated with space–time variations of dust charges [4–6]. Random charge fluctuations of macroparticles which are always present in a dust plasma may cause “anomalous heating” of the dust, but cannot explain the self-excitation of regular self-oscillations of macroparticles without additional sources compensating for dissipative losses. In this work, we will analyze regular spatial variations of macroparticle charges emerging due to inhomogeneity of the bulk plasma surrounding a dust cloud, e.g., because of the gradients of concentrations $n_{e(i)}$ or temperatures $T_{e(i)}$ of electrons (ions). Similar conditions are often realized in the plasma of an inductive high-frequency or glow discharge [7, 8].

Let us consider the motion of N_p macroparticles having the charge

$$Z(\rho, y) = Z_0 + \Delta Z(\rho, y),$$

where $\rho = (x^2 + z^2)^{1/2}$ is the radial coordinate of a particle, in the electric field

$$\mathbf{E}(\rho, y) = \mathbf{i}E(y) + \mathbf{j}E(\rho)$$

of a cylindrical trap, taking into account the pair interaction between particles (F_{int}), the force of gravity ($m_p g$), the thermophoretic force ($F_{\text{th}} \equiv F_{\text{th}}(\rho)$), and the Brownian movement of particles (F_{br}):

$$\begin{gathered} m_p \frac{d^2\mathbf{r}_k}{dt^2} = \sum_j F_{\text{int}}(r)\Big|_{r = |\mathbf{r}_k - \mathbf{r}_j|} \frac{\mathbf{r}_k - \mathbf{r}_j}{|\mathbf{r}_k - \mathbf{r}_i|} \\ - m_p \nu_{\text{fr}} \frac{d\mathbf{r}_k}{dt} + \mathbf{F}_{\text{br}} + \mathbf{F}_{\text{ext}}. \end{gathered} \tag{1}$$

Here, r is the distance between particles, m_p is the macroparticle mass, ν_{fr} is the coefficient of friction,

$$F_{\text{int}}(r) = -eZ(\rho, y)\frac{\partial \phi_D}{\partial r},$$

$$\phi_D = \frac{eZ(\rho, y)}{r}\exp\left(-\frac{r}{D}\right)$$

is the screened Coulomb potential with the screening length D, and e is the electron charge. The external force is defined by the relation

$$\begin{gathered} \mathbf{F}_{\text{ext}} = \mathbf{i}\{E(y)eZ(\rho, y) - m_p g\} \\ + \mathbf{j}\{E(\rho)eZ(\rho, y) + F_{\text{th}}\}, \end{gathered}$$

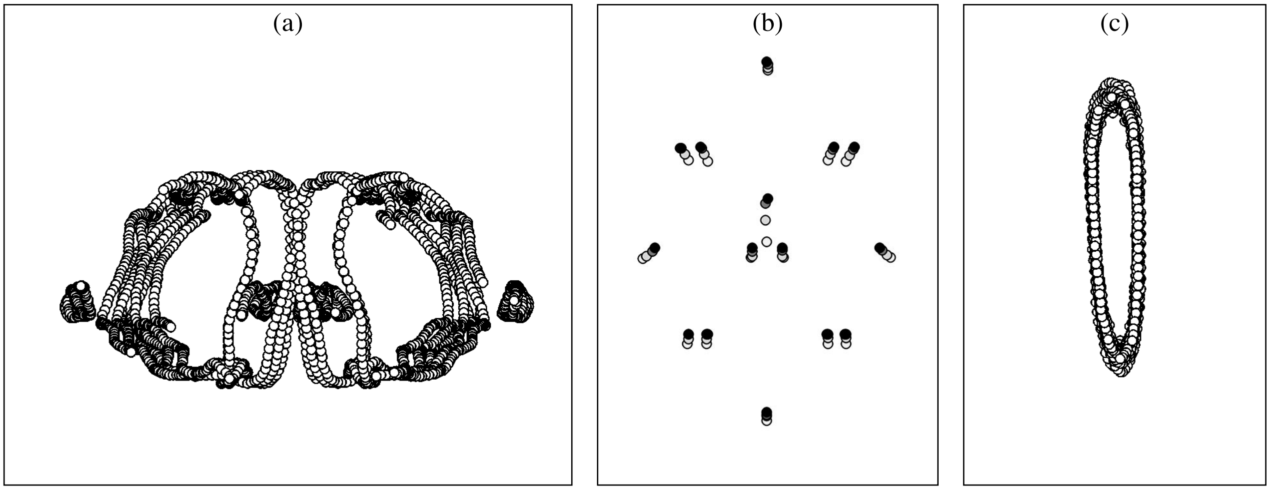


Fig. 1. Trajectories of macroparticles for the development of (a) type I and (b) type II instabilities in small dust clusters and (c) the enlarged fragment of the trajectory of an individual particle in the case (b).

where

$$F_{\text{th}} = -\frac{32}{15}\sqrt{\frac{\pi m}{8T}}a^2\lambda\frac{\partial T}{\partial \rho},$$

T and m are the temperature and mass of gas molecules, λ is the thermal conductivity, and a is the radius of a macroparticle. Thus, the force of interaction between particles and the external force acting on a particle are found to be functions of its coordinate. In the case when the curl of these forces is not equal to zero, such a system may do a positive work compensating the dissipative energy losses. This means that infinitesimal perturbations emerging in the system due to thermal or other fluctuations may increase. A detailed analysis of the linearized equations (1) is presented in [6]. This analysis proved that small perturbations in system (1) increase in the following two cases: (1) in the absence of the retrieving force (type I instability) and (2) in the vicinity of a certain characteristic resonance frequency ω_c of the system, where the friction force does not suppress oscillatory motion (type II instability).

In the case when $Z_0 \gg \Delta Z(\rho, y)$, the emerging type II instability is determined by the condition

$$\omega_c^4 < \left|\frac{\gamma_0(\beta_\rho g - \beta_y F_{th}/m_p)}{Z_0}\right|, \qquad (2)$$

where

$$\beta_\rho = \frac{\partial Z(\rho, y)}{\partial \rho}, \quad \beta_y = \frac{\partial Z(\rho, y)}{\partial y},$$

and γ_0 is the displacement parameter [6]. Since the curl of the macroparticle velocity **V** differs from zero,

$$\mathbf{\Omega} = \text{curl}\mathbf{V} \neq 0,$$

relation (2) describes the emergence of a vortex flow along a certain closed curve. The direction of this rotation in a plane parallel to the gravity field for monotonic spatial dependences $\mathbf{E}(\rho, y)$ and $Z(\rho, y)$ can be determined from the sign of

$$\Omega = \frac{g\beta_\rho^2 - \beta_y F_{\text{th}}/m_p}{Z_0\nu_{\text{fr}}}.$$

For example, macroparticles fall to the center of the dust cloud for $\Omega > 0$ and ascend for $\Omega < 0$.

Dispersive type II instability differs from dissipative type I bifurcation in a strong dependence on the frictional force whose suppression may give rise to the development of perturbations with frequencies close to a certain natural resonance frequency ω_c of the system. The criterion for the emergence of the type II instability can be written in the form

$$\nu_{fr} < \omega_c < |\Omega|/2. \qquad (3)$$

In the case of strong dispersion, virtually only one mode "survives" as a result of the evolution of the type II instability, and the steady-state motion has the form of a harmonic wave with a frequency close to the bifurcation point of the system:

$$\omega_c \approx |\Omega|/2.$$

It should be noted, however, that collective wave movements are possible only in the case of a synchronized motion of individual particles in the dust cloud [6]. Figure 1 shows the results of numerical simulation of the development of type I and type II instabilities for small dust clusters ($N_p = 15$) of charged particles.

Thus, the theoretical analysis and numerical simulation show that various self-oscillations of dust may be excited in a plasma with gradients of charges of macroparticles. Using the proposed model, we will analyze the conditions of excitation of self-oscillations observed in the strata of a dc glow discharge (see Figs. 2–4). The experiments were made with iron particles with an

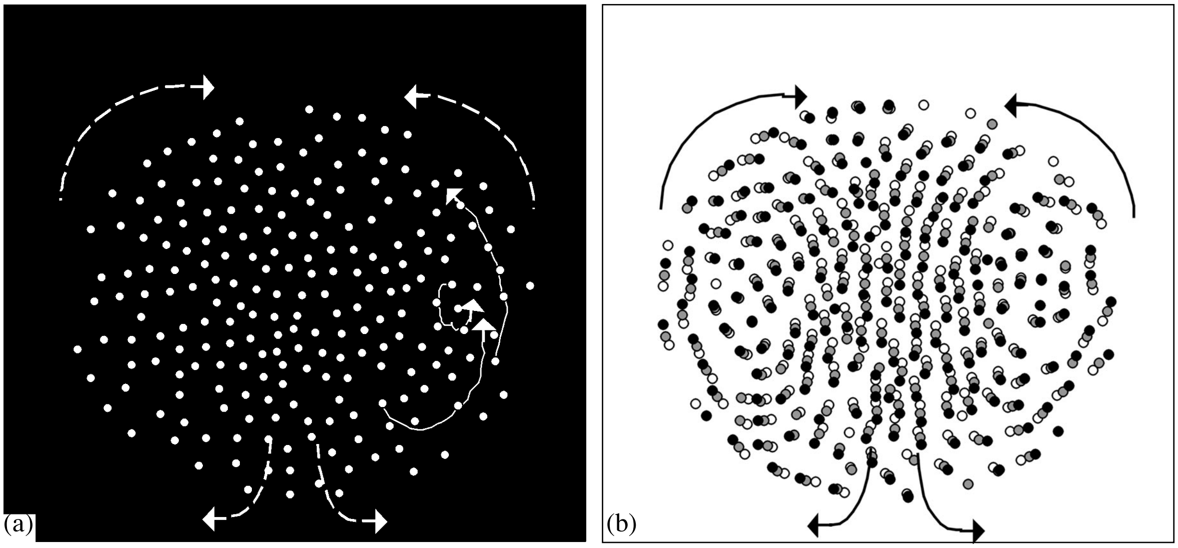


Fig. 2. (a) Schematic diagram of convection of macroparticles in the stratum of a glow discharge and (b) the results of numerical simulation.

average radius $a \approx 3.5$ μm in argon with various discharge currents $I = 0.5$–15 mA under pressures $P = 0.1$–1 torr.

In the experimental conditions under consideration, the frictional frequency can be written in the free-molecular approximation, $\nu_{fr}[c] \approx 30P$ [torr] [9], and the equilibrium charge of dust particles can be presented in the form $\langle Z\rangle \approx 7 \times 10^3 T_e$ eV for $n_e = n_i$ and $T_i \approx 0.03$ eV [10]. Assuming that the gas is cooled at the discharge tube walls and that $\partial T/\partial\rho < 0.5$ K/cm, we find that, for $I < 0.2$ A and $\rho < R/10$ (here $R = 3.25$ cm is the radius of the discharge tube) [11], the ratio

$$\frac{m_p g \beta_\rho}{\beta_y F_{th}} = \frac{125\ \text{K/cm} \times \beta_\rho}{\beta_y \partial T/\partial\rho}$$

is much greater than unity for $\beta_\rho > \beta_y/10$. Consequently,

$$\Omega = \frac{g\beta_\rho}{Z_0 \nu_{fr}},$$

and the value of β_y actually does not affect the evolution of the self-excited motion. However, the effect of $\beta_y F_{th}$ may be significant for smaller particles as well as in the case of an inductive HF or anomalous glow discharge for which higher gradients $\partial T/\partial\rho$ of gas temperatures can be observed.

Under the assumption that

$$\partial T_e/\partial\rho = 0, \quad \delta n = n_e - n_i \ll n_e \approx n_i \approx n_0,$$
$$\Delta Z(\rho, y) \ll Z_0 \approx \langle Z\rangle,$$

the value of β_ρ is determined by the gradient of δn and can be obtained from the balance equation for ion and electron flows at the surface of a dust particle [9]:

$$\beta_\rho \approx \frac{\langle Z\rangle}{n_0(1 + e^2\langle Z\rangle/aT_e)} \frac{\partial \delta n}{\partial\rho}. \qquad (4)$$

Here,

$$\delta n \equiv -\frac{\operatorname{div}\mathbf{E}(\rho, y)}{4\pi e}.$$

For a diffusion-controlled discharge, the radial polarization field is given by

$$E(\rho) \approx -\frac{T_e}{en_e}\frac{\partial n_e}{\partial\rho} \sim \frac{2.4T_e}{eR} \quad (R \gg D)$$

(see [11]). The joint solution of the Poisson and diffusion equations with Eq. (4) gives $\beta_\rho > 0$ and $\beta_\rho/Z_0 \approx 1.8D^2/R\rho^2 \approx 0.005$–$0.5$ cm^{-1} for $\rho/D \sim 1$–10. In this case, the ratio

$$\left|\frac{\Delta Z(\rho, y)}{Z_0}\right| \approx \left|-\frac{1.8D^2}{R\rho}\right| < 5\%$$

for $\rho > D$ and $D < 1000$ μm. Thus, the condition

$$\Delta Z(\rho, y) \ll Z_0$$

holds. It should be noted that the available experimental approximations

$$\phi(\rho) \propto \rho^\alpha \quad (1 < \alpha < 2)$$

of the radial electric field of a glow discharge also lead to a decrease in β_ρ with increasing ρ [5].

Figure 2a illustrates the convection of macroparticles observed in a dust cloud. The figure shows the positions of particles in the cloud and the trajectories of motion of three particles (the exposure time $t_{exp} \approx 4$ s)

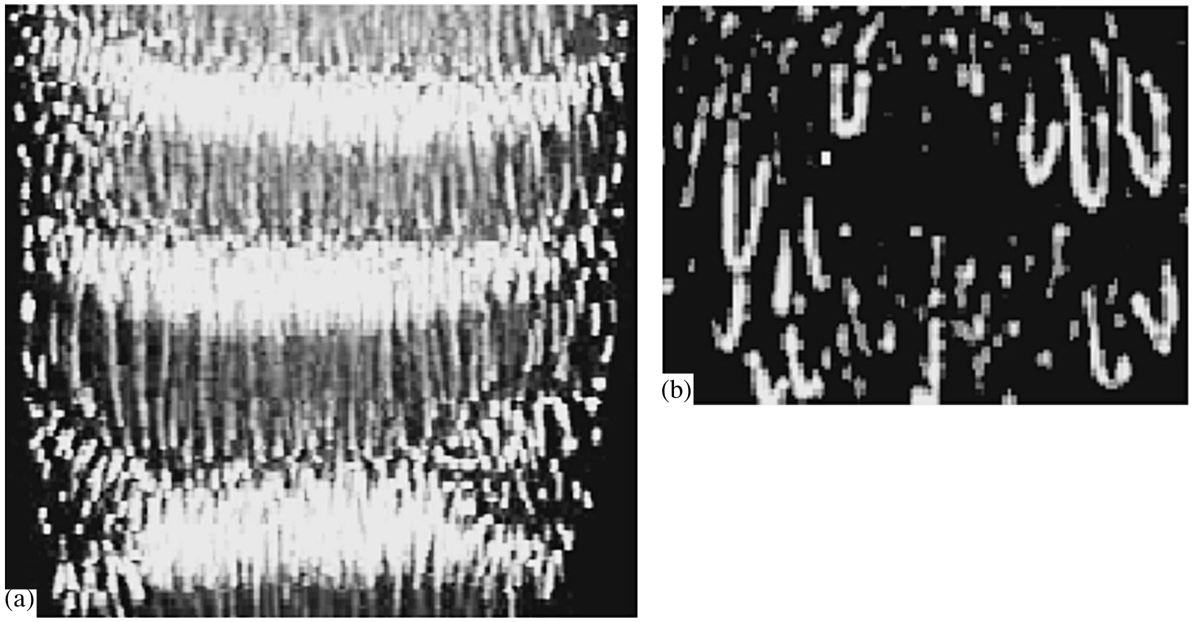


Fig. 3. (a) Videorecording of acoustic vibrations of dust and (b) magnified fragment of trajectories of motion of individual particles.

obtained as a result of computer processing of videofilms of the experiment using a special program capable of identifying the displacement of individual particles. A more detailed description of the processing of video recordings is given in [12].

The direction of dust rotation (see Fig. 2a) indicates that $\Omega > 0$ ($\beta_\rho > 0$). An experimental estimate of β_ρ can be obtained from the measurements of the velocity $V \approx A\Omega/2$ of particles and the amplitude A of their rotation. Thus, we obtain

$$\beta_\rho/Z_0 = 2V\nu_{\mathrm{fr}}/Ag \approx 0.012\ \mathrm{cm}^{-1},$$

where $\nu_{\mathrm{fr}} \approx 24\ \mathrm{s}^{-1}$ ($P = 0.8$ torr), which corresponds to theoretical predictions. The results of numerical simulation of the problem (1) with parameters close to experimental values are presented in Fig. 2b ($N_p = 3000$, $t_{\mathrm{exp}} \approx 0.5$ s). The mean kinetic values of particle energy ($\langle K\rangle \approx 4.5 \times 10^{-18}$ J) were close to the experimentally observed values.

Longitudinal vibrations of dust are presented in Fig. 3a. In spite of the fact that the collective motion of dust is a wave motion, individual particles move in the cloud in small-scale elliptical trajectories (Fig. 3b) which are similar to those being simulated (see Fig. 1c). It can be easily seen that self-oscillations are suppressed at the boundaries of the dust cloud. This can be due to the above-mentioned decrease in β_ρ in the radial direction. The frequency of oscillations at the center of the cloud was $\omega \approx 45$–$50\ \mathrm{s}^{-1}$. Assuming that $\omega \approx g\beta_\rho/2Z_0\nu_{\mathrm{fr}}$ (3) and $\nu_{\mathrm{fr}} \approx 9\ \mathrm{s}^{-1}$ ($P = 0.3$ torr), we obtain $\beta_\rho/Z_0 \approx 0.8$–$0.9\ \mathrm{cm}^{-1}$. The difference between the obtained value of β_ρ/Z_0 and numerical estimates may be due to the possible losses of the plasma component at macroparticles of the dust cloud, which may lead to an increase in the electric polarization field $E(\rho)$ (and, accordingly, in the value of β_ρ.

The rotation of macroparticles in the upper part of the dust cloud combined with oscillations or acoustic vibrations at the base of the dust structure is typical of glow discharge [8]. The proposed mechanism can easily explain such a combined motion only on the basis of the assumptions concerning different changes in the macroparticle charge in different regions of the gas discharge. For example, the observed complex motion (see Fig. 4) can be regarded as simultaneous development of two different instabilities of types I and II for different parts of the dust cloud. Assuming that

$$2\pi e^2 Z_0 \exp(-k)/l^2 \approx E,$$

where

$$E \approx 2.4T_e/eR, \quad k = l/D,$$

and l is the separation between the particles, we can eliminate the unknown T_e ($Z_0 \sim T_e$) and estimate k for different parts of the cloud. This gives $D \approx 1000\ \mu$m for the base of the dust cloud ($k \approx 0.8$ and $l_p \approx 800\ \mu$m). The observed frequency of oscillations was $\omega \sim 30$–$40\ \mathrm{s}^{-1}$. This leads to the following experimental estimate for the gradient of particle charge:

$$\beta_\rho/Z_0 \approx 2\nu_{\mathrm{fr}}\omega/g \approx 0.27\text{–}0.37\ \mathrm{cm}^{-1}$$

$$\nu_{\mathrm{fr}} \approx 4.5\ \mathrm{s}^{-1}, \quad P = 0.15\ \text{torr}.$$

The screening length D in the upper part of the cloud ($k \approx 2$, $l_p \approx 450\ \mu$m) decreases to 225 μm. It should be

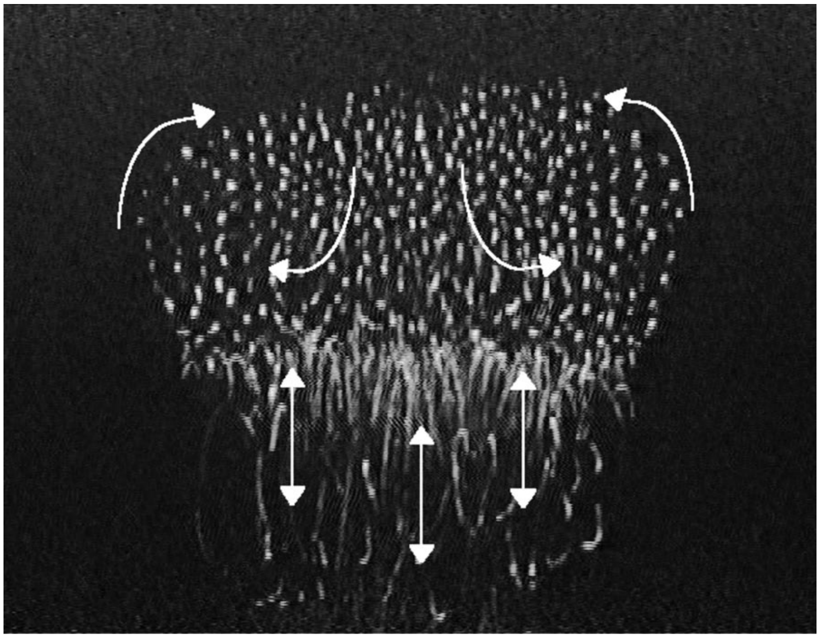


Fig. 4. Combined self-oscillations of macroparticles in the stratum of a glow discharge.

noted that the obtained values of D are in good agreement with the fact that the value of n_e increases towards the top of the cloud from 10^8 to 2×10^9 cm^{-3} for T_e = 2 eV [11]. In the latter case, the ratio $V/A \approx 0.25$ s^{-1} being measured for various amplitudes of rotation gives

$$\beta_\rho/Z_0 \approx 2V\nu_{\mathrm{fr}}/Ag \approx 0.002\ \mathrm{cm}^{-1}.$$

Thus, the ratio β_ρ for various parts of the dust cloud is not proportional to D^2. This may be due to several factors, the most obvious from which is the development of a type I instability (at the top of the structure) for extreme particles of the dust cloud, for which the value of γ_0 is quite high and satisfies condition (2) [6]. On the other hand, the development of a type II instability (at the bottom of the structure) is determined by particles from the interior of the dust cloud, where the value of β_ρ is much larger (see above).

It should be noted that all kinds of self-oscillatory motion of dust particles are also observed in experimental generators of inductive HF discharges which have inhomogeneous characteristics similar to those for a glow discharge. We can mention two basic reasons hindering the development of both types of instability in a capacitive HF discharge. The first reason is determined by the inhomogeneity of the plasma, while the second may be associated with the small number of observed layers of macroparticles and, hence, with the small displacement parameter γ_0 (2). This can explain the lack of experimental observations of vortex flows under the normal conditions. Nevertheless, the formation of bulk dust clouds (for example, under microgravitation conditions or upon the introduction of an additional electrode in the bulk of the plasma) leads to convection of macroparticles [1, 2]. The possible manifestation of a type II instability in a capacitive HF discharge is the evolution of self-oscillations for a one-dimensional layer of macroparticles [3]. Any other mechanisms proposed for explaining these oscillations have not led to reasonable quantitative estimates so far [3, 13].

The proposed mechanism for the emergence of dust self-oscillations is attractive since it is capable of explaining most of the effects (“anomalous heating” of macroparticles, vortices, acoustic vibrations, and other types of self-excited motion) observed in a laboratory dust plasma without resorting to other mechanisms or background energy sources. The formation of such self-oscillations does not require considerable spatial changes of dust charges because even a small charge gradient in a dust cloud (~1%) may serve as an effective source of the kinetic energy of macroparticles.

ACKNOWLEDGMENTS

This work was partly financed by the Russian Foundation for Basic Research (project nos. 01-02-16658 and 00-02-17520).

REFERENCES

1. D. A. Low, W. H. Steel, B. M. Annaratone, *et al.*, Phys. Rev. Lett. **80**, 4189 (1998).
2. G. Morfill, H. Thomas, U. Konopka, *et al.*, Phys. Rev. Lett. **83**, 1598 (1999).
3. S. Nunomura, T. Misawa, N. Ohno, and S. Takamura, Phys. Rev. Lett. **83**, 1970 (1999).

4. O. S. Vaulina, S. A. Khrapak, A. P. Nefedov, *et al.*, Phys. Rev. E **60**, 5959 (1999).
5. V. V. Zhakhovskiĭ, V. I. Molotkov, A. P. Nefedov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **66**, 392 (1997) [JETP Lett. **66**, 419 (1997)].
6. O. S. Vaulina, A. P. Nefedov, O. F. Petrov, *et al.*, Zh. Éksp. Teor. Fiz. **118**, 1325 (2000) [JETP **91**, 1147 (2000)].
7. Yu. V. Gerasimov, A. P. Nefedov, V. A. Sinel'shchikov, *et al.*, Pis'ma Zh. Tekh. Fiz. **24** (19), 62 (1998) [Tech. Phys. Lett. **24**, 774 (1998)].
8. V. E. Fortov, V. I. Molotkov, A. P. Nefedov, *et al.*, Phys. Plasmas **6**, 1759 (1999).
9. I. T. Yakubov and A. G. Khrapak, Sov. Technol. Rev. B, Therm. Phys. **2**, 269 (1989).
10. J. Goree, Plasma Sources Sci. Technol. **3**, 400 (1994).
11. Yu. P. Raizer, *Gas Discharge Physics* (Nauka, Moscow, 1987; Springer-Verlag, Berlin, 1991).
12. O. S. Vaulina, A. P. Nefedov, O. F. Petrov, and V. E. Fortov, Zh. Éksp. Teor. Fiz. **119**, 1129 (2001) [JETP **92**, 979 (2001)].
13. A. V. Ivlev, U. Konopka, and G. Morfill, Phys. Rev. E **62**, 2739 (2000).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 93, No. 6, 2001, pp. 1190–1208.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 6, 2001, pp. 1375–1395.
Original Russian Text Copyright © 2001 by Mel'nikov.

**PLASMA,
GASES**

The Theory of Particle Diffusion in a Strong Random Magnetic Field with an Allowance for Higher Multipole Moments of the Distribution Function

Yu. P. Mel'nikov
Rybinsk State Academy of Aviation Technology, Rybinsk, Yaroslavl oblast, 152934 Russia
e-mail: rgata@ryb.adm.yar.ru; fax:7–0855–559021

Received January 16, 2001

Abstract—The kinetic equation including a small-scale collisional integral for the particles propagating in a strong random and regular magnetic field [29] is solved by expanding the distribution function into series in spherical harmonics of the particle momentum angles. Using methods of the quantum theory of the angular moment [41], the equations for higher multipole moments of the distribution function in the space of momentum angles are derived and solved in the stationary case for the galactic cosmic rays in interplanetary space. The observed amplitudes and phases of the diurnal variation harmonics can be explained using the results of measurements of the interplanetary magnetic field performed on board the Ulysses spacecraft [12–14] and other satellites [45, 46] with an allowance for redistribution of the interplanetary and interstellar magnetic field lines. The spatial structure of the convection and diffusion fluxes of the galactic cosmic rays is refined. Formulas taking into account a change in the Earth's axis tilt relative to the direction toward the Sun are derived, which allow the annual changes in contributions to the diurnal variation harmonics to be determined. The equation of diffusion taking into account the 2nd harmonic is obtained, and the contribution of this effect to the relative particle density in the cosmic rays in a spherically symmetric case is analyzed. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The propagation of charged particles in magnetic fields is usually studied in the diffusion approximation which takes into account only the first multipole moment of the distribution function in the space of momentum angles. However, in the presence of strong variations in the particle concentration gradient N, transport range Λ, magnetic field strength $\mathbf{H}$, and magnetic field velocity $\mathbf{u}$, one must also take into account the higher multipole moments of the distribution function. An allowance for the higher moments of the distribution function is important in describing the propagation of charged particles in a turbulent medium (in particular, in a turbulent plasma with strong intermittency), in interplanetary and interstellar media, and in shear flows, shock waves, and other plasma structures where the diffusion approximation is violated [1–9].

The results of investigations of the interplanetary plasma performed on board the Ulysses spacecraft in the range of middle and high latitudes showed that the heliomagnetosphere structure and the solar wind parameters at these heliolatitudes in a solar activity minimum differ from those corresponding to the commonly accepted notions [10–16]. Such a difference may take place in a solar activity maximum as well. Thus, there is a need for continuously monitoring the parameters and configuration of the heliomagnetosphere. In circumterrestrial space, the heliomagnetosphere parameters can be determined using experimental data on the diurnal variation harmonics of the cosmic ray intensity, provided the corresponding theory is developed. The base concepts of such a theory are presented below.

We will analyze conditions for the formation of higher moments of the distribution function of the intensity of galactic cosmic rays in the inteplanetary magnetic field and consider a relationship between these moments and the diurnal variation harmonics.

The results of experimental investigations showed that diurnal variations of the cosmic ray intensity in circumterrestrial space contain the first and higher harmonics [3, 5–9, 17–24]. For the cosmic rays with $E \gtrsim$ 10 GeV, the observed diurnal variations exhibit a relative amplitude on the order of 1%, while the semidiurnal variations are on the order of 0.1%, and the 8- and 6-h variations amount to 0.03 and 0.01%, respectively. These harmonics are related to the higher multipole moments of the distribution function $F(\mathbf{r}, \mathbf{p}, t)$ in the space of momentum angles, where $\mathbf{r}$ is the coordinate, $\mathbf{p}$ is the momentum, and t is the time.

Kuz'min *et al.* [8, 9] analyzed experimental data on the 2nd harmonic of diurnal variations using the method of acceptance vectors. It was found that the contribution of the nth spherical harmonic of the distribution function to the mth diurnal variation component is proportional to the absolute value $\left|Z_n^m\right|$ of the corresponding complex component of the acceptance vector.

The contribution of second to first harmonic and the modulation of harmonics on the passage to a geographic coordinate system taking into account the Earth's axis tilt were studied as well. Since the harmonic amplitudes are close, the second to first harmonic contribution can be significant.

A screening mechanism of the 2nd harmonic formation in the diurnal variations was developed [7, 9] using an expansion of the intensity of cosmic rays (driven by a regular magnetic field layer) with respect to the pitch angle. An advantage of this model is the correct description of a break appearing in the 2nd harmonic spectrum when the Larmor radius of a particle coincides with the regular magnetic field layer half-thickness. Bieber *et al.* [21] pointed out that the 2nd harmonic appeared in the pitch angle distribution of particles in the cosmic rays as a result of the adiabatic focusing taking place during the propagation of particles in a regular component of the interplanetary magnetic field.

In addition, some researcher suggested a gradient mechanism of the semidiurnal variation of cosmic rays [9, 17], which is related to a symmetric particle concentration gradient $N(\mathbf{r}, p, t)$ in the cosmic rays relative to the helioequator plane. It was demonstrated that this mechanism accounts for a rigid energy spectrum, thus providing for a coincidence with experimental data on the harmonics of diurnal variations of the moderate-energy galactic cosmic rays. The 2nd spherical harmonic was originally studied [9, 17] in terms of the kinetic equation involving a model collisional integral written in the approximation according to which the phenomenological scattering cross section is transformed from a coordinate system moving at the magnetic field velocity to an immobile coordinate system.

A consistent theory of diffusion of cosmic rays [1, 4, 25–29] employing the collisional integral in a small-angle approximation [30, 31] was used to study equations for the 2nd harmonic and the related semidiurnal variations. It was found that, for small values of the gradients u, N, and Λ, the 2nd harmonic of the distribution function is on the order of u^2/v^2 (where $\mathbf{u}$ is the solar wind velocity and $\mathbf{v}$ is the particle velocity), which is significantly smaller as compared to the experimental value. The 2nd harmonic of diurnal variations on the order of 0.1% can be observed in the presence of a strong spatial concentration gradient $N(\mathbf{r}, p, t)$, whereas the radial gradients N considered in the above papers were small. A significant contribution to the 2nd harmonic can be provided by large transverse concentration gradients N observed in experiment [32– 34].

A magnetic-plug mechanism of the 2nd harmonic formation was proposed [35–37] to explain a semidiurnal variation in the cosmic ray intensity with a maximum in the direction of the regular magnetic field reported by Kane [18]. The interesting frequency spectra of the 2nd harmonic peaks experimentally observed with the aid of neutron monitors [18] revealed two pronounced maxima on the background of a large number of random peaks. The maximum at 3 h LT is oriented perpendicularly to the regular magnetic field $\mathbf{H}_0$, while the maximum at 9 h LT is parallel to $\mathbf{H}_0$ and related to the shock wave propagation in circumterrestrial space.

Experimental data on the diurnal 2nd harmonic, the dependence of a break in the 2nd harmonic spectrum on the solar activity cycle, and the statistical properties of the 2nd harmonic peak time and amplitude in 1965–1992 were reported in [19, 20]. According to these data, the 2nd harmonic maximum takes place at 3 h LT independently of the solar activity phase, the average amplitude is 0.05–0.1%, and the break energy is about 40 and 125 GeV at the solar activity minimum and maximum, respectively. For energies below and above the break value, the 2nd harmonic spectrum index is about 0.7 and 0.4, respectively.

Bieber *et al.* [21] analyzed the amplitudes and phases of the 2nd and 3rd third harmonics of diurnal variations of the cosmic rays intensity in the diffusion approximation using a numerical solution of the Fokker–Planck equation. Similarly to other investigations employing this approximation, the analysis yielded a 2nd harmonic spectrum increasing with the rigidity.

Experimental investigations of the diurnal 3rd harmonic were originally reported in [38–39] and continued in [22–24]. The corresponding maximum was observed about 6 h LT, while the device geometry features and the particle drift in the geomagnetic field may shift the peak position to 8–9 h LT. The observations were performed with the aid of neutron monitors. The 3rd harmonic formation was explained in terms of the proposed “loss cone” mechanism. According to the experimental data [24], the 4th harmonic of diurnal variations has an amplitude of 0.014% at $E \sim 10$ GeV with a peak at 3 h LT, is proportional to $p^{1/2}$ for energies below 100 GeV, and is zero at greater energies; the 1st, 2nd, and 3rd harmonic amplitudes for the particles with $E \sim 10$ GeV were 0.5, 0.1, and 0.04%, respectively.

Based on the whole body of data reported in [7–24], we may conclude that a characteristic feature of the high-order harmonics of the diurnal intensity variation is the $p^{0.5-1}$ law for energies E below a certain truncation level $E_c \sim 50$–130 GeV. Above this level, the 2nd harmonic amplitude decreases smoothly, while the 3rd and 4th harmonic amplitudes sharply drop. It should be noted that the diffusion approximation frequently used, together with a certain iterative procedure [9, 17, 30, 31], for the calculation of higher harmonics yields the spectrum of harmonics for the energies below E_c.

In this study, in contrast to the previous investigations, the higher moments of the distribution function (from second to fourth) are calculated using the kinetic equation of the consistent theory of diffusion of cosmic rays [1, 4, 25–27] employing the collisional integral in a small-scale approximation taking into account the high-order approximations with respect to a random magnetic field [28, 29]. Within the framework of this approach, it is possible to use an arbitrary (rather than only quadratic) relationship between the transport range Λ and the absolute momentum p. A comparison between theoretical and experimental results shows that the higher moments of the distribution function of the intensity of cosmic rays are determined primarily by the gradients of various orders of the flux of cosmic rays along the regular magnetic field. These gradients are related to the configuration of a regular interplanetary magnetic field, the characteristics of the random magnetic field, and the solar wind velocity.

2. EQUATIONS FOR MULTIPOLE MOMENTS OF THE DISTRIBUTION FUNCTION

In order to derive equations for multipole moments of the distribution function, we will use a kinetic equation with a small-scale collisional integral written in the following form:

$$\left\{\frac{\partial}{\partial t} + \mathbf{v}\cdot\frac{\partial}{\partial \mathbf{r}} - \frac{p}{R_0}(\mathbf{h}_0\cdot\mathbf{d})\right\}F(\mathbf{r},\mathbf{p},t) = \frac{1}{2}\left(d_\alpha \frac{p^2}{\Lambda|\mathbf{v}-\mathbf{u}|}d_\alpha\right)F(\mathbf{r},\mathbf{p},t). \tag{1}$$

Here, we use the notation

$$\mathbf{h}_0 = \frac{\mathbf{H}_0}{H_0},\quad R_0 = \frac{cp}{e}H_0$$

and introduce an operator

$$\mathbf{d} = (\mathbf{v}-\mathbf{u})\times\frac{\partial}{\partial\mathbf{p}}$$

facilitating subsequent calculations. Repeated Greek tensor indices imply summation.

If the transport range $\Lambda(p)$ is proportional to p^2, Eq. (1) takes into account low-order terms in the random field. According to a formula for the small-scale collisional integral StF [29, 40], an allowance for the high-order terms in the random field makes the coefficient of diffusion in the momentum space a complicated function of the eigenvalues of the square moment operator $\hat{\mathbf{L}}^2$. Nevertheless, we may still use an expansion in spherical harmonics $Y_{lm}(\vartheta, \varphi)$ of the momentum $\mathbf{p}$ angles [41]. Below, we assume that an expression for StF in Eq. (1) is valid for $\Lambda \propto p^q$, where $q < 2$. This assumption is equivalent to taking into account the high-order approximations with respect to the random magnetic field, which allows the spectrum of diurnal variations in the intensity of cosmic rays to be determined for sufficiently low energies.

Let us write the collisional integral in the right-hand part of Eq. (1) in the following form:

$$\mathrm{StF} = \sum_k\left[\frac{p^2}{2\Lambda}\left\{d^k\frac{1}{|\mathbf{v}-\mathbf{u}|}\right\}\{d_kF(\mathbf{r},\mathbf{p},t)\} + \frac{1}{2}\frac{1}{|\mathbf{v}-\mathbf{u}|}\left\{d^k\frac{p^2}{\Lambda}\right\}\{d_kF(\mathbf{r},\mathbf{p},t)\} + \frac{1}{2}\frac{p^2}{\Lambda|\mathbf{v}-\mathbf{u}|}\{d^kd_kF(\mathbf{r},\mathbf{p},t)\}\right], \tag{2}$$

where d_k are the cyclic components of the operator $\mathbf{d}$,

$$d_k = i\frac{1}{m}L_k + i\sqrt{2}\sum_{nm}C^{1k}_{1n1m}u_n\left\{\frac{\partial}{\partial\mathbf{p}}\right\}_m; \tag{3}$$

d^k are the complex conjugate cyclic components of $\mathbf{d}$ [38]; $C^{ln}_{pm\,qk}$ are the Clebsch–Gordan coefficients [41]; L_k are the cyclic components of the moment operator; and m is the particle mass. In Eq. (2) and below, the operator $\mathbf{d}$ acts only upon the functions within the same braces; the summation is implied over the repeated indices of the cyclic components. The first term in Eq. (2) describes the dynamic friction in the momentum space, the second term is related to the high-order approximations with respect to the random field, and the third term refers to the particle diffusion in the momentum space.

The distribution function can be presented in the form of a series in spherical harmonics in the space of momentum angles,

$$F(\mathbf{r},\mathbf{p},t) = \sum_l F_l(\mathbf{r},\mathbf{p},t) = \sum_l F^*_{lm}(\mathbf{r},p,t)Y_{lm}(\vartheta,\varphi),$$

where F_{lm} are the multipole moments of the distribution function. The formulas derived below will be used to study the propagation of galactic cosmic rays with the energies $E \gtrsim 10$ GeV. For this reason, taking into account the experimental data on the harmonics of diurnal variations of the intensity of cosmic rays and the solar wind velocity, we will retain the terms on the order of $(u^2/v^2)F_{00}$, $(u/v)F_{00}$, F_{00}, $(u/v)F_{1m}$, F_{1m}, F_{2m}, F_{3m}, and F_{4m}, rejecting all terms containing the $(u/v)F_{lm}$ product for $l \geq 2$ [3–9]. The particles possessing lower energies are described in a different system of approximations (see [1]).

Acting with operator d_k upon the distribution function $F(\mathbf{r}, \mathbf{p}, t)$, we obtain

$$\{d_k F(\mathbf{r}, \mathbf{p}, t)\} = i\frac{1}{m}\sum_{lm}\sqrt{l(l+1)}C^{lm+k}_{lm1k}Y_{lm+k}(\vartheta, \varphi)F^*_{lm}$$
$$+ i\sqrt{2}\sum_{lmqn}C^{1k}_{1n1m}u_n\left\{\sqrt{\frac{l}{2l+1}}(-1)^m C^{lq}_{l-1q+m1-m}\right.$$
$$\times \Phi_{lq}Y_{l-1q+m}(\vartheta, \varphi) - \sqrt{\frac{l+1}{2l+1}}(-1)^m \tag{4}$$
$$\left.\times C^{lq}_{l+1q+m1-m}\Psi_{lq}Y_{l+1q+m}(\vartheta, \varphi)\right\},$$

where

$$\Psi_{lm} = \frac{d}{dp}F^*_{lm} - \frac{l}{p}F^*_{lm}, \quad \Phi_{lm} = \frac{d}{dp}F^*_{lm} + \frac{l+1}{p}F^*_{lm}.$$

Now let us transform the term $\{d_\alpha d_\alpha F(\mathbf{r}, \mathbf{p}, t)\}$ in Eq. (2). Using operator d^k, we obtain the following expressions for the components of this term:

$$i\frac{1}{m}\left\{\sum_k L^k d_k F(\mathbf{r}, \mathbf{p}, t)\right\} = -\frac{1}{m^2}\sum_k l(l+1)F^*_{lm}Y_{lm}(\vartheta, \varphi)$$
$$+\frac{1}{m}\sum_{lmnr}(l-1)\sqrt{\frac{l}{2l+1}}C^{lm}_{1nl-1r}u_n\Phi_{ln}Y_{l-1r}(\vartheta, \varphi) \tag{5}$$
$$+\frac{1}{m}\sum_{lmnr}(l+2)\sqrt{\frac{l+1}{2l+1}}C^{lm}_{1nl+1r}u_n\Psi_{lm}Y_{l+1r}(\vartheta, \varphi).$$

Acting with the second operator of (3) upon (4), we obtain

$$\sqrt{2}\sum_{lmnqk}C^{1k}_{1n1q}u^n$$
$$\times\left\{\left\{\frac{\partial}{\partial \mathbf{p}}\right\}^q\frac{1}{m}\sqrt{l(l+1)}C^{lm+k}_{lm1k}F^*_{lm}Y_{lm+k}(\vartheta, \varphi)\right\}$$
$$= -\sum_{lmnk}(l+1)\sqrt{\frac{l}{2l-1}}C^{l-1k}_{1nlm}u^n \tag{6}$$
$$\times\left[\frac{d}{dp}\left(\frac{F^*_{lm}}{m}\right) + \frac{l+1}{pm}F^*_{lm}\right]Y_{l-1k}(\vartheta, \varphi)$$
$$-\sum_{lmnk}l\sqrt{\frac{l+1}{2l+3}}C^{l+1k}_{1nlm}u^n$$
$$\times\left[\frac{d}{dp}\left(\frac{F^*_{lm}}{m}\right) - \frac{l}{pm}F^*_{lm}\right]Y_{l+1k}(\vartheta, \varphi).$$

Expressions (5) and (6) were derived using the formulas

$$\sum_{pqr}(-1)^{p+r}C^{1r}_{1n1p}C^{lm}_{l-1q1-p}C^{l-1s}_{l-1q1-r} = -\sqrt{\frac{l-1}{2l}}C^{lm}_{1nl-1s},$$

$$\sum_{pqr}(-1)^{p+r}C^{1p}_{1n1p}C^{lm}_{l+1r1-p}C^{l+1s}_{l+1q1-p} = \sqrt{\frac{l+2}{2l+2}}C^{lm}_{1nl+1s},$$

$$\sum_{pqs}C^{1p}_{1n1q}C^{ls}_{lm1p}C^{ls}_{l-1m+p-q1q}$$
$$= -\sqrt{\frac{(l+2)(2l+1)}{2l(2l-1)}}C^{l-1m+p-q}_{1nlm},$$

$$\sum_{pqs}C^{1p}_{1n1q}C^{ls}_{lm1p}C^{ls}_{l+1m+p-q1q}$$
$$= \sqrt{\frac{l(2l+1)}{2(l+1)(2l+3)}}C^{l+1m+p-q}_{1nlm},$$

obtained using the relationships taken from [41]. Taking into account only terms of the type $(u^2/v^2)F_{00}$, $(u/v)F_{00}$, F_{00}, $(u/v)F_{1m}$, F_{1m}, F_{2m}, F_{3m}, and F_{4m} in (5) and (6), let us retain only the terms with $l > 0$ in the first component and only the terms with $l = 0, 1$ in the third component. For the above approximations, the sum in (6) reduces to

$$\sqrt{2}\sum_{lmnqk}C^{1k}_{1n1q}u^n$$
$$\times\left\{\left\{\frac{\partial}{\partial \mathbf{p}}\right\}^q\frac{1}{m}\sqrt{l(l+1)}C^{lm+k}_{lm1k}F^*_{lm}Y_{lm+k}(\vartheta, \varphi)\right\}$$
$$= -\frac{1}{\sqrt{3}}\sum_m u_m\left[\frac{d}{dp}\left(\frac{F^*_{1m}}{m}\right) + \frac{2}{pm}F^*_{1m}\right] \tag{7}$$
$$-\sqrt{\frac{2}{5}}\sum_{mnk}C^{2k}_{1m1n}u^n\left[\frac{d}{dp}\left(\frac{F^*_{1m}}{m}\right) - \frac{1}{pm}F^*_{1m}\right]Y_{2k}(\vartheta, \varphi).$$

The penultimate term in the sum entering into $\{d_\alpha d_\alpha F(\mathbf{r}, \mathbf{p}, t)\}$ can be transformed to

$$2\sum_{lmnqkrs}C^{1k}_{1r1s}u^r\left\{\left\{\frac{\partial}{\partial \mathbf{p}}\right\}^s C^{1k}_{1n1q}u_n\sqrt{\frac{l}{2l+1}}(-1)^q\right.$$
$$\left.\times C^{lm}_{l-1q+m1-q}\Phi_{lm}Y_{l-1q+m}(\vartheta, \varphi)\right\}$$

$$= \sum_{lmnpqr}\sqrt{l(l-1)}\frac{1}{2l-1}C^{l-1r}_{l-2q1p}C^{l-1r}_{lm1n}u^n u_p$$

$$\times\left[\frac{d}{dp}\Phi_{lm}+\frac{l}{p}\Phi_{lm}\right]Y_{l-2q}(\vartheta,\varphi) \qquad (8)$$

$$-\sum_{lmnpqr}\left(\frac{l}{2l+1}C^{l+1r}_{lm1n}C^{l+1r}_{lq1p}\right.$$

$$\left.+\frac{l}{2l+1}C^{lr}_{lq1n}C^{lr}_{lq1p}-2lC^{l-1r}_{lm1n}C^{l-1r}_{lq1p}\right)$$

$$\times u^n u_p\left[\frac{d}{dp}\Phi_{lm}-\frac{l-1}{p}\Phi_{lm}\right]Y_{lq}(\vartheta,\varphi).$$

This relationship was derived using the formulas

$$\sum_{aps}(-1)^s C^{1a}_{1n1p}C^{1a}_{1q1s}C^{lm}_{l-1m+s1-s}C^{l-1m+s}_{l-2m+s-p1p}$$

$$= -3\sqrt{(2l+1)(2l-1)}\sum_{kt}C^{kt}_{lm1n}C^{kt}_{l-2m+s-p1q}$$

$$\times\left\{\begin{matrix} l-1 & l-2 & 1\\ l & k & 1\\ 1 & 1 & 1\end{matrix}\right\} = \frac{1}{2}\sqrt{\frac{2l+1}{2l-1}}\sum_r C^{l-1r}_{lm1n}C^{l-1r}_{l-2m+s-p1q},$$

$$\sum_{aps}(-1)^s C^{1a}_{1n1p}C^{1a}_{1q1s}C^{lm}_{l-1m+s1-s}C^{l-1m+s}_{lm+s-p1p}$$

$$= -3\sqrt{(2l+1)(2l-1)}\sum_{k=l+1,\,l,\,l-1;\,r}C^{kr}_{lm1n}C^{kr}_{lm+s-p1q}$$

$$\times\left\{\begin{matrix} l-1 & l & 1\\ l & k & 1\\ 1 & 1 & 1\end{matrix}\right\} = -\frac{1}{2}\sqrt{\frac{2l-1}{2l+1}}\sum_r C^{l+1r}_{lm1n}C^{l+1r}_{lm+s-p1q}$$

$$-\frac{1}{2}\sqrt{\frac{2l-1}{2l+1}}\sum_r C^{lr}_{lm1n}C^{lr}_{l-2m+s-p1q}$$

$$+\sqrt{(2l+1)(2l-1)}\sum_r C^{l-1r}_{lm1n}C^{l-1r}_{l-2m+s-p1q},$$

where $\left\{\begin{matrix} l-1 & l & 1\\ l & k & 1\\ 1 & 1 & 1\end{matrix}\right\}$ are the 9j symbols obtained using the relationships taken from [41]. Also employed were the formulas for the 6j symbols, Racah coefficients, 9j symbols, and Fano factors [41]. Retaining only terms of the type $(u^2/v^2)F_{00}$ in (8), we conclude that the contribution of this term to the collisional integral is zero.

The last term in $\{d_\alpha d_\alpha F(\mathbf{r},\mathbf{p},t)\}$ can be transformed to

$$-2\sum_{lmnqkrs}C^{1k}_{1r1s}u^r\left\{\left\{\frac{\partial}{\partial\mathbf{p}}\right\}^s\sqrt{\frac{l+1}{2l+1}}(-1)^q C^{1k}_{1n1q}\right.$$

$$\left.\times C^{lm}_{l+1q+m1-q}u_n\Psi_{lm}Y_{l+1q+m}(\vartheta,\varphi)\right\}$$

$$=\sum_{lmnpqr}\sqrt{(l+1)(l+2)}\frac{1}{2l+3}C^{l+1r}_{lm1n}C^{l+1r}_{l+1q1p}u^n u_p$$

$$\times\left[\frac{d}{dp}\Psi_{lm}-\frac{l+1}{p}\Psi_{lm}\right]Y_{l+2q}(\vartheta,\varphi) \qquad (9)$$

$$+\sum_{lmnpqr}(l+1)\left\{\frac{2}{(2l+1)(2l+3)}C^{l+1r}_{lm1n}C^{l+1r}_{lq1p}\right.$$

$$\left.+\frac{1}{2l+1}C^{lr}_{lm1m}C^{lr}_{lq1p}+\frac{1}{2l+1}C^{l-1r}_{lm1n}C^{l-1r}_{1q1p}\right\}$$

$$\times u^n u_p\left[\frac{d}{dp}\Psi_{lm}+\frac{l+2}{p}\Psi_{lm}\right]Y_{lq}(\vartheta,\varphi).$$

This relationship was derived using the relationships

$$\sum_{aps}(-1)^s C^{1a}_{1n1p}C^{1a}_{1q1s}C^{lm}_{l+1m+s1-s}C^{l+1m+s}_{lm+s-p1p}$$

$$= -3\sqrt{(2l+1)(2l-1)}\sum_{k=l+1,\,l,\,l-1;\,r}C^{kr}_{lm1n}C^{kr}_{lm+s-p1q}$$

$$\times\left\{\begin{matrix} l+1 & l & 1\\ l & k & 1\\ 1 & 1 & 1\end{matrix}\right\} = -\sqrt{\frac{1}{(2l+1)(2l+3)}}$$

$$\times\sum_r C^{l+1r}_{lm1n}C^{l+1r}_{lm+s-p1q}-\frac{1}{2}\sqrt{\frac{2l+3}{2l+1}}\sum_r C^{lr}_{lm1n}C^{lr}_{lm+s-p1q}$$

$$-\frac{1}{2}\sqrt{\frac{2l+3}{2l+1}}\sum_r C^{l-1r}_{lm1n}C^{l-1r}_{lm+s-p1q},$$

$$\sum_{aps}(-1)^s C^{1a}_{1n1p}C^{1a}_{1q1s}C^{lm}_{l+1m+s1-s}C^{l+1m+s}_{l+2m+s-p1p}$$

$$= -3\sqrt{(2l+1)(2l+3)} \sum_{k = l+1,\, l,\, l-1;\, r} C^{kr}_{lm1n} C^{kr}_{l+2m+s-p1q}$$

$$\times \begin{Bmatrix} l+1 & l+2 & 1 \\ l & k & 1 \\ 1 & 1 & 1 \end{Bmatrix} = \frac{1}{2}\sqrt{\frac{2l-1}{2l+1}} \sum_{r} C^{l+1r}_{lm1n} C^{l+1r}_{l+2m+s-p1q},$$

following from the formulas obtained in [41]. Since the value of (9) is proportional to u^2/v^2, we retain only terms of the type $(u^2/v^2)F_{00}$ in this expression. This yields

$$-2 \sum_{lmnqkrs} C^{1k}_{1r1s} u^r \left\{ \left\{ \frac{\partial}{\partial \mathbf{p}} \right\}^s \sqrt{\frac{l+1}{2l+1}} (-1)^q C^{1k}_{1n1q} \right.$$

$$\left. \times C^{lm}_{l+1q+m1-q} u_n \Psi_{lm} Y_{l+1q+m}(\vartheta, \varphi) \right\} \tag{10}$$

$$= \frac{2}{3} u^2 \left[\frac{d}{dp} \Psi_{00} + \frac{2}{p} \Psi_{00} \right] Y_{00} - \sqrt{\frac{2}{15}} \sum_{snq} C^{2s}_{1n1q} u^n u^q$$

$$\times \left[\frac{d}{dp} \Psi_{00} - \frac{1}{p} \Psi_{00} \right] Y_{2s}(\vartheta, \varphi).$$

Using expressions (4)–(10), we eventually obtain the following expression for $\{d_\alpha d_\alpha F(\mathbf{r}, \mathbf{p}, t)\}$:

$$\sum_k \{d^k d_k F\} = -\frac{1}{m^2} \sum_{lm} l(l+1) F^*_{lm} Y_{lm}(\vartheta, \varphi)$$

$$-\frac{2}{\sqrt{3}} \frac{1}{m} \frac{dF_{00}}{dp} \sum_n u^n Y_{1n}(\vartheta, \varphi)$$

$$-3\sqrt{\frac{2}{5}} \frac{1}{m} \sum_{nms} C^{2s}_{1n1m} u^n \Psi_{1m} Y_{2s}(\vartheta, \varphi)$$

$$+\frac{1}{\sqrt{3\pi}} \sum_m u_m \left[\frac{d}{dp} \left(\frac{F^*_{1m}}{m} \right) + \frac{F^*_{1m}}{pm} \right] \tag{11}$$

$$-\sqrt{\frac{2}{5}} \sum_{nms} C^{2s}_{1n1m} u^n \left[\frac{d}{dp} \left(\frac{\overset{*}{F}_{1m}}{m} \right) - \frac{\overset{*}{F}_{1m}}{pm} \right] Y_{2s}(\vartheta, \varphi)$$

$$+\frac{1}{3\sqrt{\pi}} u^2 \left[\frac{d}{dp} \Psi_{00} + \frac{2}{p} \Psi_{00} \right]$$

$$-\sqrt{\frac{2}{15}} \sum_{nms} C^{2s}_{1n1m} u^n u^m \left[\frac{d}{dp} \Psi_{00} - \frac{2}{p} \Psi_{00} \right] Y_{2s}(\vartheta, \varphi)$$

$$-\sqrt{\frac{2}{5}} \sum_{nms} C^{2s}_{1n1m} u^n \left[\frac{d}{dp} \left(\frac{F^*_{1m}}{m} \right) - \frac{1}{pm} F^*_{1m} \right] Y_{2s}(\vartheta, \varphi).$$

Now, let us determine the term

$$\left\{ d_\alpha \frac{1}{|\mathbf{v} - \mathbf{u}|} \right\} \{ d_\alpha F(\mathbf{r}, \mathbf{p}, t) \}$$

in Eq. (2). Since $\{d_\alpha|\mathbf{v} - \mathbf{u}|^{-1}\}$ is multiplied by $\{d_\alpha F(\mathbf{r}, \mathbf{p}, t)\}$ given by (4), we take into account only the first two terms in u/v in $|\mathbf{v} - \mathbf{u}|^{-1}$. Using Eq. (3), we obtain

$$\left\{ d^k \frac{1}{|\mathbf{v} - \mathbf{u}|} \right\} = i\sqrt{\frac{2\pi}{3}} \frac{2}{pv} \left(-\frac{v^2}{c^2} \right)$$

$$\times \sum_{nm} (-1)^k C^{1m}_{1n1-k} u^n Y_{1m}(\vartheta, \varphi) - i\frac{2\sqrt{2}}{3p} \frac{v^2}{c^2} \sum_{nm} C^{1k}_{1n1m} u^n u^m$$

$$-i\frac{4\sqrt{\pi}}{3pv^2} \left(3 - 2\frac{v^2}{c^2} \right) \sum_{nmpq} C^{1k}_{1n1p} C^{1m}_{2q1p} u^n u^m Y_{2q}(\vartheta, \varphi).$$

Retaining only the first-order terms in u/v in this expression, we arrive at

$$\left\{ d^k \frac{1}{|\mathbf{v} - \mathbf{u}|} \right\} = -i\sqrt{\frac{2\pi}{3}} \frac{2}{pv} \frac{v^2}{c^2}$$

$$\times \sum_{nm} (-1)^m C^{1k}_{1n1m} u^n Y_{1-m}(\vartheta, \varphi). \tag{12}$$

Multiplying (12) by $\{d_\alpha F(\mathbf{r}, \mathbf{p}, t)\}$ and taking into account only terms of the type $(u^2/v^2)F_{00}$, $(u/v)F_{00}$, F_{00}, $(u/v)F_{1m}$, F_{1m}, and F_{2m}, we obtain

$$\sum_k \left\{ d^k \frac{1}{|\mathbf{v} - \mathbf{u}|} \right\} \{ d_k F(\mathbf{r}, \mathbf{p}, t) \} = \frac{1}{\sqrt{3\pi}} \frac{1}{mE} \sum_m u_m F^*_{1m}$$

$$-\sqrt{\frac{2}{5}} \frac{1}{mE} \sum_{mnq} C^{2q}_{1n1m} u^n F^*_{1m} Y_{2q}(\vartheta, \varphi) + \frac{1}{3\sqrt{\pi}} \frac{u^2}{E} \frac{dF_{00}}{dp} \tag{13}$$

$$-\sqrt{\frac{2}{15}} \frac{1}{E} \frac{dF_{00}}{dp} \sum_{mnq} C^{2q}_{1n1m} u^n u^m Y_{2q}(\vartheta, \varphi).$$

This relationship was derived using the formulas

$$\sum_{abc} (-1)^c C^{1a}_{1n1c} C^{1b}_{1m1a} C^{2q}_{1-c1b} = -\frac{1}{2} C^{2q}_{1n1m},$$

$$\sum_{abp} (-1)^b C^{1a}_{1n1b} C^{1a}_{1m1p} C^{2q}_{1-b1p} = -\frac{1}{2} (-1)^m C^{2q}_{1n1m},$$

obtained using relationships for the sums of products of the Clebsch–Gordan coefficients taken from [41].

The term related to a nonquadratic dependence of the transport range Λ on p can be transformed as follows:

$$\sum_k \left\{ d^k \frac{p^2}{\Lambda} \right\} \{ d_k F(\mathbf{r}, \mathbf{p}, t) \} = \left[\frac{1}{\sqrt{3}} \frac{1}{m} \sum_m u_m F^*_{1m} \right.$$

$$+ \frac{1}{3\sqrt{\pi}} u^2 \frac{dF_{00}}{dp} - \sqrt{\frac{2}{5}} \frac{1}{m} \sum_{mnq} C^{2q}_{1n1m} u^n F^*_{1m} Y_{2q}(\vartheta, \varphi) \qquad (14)$$

$$\left. - \sqrt{\frac{2}{15}} \frac{dF_{00}}{dp} \sum_{mnq} C^{2q}_{1n1m} u^n u^m Y_{2q}(\vartheta, \varphi) \right] \left\{ \frac{d}{dp} \frac{p^2}{\Lambda} \right\}.$$

Expanding the left-hand part of Eq. (1) in spherical harmonics and taking into account the differentiation formulas from [41], we obtain

$$\sum_{lm} \frac{\partial}{\partial t} F^*_{lm} Y_{lm}(\vartheta, \varphi)$$

$$+ v \sum_{lmnq} \sqrt{\frac{l+1}{2l+3}} C^{l+1}_{lm1q} \nabla^q F^*_{lm} Y_{l+1n}(\vartheta, \varphi)$$

$$- v \sum_{lmnq} \sqrt{\frac{l}{2l-1}} C^{l-1}_{lm1q} \nabla^q F^*_{lm} Y_{l-1n}(\vartheta, \varphi)$$

$$- i \frac{v}{R_0} \sum_{lmnq} \sqrt{l(l+1)} C^{lm+k}_{lm1k} h^k_0 F^*_{lm} Y_{lm+k}(\vartheta, \varphi) \qquad (15)$$

$$+ \frac{p}{R_0} \sum_{lmnq} \sqrt{\frac{l}{2l-1}} C^{l-1n}_{lm1q} \Phi_{lm} [\mathbf{u} \times \mathbf{h}_0]^q Y_{l-1n}(\vartheta, \varphi)$$

$$- \frac{p}{R_0} \sum_{lmnq} \sqrt{\frac{l+1}{2l+3}} C^{l+1}_{lm1q} \Psi_{lm} [\mathbf{u} \times \mathbf{h}_0]^q Y_{l+1n}(\vartheta, \varphi),$$

where

$$[\mathbf{q} \times \mathbf{h}_0]^k = i\sqrt{2} \sum_{mn} C^{1k}_{1n1m} q^m h^n_0, \quad \nabla^m = \left\{ \frac{\partial}{\partial \mathbf{r}} \right\}^m.$$

Substituting formulas (11), (12), (14), and (15) into expression (2) for StF and into Eq. (1), we obtain the latter equation written in terms of the expansion of the distribution function in spherical harmonics. Multiplying this equation sequentially by the first five spherical functions $Y_{lm(\vartheta, \varphi)}$ and integrating over the solid angle, we obtain equations for the first five moments of the distribution function.

The first two of these equations are as follows:

$$\frac{\partial}{\partial t} F^*_{00} + v \frac{1}{\sqrt{3}} \sum_m \nabla_m F^*_{1m}$$

$$= \frac{1}{\sqrt{3} R_0} \sum_m [\mathbf{u} \times \mathbf{h}_0]_m \left[p \frac{d}{dp} F^*_{1m} + 2F^*_{1m} \right]$$

$$+ \frac{1}{\sqrt{3}\Lambda} \sum_m u_m \left[p \frac{d}{dp} F^*_{1m} + F^*_{1m} + \frac{\Lambda}{p} \left\{ \frac{d}{dp} \frac{p^2}{\Lambda} \right\} F^*_{1m} \right] + \frac{u^2}{3v\Lambda} \qquad (16)$$

$$\times \left[p^2 \frac{d^2}{dp^2} F_{00} + \left(1 + \frac{v^2}{c^2} \right) p \frac{d}{dp} F_{00} + \Lambda \frac{d}{dp} F_{00} \left\{ \frac{d}{dp} \frac{p^2}{\Lambda} \right\} \right],$$

$$\frac{\Lambda}{v} \frac{\partial}{\partial t} F^*_{1k} + \frac{\Lambda}{\sqrt{3}} \nabla^k F_{00} + \sqrt{\frac{2}{5}} \Lambda \sum_{mn} C^{2m}_{1k1n} \nabla^n F^*_{2m}$$

$$+ \frac{\Lambda}{R_0} [\mathbf{h}_0 \times \mathbf{F}_1]^k - \frac{\Lambda}{\sqrt{3} v R_0} [\mathbf{u} \times \mathbf{h}_0]^k \frac{d}{dp} F_{00} \qquad (17)$$

$$= -F^*_{1k} - \frac{1}{\sqrt{3} v} u^k p \frac{d}{dp} F_{00}.$$

A similar system of equations for the first two moments of the distribution function with an arbitrary dependence of Λ on p was previously obtained in [1, 25] (without taking into account the 2nd harmonic) and in [30, 31] (with an allowance for the 2nd harmonic, but only for a quadratic dependence of Λ on p).

In the consistent theory of diffusion of cosmic rays [1, 30, 31], it is a usual practice to employ the quantities $N(\mathbf{r}, p, t)$, $\mathbf{J}(\mathbf{r}, p, t)$, and $M_m(\mathbf{r}, p, t)$ that are related to the coefficients F_{lm} by the expansion

$$F(\mathbf{r}, \mathbf{p}, t) = \frac{N(\mathbf{r}, p, t)}{4\pi} + \sqrt{\frac{3}{4\pi}} \sum_m \frac{J^*_m(\mathbf{r}, p, t)}{v} Y_{1m}(\vartheta, \varphi)$$

$$+ \sqrt{\frac{3}{10\pi}} \sum_m \frac{M^*_m(\mathbf{r}, p, t)}{v^2} Y_{2m}(\vartheta, \varphi).$$

Equations for the second, third, and fourth moments of the distribution function are as follows:

$$\frac{\Lambda}{3v} \frac{\partial}{\partial t} F^*_{2k} + F^*_{2k} - i \sqrt{\frac{2}{3}} \frac{\Lambda}{R_0} \sum_{mn} C^{2k}_{2m1n} F^*_{2m} h^n_0$$

$$= -\frac{\Lambda}{3} \sqrt{\frac{2}{5}} \sum_{mn} C^{2k}_{1n1m} \nabla^n F^*_{1m} + \frac{1}{\sqrt{15}} \Lambda \sum_{mn} C^{2k}_{1n3m} \nabla^n F^*_{3m}$$

$$+\frac{\Lambda}{3vR_0}\sqrt{\frac{2}{5}}\sum_{mn}C^{2k}_{1m1n}[\mathbf{u}\times\mathbf{h}_0]^n\left[p\frac{d}{dp}F^*_{1m}-F^*_{1m}\right]$$

$$-\frac{\Lambda}{3v}\sqrt{\frac{2}{5}}\sum_{mn}C^{2k}_{1m1n}u^n\left[2p\frac{d}{dp}F^*_{1m}-F^*_{1m}+\frac{\Lambda}{2p}\left\{\frac{d}{dp}\frac{p^2}{\Lambda}\right\}F^*_{1m}\right] \tag{18}$$

$$-\frac{\Lambda}{3v^2}\frac{1}{\sqrt{30}}\sum_{mn}C^{2k}_{1n1m}u^nu^m$$

$$\times\left[p^2\frac{d^2F_{00}}{dp^2}+\left(1+\frac{v^2}{c^2}\right)p\frac{dF_{00}}{dp}+\Lambda\frac{dF_{00}}{dp}\left\{\frac{d}{dp}\frac{p^2}{\Lambda}\right\}\right],$$

$$\frac{\Lambda}{6v}\frac{\partial F^*_{3k}}{\partial t}+F^*_{3k}-i\frac{1}{\sqrt{3}}\frac{\Lambda}{R_0}\sum_{mn}C^{3k}_{3m1n}F^*_{3m}h^n_0$$

$$=-\frac{\Lambda}{6}\sqrt{\frac{3}{7}}\sum_{mn}C^{3k}_{1n2m}\nabla^nF^*_{2m}+\frac{1}{3\sqrt{7}}\Lambda\sum_{mn}C^{3k}_{1n4m}\nabla^nF^*_{4m}, \tag{19}$$

$$\frac{\Lambda}{10v}\frac{\partial F^*_{4k}}{\partial t}+F^*_{4k}-i\frac{1}{\sqrt{5}}\frac{\Lambda}{R_0}\sum_{mn}C^{4k}_{4m1n}h^n_0F^*_{4m}$$

$$=-\frac{\Lambda}{15}\sum_{mn}C^{4k}_{3m1n}\nabla^nF^*_{3m}+\frac{\Lambda}{6\sqrt{5}}\sum_{mn}C^{4k}_{5m1n}\nabla^nF^*_{5m}. \tag{20}$$

Equations (18)–(20) were obtained by ignoring terms of the type $(u/v)F_{lm}$ with $l \geq 2$. This implies that the ratio R_0/Λ must not be very small, $R_0/\Lambda \gtrsim 0.1$, a condition which is satisfied when the galactic cosmic rays propagate in interplanetary space [42].

3. MULTIPOLE MOMENTS OF THE DISTRIBUTION FUNCTION AND DIURNAL VARIATION HARMONICS: DISCUSSION OF RESULTS

The solution of equations for the multipole moments of the distribution function is facilitated by passing to the fluxes, including the total flux $\mathbf{j}(\mathbf{r}, p, t)$ and the diffusion flux $\mathbf{i}(\mathbf{r}, p, t)$, These fluxes have a dimensionality of the particle number density in the phase space:

$$j^k = \sqrt{\frac{4\pi}{3}}F^*_{1k} = i^k - \frac{u^k}{v}\frac{p}{3}\frac{\partial N}{\partial p}.$$

An equation for the particle concentration $N(\mathbf{r}, p, t)$ can be written in the following form:

$$\frac{\partial N}{\partial t}+v\,\mathrm{div}\,\mathbf{j} = \frac{\mathbf{u}\times\mathbf{h}_0}{R_0}\cdot\left\{p\frac{d}{dp}\mathbf{i}+2\mathbf{i}\right\}$$

$$+\frac{\mathbf{u}}{\Lambda}\cdot\left[p\frac{d}{dp}\mathbf{i}+\mathbf{i}+\mathbf{i}\frac{\Lambda}{p}\left\{\frac{d}{dp}\frac{p^2}{\Lambda}\right\}\right], \tag{21}$$

which is analogous to the expressions derived in [1, 25]. Note that, despite the fact that Eq. (21) does not explicitly contain higher moments of the distribution function related to the 2nd and higher harmonics in the momentum space, these higher moments enter into this equation implicitly via the diffusion flux $\mathbf{i}(\mathbf{r}, p, t)$.

Let us ignore the term $\partial\mathbf{j}/\partial t$ in Eq. (17), assuming that the characteristic time of variation of the galactic cosmic ray flux over the observation time is large ($\tau \gg L/v$), and introduce the 2nd harmonic density

$$f^m = \sqrt{\frac{10\pi}{3}}F^m_2.$$

Then, the equation for the flux can be transformed as follows:

$$i^k+\frac{\Lambda}{R_0}[\mathbf{h}_0\times\mathbf{i}]^k$$

$$= -\frac{\Lambda}{3}\nabla^kN-\frac{2\Lambda}{5}\sum_{mn}C^{2m}_{1k1n}\nabla^nf^m. \tag{22}$$

Solving this equation with respect to the diffusion flux $\mathbf{i}$ yields

$$i^k = -\frac{\Lambda}{3}\left(1+\frac{\Lambda^2}{R_0^2}\right)^{-1}$$

$$\times\left\{q^k+\frac{\Lambda^2}{R_0^2}h^k_0(\mathbf{h}_0\cdot\mathbf{q})-\frac{\Lambda}{R_0}[\mathbf{h}_0\times\mathbf{q}]^k\right\}, \tag{23}$$

where

$$q^k = \nabla^kN+\frac{6}{5}\sum_{mn}C^{2m}_{1k1n}\nabla^nf^m. \tag{24}$$

In expressions (23) and (24), we have to take into account the second moment f^m because the experimental observations for galactic cosmic rays with an energy of 50–100 GeV showed that the relative values of the 2nd and 1st harmonics can be comparable [8]. Higher moments of the distribution function enter into the equation (21) for $N(\mathbf{r}, p, t)$ via expressions for the second moment and the fluxes (23) and (24).

Equations for the higher moments (second, third, and fourth) will be considered in a coordinate system with $z \parallel \mathbf{h}_0$. This is related to the fact that formulas obtained in an arbitrary coordinate system are rather cumbersome (see Appendix 5 in [40]). Another reason is that the angular dependence of $\mathbf{h}_0$ is conveniently expressed through the Wigner D functions [41].

In equation (18) for the second moment of the distribution function, we neglect terms of the type $\partial F^k_2/\partial t$ and F^m_3. In the right-hand part of Eq. (18), we retain only the first term involving ∇_nF_{1m}. Experimental data on the transport range Λ, solar wind velocity u, and Larmor radius R_0 in the regular component of the interplanetary magnetic field $\mathbf{H}_0$ for particles with 10–100 GeV energies showed that this term provides for the maximum contribution. Thus, Eq. (18) can be written in the following form:

$$f^k - i\sqrt{\frac{2}{3}}\frac{\Lambda}{R_0}\sum_{mn} C^{2k}_{2n1m} f^n h_0^m = -\frac{\Lambda}{3}\sum_{mn} C^{2k}_{1m1n}\nabla^m j^n. \quad (25)$$

A solution to this equation is

$$f^k = -\frac{\Lambda}{3}\left\{1 - ik\frac{\Lambda}{3}\frac{h_0}{R_0}\right\}^{-1}\sum_{mn} C^{2k}_{1m1n}\nabla^m j^n. \quad (26)$$

Using this solution, we obtain the total term related to the second moment of the distribution function:

$$F_2(\mathbf{r}, \mathbf{p}, t) = \sum_m F_2^m(\mathbf{r}, p, t) Y_{2m}(\vartheta, \varphi).$$

In order to determine the terms entering into the expression for F_2, let us orient the x axis perpendicularly to $\mathbf{h}_0$ and parallel to the helioequator plane, while the y axis is directed toward the North Pole and perpendicular to $\mathbf{h}_0$. In the coordinates, the second moment F_2 is expressed as

$$\begin{aligned} F_2(\mathbf{r}, \mathbf{p}, t) = {} & -\frac{\Lambda}{16\pi}\left[\left(I_1 + \frac{2\Lambda}{3R_0}h_0 I_2\right)\cos(2\varphi) \right. \\ & \left. - \left(\frac{2\Lambda}{3R_0}h_0 I_1 - I_2\right)\sin(2\varphi)\right][1 - \cos(2\vartheta)] \\ & + \frac{\Lambda}{8\pi}\lambda_2\left[\left(-I_3 - \frac{\Lambda}{3R_0}h_0 I_4\right)\cos\varphi \right. \\ & \left. - \left(I_4 + \frac{\Lambda}{3R_0}h_0 I_3\right)\sin\varphi\right]\sin(2\vartheta) - \frac{\Lambda}{48\pi}I_0[1 + 3\cos(\vartheta)], \end{aligned} \quad (27)$$

where

$$\lambda_1 = \left(1 + \frac{4}{9}\frac{\Lambda^2}{R_0^2}\right)^{-1}, \quad \lambda_2 = \left(1 + \frac{1}{9}\frac{\Lambda^2}{R_0^2}\right)^{-1},$$

$$\begin{aligned} I_1 &= \frac{\partial}{\partial x}j_x - \frac{\partial}{\partial y}j_y, \quad I_2 = \frac{\partial}{\partial x}j_y + \frac{\partial}{\partial y}j_x, \\ I_3 &= \frac{\partial}{\partial x}j_z + \frac{\partial}{\partial z}j_x, \quad I_4 = \frac{\partial}{\partial y}j_z + \frac{\partial}{\partial z}j_y, \end{aligned} \quad (28)$$

$$I_0 = 3\frac{\partial}{\partial z}j_z - \operatorname{div}\mathbf{j},$$

and $h_0 \equiv h_{0z} = 1$; ϑ and φ are the polar and azimuthal angles in the spherical coordinate system.

As can be seen, the contribution due to the 2nd harmonic (27) in the expression for $N(\mathbf{r}, p, t)$ can be ignored. It is very convenient to analyze the observations of diurnal harmonics from registrators receiving radiation in the helioequator plane. Ignoring the Earth’s axis tilt to the helioequator plane, as well as the annual variations in this tilt, we can analyze the results of observations from muon telescopes directed 30° southward and (with somewhat worse results) from those oriented in the vertical direction [8, 43].

According to the experimental data, $R_0/\Lambda \approx 0.1$ [42, 44]. Therefore, assuming that the currents and gradients (N, R_0, Λ, u) are not anomalous, the relative value of the second moment in the zeroth approximation with respect to R_0/Λ is

$$\delta_2(\vartheta) = \frac{F_2}{F_0} = -\frac{3}{4}\frac{\Lambda}{N}\frac{\partial}{\partial z}j_z\cos(2\vartheta). \quad (29)$$

In deriving this formula, we used an estimate

$$\operatorname{div}\mathbf{j} \approx \frac{u}{v}\frac{j}{R_0} \ll \frac{\partial}{\partial z}j_z,$$

following from Eq. (21), and assumed that the flux gradient $\partial j_z/\partial z$ is sufficiently large (e.g., due to divergence of the magnetic field lines).

Taking into account the convection flux, we can express j_z through the 1st harmonic amplitude,

$$j_z = \frac{1}{3}\delta_1 N,$$

and assume that

$$j_z \approx -0.002N\frac{1\text{ a.u.}}{z}, \quad \Lambda \approx 1\text{ a.u.},$$

for particles with $E = 10$ GeV [1, 3–9]. As will be demonstrated below, this assumption provides for the 2nd, 3rd, and 4th harmonics of diurnal variation of the galactic cosmic rays and the 2nd and 3rd harmonic phases coinciding with experiment. According to the experimental data [45–47], the azimuthal component of the regular interplanetary magnetic field is proportional to $1/r^{1.2}$ and the radial component is proportional to $1/r^2$, where r is the radial distance to the Sun. Taking into account that the particles are driven by the magnetic field, we will assume that, to a first approximation, the cosmic ray flux parallel to the regular magnetic field and directed toward the Sun represents a projection of the azimuthal flux. Another assumption is that the radial fluxes—diffusion and convection—are mutually compensating [4]. This yields the estimate $\delta_2 \approx 0.2\%$ (29) and gives the maximum of $\delta_2(\vartheta)$ at 3 h LT oriented perpendicularly to the regular magnetic field—in agreement with the experimental results [5, 18–23].

It was found [3, 18] that, besides the $\delta_2(\vartheta)$ maximum at 3 h LT corresponding to $\partial j_z/\partial z > 0$ in Eq. (29), the experimental data indicate another maximum at 9 h LT (related to the shock wave) corresponding to $\partial j_z/\partial z < 0$ in Eq. (29). The change in the sign of $\partial j_z/\partial z$ can be related to a change in the direction of either the cosmic ray flux or the current gradient that may take place when the Earth passes the regions of shock waves or magnetic plugs.

Based on the experimental data gained in 1971–1975 for cosmic rays with energies on the order of 50 GeV, it was demonstrated [43] that the j_z maxima took place in the 6th ($j_z < 0$) and 12th ($j_z > 0$) months, coinciding with the 2nd harmonic peaks always observed at 3 h LT. According to Eq. (29), the constant time of the $\delta_2(\vartheta)$ maximum is related to a change in the direction of decreasing flux density of the cosmic rays. Taking in to account that the cosmic ray particles are driven by the magnetic field, we may suggest that (in 1971–1975) the magnetic field lines near the helioequator plane converged (on the average) toward this plane on one side of the Sun, while diverging on the other side of the Sun.

Using formula (29) for the second moment, we can refine the average mechanism of the 2nd harmonic formation for the cosmic ray particles moving in the magnetic field. Propagating in a regular component of the interplanetary magnetic field, these particles retain the adiabatic invariant $\mathbf{p}_\perp^2/H_0$. Assuming that the magnetic field lines in the regular component of the interplanetary magnetic field diverge with increasing distance to the Sun, the number of particles with the pitch angle $\pi/2$ in the cosmic rays propagating toward the Sun increases and the $\delta_2(\vartheta)$ maximum is observed at 3 h LT. On the contrary, the number of such particles in the cosmic rays propagating in the opposite direction decreases and the $\delta_2(\vartheta)$ maximum corresponds to 9 h LT.

In Eq. (19) for the third moment $F_3^m(\mathbf{r}, \mathbf{p}, t)$, we will neglect the terms $\partial F^*_{3m}/\partial t$ and F_{4m}, which are smaller than F_3^m according to the experimental data [24]. Then, Eq. (19) simplifies to

$$F_3^k - i\frac{1}{\sqrt{3}}\frac{\Lambda}{R_0}\sum_{mn} C^{3k}_{3m1n} F_3^m h_0^k = -\sqrt{\frac{3}{7}}\frac{\Lambda}{6}\sum_{mn} C^{3k}_{1n2m}\nabla^k F_2^m.$$

In the coordinate system with $\mathbf{h}_0 \parallel z$, a solution to this equation is as follows:

$$F_3^k = -\frac{1}{\sqrt{70\pi}}\frac{\Lambda}{2}\left\{1 - ik\frac{\Lambda}{6}\frac{h_0}{R_0}\right\}^{-1}\sum_{mn} C^{3k}_{1n2m}\nabla^n f^m. \quad (30)$$

Substituting f^m from (26), we can determine F_3^k in terms of the current gradient $\mathbf{j}$ and the gradients of Λ and R_0. Using these F_3^k values, we can determine the total term related to the third spherical harmonic:

$$F_3(\mathbf{r}, \mathbf{p}, t) = \sum_m F_3^m(\mathbf{r}, p, t) Y_{3m}(\vartheta, \varphi).$$

Taking into account that the galactic cosmic rays propagating in interplanetary space obey the relationship $\operatorname{div}\mathbf{j} \approx uj/vR_0$ and taking $\varphi = 0$ (i.e., assuming the registrators to be oriented in the helioequator plane), we obtain (to within the second-order terms in R_0/Λ)

$$\begin{aligned} F_3(\mathbf{r}, \mathbf{p}, t) = {} & -\frac{3}{16}R_0\left[\frac{R_0}{\Lambda}\frac{\partial}{\partial x}(\Lambda I_0) + \frac{1}{6}h_0\frac{\partial}{\partial y}(\Lambda I_0)\right. \\ & \left. + 6\frac{R_0}{\Lambda}\frac{\partial}{\partial z}(R_0 I_4) + \frac{\partial}{\partial z}(R_0 I_3)\right]\sin(3\vartheta) \\ & + \frac{\Lambda}{192\pi}\frac{\partial}{\partial z}(\Lambda I_0)\cos(3\vartheta). \end{aligned}$$

For particles with $E = 10$ GeV, we may neglect the terms proportional to R_0/Λ and R_0^2/Λ^2 and obtain an estimate for the relative value of the third moment:

$$\delta_3(\vartheta) = \frac{F_3}{F_0} = \frac{\Lambda}{16N}\frac{\partial}{\partial z}\left(\Lambda\frac{\partial}{\partial z}j_z\right)\cos(3\vartheta). \quad (31)$$

Substituting

$$j_z \approx -0.0002N\frac{1\text{ a.u.}}{z},$$

we find that $\vartheta_{\max} = \pi/3$ and the maximum time is 5 h LT, in general agreement with the experimental data [21–24]. The relative amplitude of the 3rd harmonic for particles with the energies $E = 10$ GeV is on the order of $\delta_3 \approx 0.025\%$. This estimate is somewhat smaller than the values given in [21–24], but, taking into account a large scatter of the experimental data, the agreement can be considered as quite satisfactory.

In Eq. (20) for the fourth moment, we neglect the terms $\partial F^*_{4m}/\partial t$ and F_5^m, which are apparently smaller than F_4^m [24]. Then, Eq. (20) yields

$$\begin{aligned} & F_4^k - i\frac{1}{\sqrt{5}}\frac{\Lambda}{R_0}\sum_{mn} C^{4k}_{4m1n} F_4^m h_0^n \\ & = -\frac{\Lambda}{15}\sum_{mn} C^{4k}_{3m1n}\nabla^n F_3^m. \end{aligned} \quad (32)$$

In the coordinate system with $\mathbf{h}_0 \parallel z$, a solution to this equation is as follows:

$$F_4^k = -\frac{\Lambda}{15}\left\{1 - ik\frac{\Lambda}{10}\frac{h_0}{R_0}\right\}^{-1}\sum_{mn} C^{4k}_{3m1n}\nabla^n F_3^m. \quad (33)$$

Substituting F_3^m from (30) and f^m from (26), we can determine F_4^k in terms of the flux $\mathbf{j}$. Using these values, we can determine the total term related to the fourth spherical harmonic:

$$F_4(\mathbf{r}, \mathbf{p}, t) = \sum_m F_4^m(\mathbf{r}, \mathbf{p}, t) Y_{4m}(\vartheta, \varphi).$$

In the first approximation with respect to R_0/Λ for $\varphi = 0$, this yields

$$F_4(\mathbf{r}, \mathbf{p}, t) = \frac{9}{1340\pi}\Lambda\left\{\frac{\partial}{\partial y}\left(R_0\frac{\partial}{\partial x}\Lambda\frac{\partial j_z}{\partial z}\right) - \frac{\partial}{\partial x}\left(R_0\frac{\partial}{\partial y}\Lambda\frac{\partial j_z}{\partial z}\right) - \frac{\partial}{\partial z}\left[\frac{\Lambda}{3}\left(\frac{\partial}{\partial y}R_0\left(\frac{\partial}{\partial x}j_z + \frac{\partial}{\partial z}j_x\right) - \frac{\partial}{\partial x}R_0\left(\frac{\partial}{\partial y}j_z + \frac{\partial}{\partial z}j_y\right) + \frac{\partial}{\partial z}\left(\Lambda\frac{\partial j_z}{\partial z}\right)\right)\right]\right\}\cos(4\vartheta). \quad (34)$$

In the zeroth approximation with respect to R_0/Λ, the relative value of the fourth moment is

$$\delta_4(\vartheta) = \frac{F_4}{F_0} = -\frac{3\Lambda}{335N}\frac{\partial}{\partial z}\left(\Lambda\frac{\partial}{\partial z}\left(\Lambda\frac{\partial j_z}{\partial z}\right)\right)\cos(4\vartheta). \quad (35)$$

The relative amplitude of the 4th harmonic for the particles with $E = 10$ GeV and the flux

$$j_z \approx -0.002N\frac{1 \text{ a.u.}}{z}$$

can be estimated as $\delta_4 \approx 0.01\%$, in agreement with the experimental data [24]. The phase of the 4th harmonic estimated by formula (35) corresponds to a $\delta_4(\vartheta)$ maximum at 0 h LT, which was not reported in [24]. The maximum in $\delta_4(\vartheta)$ at 3 h LT observed in [24] is attained at

$$\frac{\partial}{\partial z}\left(\Lambda\frac{\partial}{\partial z}\left(\Lambda\frac{\partial}{\partial z}j_z\right)\right) < 0, \quad (36)$$

which corresponds to the flux

$$j_z \approx +0.002N\frac{1 \text{ a.u.}}{z}$$

directed outward from the Sun. Such fluxes were described in [43].

Phases of the higher harmonics $\delta_2(\vartheta)$, $\delta_3(\vartheta)$, and $\delta_4(\vartheta)$ coinciding with the experimental data (i.e., at 3, 5, and 3 h LT) [21–24] correspond to the following flux determined using the MATHCAD program package:

$$j_z = \left(0.12\left(\frac{1 \text{ a.u.}}{z}\right)^2 - 0.44\left(\frac{1 \text{ a.u.}}{z}\right)^{1.2}\right)0.01N. \quad (37)$$

In this case, the estimate for $\delta_1 \approx 1\%$ is somewhat greater than the experimental value, while the amplitudes of other harmonics ($\delta_2 \approx 0.2\%$, $\delta_3 \approx 0.03\%$, and $\delta_4 \approx 0.01\%$) are close to the experimental results [21–24]. The first positive term in (37) is mostly due to the projection of the convection flux directed radially outward from the Sun and parallel to the solar wind velocity. In addition, this term contains a contribution due to the radial diffusion component directed toward the Sun. The second (negative) term is due to the projection of the azimuthal diffusion flux of particles directed toward the Sun along the lines of a regular interplanetary magnetic field; the power z in this term reflects the radial dependence of the azimuthal component of the regular interplanetary magnetic field [45–47]. This separation of the fluxes into convection and diffusion is approximate, with the cross-term contribution being proportional to the 2nd harmonic amplitude.

The above structures of the solar wind and the magnetic field in the heliomagnetosphere agree with the concepts developed in [16] based on the direct measurements of the interplanetary plasma and magnetic field parameters performed by Ulysses spacecraft [10– 15].

Upon selecting the particle flux in the form (37), we conclude that the first term produces the main contribution to the third-order gradient of the j_z flux along z and to $\delta_4(\vartheta)$, indicating the correct time of the $\delta_4(\vartheta)$ maximum at 3 h LT. The main contribution to the gradients of lower orders of the j_z flux along z and the $\delta_2(\vartheta)$ and $\delta_3(\vartheta)$ values is from the second term, which also yields correct times of the $\delta_2(\vartheta)$ and $\delta_3(\vartheta)$ maxima at 3 and 5 h LT, respectively.

Thus, according to the proposed mechanism of the formation of the $\delta_3(\vartheta)$ harmonic and the $\delta_4(\vartheta)$ harmonic, the terms with $\cos(3\vartheta)$ in Eq. (31) and $\cos(4\vartheta)$ in Eq. (35) may contain either positive factors (corresponding to the maxima at 1 and 3 h LT, respectively) or negative ones (corresponding to the $\delta_3(\vartheta)$ and $\delta_4(\vartheta)$ maxima at 5 and 0 h LT, respectively). The two maxima in $\delta_3(\vartheta)$ were experimentally observed in [22], where the neutron monitors showed the 3rd harmonic maxima at 5 h LT for high solar activity and at 1 h LT for low solar activity. We may suggest that the flux of cosmic rays in the helioequator plane, contributing to the 3rd harmonic, is directed toward the Sun during the high activity period and outward from the Sun during the low activity period. It should be noted that the Hall current perpendicular to $\mathbf{h}_0$ does not contribute, even despite a considerable magnitude, to $\delta_2(\vartheta)$, $\delta_3(\vartheta)$, and $\delta_4(\vartheta)$ in this approximation [40].

In this study, in contrast to [30, 31], devoted to the 2nd harmonic, $\delta_2(\vartheta)$, $\delta_3(\vartheta)$, and $\delta_4(\vartheta)$ are believed to appear mostly as a result of variations in the geometry of the regular interplanetary magnetic field and the related adiabatic focusing of the particle flux of cosmic rays [21]. We consider the current j_z as given; that is, we believe that factors determining this current are outside the scope of considerations concerning the harmonics under study. This approach differs from the mechanism of the 2nd harmonic formation considered in [30, 31], where the flux $\mathbf{j}$ determined from the equation for the 1st and zeroth harmonics of the distribution function was substituted into the expression for $\delta_2(\vartheta)$. Thus, the appearance of $\delta_2(\vartheta)$ in [30, 31] was related to the spatial derivatives of zeroth to second order of the quantities u, Λ, $N(\mathbf{r}, p, t)$ multiplied by the ratio u/v (with the total power of the gradients and the u/v ratio equal to two). Owing to the smallness of these factors, the δ_2 value (on the order of u^2/v^2) obtained in [30, 31] was significantly smaller as compared to the experimental values.

In deriving formulas (29), (31), and (35) for the harmonics $\delta_2(\vartheta)$, $\delta_3(\vartheta)$, and $\delta_4(\vartheta)$, we ignored the higher multipole moments of the distribution function. This iterative procedure is based on the smallness of the 2nd, 3rd, and 4th harmonics of the diurnal variation relative to the first harmonic. According to the experimental data, each next higher harmonic is 2–5 times smaller than the preceding one [24]. In this study, the equations for multipole moments of the distribution function are treated in a somewhat broader sense than was done in [30, 31], since the lower moments in equations for the higher harmonics can be preset, for example, from experiment.

Let us consider the spectral energy characteristics of the diurnal variation harmonics using the data for 1971–1975 as well as the previous data [22–24, 48]. Taking into account the experimental data on the 1st harmonic [49], we assume $j_z/N \propto$ const to be constant and independent of p. Then, Eq. (31) yields

$$\delta_2 \propto \Lambda(p) \propto p^{0.5-2},$$

which approximately coincides with the existing notions of the 2nd harmonic spectrum [9, 48] and is close to the experimental data for cosmic ray particle energies $E < E_{cr}$, where E_{cr} is a certain critical energy determined at the point where the 2nd harmonic spectrum exhibits a break (the power of momentum p becomes negative). For the screening mechanism, E_{cr} is the energy of particles in the cosmic rays for which the Larmor radius becomes equal to half of the regular magnetic field layer thickness near the helioequator plane. Note that this mechanism describes well the δ_2 spectrum for particles with both $E < E_{cr}$ and $E > E_{cr}$.

Using Eqs. (31) and (35) and assuming $j_z/\mathrm{N} \propto$ const, we obtain the following estimates for the 3rd and 4th harmonics of the diurnal variation of cosmic rays for $E < E_{cr}$:

$$\delta_3 \propto \Lambda^2(p) \propto p^{1-4}, \quad \delta_4 \propto \Lambda^3(p) \propto p^{2-6}. \tag{38}$$

This spectrum is more rigid as compared to the experimental data, according to which $\delta_3 \propto p$ and $\delta_4 \propto p^{0.5}$ for $E < E_{cr}$ [21–24]. The difference of the predicted δ_2, δ_3, and δ_4 spectra from the experimental results and the inapplicability of this analysis to the case of $E > E_{cr}$ is apparently related to the fact that we employed an analog of the diffusion approximation for the higher harmonics. This approximation consists in ignoring a higher harmonic in the equation for the given harmonic. The applicability of the small-angle approximation, even with an allowance for the higher approximations with respect to the random magnetic field in the collisional integral, is also not quite clear.

For $E > E_{cr}$, the Larmor diameter becomes greater than the regular magnetic field layer thickness and the amplitudes of δ_2, δ_3, and δ_4 harmonics exhibit a sharp drop. The formulas for $\delta_2(\vartheta)$, $\delta_3(\vartheta)$, and $\delta_4(\vartheta)$ harmonics derived in this section can be refined using the gradients of R_0, Λ, and $\mathbf{j}$. However, the lack of reliable experimental data on these gradients hinders the use of such additional refining terms in the formulas for $\delta_2(\vartheta)$, $\delta_3(\vartheta)$, and $\delta_4(\vartheta)$.

The allowance for an arbitrary relationship of the type $\Lambda \propto p^\mu$ ($\mu < 2$) in the kinetic equation (1) yields additional terms in Eq. (18) for F_{2m} and, hence, in related Eqs. (19) and (20) for F_{3m} and F_{4m}. However, smallness of the ratios $u/v \sim 10^{-3}$ and $R_0/\Lambda \sim 10^{-1}$ allows these corrections in Eq. (18) to be omitted. This implies that the equation for F_{2m} without the last two terms, as well as Eqs. (19) and (20) for F_{3m} and F_{4m}, is valid for $\Lambda \propto p^\mu$ with $\mu < 2$.

4. ANNUAL CHANGES IN THE DIURNAL VARIATION HARMONICS RELATED TO THE EARTH'S AXIS TILT

Let us take into account the tilt of the Earth's axis relative to the helioequator plane and the annual changes in the phase and amplitude of the diurnal variation harmonics of the galactic cosmic rays. These annual changes are related to the change in the Earth's axis tilt relative to the direction toward the Sun in the plane of the neutral surface of the averaged large-scale interplanetary magnetic field, which is assumed to coincide with the helioequator plane. This implies that we will ignore the terms on the order of 10^{-2} in the multipole moments. Apparently, these terms have to be taken into account simultaneously with refining some other conditions, for example, concerning the deviation of the large-scale interplanetary magnetic field direction from the average one.

Mathematically, the problem reduces to separating factors depending only on the direction of vector $\mathbf{h}_0$ in each multipole moment F_{lm}. The coordinate systems used in this study coincide with those introduced in [3]. The transformation formulas employed below allow the conversion of multipole moments (second to fourth) to be described in a general manner and performed more simply as compared to the method used in [3, 8]. All calculations will be performed analytically, which allows the influence of particular physical parameters in each calculation stage to be followed more accurately and the results of calculations with modified physical parameters and distribution functions to be analyzed.

Let us conventionally assume $\mathbf{h}_0$ to be directed outward from the Sun, the opposite case being readily derived. Upon the coordinate system rotation, the function F_{lm} transforms as described by the Wigner D functions [41]:

$$F''_{lm}(\mathbf{r}'', p'', t) = \sum_k F^k_l(\mathbf{r}, p, t) D^{l*}_{km}(\alpha, \beta, \gamma), \tag{39}$$

where $F''_{lm}(\mathbf{r}'', p'', t)$ are the multipole moments of the distribution function multiplied by $Y_{lm}(\vartheta'', \varphi'')$ in the rotated coordinate system; α, β, and γ are the Euler angles [41]. The coordinate system rotation will be performed according to scheme A (see [41]). The transition from the coordinate system with the axis $z \parallel \mathbf{h}_0$ to the geographic system of coordinates x'', y'', z'' related to the Earth is conveniently performed in two steps as described in [3]. It is assumed that the regular interplanetary magnetic field is oriented at an angle of ϑ_0 relative to the direction toward the North Pole and at an angle of φ_0 relative to the direction outward from the Sun in the helioequator plane.

In the first step, we perform the transformation from the coordinate system with the axis $z \parallel \mathbf{h}_0$ (x axis facing the South Pole, y axis lying in the helioequator plane) to the coordinate system with the z' axis directed toward the North Pole (parallel to the Sun's rotation axis) and the x' axis directed radially outward from the Sun in the helioequator plane.

The second rotation transforms the coordinate system to that with the z'' axis coinciding with the Earth's axis (tilted at an angle $\delta = 23.5°$ relative to the z' axis). The x'' axis is directed outward from the Sun and the y'' axis is perpendicular to the direction toward the Sun.

A matrix of the resultant complex rotation is as follows:

$$\begin{aligned} & D^{l*}_{mn}(\alpha, \beta, \gamma) \\ & = \sum_k D^{l*}_{mk}(0, -\vartheta_0, -\varphi_0) D^{l*}_{kn}(\Phi, \delta, -\Phi), \end{aligned} \tag{40}$$

where ϑ_0 and φ_0 are the polar and azimuthal angles, respectively, determining the direction of vector $\mathbf{h}_0$ in the x', y', z system, and Φ is the angle determining the direction of the axis of rotation of the z' axis by the angle δ ($\Phi = -2\pi t/T$, where T is the month number). Using these definitions and the formulas from [41], the Euler angles α, β, and γ can be determined from the following relationships:

$$\begin{aligned} \cot\alpha &= \cos\vartheta_0 \cot(\Phi - \vartheta_0) - \cot\delta \frac{\sin\vartheta_0}{\sin(\Phi - \varphi_0)}, \\ \cos\beta &= \cos\delta\cos\vartheta_0 + \sin\delta\sin\vartheta_0\cos(\Phi - \varphi_0), \\ \cot(\gamma + \Phi) &= \cos\delta\cot(\Phi - \varphi_0) - \cot\vartheta_0 \frac{\sin\delta}{\sin(\Phi - \varphi_0)}. \end{aligned} \tag{41}$$

The function $D^{l*}_{mn}(\alpha, \beta, \gamma)$ in matrix (40) can be represented as a product of three factors, each depending on a single Euler angle [41]:

$$D^{l*}_{mn}(\alpha, \beta, \gamma) = \exp(im\alpha) d^l_{mn}(\beta) \exp(in\gamma), \tag{42}$$

where $d^l_{mn}(\beta)$ are real functions derived in the explicit form in [41] and in [40, Appendix 9].

The reverse transformations of the multipole moments $F''_{lm}(\mathbf{r}'', p'', t)$ from the geographic coordinate system to the system with $z \parallel \mathbf{h}_0$ are described by formula (39) using the inverse Wigner D functions:

$$[D^{l*}_{mn}(\alpha, \beta, \gamma)]^{-1} = D^l_{nm}(\alpha, \beta, \gamma). \tag{43}$$

Using this method, the transition to reverse transformations is performed much more simply as compared to the procedure employing Cartesian coordinates [3, 9]. Obviously, the sequential study of harmonics of the diurnal variation in the intensity of cosmic rays should be based on determination of the multipole moments F''_{lm} from the experimental data in the coordinate system x'', y'', z'' followed by recalculation to the coordinate system with $z \parallel \mathbf{h}_0$ using the inverse Wigner functions (43).

Upon solving the kinetic equation in the coordinate system with $z \parallel \mathbf{h}_0$ jointly with the transformation formula (39) using the Wigner functions $D^{l*}_{mk}(0, -\vartheta_0, -\varphi_0)$, we obtain a solution to the kinetic equation (1) with an arbitrary orientation of the regular interplanetary magnetic field $\mathbf{h}_0(\vartheta_0, \varphi_0)$. Using the above formulas, we can find the annual changes in the amplitude and phase of the 2nd, 3rd, and 4th harmonics of the diurnal variation in the intensity of galactic cosmic rays, which are related to the Earth's axis tilt relative to the helioequator plane. For this purpose, let us assume that $\mathbf{h}_0$ is oriented so as to make the angles $\vartheta_0 = \pi/2$ and $\varphi_0 = -\pi/4$ in the coordinate system x', y', z' related to the Sun.

The 2nd harmonic of diurnal variations in the coordinate system with $z \parallel \mathbf{h}_0$ can be expressed in the form of Eq. (29):

$$F_2 = F_{20} Y_{20}(\vartheta, \varphi),$$

where

$$F_{20} = -a_2 \frac{2}{3} \frac{1}{\sqrt{5\pi}} N, \quad a_2 = \frac{3}{4} \frac{\Lambda}{N} \frac{\partial j_z}{\partial z}, \quad a_2 > 0,$$

and a_2 is the relative amplitude of the 2nd harmonic. The constant term (independent of angles) appearing in this representation yields a small contribution to the isotropic component and is ignored. In the geographic coordinate system x'', y'', z'', the 2nd harmonic of diurnal variations acquires the following form:

$$F_2''(\vartheta'', \varphi'') = -a_2\frac{N}{4\pi}[\sin^2\beta_0\sin^2\vartheta''\cos 2(\varphi''+\gamma_0) - 4\sin\beta_0\cos\beta_0\sin\vartheta''\cos\vartheta''\cos(\varphi''+\gamma_0) + f_2], \tag{44}$$

where f_2 is a term independent of the angle φ'' and, hence, not contributing to the diurnal variation harmonics. The second term in (44) contributes to the 1st harmonic of diurnal variations. The angles β_0 and γ_0 are as follows (see Eqs. (A8.18) and (A8.19) in [40]):

$$\begin{gathered}\beta_0 = -\frac{\pi}{2} + \sin\delta\cos\left(\Phi + \frac{\pi}{2}\right),\\ \gamma_0 = \frac{\pi}{4} + \frac{1}{4}\cos\delta\sin^2\delta\sin\left(2\left(\Phi + \frac{\pi}{4}\right)\right).\end{gathered} \tag{45}$$

The 3rd harmonic of diurnal variations in the coordinate system with $z \parallel \mathbf{h}_0$ can be presented in the form of Eq. (31):

$$F_3 = -a_3\frac{N}{4\pi}\cos(3\vartheta) \approx F_{30}Y_{30}(\vartheta, \varphi), \tag{46}$$

where

$$F_{30} = -a_3\frac{4}{5}\frac{1}{\sqrt{7\pi}}N, \quad a_3 = -\frac{1}{16}\frac{\Lambda}{N}\frac{\partial}{\partial z}\left(\Lambda\frac{\partial j_z}{\partial z}\right),$$

$$a_3 > 0.$$

The contribution of 3rd to 1st harmonic will be ignored. Using the formulas (39) describing the transformation of multipole moments $F_{lk}(\mathbf{r}, p, t)$, we obtain an expression for the 3rd harmonic in the geographic coordinate system:

$$\begin{aligned}F_3''(\vartheta'', \varphi'') = -a_3\frac{N}{4\pi}\Big[&-\sin^3\beta_0\sin^3\vartheta''\cos(3(\varphi''+\gamma_0))\\ &+ 6\sin^2\beta_0\cos\beta_0\sin^2\vartheta''\cos\vartheta''\cos(2(\varphi''+\gamma_0))\\ &+ \frac{3}{5}\sin\beta_0(1-5\cos^2\beta_0)\\ &\times(5\cos\vartheta''-1)\sin\vartheta''\cos(\varphi''+\gamma_0) + f_3\Big],\end{aligned} \tag{47}$$

where f_3 is a term independent of the angle φ'', which makes only a small contribution to the isotropic component and, hence, can be ignored. The second term contributes to the 2nd harmonic, and the third term, to the 1st harmonic.

The 4th harmonic of diurnal variations in the coordinate system with $z \parallel \mathbf{h}_0$ can be presented in the form of Eq. (35):

$$F_4 = a_4\frac{N}{4\pi}\cos(4\vartheta) \approx F_{40}Y_{40}(\vartheta, \varphi), \tag{48}$$

where $a_4 > 0$. With an allowance for all the approximations introduced in Section 3, we obtain

$$F_{40} = a_4\frac{32}{105\sqrt{\pi}}N,$$

$$a_4 = -\frac{3}{335}\frac{\Lambda}{N}\frac{\partial}{\partial z}\left(\Lambda\frac{\partial}{\partial z}\left(\Lambda\frac{\partial j_z}{\partial z}\right)\right).$$

In this representation, the contribution of the 4th harmonic to the 2nd harmonic is on the order of $0.1a_4N$, and that to the isotropic component, on the order of $0.05a_4N$. Taking into account small accuracy of determination of the 2nd harmonic [18–24], these contributions can be ignored.

Using formulas (39) and (40) describing a transition to the geographic coordinate system, we obtain

$$\begin{aligned}F_4''(\vartheta'', \varphi'') = a_4\frac{N}{4\pi}\Big[&\sin^4\beta_0\sin^4\vartheta''\cos(4(\varphi''+\gamma_0))\\ &- 8\sin^3\beta_0\cos\beta_0\sin^3\vartheta''\cos\vartheta''\cos(2(\varphi''+\gamma_0))\\ &- \frac{4}{7}\sin^2\beta_0(1-7\cos^2\beta_0)(7\cos^2\vartheta''-1)\\ &\times\sin^2\vartheta''\cos(2(\varphi''+\gamma_0)) + \frac{8}{7}\sin\beta_0\cos\beta_0(3-7\cos^2\beta_0)\\ &\times(7\cos^2\vartheta''-3)\sin\vartheta''\cos\vartheta''\cos(\varphi''+\gamma_0) + f_4\Big],\end{aligned} \tag{49}$$

where the term f_4 contributes only slightly to the isotropic component of N and can be ignored. The second, third, and fourth terms contribute to the 3rd, 2nd, and 1sr harmonics, respectively.

As can be see from the above formulas (44), (47), and (49) for the 2nd, 3rd, and 4th harmonics, respectively, the highest harmonics of diurnal variations in the coordinate system with $z \parallel \mathbf{h}_0$ contribute to the lower harmonics because of the Earth's axis tilt relative to the helioequator plane. A change in the Earth's axis tilt relative to the direction toward the Sun accounts for annual changes in the relative contributions to harmonics.

An especially large contribution, on the order of a_2, is from 2nd to 1st harmonic. Taking into account close values of the 1st and 2nd harmonics, this contribution can be very significant [8, 27]. In addition, the phase of diurnal harmonics changes by $0.15m''$ h (where m'' is the harmonic number) with a period of 0.5 year, and the amplitude of diurnal harmonics is modulated with a period of one year. The sign of the terms contributing to some lower harmonics also exhibits a change with a period of one year. A modulation of the uppermost harmonics related to a change in the Earth's axis tilt relative to the direction toward the Sun is rather small, amounting to 1/5 (or less) of the higher harmonic amplitude.

5. A DIFFUSION EQUATION FOR THE PARTICLE CONCENTRATION IN COSMIC RAYS WITH AN ALLOWANCE FOR THE SECOND HARMONIC

It was of interest to derive the equation of diffusion for the concentration of particles $N(\mathbf{r}, p, t)$ with an allowance for the second multipole moment in the expansion of the distribution function $F(\mathbf{r}, \mathbf{p}, t)$ in series with respect to the momentum angles. Previously [1, 42], this was done so as to take into account a small anisotropic additive to the distribution function. However, the method of averaging used in [1, 42] was applicable mostly to low-energy particles, for which the Larmor radius in a regular magnetic field is small as compared to the correlation length of the random magnetic field ($R_0 \ll L_c$) and the particle motion is one-dimensional. We will consider the case of $R_0 > L_c$, in which an allowance for the second harmonic corresponds to taking into account the following orders in the parameter $\Lambda/\Delta L_1$ (where ΔL_1 is the characteristic scale of variation of the system parameters for $R_0 \longrightarrow \infty$) or in the parameter $R_0/\Delta L_1$ for $R_0 < \Lambda$.

In order to derive the required diffusion equation taking into account the 2nd harmonic, it is convenient to use Eq. (21). Note, however, that both right and left parts of Eq. (21) contain terms of different order of smallness. If we ignore the term $\partial N/\partial t$, the remaining term in the left-hand part is on the order of

$$v \operatorname{div} \mathbf{j} \sim \frac{cj}{1\ \text{a.u.}},$$

which is related to the divergence of the $\mathbf{H}_0$ field lines in the helioequator plane and to the effect of the magnetic field upon the charged particles. At the same time, the total left-hand part is

$$v \operatorname{div} \mathbf{j} \sim \frac{uj}{R_0},$$

which follows from the right-hand part of Eq. (21). Thus, the right-hand part of Eq. (21) contains terms on the order of 10^{-2}–10^{-3} relative to the maximum terms in the left-hand part of this equation. For this reason, we substitute the diffusion current $\mathbf{i}$ into the term $v \operatorname{div} \mathbf{j}$ with an allowance for the 2nd harmonic given by formulas (23) and (24), while using the same diffusion current $\mathbf{i}$ without the 2nd harmonic correction in the right-hand part (that is, we omit the second term in formula (24) for q^k).

It can be shown that the law of conservation of the number of particles in this approximation is retained. For this purpose we will use the method employed in [1, 50], extending this approach to the case of an arbitrary dependence of Λ on p. Substituting the diffusion current $\mathbf{i}$ into the right-hand part of Eq. (21) with neglect of the 2nd harmonic and using the relationship

$$\begin{aligned} p\frac{\partial}{\partial p}\chi_{km} &= -\chi_{km}\frac{p^2}{\Lambda}p\frac{\partial}{\partial p}\left(\frac{\Lambda}{p^2}\right) \\ &+ 2\chi_{km}\left(1+\frac{\Lambda^2}{R_0^2}\right)^{-1}\left(1+\frac{p^2}{\Lambda}p\frac{\partial}{\partial p}\left(\frac{\Lambda}{p^2}\right)\right) \\ &- \frac{\Lambda}{3}\left(1+\frac{\Lambda^2}{R_0^2}\right)^{-1}\left[2\frac{\Lambda^2}{R_0^2}h_{0k}h_{0m}\left(1+\frac{p^2}{\Lambda}p\frac{\partial}{\partial p}\left(\frac{\Lambda}{p^2}\right)\right)\right. \\ &\left. - \frac{\Lambda}{R_0}\varepsilon_{knm}h_{0n} - \frac{p}{\Lambda}p\frac{\partial}{\partial p}\left(\frac{\Lambda}{p^2}\right)\varepsilon_{knm}h_{0n}\right], \end{aligned} \tag{50}$$

where ε_{knm} is the fully symmetric Levi-Civita unit tensor and

$$\begin{aligned} \chi_{km} &= -\frac{\Lambda}{3}\left(1+\frac{\Lambda^2}{R_0^2}\right)^{-1} \\ &\times\left(\delta_{km}+\frac{\Lambda^2}{R_0^2}h_{0k}h_{0m}+\frac{\Lambda}{R_0}\varepsilon_{kmn}h_{0n}\right), \end{aligned} \tag{51}$$

we can represent the right-hand part of Eq. (21) in the following form:

$$\begin{aligned} &-\left(\mathbf{u}\cdot\frac{\partial}{\partial \mathbf{r}}\right)N - \frac{p}{3}\left(\mathbf{u}\cdot\frac{\partial}{\partial \mathbf{r}}\right)\frac{\partial}{\partial p}N \\ &= -\frac{1}{3p^2}\frac{\partial}{\partial p}\left(p^3\left(\mathbf{u}\cdot\frac{\partial}{\partial \mathbf{r}}\right)N\right). \end{aligned} \tag{52}$$

Substituting this expression into Eq. (21), multiplying both parts by p^2, and integrating over the entire momentum space, we obtain

$$\begin{aligned} \frac{\partial N^*}{\partial t} + v \operatorname{div} \mathbf{j}^* &= -\frac{1}{3} \\ \times\left[p^3\left(\mathbf{u}\cdot\frac{\partial}{\partial \mathbf{r}}\right)N^*\right|_{p=\infty} &- \left.p^3\left(\mathbf{u}\cdot\frac{\partial}{\partial \mathbf{r}}\right)N^*\right|_{p=0}\Bigg], \end{aligned} \tag{53}$$

where

$$N^*(\mathbf{r}, t) = \int_0^\infty dp\, p^2 N(\mathbf{r}, p, t),$$

$$\mathbf{j}^*(\mathbf{r}, t) = \int_0^\infty dp\, p^2 \mathbf{j}(\mathbf{r}, p, t).$$

Taking into account data on the spectrum of cosmic rays, we find that the right-hand part of Eq. (53) is zero irrespective of the current substituted into the left-hand part of this equation. As can be seen, the asymptotic spectrum of concentration for the cosmic rays $N \propto p^{-\gamma}$ is limited by the condition $\gamma > 3$.

Substituting into $v\operatorname{div}\mathbf{j}$ an expression for the current taking into account the 2nd harmonic according to formulas (23) and (24) and expressing in (21) the 2nd harmonic through the 1st one as in (26), we obtain the diffusion equation in the coordinate system with $z \parallel \mathbf{h}_0$:

$$\begin{aligned}\frac{\partial N}{\partial t} - \frac{\partial}{\partial x_\alpha} v\chi_{\alpha\beta}\frac{\partial N}{\partial x_\beta} + \mathbf{u}\cdot\frac{\partial N}{\partial \mathbf{r}} - \operatorname{div}\mathbf{u}\frac{p}{3}\frac{\partial N}{\partial p} \\ = \left[-\frac{2}{5}v\sum_{mpq} C^{2n}_{1q1p}\frac{\partial}{\partial x_p^*}\left(1+\frac{\Lambda^2}{R_0^2}\right)^{-1}\right. \\ -\frac{2}{5}v\sum_{mq} C^{2m}_{1q10}\frac{\partial}{\partial x_0}\left(1+\frac{R_0^2}{\Lambda^2}\right)^{-1} \\ \left. + i\frac{2}{5}v\sum_{mpn} C^{2m}_{1q1n}\frac{\partial}{\partial x_n^*}\frac{n\Lambda}{\Lambda/R_0 + R_0/\Lambda}\right]\frac{\partial}{\partial x_q^*}b_m,\end{aligned} \tag{54}$$

where the coefficients b_m are determined by the formula

$$b_m = \frac{\Lambda}{3}\left(1 - im\frac{\Lambda}{R_0}\right)^{-1}\sum_{qn} C^{2m}_{1q1n}\nabla^q j^n,$$

with neglect of the second moment of the current. Thus, we obtain a closed system of equations with respect to $N(\mathbf{r}, p, t)$. Let us study a solution to this equation in a simple spherically symmetric case for the particles of cosmic rays of sufficiently large energies.

According to [5–9, 51, 52], a strong regular magnetic field with a wavy zero surface possesses a nearly spherically symmetric shape. The zero layer thickness is apparently very small. Therefore, the main contribution to a change in the concentration of cosmic rays $N(\mathbf{r}, p, t)$ for $E > 100$ GeV is related to the regions of the heliomagnetosphere at sufficiently high latitudes, where the interplanetary magnetic field possesses a spherically symmetric structure for high-energy particles [12, 16]. This approximation is even more valid during the maximum solar activity [8], when the interplanetary magnetic field is even more “disordered.”

We will take $H_0 = 0$ and assume that $\Lambda = \text{const}$ and $u = \text{const}$ in the modulation region. Let us perform a transformation from the Clebsch–Gordan coefficients to the 3*jm* symbols by the formulas [41]

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = (-1)^{j_3+m_2+2j_1}\frac{1}{\sqrt{2j_3+1}}C^{j_3 m_3}_{j_1 - m_1 j_2 - m_2}.$$

We use a relationship for the sum containing products of the 3*jm* symbols [41],

$$\begin{aligned}\sum_k (-1)^{q-k}\begin{pmatrix} a & b & q \\ \alpha & \beta & -k \end{pmatrix}\begin{pmatrix} q & d & c \\ k & \delta & \gamma \end{pmatrix} \\ = (-1)^{2a}\sum_{x\xi}(-1)^{x-\xi}(2x+1) \\ \times\begin{pmatrix} a & b & x \\ \alpha & \gamma & -\xi \end{pmatrix}\begin{pmatrix} x & d & b \\ \xi & \delta & \beta \end{pmatrix}\begin{Bmatrix} b & d & x \\ c & a & q \end{Bmatrix},\end{aligned} \tag{55}$$

where $\begin{Bmatrix} b & d & x \\ c & a & q \end{Bmatrix}$ is the 6*jm* symbol [41], and represent the right-hand part of Eq. (54) in the form

$$\frac{4}{45}v\Lambda^2\Delta\operatorname{div}\mathbf{j},$$

where Δ is the Laplace operator. The complete diffusion equation (54) for $N(\mathbf{r}, p, t)$ in the spherically symmetric case is as follows:

$$\begin{aligned}\frac{3}{v}\frac{\partial N}{\partial t} - \Lambda\frac{\partial^2 N}{\partial r^2} - \frac{2\Lambda}{r}\frac{\partial N}{\partial r} + \frac{3u}{v}\frac{\partial N}{\partial r} \\ -\frac{2up}{vr}\frac{\partial N}{\partial p} = \frac{4}{45}\Lambda^3\frac{\partial^4 N}{\partial r^4} + \frac{16}{45r}\Lambda^3\frac{\partial^3 N}{\partial r^3}.\end{aligned} \tag{56}$$

Let us consider the stationary case and set

$$\frac{\partial N}{\partial t} = 0, \quad N \propto p^{-\gamma}, \quad \gamma = 4.5.$$

We also assume that the concentration of particles in the cosmic rays at the boundary of the modulation region at a distance r_0 from the Sun is N_0 and introduce the quantities

$$y' = \tau - \tau_0, \quad \tau = \frac{3ur}{v\Lambda}, \quad \tau_0 = \frac{3ur_0}{v\Lambda}.$$

The N/N_0 ratio can be presented in the following form:

$$\frac{N}{N_0} = 1 + q_1 y' + q_2 y'^2 + q_3 y'^3 + q_4 y'^4. \tag{57}$$

Substituting this expression into Eq. (56) and equating the coefficients at equal powers of y', we can write a solution to Eq. (56) with an allowance for $u/v \ll 1$. Retaining only the first three terms in expansion (57), we obtain

$$\begin{aligned}\frac{N}{N_0} = 1 + \frac{2\gamma}{3}(\tau - \tau_0) \\ \times\left[2 - \tau_0 + \frac{\tau_0}{3}\left(1 + \frac{2\gamma}{3}\right) + \frac{16u^2}{3v^2}\right]^{-1}\end{aligned}$$

$$+\frac{\gamma}{3}\left(1+\frac{2\gamma}{3}\right)(\tau-\tau_0)^2 \tag{58}$$

$$\times\left[2-\tau_0+\frac{\tau_0}{3}\left(1+\frac{2\gamma}{3}\right)+\frac{16u^2}{3v^2}\right]^{-1}$$

$$\times\left(3-\frac{\gamma-3}{6}\tau_0+\frac{6u^2}{v^2}\right)^{-1}.$$

As can be readily seen, the cosmic rays with $E \gtrsim$ 100 GeV are characterized by $\tau \ll 1$, and, hence, there is no need in writing the following terms in this expansion.

Making allowance for the 2nd harmonic increases the relative particle concentration in the cosmic rays by an amount on the order of the relative value of the 2nd harmonic, which, for the spherically symmetric modulation region and $\Lambda = \text{const}$, is on the order of u^2/v^2 [30, 31].

6. DISCUSSION OF RESULTS AND CONCLUSION

Using a kinetic equation with a small-scale collisional integral taking into account higher approximations with respect to the random magnetic field and applying methods of the quantum theory of the angular moment [41], we have derived and solved the equations for higher multipole moments of the distribution function in the space of momentum angles. The final formulas for the relative values of the higher moments were obtained in the iteration approximation with a smallness parameter equal to the ratio of amplitudes of a given higher harmonic and the preceding (lower) harmonic. According to the experimental data [21–24], this ratio typically varies from 1/5 to 1/2.

It was demonstrated that the experimentally observed higher harmonics of diurnal variations of the intensity of galactic cosmic rays can be explained by changes in the total density of two fluxes of the cosmic rays. The first is the diffusion flux of cosmic rays exhibiting adiabatic focusing in the divergent lines of a regular interplanetary magnetic field. This flux produces the main contribution to the second and third harmonic of diurnal variations and can form outside the space region under consideration. The second is the radial convection flux of cosmic rays directed outward from the Sun, which mostly contributes to the fourth harmonic of diurnal variations.

The radial fluxes are mutually compensating, and the resultant azimuthal flux yields the diurnal variation at 18 h LT. This pattern is generally qualitative, since the conditions of diffusion approximation for the galactic cosmic rays are violated in circumterrestrial space. The pattern is in agreement with the previously existing notions and with the structure of fluxes derived from experimental data on the spectrum, amplitude, phase, and long-period changes in the 1st harmonic of diurnal variations of the galactic cosmic rays related to the solar activity phase [3, 4, 8, 9].

It was shows that the galactic cosmic rays can frequently be considered using the approximation of the particles moving in a regular magnetic field. The radial variation of the radial and azimuthal components of the regular interplanetary magnetic field and the approximated spherically symmetric structure of the solar wind velocity agree with the experimental data on the harmonics of diurnal variations. The relative stability of the observed amplitudes and phases of the higher harmonics and a strong sensitivity of harmonics calculated by formulas (29), (31), and (35) with respect to the relative amplitudes and radial dependence of the galactic cosmic ray fluxes are evidence of a certain stationary character of the cosmic ray fluxes in the heliomagnetosphere and of the presence of a turbulent zone in the transition region between the heliomagnetosphere and interstellar medium [53]. A sufficiently large radial convection flux of cosmic rays near the Earth's orbit is indicative of an approximate validity of the conditions of diffusion approximation in the radial direction, that is, of the smallness of the transverse component relative to the regular diffusion field, which can be related to the anisotropy and fibrous structure of the random interplanetary magnetic field [3, 54].

Data on the variation of muonic activity at a depth of 0 m, related to variations in the intensity of cosmic rays with energies on the order of 50 GeV for the minimum solar activity period of 1971–1975 were reported in [43]. It was demonstrated that the maxima of j_z in that period fall within the 6th ($j_z < 0$) and 12th ($j_z > 0$) months, as well as the maxima in the 2nd harmonic observed at the unchanged time of 3 h LT. The constancy of the time of maximum in $d_2(\theta)$ implies a change in the direction of decrease in the density of cosmic rays. Assuming the charged particles in the cosmic rays to be trapped in a magnetic field, we may suggest that, on one side of the Sun, the magnetic field lines near the helioequator plane converged on the average to this plane, whereas, on the other side of the Sun, these lines diverged on the scale of the order of 0.4 a.u. A possible reason for this can be the interaction between the heliomagnetosphere and the magnetic field of the Galactic, with redistribution of the interplanetary and interstellar magnetic field lines [53]. The 3rd harmonic maxima, determined from the results of observations using neutron monitors [22] for the transverse (relative to the regular magnetic field) scales on the order of 0.05 a.u., correspond to 5 h LT in the period of high solar activity and 1 h LT during the low solar activity period. Thus, we may suggest that, in the former case, the cosmic ray flux producing a maximum contribution to the 3rd harmonic was directed toward the Sun (that is, the main contribution to the 3rd harmonic was described by the second term in Eq. (37)). In the latter case, the cosmic ray flux was directed predominantly outward from the Sun and the main contribution to the

3rd harmonic was that described by the first term in Eq. (37). This can be related to a change in geometry of the regular interplanetary magnetic field.

A correct time of the maximum in the 4th harmonic of diurnal variations in the intensity of galactic cosmic rays is provided by the radial component of the convection current of cosmic rays directed outward from the Sun.

As can be seen from the formulas (44), (47), and (49) describing (in the geographic coordinate system) the second, third, and fourth moments of the distribution function, the highest moment of the distribution function in the system with $z \parallel \mathbf{h}_0$ contributes, due to the Earth's axis tilt relative to the helioequator plane, to the lower moments in the geographic coordinate system. A change in the Earth's axis tilt relative to the direction toward the Sun accounts for the annual variations in the relative contribution to the moments of the distribution function and to the harmonics of diurnal variations. Taking into account smallness of the angle between the Earth's axis and the direction perpendicular to the ecliptic plane ($\delta = 23.5°$), we can determine the character of the main contribution to the moments of the distribution function in the geographic coordinate system. According to this, the phase of the diurnal harmonics exhibits a change on the order of 0.4 h with a period of 0.5 year and a small amplitude modulation with a period of one year. The sign of the terms contributing to some lower harmonics also exhibits a change with a period of one year. A sufficiently large contribution to the 1st harmonic is from the 2nd harmonic. A modulation of the uppermost harmonics related to a change in the Earth's axis tilt relative to the direction toward the Sun is rather small, amounting to 1/5 (or less) of the higher harmonic amplitude.

We have derived a diffusion equation using a solution to the equation for the second moment of the distribution function in a system of coordinates with $z \parallel \mathbf{h}_0$. The new equation was solved in the stationary case for a spherically symmetric field with $\mathbf{H}_0 = 0$. This approximation is applicable to cosmic rays with energies above 100 GeV in the period of maximum solar activity. An allowance for the second moment of the distribution function in a spherically symmetric case leads to an increase in the relative concentration of particles by an amount on the order of the relative magnitude of the 2nd harmonic, which is on the order of u^2/v^2 for $\Lambda =$ const in a spherically symmetric modulation region.

ACKNOWLEDGMENTS

The author is grateful to A.Z. Dolginov, D.A. Varshalovich, and I.N. Toptygin for fruitful discussions.

This study was supported by the Russian Foundation for Basic Research, project no. 00-02-17553.

REFERENCES

1. I. N. Toptygin, *Cosmic Rays in Interplanetary Magnetic Fields* (Nauka, Moscow, 1983; Reidel, Dordrecht, 1985).
2. A. M. Bykov and I. N. Toptygin, Usp. Fiz. Nauk **163** (11), 19 (1993) [Phys. Usp. **36**, 1020 (1993)].
3. G. F. Krymskiĭ, A. I. Kuz'min, P. A. Krivoshapkin, *et al.*, *Cosmic Rays and Solar Wind* (Nauka, Novosibirsk, 1981).
4. L. I. Dorman, *The Experimental and Theoretical Principles of the Astrophysics of Cosmic Rays* (Nauka, Moscow, 1975).
5. J. G. Ables, K. G. Mc Cracen, and U. R. Rao, in *Proceedings of the 9th International Conference on Cosmic Rays, ICCR, London, 1965*, Vol. 1, p. 208.
6. A. I. Kusmin, G. F. Krymsky, A. M. Altukhov, *et al.*, in *Proceedings of the 9th International Conference on Cosmic Rays, ICCR, London, 1965*, Vol. 1, p. 501.
7. P. A. Krivoshapkin, G. F. Krymskiĭ, A. I. Kuz'min, and G. V. Skripin, Geomagn. Aéron. **9**, 228 (1969).
8. A. I. Kuz'min, *Variation of Cosmic Rays and Solar Activity* (Nauka, Moscow, 1968).
9. *Distribution of Galactic Cosmic Rays and Dynamics of Structural Formations in Solar Wind*, Ed. by A. I. Kuz'min (Yakutskiĭ Fil. Sib. Otd. Akad. Nauk SSSR, Yakutsk, 1973).
10. J. L. Phillips, S. J. Bame, A. Barnes, *et al.*, Geophys. Res. Lett. **22**, 3301 (1995).
11. J. L. Phillips, S. J. Bame, A. Barnes, *et al.*, Geophys. Res. Lett. **22**, 3305 (1995).
12. E. J. Smith and A. Balogh, Geophys. Res. Lett. **22**, 3317 (1995).
13. E. J. Smith and R. G. Marsden, Geophys. Res. Lett. **22**, 3297 (1995).
14. E. J. Smith, R. G. Marsden, and D. E. Page, Science **268**, 1005 (1995).
15. J. L. Phillips, S. J. Bame, W. C. Feldman, *et al.*, Science **268**, 1030 (1995).
16. I. S. Veselovskiĭ and O. A. Panasenko, Izv. Ross. Akad. Nauk, Ser. Fiz. **62**, 1819 (1998).
17. J. J. Quenbi and B. Lietti, Planet. Space Sci. **16**, 1209 (1968).
18. R. P. Kane, J. Geophys. Res. **80**, 470 (1975).
19. H. S. Ahluwalia and M. M. Fikani, J. Geophys. Res. **101**, 11 075 (1996).
20. H. S. Ahluwalia and M. M. Fikani, J. Geophys. Res. **101**, 11 087 (1996).
21. J. W. Bieber, M. A. Pomerantz, and G. H. Tsao, in *Proceedings of the 18th International Cosmic Ray Conference, ICRC, Bangalore, 1983*, Vol. 3, p. 289.
22. T. Kanno, Y. Ishida, and T. Saito, in *Proceedings of the 14th International Cosmic Ray Conference, ICRC, Munich, 1975*, Vol. 4, p. 1231.
23. K. Nagashima, Z. Fujii, K. Fujimoto, *et al.*, in *Proceedings of the 15th International Cosmic Ray Conference, ICRC, Plovdiv, 1977*, Vol. 4, p. 72.
24. K. Nagashima, J. Kondo, Z. Fuyii, and K. Fujimoto, in *Proceedings of the 15th International Cosmic Ray Conference, ICRC, Plovdiv, 1977*, Vol. 4, p. 78.
25. A. Z. Dolginov and I. N. Toptygin, Icarus **3**, 54 (1968).

26. L. I. Dorman and M. E. Katz, Space Sci. Rev. **20**, 529 (1977).
27. Yu. P. Mel'nikov, Geomagn. Aéron. **24**, 371 (1984).
28. Yu. P. Mel'nikov, Geomagn. Aéron. **33**, 18 (1993).
29. Yu. P. Mel'nikov, Zh. Éksp. Teor. Fiz. **109**, 1599 (1996) [JETP **82**, 860 (1996)].
30. L. I. Dorman, M. E. Kats, and Yu. I. Fedorov, in *Proceedings of the IX Leningrad Workshop on Cosmophysics, Leningrad, 1977* (Leningr. Inst. Yad. Fiz., Leningrad, 1978), p. 338.
31. L. I. Dorman, M. E. Kats, and Yu. I. Fedorov, Izv. Akad. Nauk SSSR, Ser. Fiz. **43**, 2558 (1979).
32. L. I. Dorman and S. Fisher, Kosm. Luchi, No. 8, 88 (1967).
33. L. I. Dorman, A. A. Luzov, and V. P. Mamrukova, Dokl. Akad. Nauk SSSR **172**, 833 (1967).
34. E. Antonucci and D. Marocci, J. Geophys. Res. **81**, 4627 (1976).
35. P. A. Krivoshapkin, G. F. Krymskiĭ, A. I. Kuz'min, and G. V. Skripin, in *Distribution of Galactic Cosmic Rays and Dynamics of Structural Formations in Solar Wind*, Ed. by A. I. Kuz'min (Yakutskiĭ Fil. Sib. Otd. Akad. Nauk SSSR, Yakutsk, 1973), p. 105.
36. S. M. Komoldinov, V. P. Mamrukova, A. M. Altukhov, *et al.*, in *Proceedings of the 14th International Cosmic Ray Conference, ICRC, Munich, 1975*, Vol. 3, p. 1102.
37. S. M. Komoldinov, Candidate's Dissertation in Physics and Mathematics (Inst. Yad. Fiz. Mosk. Gos. Univ., Moscow, 1983).
38. K. Nagashima, K. Fujimoto, Z. Fujii, *et al.*, Rep. Ionos. Space Res. Jpn. **26**, 1 (1972).
39. K. Nagashima, K. Fujimoto, Z. Fujii, *et al.*, Rep. Ionos. Space Res. Jpn. **26** (1, 2), 31 (1972).
40. Yu. P. Mel'nikov, Candidate's Dissertation in Physics and Mathematics (LPI, Leningrad, 1989).
41. D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskiĭ, *Quantum Theory of Angular Moment* (Nauka, Leningrad, 1975; World Scientific, Singapore, 1988).
42. L. I. Dorman, V. N. Malyshkin, and N. P. Milovidova, Izv. Akad. Nauk SSSR, Ser. Fiz. **43**, 2566 (1979).
43. G. F. Krymskiĭ, A. I. Kuz'min, P. A. Krivoshapkin, *et al.*, Izv. Akad. Nauk SSSR, Ser. Fiz. **40**, 604 (1976).
44. L. I. Dorman, Kosm. Luchi, No. 13, 5 (1972).
45. A. D. Chertkov, *Solar Wind and Internal Structure of the Sun* (Nauka, Moscow, 1985).
46. K. W. Behannon, in *Physics of Solar Planetary Environments: Proceedings of International Symposium on Solar-Terrestrial Physics, Colombia, Boulder, 1976*, Vol. 1, p. 332.
47. R. L. Rosenberg, M. G. Kivelson, R. J. Coleman, Jr., and E. J. Smith, J. Geophys. Res. **83**, 4165 (1978).
48. P. A. Krivoshapkin, G. F. Krymskiĭ, A. I. Kuz'min, and G. V. Skripin, in *Distribution of Galactic Cosmic Rays and Dynamics of Structural Formations in Solar Wind*, Ed. by A. I. Kuz'min (Yakutskiĭ Fil. Sib. Otd. Akad. Nauk SSSR, Yakutsk, 1973), p. 43.
49. A. I. Gavril'ev, I. P. Karmadonov, P. A. Krivoshapkin, *et al.*, Izv. Akad. Nauk SSSR, Ser. Fiz. **42**, 1018 (1978).
50. L. I. Dorman, M. E. Kats, Yu. I. Fedorov, and B. A. Shakhov, Zh. Éksp. Teor. Fiz. **79**, 1267 (1980) [Sov. Phys. JETP **52**, 640 (1980)].
51. E. J. Smith, B. T. Tsurutani, and R. L. Rosenberg, EOS Trans. Am. Geophys. Union **24**, 997 (1976).
52. H. Alfven, Rev. Geophys. Space Phys. **15**, 271 (1977).
53. G. F. Krymskiĭ, P. A. Krivoshapkin, V. P. Mamrukova, and G. V. Skripin, Geomagn. Aéron. **21**, 923 (1981).
54. K. G. Ivanov, Geomagn. Aéron. **38**, 1 (1998).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 93, No. 6, 2001, pp. 1209–1215.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 6, 2001, pp. 1396–1403.
Original Russian Text Copyright © 2001 by Klochkov, Rukhadze.

**PLASMA,
GASES**

On the Types of Instability of Spatially Restricted Systems

D. N. Klochkov and A. A. Rukhadze*
General Physics Institute, Russian Academy of Sciences, ul. Vavilova 38, Moscow, 119991 Russia
**e-mail: rukh@fpl.gpi.ru*

Received May 23, 2001

Abstract—A classification of instabilities in spatially restricted systems is presented, which generalizes a classification considered in book [1]. It is shown that, if a system has no active boundaries and the waves are not amplified in an infinite homogeneous medium, which corresponds to the absence of solutions of the dispersion equation with the negative imaginary part of the wave vector at the real frequency, then only nonamplified instabilities with a nonlocal resonance can be developed. The development of nonamplified instability is considered in a spatially restricted system through which a flux propagates, when along with natural waves the excitation of the waves of fluxes playing a key role in the development of the instability is taken into account. *© 2001 MAIK “Nauka/Interperiodica”.*

It is commonly accepted that an oscillator is an amplifier with a feedback [2]. In the case of distributed spatially restricted finite systems, the feedback is provided by a natural wave reflected from the system boundaries, and the term “amplifier” means the amplification of at least one of the natural waves in the medium. This means that the dispersion equation $D(\omega, k) = 0$ for natural waves in an infinite homogeneous medium has complex roots $k = k_\nu(\omega)$ for real values of the frequency ω, some of the roots lying in the lower part of the complex k plane. It follows from this that a convective or absolute instability also takes place in an infinite medium. An example for beam systems is a well-known Cherenkov instability [3, 4].

Nevertheless, at least one more class of instabilities exists in spatially restricted systems, when all the roots $k = k_\nu(\omega)$ are real for real ω; moreover, the branching point of solutions $k_\nu(\omega)$ lies on the real axis of the complex ω plane. This excludes the development of global instability in a finite system and of convective and absolute instability in an infinite system [1, 5]. Moreover, the increment of these instabilities is inversely proportional to the system length L, i.e.,

$$\gamma_\omega^+ \propto L^{-1}.$$

Because the analysis of instabilities in a finite spatial region was commonly performed by considering the asymptotic form of the instability condition for $L \longrightarrow \infty$, some instabilities encountered in finite plasma media were simply ignored. In this connection, it is necessary to generalize the adopted classification of instabilities in finite systems presented in books [1, 4] and in original paper [5]. First, we will study two-wave instabilities and obtain criteria for treating an oscillator as an amplifier with a feedback. Then, we will investigate, by the example of four waves, a class of multiwave instabilities, which do not fit in the conventional scheme, and show that lasing at four waves is possible in a finite system in the absence of wave amplification.

Consider a spatially restricted system in which waves propagate along the z axis and undergo transformations at the boundaries $z = 0$ and $z = L$. Let us assume that a linearized wave equation for the corresponding homogeneous medium contains a polynomial of the power N with respect to the spatial derivative $\partial/\partial z$. Then, the wave dispersion equation $D(\omega, k) = 0$ determines the laws of wave dispersion:

$$k_\nu = k_\nu(\omega), \quad \nu = 1, \ldots, N.$$

In this case, the wave perturbations in a system of a finite length can be represented as a superposition of normal waves ($\nu = 1, \ldots, N$) in an infinite homogeneous medium and of the transient field near the system boundaries:

$$A(t, z) = \sum_{\nu=1}^{N} A_\nu \exp[-i\omega t + ik_\nu(\omega)z] + \sum_s B_s \exp[-i\omega t + ik_s(\omega)z]. \tag{1}$$

Here, the sum of decaying modes B_s is separated. These modes determine the transient field near the boundaries $z = 0$ and $z = L$ for the case of transversely inhomogeneous systems and are neglected in the case of an infinite system. As examples, we consider the cases when a thin-wall tubular electron beam is injected into a resonator or a tubular plasma is contained in a resonator. In these cases, the transient field near the boundaries $z = 0$ and $z = L$ is described by the second sum in (1), where the index s numbers transverse wave numbers $k_{\perp s}$. The consideration of these waves leads to the so-called one-sided solutions [5]. Below, we will omit waves $B_s\exp(-i\omega t + ik_s z)$ in the dispersion equation for a spatially restricted medium, by assuming that the system is long enough and the modes B_s decay so rapidly that the amplitudes of these waves reflected from another end of the system tend to zero in the vicinity of the boundary under study. Mathematically, the condition for neglecting the modes decaying inside the system is described by the inequality

$$\min\{L\mathrm{Im}k_s\} \gg 1, \tag{2}$$

which is assumed always valid below.

We will call the wave of the form $\exp[-i\omega t + ik_\nu(\omega)z]$ specified by the branch $k_\nu(\omega)$ propagating in a certain direction if it decays for $\mathrm{Im}\omega \longrightarrow \infty$ when z changes in this direction. Thus, if we have $\mathrm{Im}\omega \longrightarrow \infty$ for $\mathrm{Im}k_\nu(\omega) \longrightarrow \infty$, then the wave propagates in the positive direction along the z axis; if the former holds for $\mathrm{Im}k_\nu(\omega) \longrightarrow -\infty$, then the wave propagates in the opposite direction. We can show that, for waves close to harmonic waves, this definition of the wave propagation coincides with the definition using their group velocity.

Let us separate the waves into two groups: the group of waves propagating in the positive direction along the z axis (with the index "+") and the group of waves propagating in the opposite direction (with the index "–"). At the boundaries $z = 0$ and $z = L$, each wave of one group is transformed to a wave of the other group. We assume that the corresponding diffraction problem is solved, i.e., the transformation coefficients of waves $A^\pm_\nu$ are found taking into account the waves B_s of the transient field. Then, the relation

$$A^+_p = \sum_{j=1}^{n^-} T^{(\mathrm{in})}_{pj}A^-_j, \quad p = 1, \ldots, n^+, \tag{3}$$

will be valid at the boundary z = 0, which describes the transformation of waves k^-_j into waves k^+_p. Similarly, the transformation of waves k^+_μ into waves k^-_n at the boundary $z = L$ is described by the equation

$$A^-_n\exp(ik^-_nL) = \sum_{j=1}^{n^+} T^{(\mathrm{out})}_{nj}A^+_j\exp(ik^+_jL), \quad n = 1, \ldots, n^-. \tag{4}$$

Here, $n^+ + n^- = N$.

A system of homogeneous linear equations (3), (4) has a nontrivial solution if its determinant is zero. This yields the dispersion equation

$$\det\left\{\sum_{p=1}^{n^-} T^{(\mathrm{in})}_{jp}T^{(\mathrm{out})}_{pn}\exp(i(k^+_n - k^-_p)L) - \delta_{jn}\right\} = 0 \tag{5}$$

for a spatially restricted system, which determines the discrete spectrum of frequencies ω of the system.

In a particular case, a situation is possible when only waves k^+_1 and k^-_1 are present in the system, i.e., $N = 2$. If $N > 2$, we will separate, for ω lying in the upper part of the complex plane, such waves k^+_{n*} and k^-_{p*} for which the imaginary part of the difference is minimal, i.e.,

$$\min_{\mathrm{Im}\omega > 0}\{\mathrm{Im}(k^+_n - k^-_p)\} = \mathrm{Im}(k^+_{n*} - k^-_{p*}). \tag{6}$$

We will call the state corresponding to excitation of waves k^+_{n*} and k^-_{p*} "the ground wave state." For each wave state described by the term $\{k^+_n - k^-_p\}$ in the exponential factor, we will determine "the wave gap" Δk_{np} with respect to the ground wave state, i.e., calculate the value

$$\Delta k_{np} = \mathrm{Im}(k^+_n - k^-_p) - \mathrm{Im}(k^+_{n*} - k^-_{p*}). \tag{7}$$

If the system length L is large enough, i.e., if the inequality

$$L\min\{\Delta k_{np}\} > 1 \tag{8}$$

is fulfilled, then the term $\exp[i(k^+_{n*} - k^-_{p*})L]$ describing the ground wave state is dominant in Eq. (5). Note that the condition (8) is not equivalent to the condition (2). The condition (8) may be simply invalid, or vice versa. Thus, for resonance instabilities (for example, the Cherenkov instability), the condition (8) is satisfied and proves to be stronger than the condition (2). We will assume that the condition (8) is satisfied. In this case, Eq. (5) is substantially simplified and takes the form

$$T^{(\mathrm{in})}_{n*p*}T^{(\mathrm{out})}_{p*n*}e^{i(k^+_{n*} - k^-_{p*})L} = 1. \tag{9}$$

Let us now classify possible instabilities described by Eq. (9). For this purpose, consider the modulus of Eq. (9):

$$\left|T^{(\mathrm{in})}_{n*p*}T^{(\mathrm{out})}_{p*n*}\right|\exp[-\mathrm{Im}(k^+_{n*} - k^-_{p*})L] = 1. \tag{10}$$

If $\mathrm{Im}(k^+_{n*} - k^-_{p*}) > 0$, then the exponential factor in Eq. (10) is less than unity. In this case, the instability can be developed only if

$$\left|T^{(\mathrm{in})}_{n*p*}\right|\left|T^{(\mathrm{out})}_{p*n*}\right| > 1,$$

i.e., if superreflection occurs from at least one of the system boundaries. An example is the development of acoustic vibrations in the Helmholtz resonator, with a plane-parallel air flow moving over its open orifice [6].

If $\mathrm{Im}(k^+_{n*} - k^-_{p*}) < 0$, then the exponential factor in Eq. (10) is greater than unity. In this case, the instability can be developed even in the presence of losses at the system boundaries.

Consider the asymptotics $L \longrightarrow \infty$. Depending on the sign of the difference $\operatorname{Im}(k^+_{n*} - k^-_{p*})$, the exponential factor in Eq. (10) tends either to zero or infinity. For this reason, Eq. (10) is satisfied for $L \longrightarrow \infty$ only under the condition

$$\operatorname{Im}(k^+_{n*} - k^-_{p*}) = 0. \tag{11}$$

Equation (11) written with respect to ω is a particular case of Eq. (10), when the effect of the system boundaries is excluded. In fact, we have written the condition for the development of instability in an infinite medium and have taken the finite size of the system into account through the reflected wave. Instabilities of this type were called global [1, 4, 5].

Consider now the question of when the inequality $\operatorname{Im}(k^+_{n*} - k^-_{p*}) \le 0$ takes place for $\operatorname{Im}\omega > 0$, i.e., when the instability is developed in the absence of active boundaries. Let the wave number has the form

$$k(\omega) = k'(\omega) + ik''(\omega)$$

for real ω ($\operatorname{Im}\omega = 0$).

We assume that there is no aperiodic amplification in the system, i.e., the condition

$$k''(\omega) \ll k'(\omega)$$

is fulfilled. Let the frequency be shifted to the upper part of the complex plane:

$$\omega = \omega' + i\omega'', \quad \omega'' > 0.$$

Assuming that $\omega'' \ll \omega'$, we obtain for the wave packet

$$k(\omega) = k'(\omega') + i\left(k'' + \frac{\omega''}{v_{gr}}\right). \tag{12}$$

Here,

$$v_{gr} = \frac{d\omega'}{dk'} = \left(\frac{dk'(\omega')}{d\omega'}\right)^{-1}$$

is the group velocity of the wave. The condition $\operatorname{Im}k < 0$ of the spatial amplification of the wave takes the form

$$k'' < \frac{\omega''}{v_{gr}}. \tag{13}$$

This relation shows that the amplification can occur in the medium if the inverse dispersion wave with $k = \alpha/\omega$ and the group velocity $v_{gr} = -\omega^2/\alpha < 0$ propagates, or the value of $|k''|$ is so large that the condition (13) is satisfied for the negative right part.

Let the dispersion law (12) be satisfied for the waves k^+_{n*} and k^-_{p*}. Omitting the indices $n*$ and $p*$, which are now unnecessary, we can write

$$k^+ - k^- = k'_+(\omega') - k'_-(\omega') + i\left[k''_+ - k''_- + \omega''\left(\frac{1}{v^+_{gr}} - \frac{1}{v^-_{gr}}\right)\right], \tag{14}$$

which yields the required inequality

$$k''_+ - k''_- + \omega''\left(\frac{1}{v^+_{gr}} - \frac{1}{v^-_{gr}}\right) \le 0. \tag{15a}$$

Therefore, in the absence of active boundaries, the instability is developed at two waves in a finite system when the inequality (15a) is satisfied, the equality sign corresponding to the development of global instability with the increment

$$\omega'' = \frac{k''_- - k''_+}{1/v^+_{gr} - 1/v^-_{gr}}.$$

The appearance of the group velocity in the expression for the increment is explained by the fact that the feedback in the system occurs with the group velocity because the energy transfer by the wave proceeds at this velocity. If no amplification occurs in the system but the active boundaries are present, then Eq. (10) has a solution for

$$\left|T^{(\mathrm{in})}_{n*p*}\right|\left|T^{(\mathrm{out})}_{p*n*}\right| > 1. \tag{15b}$$

In this case, the wave amplification takes place outside the system under study. By expanding the system to include the region where the amplification occurs, we obtain the system with amplification.

Thus, the development of instability at two waves in a spatially restricted system can occur if, along with the condition (8), at least one of the conditions (15) is satisfied. In this case, the instability is an amplifying instability with a feedback in the sense that at least one wave is amplified in the system, while one or two waves realize the feedback at the group velocity. All these instabilities correspond to the conventional concept of an oscillator as an amplifier with a feedback [2].

If both conditions (15) are not satisfied and there are no more waves in the system except the two waves under study, then the system will be stable with respect to initial perturbations. If, however, there are more than two waves in the system, the question about the system stability cannot be solved based on criteria (15). Moreover, the wave gap Δk_{np} may prove to be small or completely absent, or the system will be so short that the condition (8) is not satisfied in all the cases considered above. Then, the two waves cannot be separated according to the condition (6), and, therefore, Eq. (5) cannot be reduced to the form (10). In this case, all waves for which the expression $L\operatorname{Im}(k^+_n - k^-_p)$ has approximately the same value enter Eq. (5) on equal terms. The substitution of dispersion laws $k_\nu = k_\nu(\omega)$ into Eq. (5) gives a transcendental complex equation for ω. If the solution to this equation shows that ω has a positive imaginary part ($\operatorname{Im}\omega > 0$), then the system is unstable. In this case, conditions (15) are not necessarily satisfied; i.e., neither active boundaries nor inverse dispersion waves nor local spatial wave amplification ($k'' = 0$) can be present in the system. This class of instabilities, which we will call “nonamplified” instabilities, is beyond the conventional concept of an oscillator as an amplifier with a feedback. In this case, the number N of waves in the system can be arbitrary ($N > 2$).

The simplest and often encountered situation is the case of four waves. This case is observed when a particle flux propagates through a system (for example, the electron or ion beam stabilized in a magnetic field). We assume that the flux propagates in the positive direction along the z axis. Then, the three waves in the system propagate in the flux direction (the natural wave of the system with the wave vector k_1 and two waves of the beam with the wave vectors k_3 and k_4), and only one reflected natural wave with $k_2 = k^-$ propagates in the opposite direction.

If at the right boundary $z = L$ of the system the waves A_1, A_3, and A_4 propagating in the flux direction are partially transformed to the counterpropagating wave A_2 and partially to the waves leaving the system, this process can be written in the form

$$\exp(ik_2L) = \sum_{\substack{\nu=1\\ \nu\neq 2}}^{4} T_{2\nu}^{(\text{out})}T_{\nu 2}^{(\text{in})}\exp(ik_\nu L). \qquad (16)$$

Equation (16) is a particular case of Eq. (5) for four waves.

We assume that in the absence of the flux the system medium is homogeneous (symmetric with respect to the inversion $z \longrightarrow -z$). Then, the wave numbers of the natural waves of the system without the flux have the form $k_{1,2} = \pm a$, where a takes a discrete set of values because the system (resonator) is finite. Because the condition (8) is not fulfilled, no local resonance is observed in the system, i.e., $\omega \neq au$ where u is the flux velocity. We assume that the flux density is so low that the flux can be treated using perturbation theory. In this case, we have for the natural waves of the system

$$k_{1,2} = \pm a + \delta k_{1,2} \qquad (17a)$$

and for waves of the flux

$$k_{3,4} = \frac{\omega}{u} \pm \delta k_3. \qquad (17b)$$

Here,

$$\delta k_1 = O(\omega_b^2), \quad \delta k_2 = O(\omega_b^2),$$
$$\delta k_3 = -\delta k_4 = O(\omega_b),$$

where ω_b is a small parameter characterizing the unperturbed flux density. For charged particle fluxes, ω_b is the Langmuir frequency of flux particles, and a small dimensionless parameter in this case is the ratio ω_b/ω_0, where ω_0 is the characteristic frequency of the system. The presence of the particle flux in plasmalike media results in the correction $\delta\varepsilon$ to the permittivity tensor, which is proportional to $\omega_b^2/(\omega - ku)^2$. The appearance of the flux expression $\omega - ku$ in the denominator follows from the equation of motion for flux particles. As a result, we obtain corrections δk_i in the dispersion law for the wave from the perturbed part $\delta\varepsilon$ of the permittivity tensor. Thus, if a homogeneous monoenergetic electron beam is injected into an empty metal resonator with a circular cross section, then

$$\varepsilon_\parallel = -\frac{\omega_b^2\gamma^{-3}}{(\omega - ku)^2},$$

and

$$\delta k_{1,2} = \mp\frac{k_\perp^2\gamma^{-3}}{2a(\omega \mp au)^2}\omega_b^2 \qquad (18a)$$

$$\delta k_{3,4} = \frac{\omega}{u}\frac{\gamma^{-5/2}}{\sqrt{\omega^2 - a^2u^2}}\omega_b. \qquad (18b)$$

Here, $k_\perp$ is the transverse wave number and

$$\gamma = (1 - u^2/c^2)^{-1/2}$$

is the relativistic factor.

We assume that neither the velocity nor the density of the incident flux at the entrance boundary $z = 0$ is perturbed. To satisfy this condition, we will assume that the waves undergo mirror reflection from this boundary, i.e., the waves do not leave the system through the boundary $z = 0$. The condition of mirror reflection of the waves can be written in the form

$$\sum_{\nu=1}^{4} k_\nu A_\nu + \sum_s k_s B_s = 0. \qquad (19a)$$

The absence of the flux velocity perturbations gives [7]

$$\sum_{\nu=1}^{4}\frac{A_\nu}{\omega - k_\nu u} + \sum_s \frac{B_s}{\omega - k_\nu u} = 0, \qquad (19b)$$

while the absence of the flux density perturbations gives

$$\sum_{\nu=1}^{4}\frac{A_\nu}{(\omega - k_\nu u)^2} + \sum_s \frac{B_s}{(\omega - k_\nu u)^2} = 0. \qquad (19c)$$

For low-density fluxes, when $B_s = O(\omega_b^2)$, or for transversely homogeneous systems, the transformation coefficients have the form [7]

$$T_{12}^{(\text{in})} = 1 + O(\omega_b^2), \qquad (20a)$$

$$T_{32}^{(\text{in})} = \alpha\omega_b, \qquad (20b)$$

$$T_{42}^{(\text{in})} = -\alpha\omega_b. \qquad (20c)$$

The transformation coefficients $T_{2\nu}^{(\mathrm{out})}$ can be represented in the most general form:

$$T_{21}^{(\mathrm{out})} = \kappa_1 + g_1\omega_b^2, \tag{21a}$$

$$T_{23}^{(\mathrm{out})} = \kappa_b + g_b\omega_b^2, \tag{21b}$$

$$T_{24}^{(\mathrm{out})} = \kappa_b - g_b\omega_b^2. \tag{21c}$$

Here, κ_1 is the coefficient of reflection of the copropagating wave from the plane $z = L$ in the absence of the flux and κ_b is transformation coefficient for flux waves. As a result, we can write in the general form

$$T_{21}^{(\mathrm{out})}T_{12}^{(\mathrm{in})} = \kappa_1 + q_1\omega_b^2, \tag{22a}$$

$$T_{23}^{(\mathrm{out})}T_{32}^{(\mathrm{in})} = -b\omega_b + F\omega_b^2, \tag{22b}$$

$$T_{24}^{(\mathrm{out})}T_{42}^{(\mathrm{in})} = b\omega_b + F\omega_b^2. \tag{22c}$$

The quantities q, b, and F depend on the system geometry and are functions of the frequency and wave numbers (see details in [7]). For example, if a homogeneous monoenergetic electron beam is injected into an ideal vacuum metal resonator of a circular section (without ohmic and radiative losses, $\kappa_1 = 1$), then

$$b = \frac{\omega}{au}\frac{k_\perp^2 u^2 \gamma^{-1/2}}{(\omega^2 - a^2u^2)^{3/2}}, \tag{23a}$$

$$F = 2\frac{\omega}{au}\frac{\omega^2 + a^2u^2}{(\omega^2 - a^2u^2)^3}k_\perp^2 u^2\gamma^{-3}, \tag{23b}$$

$$q_1 = -2F. \tag{23c}$$

By assuming that

$$L\delta k_1 \ll 1, \quad L\delta k_2 \ll 1, \quad L\delta k_3 \sim 1,$$

we obtain the final expression for the dispersion instability equation linearized in a small parameter ($\sim\omega_b/\omega$)

$$\mathcal{D}(a) \equiv \mathcal{D}_0(a) + \mathcal{D}_1(a) = 0. \tag{24}$$

Here,

$$\mathcal{D}_0(a) = e^{-iaL} - \kappa_1 e^{iaL} \tag{25}$$

is the unperturbed part of the dispersion equation, which determines the discrete set of values of the unperturbed parameter a,

$$a = \frac{\pi n}{L} - \frac{1}{2L}\arg\kappa_1 - \frac{i}{2L}\ln\frac{1}{|\kappa_1|}, \tag{26}$$

and, hence, the discrete set of natural oscillations of the system $\omega = \omega(\mathrm{Re}\,a)$. The imaginary part of the parameter a determines the radiative losses:

$$\gamma_\omega^- = -\frac{d\omega}{da}\mathrm{Im}\,a = \frac{1}{2L}\frac{d\omega}{da}\ln\frac{1}{|\kappa_1|}. \tag{27}$$

Here,

$$\frac{d\omega}{da} = v_{gr}$$

is the group velocity of the natural wave of the system.

Consideration of the flux perturbation gives the second term $\mathcal{D}_1(a)$ in the dispersion equation:

$$\begin{aligned}\mathcal{D}_1(a) &= iL\delta k_2 e^{-iaL} - (iL\delta k_1\kappa_1 + q_1\omega_b^2)e^{iaL} \\ &+ 2ib\omega_b\sin(L\delta k_3)e^{i\theta} - 2F\omega_b^2\cos(L\delta k_3)e^{i\theta}.\end{aligned} \tag{28}$$

Here,

$$\theta = \frac{\omega L}{u}$$

is the transit angle (phase) of flux particles. The perturbation $\delta\omega$ of the natural frequency of the system corresponding to this term is

$$\begin{aligned}\delta\omega &= -\frac{d\omega}{da}\frac{\mathcal{D}_1(a)}{\partial\mathcal{D}_0(a)/\partial a} \\ &= \frac{1}{2}\frac{d\omega}{da}(\delta k_2 - \delta k_1) + (-1)^n\frac{\omega_b b}{L\sqrt{|\kappa_1|}} \\ &\times \frac{d\omega}{da}\sin(L\delta k_3)\exp\left(i\theta - \frac{i}{2}\arg\kappa_1\right) \\ &+ i\frac{\omega_b^2}{L}\frac{d\omega}{da}\left[\frac{q_1}{2\kappa_1} + (-1)^n\frac{F}{\sqrt{|\kappa_1|}}\cos(L\delta k_3)\right. \\ &\left.\times\exp\left(i\theta - \frac{i}{2}\arg\kappa_1\right)\right].\end{aligned} \tag{29}$$

From this, we find the instability increment taking into account the radiative losses:

$$\begin{aligned}\gamma_\omega &= \mathrm{Im}\,\delta\omega - \gamma_\omega^- \\ &= \frac{1}{2}v_{gr}\mathrm{Im}(\delta k_2 - \delta k_1) + (-1)^n\frac{\omega_b v_{gr}}{L} \\ &\times\mathrm{Im}\left[\frac{b}{\sqrt{\kappa_1}}\sin(L\delta k_3)\exp\left(i\theta - \frac{i}{2}\arg\kappa_1\right)\right] \\ &+ \frac{\omega_b^2 v_{gr}}{L}\mathrm{Re}\left[\frac{q_1}{2\kappa_1} + (-1)^n\frac{F}{\sqrt{|\kappa_1|}}\cos(L\delta k_3)\right. \\ &\left.\times\exp\left(i\theta - \frac{i}{2}\arg\kappa_1\right)\right] - \frac{v_{gr}}{2L}\ln\frac{1}{|\kappa_1|}.\end{aligned} \tag{30}$$

If the system exhibits amplification, i.e., the condition (15a) is fulfilled, then a nonzero wave gap Δk_{np} exists. In this case, beginning from some value of the system length L^*, the condition (8) will be fulfilled for all $L > L^*$. Therefore, in long systems with $L > L^*$, two-wave lasing will occur, which is described by Eq. (9).

Indeed, for the system length $L \longrightarrow \infty$ in expression (30), we obtain the increment of the global instability

$$\gamma_\omega = \frac{1}{2}v_{gr}\mathrm{Im}(\delta k_2 - \delta k_1). \tag{31}$$

If all δk_i are real [in this case, only $T_{np}^{(\mathrm{out})}$ can be complex], then the condition (8) is not satisfied for any lengths L, and the instability increment (decrement) is

$$\begin{aligned}\gamma_\omega &= (-1)^n\frac{\omega_b v_{gr}}{L}\frac{|b|}{\sqrt{|\kappa_1|}}\sin(L\delta k_3)\\ &\times\sin\left(\theta - \frac{1}{2}\arg\kappa_1 + \arg\kappa_b\right)\\ &+ \frac{\omega_b^2 v_{gr}}{L}\left[\frac{|q_1|}{2|\kappa_1|}\cos(\arg q_1 - \arg\kappa_1)\right.\\ &+ (-1)^n\frac{|F|}{\sqrt{|\kappa_1|}}\cos(L\delta k_3)\\ &\left.\times\cos\left(\theta - \frac{1}{2}\arg\kappa_1 + \arg F\right)\right] - \frac{v_{gr}}{2L}\ln\frac{1}{|\kappa_1|}.\end{aligned} \tag{32}$$

If $\gamma_\omega > 0$, then a nonamplified instability appears in the system because none of the conditions (8) and (15) are satisfied for real $k_\nu = k_\nu(\omega)$ at $\mathrm{Im}\,\omega = 0$. For $L \longrightarrow \infty$, the increment of the nonamplified instability tends to zero, and, therefore, this instability appears only in spatially restricted media.

The systems can be divided into short and long ones according to the value of $L\delta k_3$. In short systems, $L\delta k_3 \ll 1$, so that

$$\sin(L\delta k_3) \sim L\delta k_3 \sim \omega_b L.$$

As a result, both of the first terms in (32) are of the same order of magnitude, and the increment in the system without losses is $\gamma_\omega \propto \omega_b^2$; i.e., it is proportional to the flux density. Thus, if a homogeneous monoenergetic electron beam is injected into an ideal vacuum metal resonator of a circular section (without ohmic and radiative losses), then the increment is

$$\begin{aligned}\gamma_\omega^+ &= \frac{k_\perp^2 c^2\gamma^{-3}}{(\omega^2 - a^2u^2)^2}\frac{u}{L}\omega_b^2\\ &\times\left[(-1)^n\theta\sin\theta + 2\frac{\omega^2 + a^2u^2}{\omega^2 - a^2u^2}((-1)^n\cos\theta - 1)\right].\end{aligned} \tag{33}$$

The condition for the instability development has the form

$$(-1)^n\theta\sin\theta + 2\frac{\omega^2 + a^2u^2}{\omega^2 - a^2u^2}((-1)^n\cos\theta - 1) > 0. \tag{34}$$

In long systems, when $L\delta k_3 \geq 1$, the second term in (32) has a higher smallness order over the parameter ω_b compared to the first term. For this reason, the expression for the increment in long systems is simplified to be

$$\begin{aligned}\gamma_\omega &= \frac{v_{gr}}{L}\left[(-1)^n\omega_b\frac{|b|}{\sqrt{|\kappa_1|}}\sin(L\delta k_3)\right.\\ &\left.\times\sin\left(\theta - \frac{1}{2}\arg\kappa_1 + \arg\kappa_b\right) - \frac{1}{2}\ln\frac{1}{|\kappa_1|}\right].\end{aligned} \tag{35}$$

The condition $\gamma_\omega > 0$ for the instability development in long systems takes the form

$$\begin{aligned}&(-1)^n\omega_b|b|\sin(L\delta k_3)\\ &\times\sin\left(\theta - \frac{1}{2}\arg\kappa_1 + \arg\kappa_b\right) > \frac{\sqrt{|\kappa_1|}}{2}\ln\frac{1}{|\kappa_1|}.\end{aligned} \tag{36}$$

One can see from (36) that the fulfillment of the inequality strongly depends on the transit angle θ. In this sense, we can say that nonamplified instabilities are determined by the nonlocal resonance, which depends on the flux parameters and linear dimensions of the system. In addition, analysis of expression (36) shows that the nonamplified instability develops in sufficiently high-Q systems when

$$2\omega_b|b| > \sqrt{|\kappa_1|}\ln|\kappa_1|^{-1}.$$

For an ideal resonator, the condition (36) is substantially simplified:

$$(-1)^n\sin(L\delta k_3)\sin\theta > 0. \tag{37}$$

We derived the expression for the increment γ_ω assuming an arbitrary dispersion law $a = a(\omega)$ for the natural waves of the system. For this reason, the nonamplified instability is universal in a sense that it is manifested for waves with any dispersion law and nonzero group velocity. Unlike resonance instabilities, such as the Cherenkov instability, the nonamplified instability can develop when the natural waves of the system are not retarded. Some examples are excitation of various whistles with resonators [6] (flutes, organ tubes, etc.) and excitation of an acoustoelectric generator at electron-flux velocities lower than the sound velocity in a ferroelectric. In electronics, this is excitation of a monotron and generation of microwave radiation on an oblique Langmuir wave in a plasma resonator at plasma densities lower than the critical one, when the resonance Cherenkov instability cannot develop [8], as well as excitation of potential Langmuir waves in a plane plasma layer upon continuous injection of a monoenergetic electron beam [9].

The nonradiative instabilities, such as the aperiodic Pierce instability [1], belong to the type of instabilities described above. The dispersion law in the most general form can be written as

$$k_1 = \frac{\omega}{v_1}, \quad k_2 = -\frac{\omega}{v_2}, \tag{38a}$$

$$k_{3,4} = \frac{\omega \pm \eta}{v_{gr}}. \tag{38b}$$

Here, the parameter η characterizing the flux is no longer a small quantity, and v_{gr} is the group velocity of the beam waves. Under the condition $\omega < \eta$, the phase velocity of the slow beam wave is directed oppositely to the flux, whereas its group velocity is directed along the flux. Although in this case the slow beam wave behaves as a wave with inverse dispersion, the condition (15a) is not satisfied because $v_{gr} > 0$ and $k_4'' = 0$.

The potential wave k_1 and two beam waves k_3 and k_4 propagate in the flux direction, while the wave k_2 propagates in the opposite direction, thereby performing a feedback in the system. Let us introduce the notation

$$T_{2\nu}^{(\text{out})} T_{\nu 2}^{(\text{in})} = \alpha_\nu$$

and rewrite Eq. (16) for this case in the form

$$\alpha_1 \exp\left(i\frac{\omega L}{v_1}\right) - \exp\left(-i\frac{\omega L}{v_2}\right)$$
$$+ \left[(\alpha_3 + \alpha_4)\cos\left(\frac{\eta L}{v_{gr}}\right) + i(\alpha_3 - \alpha_4)\sin\left(\frac{\eta L}{v_{gr}}\right)\right] \tag{39}$$
$$\times \exp\left(i\frac{\omega L}{v_{gr}}\right) = 0.$$

This equation cannot be solved in the general case because there is no small parameter in which the solution could be expanded. Because particular cases of Eq. (39) were analyzed in detail in the literature (see, for example, [10–12]), we will not consider them here. Note only two important circumstances.

It is usually assumed that metal electrodes are located at the boundaries $z = 0$ and $z = L$. At the boundary $z = L$, more general conditions can be set without the loss of the self-consistence. Assuming that a homogeneous dielectric with the permittivity

$$\varepsilon(\omega) = \varepsilon'(\omega) + \varepsilon''(\omega)$$

is located in the region $z > L$, we can obtain from the boundary conditions for the field at $z = L$ the coefficients $T_{2\nu}^{(\text{out})}$, although in this case we can only conditionally speak about the waves leaving the system. For $\varepsilon \longrightarrow \infty$, we arrive at a particular case of mirror boundary conditions on the metal.

Consider the asymptotics $L \longrightarrow \infty$ for very long systems. In this case, the second term in Eq. (39) can be neglected under the condition $\omega'' > 0$. The remaining two terms will be of the same order of magnitude if the increment is inversely proportional to the system length, i.e., if $\omega'' \propto 1/L$. This means that this instability is absent in an infinite system.

Therefore, a feedback amplifier is a particular case of an oscillator. In plasmalike spatially restricted media, there exists a broad class of instabilities that cannot fit in conventional concepts. The conditions for the development of these instabilities are determined by a nonlocal resonance. This precludes the creation of amplifiers operating on this type of instabilities; however, it permits the fabrication of a broad class of oscillators.

In this paper, we have considered the multiwave instability of a system with respect to the ground unexcited state in which the nonequilibrium is caused by the flux. Therefore, we have not analyzed here parametric instabilities in a finite space [13], such as stimulated Raman scattering or stimulated Brillouin scattering. The nonequilibrium in these processes is caused by the nonlinear distortion of the unperturbed state in the presence of the pump wave with a large amplitude.

ACKNOWLEDGMENTS

The authors thank V. P. Silin for valuable comments and advice.

REFERENCES

1. E. M. Lifshitz and L. P. Pitaevskii, *Physical Kinetics* (Nauka, Moscow, 1979; Pergamon, Oxford, 1981).
2. P. S. Landa, *Nonlinear Oscillations and Waves* (Nauka, Moscow, 1997).
3. A. F. Alexandrov, L. S. Bogdankevich, and A. A. Rukhadze, *Principles of Plasma Electrocdynamics* (Vysshaya Shkola, Moscow, 1978; Springer-Verlag, Berlin, 1984).
4. A. I. Akhiezer, I. A. Akhiezer, R. V. Polovin, *et al.*, *Plasma Electrodynamics* (Nauka, Moscow, 1974; Pergamon, Oxford, 1975).
5. A. G. Kulikovskiĭ, Prikl. Mat. Mekh. **30**, 148 (1966).
6. Yu. A. Stepanyants and A. L. Fabrikant, *Wave Propagation in Shear Flows* (Nauka, Moscow, 1996).
7. D. N. Klochkov, M. Yu. Pekar, and A. A. Rukhadze, Phys. Plasmas **7**, 4707 (2000).
8. D. N. Klochkov and M. Yu. Pekar, Fiz. Plazmy **23**, 650 (1997) [Plasma Phys. Rep. **23**, 602 (1997)].
9. S. S. Kalmykova, Zh. Éksp. Teor. Fiz. **65**, 2250 (1973) [Sov. Phys. JETP **38**, 1124 (1973)].
10. M. V. Kuzelev and A. A. Rukhadze, *Electrodynamics of Dense Electromagnetic Beams in a Plasma* (Nauka, Moscow, 1990).
11. S. Kuhn, Contrib. Plasma Phys. **34**, 495 (1994).
12. V. V. Vladimirov, A. N. Mosiyuk, and M. A. Mukhtarov, Fiz. Plazmy **9**, 992 (1983) [Sov. J. Plasma Phys. **9**, 578 (1983)].
13. L. M. Gorbunov, Zh. Éksp. Teor. Fiz. **67**, 1386 (1974) [Sov. Phys. JETP **40**, 689 (1974)].

Translated by M. Sapozhnikov

Journal of Experimental and Theoretical Physics, Vol. 93, No. 6, 2001, pp. 1216–1221.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 6, 2001, pp. 1404–1410.
Original Russian Text Copyright © 2001 by Artamonov, Krasov, Paperny.

PLASMA, GASES

Registration of Accelerated Multiply Charged Ions from the Cathode Jet of a Vacuum Discharge

M. F. Artamonov, V. I. Krasov, and V. L. Paperny*
Irkutsk State University, Irkutsk, 664003 Russia
**e-mail: paperny@math.isu.runnet.ru*

Received June 18, 2001

Abstract—The parameters of Cu^{n+} and Ta^{n+} ions from the plasma of a vacuum spark with a voltage up to 2.5 kV and a current rise rate up to 2×10^{10} A/s are studied using the time-of-flight method. At the initial stage of the discharge, bursts of beams of accelerated multiply charged ions from the cathode flame have been detected. It is established that the charge state distribution and energy of a beam are controlled by the initial voltage U_0 of the capacitor. Upon an increase in this voltage, the average charge of copper ions attains the value +9, and the average charge of tantalum ions can be as high as +20, while the energy attains values of 150 and 350 keV, respectively. It is found that the average energy of ions with charge Z increases in proportion to the charge and is close to the energy eZU_0 which would have been acquired by ions accelerated in the electric field of the discharge gap. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In recent years, considerable efforts were made for creating sources of multiply charged metal ions, which are of interest for projects of heavy-ion accelerators intended for research purposes, for solving material science problems, and for medical applications. One of such sources is a plasma emerging as a result of exposure of a target to a high-power laser pulse, which makes it possible to obtain, for example, copper ions with a charge up to +25 and an energy up to 130 keV [1].

Another promising source of ions is a vacuum spark producing a high energy density liberated in the plasma column, which is required for generating multiply charged ions and characterized by a relatively low total energy consumption and a simple construction. Indeed, bursts of X rays emitted by hydrogen- or helium-like ions of the cathode material (iron or titanium) from micropinch structures formed in the plasma of a high-current spark were detected even in early experiments (see, for example, [2, 3]), but direct observations of these ions have not been reported so far [4]. The directly measured maximum charge of ions in the plasmas of various types of vacuum spark did not exceed +7 (W^{7+} [5], Cu^{7+} [6]), and the energy of ions attains values of 10–15 MeV for the discharge gap voltage U_0 = 300 kV [5].

In the present work, we report on the registration of multiply charged accelerated ion beams of the cathode material (Cu^{n+}, Ta^{n+}), generated at the initial stage of a vacuum spark with relatively low values of current amplitude (up to 10 kA) and capacitor energy (up to 7 J). It will be proved that the charge state distribution and the energy distribution of ions are close to the parameters observed in a laser plasma, while the typical values of supplied energy are almost two orders of magnitude lower. This indicates that a spark discharge can be used in principle as an effective source of accelerated multiply charged metal ions.

2. EXPERIMENTAL RESULTS

The experiments were carried out in a vacuum discharge with a capacitor voltage U_0 up to 2.5 kV. The electrode system consisted of a copper or tantalum cathode of diameter 1 mm and a plane grounded grid-type anode separated by 9 mm from the end face of the cathode and was placed in a chamber with a vacuum not worse than (5–8) × 10^{-6} torr. In contrast to the high-voltage discharges used in [5, 6], the discharge was initiated at the end face of the cathode through the breakdown over the surface of an insulator insert between the cathode and the igniting electrode. The discharge current was sustained by a capacitor (C = 2 μF) and was measured by a Rogowki coil directly in the cathode circuit. The total inductance of the discharge circuit did not exceed 40 nH. Before measurements, the cathode surface was cleaned by "training" during 10^3 "shots," after which the variations of discharge parameters in various shots did not exceed 20%.

The energy and charge state distributions of ions were measured by the time-of-flight method with the help of an electrostatic analyzer of the "plane capacitor" type with an the energy resolution $\Delta E/E \approx 2 \times 10^{-2}$ and a time resolution of the registering circuit of about 40 ns. We used a microchannel plate as a detector of ions. The gain of the plate was adjusted for different ion charge states and impact energies using the data obtained in [7]. The analyzer was placed behind the grid anode and the drift gap so that the ions emitted by the cathode flame and moving along the discharge axis towards the anode were registered.

The time-of-flight method makes it possible to determine the value of μ/Z (μ is the atomic weight and eZ is the charge of an ion) for each species of ions with an energy specified by the bias voltage at the analyzer plate from the delay of the corresponding signal of the detector. For fixed values of energy resolution of the analyzer, the temporal resolution of the detecting circuit, and the duration of the ion emission, the resolution $\alpha = Z/\Delta Z$ of the method in the charge states of ions is determined by the time of flight through the gap between the emission region and the detector. Under our experimental conditions and for a gap length of 60 cm, the limiting resolution for Cu^{n+} ions was $\alpha(E) = 20$; the corresponding value of Ta^{n+} ions was $\alpha(E) = 30$ in view of their larger mass and, hence, lower velocity.

Figure 1a shows a typical oscillogram of the discharge current in experiments with a copper cathode; the corresponding signal from the analyzer for a fixed value of E/Z is shown in Fig. 1b. The signal was processed using the following procedure. Assuming that the last peak in Fig. 1b corresponds to Cu^{+} ions (having the maximum value of $\mu/Z = 64$, i.e., the maximum time of flight), we can determine the starting instant t_1 for these ions, which is marked by arrows in Figs. 1a and 1b. The values of μ/Z for the ions generating these peaks can be determined from the delays of the remaining peaks relative to t_1. The obtained values of $\mu/Z = 32$, 21, etc., can be naturally ascribed to Cu^{2+}, Cu^{3+}, etc., ions. This means that ions in all charge states are produced simultaneously. The time width of ion signals corresponding to the duration of the emission process amounts to $\Delta t \approx 0.05$–0.1 μs (for the given value of $E/Z = 2.1$ keV). According to the results of measurements, an increase in the energy of the ions being registered leads to a decrease in the value of Δt, which tends to the temporal resolution of the measuring circuit. It can also be seen from Fig. 1a that ions are produced at the initial stage of the discharge at $t_1 = 300$–400 ns after its beginning for a current approximately equal to 3 kA, which is much smaller than the amplitude value. The instant of ion emission was found to be close to the instant of the maximum current rise rate.

It can be seen from Fig. 1b that a noticeable peak of H^{+} ions was also registered in experiments in addition to the peaks associated with ions of the cathode material. In addition, ions of other light impurities desorbed from the surface of the cathode and the insulator insert, O^{n+}, C^{n+}, N^{n+} ($n = 1, 2$), and having values of μ/Z close to those for corresponding copper ions (e.g., Cu^{4+}, Cu^{5+}, Cu^{8+}, ions, etc.) may also contribute to the signals. An analysis of the obtained results proved that such ions make an insignificant contribution to the charge composition of the beam.

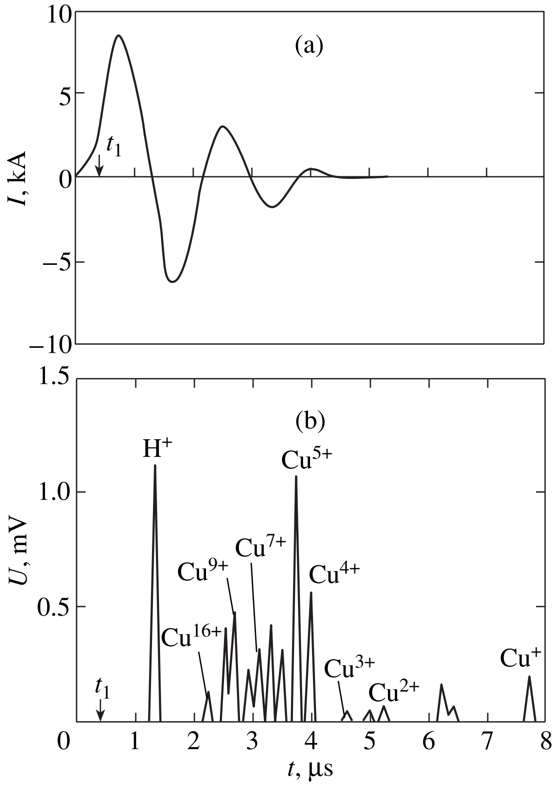


Fig. 1. (a) Oscillogram of the discharge current for a storage voltage of 2.0 kV; (b) analyzer signal for energy E/Z = 2.1 keV of Cu^{n+} ions.

Ion signals detected in various discharges for fixed discharge parameters had a considerable spread in amplitude; for this reason, these signals were averaged for subsequent processing over a series of 12 discharges. The averaged amplitudes of the signals corresponding to ions of a given charge state and measured for various values of E/Z were used for constructing the energy spectrum of such ions. Figure 2a shows the ion spectra obtained in this way for different charge states of Cu^{n+} for two values of the storage voltage U_0. It can be seen that a typical spectrum consists of a core in which the main fraction of ions with energies not exceeding a few kiloelectronvolts is concentrated and a "tail" of accelerated ions whose fraction is a few percent of the total number of particles and whose energy can be as high as 150 keV. Another peculiarity observed in Fig. 2a is a nearly linear increase in the maximum energy of the ions being registered with their charge. In addition, a comparison of spectra *1* and *2* obtained for different values of storage voltage U_0 shows that the spectra of all ion components are broadened towards higher energies upon an increase in this voltage.

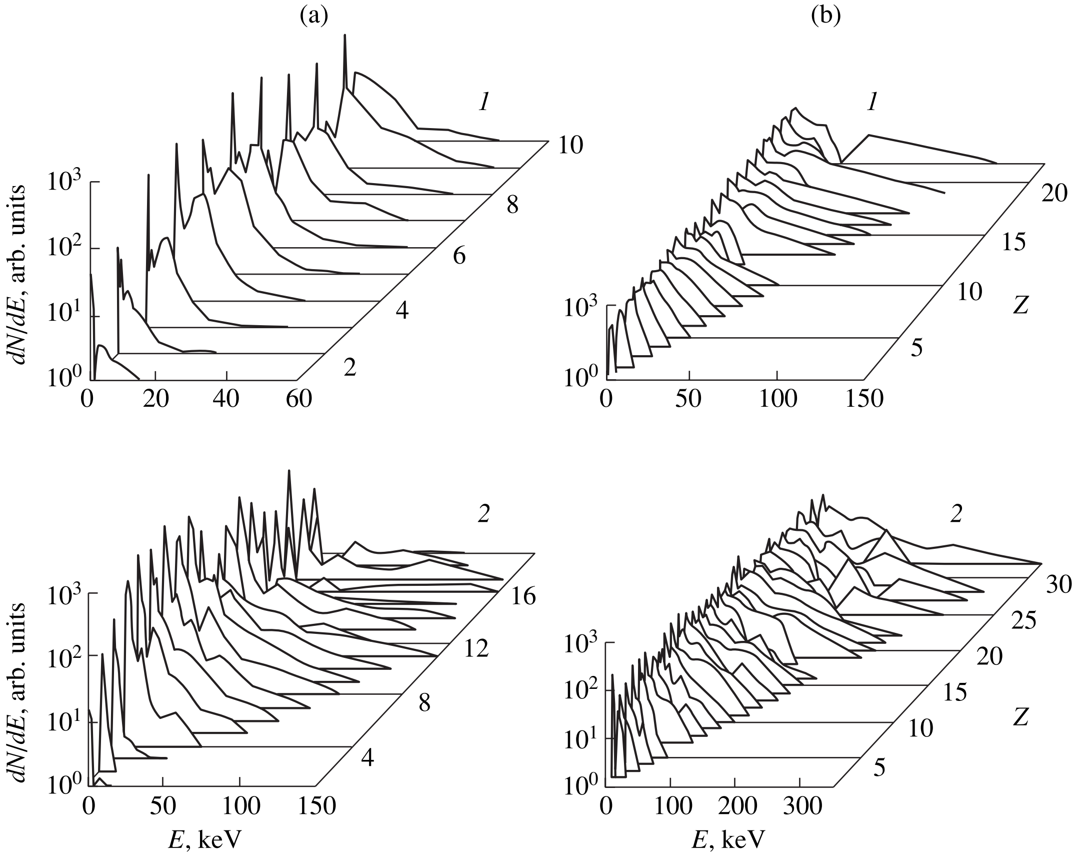


Fig. 2. (a) Energy spectra of Cu^{n+} ions for a storage voltages of 1.0 (*1*) and 2.5 keV (*2*); (b) spectra of Ta^{n+} ions for voltages of 0.2 (*1*) and 1.5 kV (*2*).

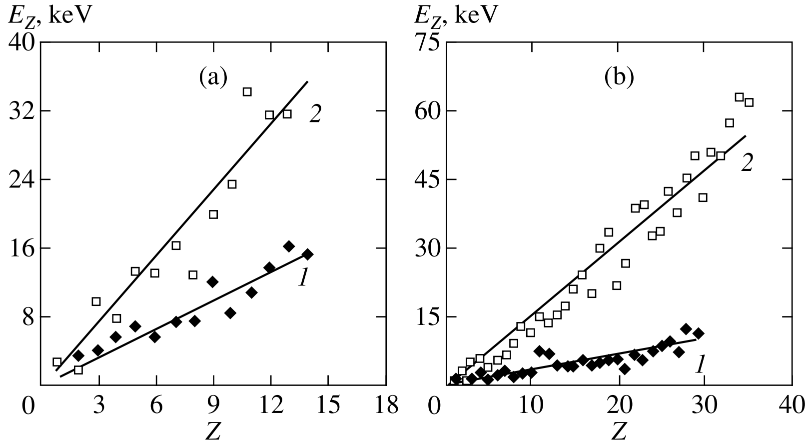


Fig. 3. (a) Dependence of the mean energy E_Z on the charge multiplicity of Cu^{n+} ions for the storage voltage $U_0 = 1.0$ (*1*) and 2.5 keV (*2*); the equation of the straight lines is $\langle E_Z/e \rangle = kZ$, where coefficient $k = U_0$; (b) the same for Ta^{n+} ions for $U_0 = 0.3$ (*1*) and 1.5 kV (*2*); coefficient $k = U_0$.

A quantitative parameter characterizing the observed features of the acceleration process is the mean energy E_Z of ions having a given charge, which was obtained by integrating their energy spectrum. Figure 3a shows the dependence of E_Z on the charge of a Cu^{n+} ion for two values of the storage voltage, i.e., for two initial voltages across the discharge gap. It can be seen from the figure that the mean energy of ions increases almost linearly with the charge over a wide range of its variation.

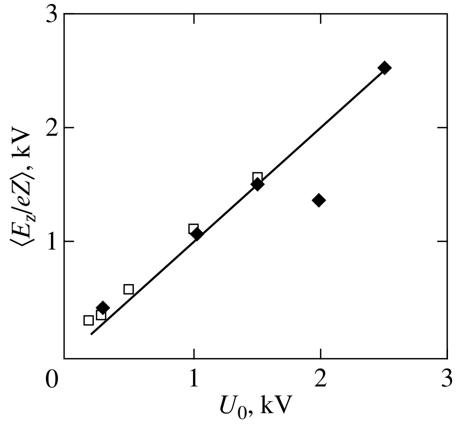


Fig. 4. Energies per unit charge $\langle E_Z/Z \rangle$ averaged over the charge state distribution of Cu^{n+} (◆) and Ta^{n+} (□) ions as functions of the storage voltage. The equation of the straight line is $\langle E_Z/Z \rangle = eU_0$.

In order to find out whether the acceleration process depends on the ion mass, we made similar measurements with a tantalum cathode. In these experiments, we observed Ta^{n+} ions whose charge composition turned out to be much wider than in the case of Cu^{n+} ions. Tantalum ions in various charge states were also produced simultaneously at the initial stage of the discharge; the duration of this process was found to be close to the corresponding value for Cu^{n+} ions. Using the procedure of signal processing described above, we constructed the energy spectra of Ta^{n+} ions at different storage voltages. These spectra are shown in Fig. 2b. It follows that the spectra for tantalum and copper ions have similar structures. The maximum detected energy of tantalum ions also increases with the charge almost linearly, attaining a value of 350 keV for Ta^{+32} ions at a storage voltage $U_0 = 1.5$ kV. In addition, it follows from Fig. 3b that, as in the case of Cu^{n+} ions, there exists a linear dependence between the charge state and the mean energy E_Z of various components of Ta^{n+} ions, but in a considerably wider range of charge states as compared to copper ions.

Figure 3 also shows that the ratio E_Z/Z for accelerated ions is virtually independent of the charge state and is determined only by the storage voltage. Under the assumption that the mean energy E_Z of ions is a linear function of charge Z, we approximated the experimental results (Fig. 3) for each value of storage voltage by a straight line plotted by the least squares technique. The slopes of the straight lines give the values of energy per unit charge $\langle E_Z/Z \rangle$ averaged over the charge state distribution. Figure 4 presents the results of data processing for the entire range of the discharge parameters and for both materials of the cathode. It can be seen from the figure that the mean energy acquired by ions as a result of acceleration is independent of the ion mass and is close to the value of energy eZU_0 which would be acquired by ions in the case of direct acceleration in the electric field of the discharge gap.

By integrating the energy spectra, we obtained the charge state distribution of ion beams of tantalum and copper for different values of storage voltage (Fig. 5). It was mentioned above that the resolution of the method used does not allow us to single out the signals of tantalum ions with charges differing by unity in each discharge in the range of high degree of ionization ($Z >$ +30). For this reason, the charge state distribution of the ion beam in this region was plotted not for an individual component, but for groups of unresolved components, and demonstrates only the general tendency of ion distribution. It can be seen from the figure that the charge state distributions of the beams of Ta^{n+} and Cu^{n+} ions are

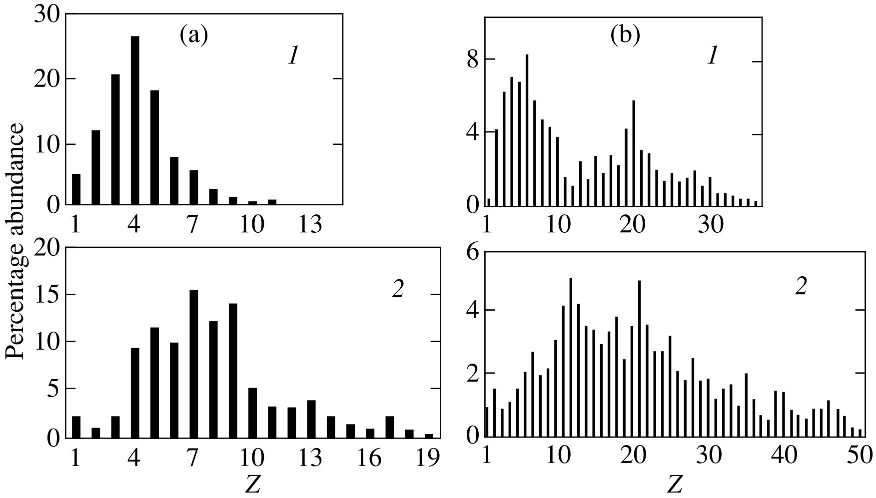


Fig. 5. (a) Charge compositions of the beam of Cu^{n+} ions for the initial storage voltage $U_0 = 0.3$ (*1*) and 2.0 kV (*2*); (b) the same for Ta^{n+} ions for $U_0 = 0.2$ (*1*) and 1.5 keV (*2*).

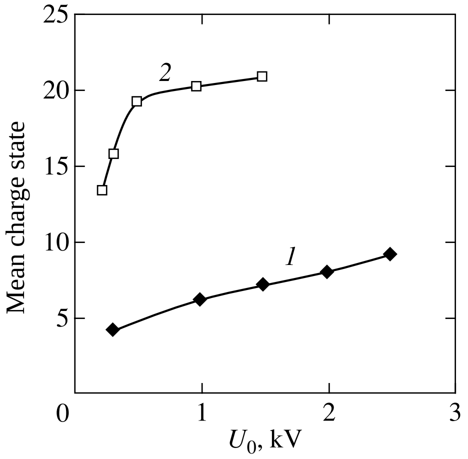


Fig. 6. The mean charge state for cooper (*1*) and tantalum (*2*) ions versus the storage voltage.

similar, but the maximum detectable charge of copper ions is +19, while for tantalum it lies near +50. In addition, it can be seen that the range of the charge states of the ions being detected becomes wider upon an increase in the capacitor voltage.

An important parameter for estimating the effectiveness of the ionization mechanism is the mean charge of the ion beam, which can be obtained from the charge distribution of ions. The dependence of this quantity on the storage voltage is shown in Fig. 6 for both species of ions. It can be seen that the mean charge of the copper ion beam increases monotonically with U_0, while the mean charge of tantalum ions attains its maximum value for U_0 = 500 V and then remains virtually unchanged.

3. DISCUSSION

The results of measurements show that the plasma jet of a vacuum discharge produces beams of accelerated multiply charged ions of the cathode material. The obtained values of the ion charge are much higher than those observed earlier in various types of vacuum spark [6, 7]. Multiply charged ions can be produced only in a plasma with a high electron temperature. The simplest estimate of this temperature was obtained in the model of ionization equilibrium from the relative concentrations of ions with different charge states in the vicinity of the peak of the distribution presented in Fig. 5a, i.e., for the major part of ions. The obtained value of $T_e \sim$ 100 eV is the order of the temperature measured earlier in the cathode jet of a laser-induced low-current vacuum discharge [8].

An important feature of the production of multiply charged ions is the short-term “burst” nature of the process. Ions are produced at the initial stage of the discharge 400–600 ns prior to the attainment of the peak value of current. The duration Δt of this process is several tens of nanoseconds and is much smaller than the time of the discharge current rise. This indicates a small spatial scale δ of the ion production region, which can be estimated using the measured value of Δt: $\delta \geq \Delta t V \approx 6 \times 10^{-2}$ cm (here, $V = 1.2 \times 10^6$ cm/s is the velocity of the cathode jet in spark discharges [9]).

It is well known that the vacuum discharge plasma is generated in microscopic regions on the surface of the cathode and expands towards the anode in the form of a spatially localized cathode flame. The size of the flame at the instant t_1 of ion production can be estimated at $\delta_1 = t_1 V \approx 0.4$–$0.5$ cm, which is smaller than the length of the discharge gap, but considerably exceeds the scale δ of the ion production region. It follows hence that multiply charged ions are produced in a local region of hot plasma in the front of the cathode flame moving towards the anode, i.e., before the plasma fills the discharge gap.

The reason for the formation of the local region in the hot plasma may be compression by the intrinsic magnetic field of the current. This phenomenon is well known for high-current discharges for which the formation of such microscopic regions at the constriction of the plasma filament due to the pinch effect was observed at a late stage of the discharge for currents of the order of 100 kA and higher [2, 3]. However, the results of recent experiments [8] and theoretical calculations [10] indicate the possibility of formation of microscopic structures with $T_e \sim 100$ eV in a cathode flame plasma at the initial stage of a vacuum discharge for currents from hundreds [10] to several amperes [8]. The process occurs near the boundary of the plasma cloud expanding into the vacuum, where the density of the inhomogeneous plasma decreases and the magnetic pressure becomes higher than the pressure in the plasma even for such weak currents. Hence, we can naturally assume that the pinch effect leading to local heating of the cathode flame plasma is possible in our case also for currents of the order of several kiloamperes, i.e., almost two orders of magnitude lower than in “classical” systems. The necessary condition for this effect is the formation of a plasma flame with a high density gradient in its front. The formation of such a flame is ensured by a high rate of the discharge current rise, which is accompanied by the formation of corresponding emission centers on the surface of the cathode [9] and by a rapid increase in the plasma density in the vicinity of the cathode.

Let us now compare our results with those obtained in [6], where the ion composition of the plasma of a vacuum spark with currents close to values typical of our experiments (about 10 kA) was investigated. However, in contrast to our experiments, the measurements were made for time intervals considerably exceeding the time of filling of the discharge gap with the plasma and led to the maximum charge +7 for copper ions, which is much lower than the values obtained by us. This indicates the important role of the inhomogeneous structure of the cathode jet in the production of multiply charged ions at the initial stage of the discharge.

The characteristic feature of our experiments is that, in contrast to the high-voltage discharge studied in [5], the cathode flame moves towards the anode not in a vacuum, but in the plasma generated by the igniter. The experimental technique is similar to that used for a laser-induced discharge [8] and ensures a high rate of current rise (which is a necessary condition for the emergence of the pinch effect in the jet) for a relatively low voltage at the capacitor.

Let us consider the peculiarities of the process of acceleration of multiply charged ions in the given experiment. The above analysis proved that the production and acceleration of ions occur at the front of the cathode flame expanding into a vacuum. In this case, the observed linear dependence of the mean energy of ions on the charge in a wide range of its variation indicates the electrostatic mechanism of acceleration of ions in the self-consistent field emerging at the flame front and directed oppositely to the external electric field applied to the electrodes. Another important feature of the acceleration process is the closeness of the mean energy of ion per unit charge $\langle E_Z/Z \rangle$ to the energy eU_0 acquired by an electron in an external electric field. These results are in qualitative agreement with the conclusions drawn on the basis of the model proposed in [11], where ions at the front of the cathode flame are accelerated by the electric field of the space charge of an electron beam. The beam is emitted at the boundary of the plasma flame and ensures the closure of the rapidly increasing current between the plasma boundary and the anode. The energy of the ions is determined by the energy of the electrons in the beam and is close to $E \sim ZeU_0$.

Let us now compare our results with the parameters of the ions of the plasma formed as a result of exposure of the target to a high-power laser pulse. The measurements of the energy and charge distributions of ions of the laser plasma were made using a technique similar to that used in our experiments. In addition to the signals from multiply charged ions of metals, a signal from H^+ ions was registered, which was 5–10 times stronger than the signals from ions of other light impurities whose effect on the total charge distribution was disregarded in these experiments [12].

Ion beams in the experiment described above and the ions of laser plasma have similar characteristics. For example, the charge state distribution and the form of its variation are the same for different materials of the cathode (target), and the maximum detected values of charge and energy of the ions of a given species are close. The measurements also proved that the mean charge of the copper ion beam increases monotonically with the storage voltage, while the charge state distribution of tantalum ions remains virtually unchanged upon the variation of the storage voltage (i.e., the energy supplied to the discharge) in a wide range. The plasma emerging as a result of irradiation of the tantalum target by a high-power laser pulse also exhibits a relatively weak dependence of the charge state of the ion component on the pulse energy [12].

A structure of the energy spectra of ions similar to that depicted in Fig. 2 was observed for multiply charged ions of a laser plasma [1]. As in our experiments, the maximum and mean energies of different ion components were almost linear functions of their charge states. In addition, the energy spectra of accelerated ions, which were obtained as a result of computer simulation of the expansion of a laser plasma bunch into a vacuum [13], also have a structure similar to that observed in the given experiment (i.e., a relatively cold core and a hot “tail”). According to the existing models of this process, plasma ions are accelerated by the electric field of charge separation, emerging at the plasma–vacuum interface as a result of the departure of hot electrons from the plasma [12, 14]. Thus, the mechanisms of acceleration of ions in the current-free laser plasma expanding into the vacuum and in the current-carrying plasma of a vacuum spark are of similar origins.

REFERENCES

1. W. Mroz, P. Paris, J. Wolowski, *et al.*, Fusion Eng. Des. **32–33**, 425 (1996).
2. K. N. Koshelev and N. R. Pereira, J. Appl. Phys. **69**, R21 (1991).
3. C. R. Negus and N. J. Peacock, J. Phys. D **12**, 91 (1979).
4. A. A. Gorbunov, M. A. Gulin, A. N. Dolgov, *et al.*, Preprint No. 024-88, MIFI (Moscow Engineering Physics Institute, Moscow, 1988).
5. E. D. Korop and A. A. Plyutto, Zh. Tekh. Fiz. **37**, 72 (1970) [Sov. Phys. Tech. Phys. **12**, 53 (1970)].
6. A. Anders, I. G. Brown, R. A. MacGill, *et al.*, IEEE Trans. Plasma Sci. **25**, 718 (1997).
7. M. P. Stockli and D. Fry, Rev. Sci. Instrum. **68**, 3053 (1997).
8. N. Vogel’, Pis’ma Zh. Éksp. Teor. Fiz. **67**, 622 (1998) [JETP Lett. **67**, 647 (1998)].
9. G. A. Mesyats, *Ectons* (Nauka, Yekaterinburg, 1994), Part 2.
10. E. A. Zverev and I. A. Krinberg, Pis’ma Zh. Tekh. Fiz. **24**, 486 (1998) [Tech. Phys. Lett. **24**, 728 (1998)].
11. S. A. Barengol’ts, G. A. Mesyats, and É. A. Perel’shteĭn, Zh. Éksp. Teor. Fiz. **118**, 1358 (2000) [JETP **91**, 1176 (2000)].
12. L. Laska, J. Krasa, K. Masek, *et al.*, Czech. J. Phys. **46**, 1099 (1996).
13. A. V. Kuznetsov, T. Zh. Esirkepov, F. F. Kamenets, and S. V. Bulanov, Fiz. Plazmy **27**, 225 (2001) [Plasma Phys. Rep. **27**, 211 (2001)].
14. S. V. Bulanov, T. Zh. Esirkepov, F. Califano, *et al.*, Pis’ma Zh. Éksp. Teor. Fiz. **71**, 593 (2000) [JETP Lett. **71**, 407 (2000)].

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 93, No. 6, 2001, pp. 1222–1230.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 6, 2001, pp. 1411–1420.
Original Russian Text Copyright © 2001 by Makarov.

PLASMA, GASES

The Formation of an Intense Secondary Pulsed Molecular Beam and Obtaining Accelerated Molecules and Radicals in It

G. N. Makarov*
Institute of Spectroscopy, Russian Academy of Sciences, Troitsk, Moscow oblast, 142190 Russia
**e-mail: g.makarov@isan.troitsk.ru*

Received July 6, 2001

Abstract—A method for obtaining an intense secondary pulsed molecular beam is described. The kinetic energy of molecules in the beam can be controlled by vibrational excitation of the molecules in the source under high-power IR laser radiation. A compression shock (shock wave) is used as a source of secondary beams. The shock wave is formed in interaction between an intense pulsed supersonic molecular beam (or flow) and a solid surface. The characteristics of the secondary beam were studied. Its intensity and the degree of gas cooling in it were comparable with the corresponding characteristics of the unperturbed primary beam. Vibrational excitation of molecules in the shock wave and subsequent vibrational–translational relaxation, which occurs when a gas is expanded in a vacuum, allow the kinetic energy of molecules in the secondary beam to be substantially increased. Intense [$\geq 10^{20}$ molecules/(sr s)] beams of SF_6 and CF_3I molecules with kinetic energies approximately equal to 1.5 and 1.2 eV, respectively, were generated in the absence of carrier gases, and SF_6 molecular beams with kinetic energies approximately equal to 2.5 and 2.7 eV with He (SF_6/He = 1/10) and H_2 (SF_6/H_2 = 1/10) as carrier gases, respectively, were obtained. The spectral and energy characteristics of acceleration of SF_6 molecules in the secondary beams were studied. The optimal conditions were found for obtaining high-energy molecules. The possibility of accelerating radicals in secondary molecular beams was demonstrated. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Currently, molecular (atomic) beams [1, 2], including high-energy beams [1], are extensively used in research work. The kinetic energy of molecules (atoms) in high-energy beams far exceeds their thermal energy, which equals $E_{\text{kin}} \leq 0.05$ eV at ≈300 K. Intense [$\geq 10^{20}$ molecules/(sr s)] beams of accelerated molecules with kinetic energies in the range of approximately one to several electronvolts are used in various fundamental and applied studies (in studying chemical reactions with energy barriers, elastic and inelastic collisions, interactions of molecules with surfaces, etc.) [1]. The most extensively used technique for the preparation of intense molecular beams is their isolation from gas dynamically cooled flows with the use of skimmers; the initial flows are generated by pulsed nozzles [2]. The most important characteristics of pulsed beams are their intensity, duration, velocity, and the spread of molecular velocities (the degree of gas cooling).

The specified energy range is difficult to work with, and we do not known a universal method for generating the corresponding molecular (atomic) beams. There are several techniques for obtaining such beams [1] (also see [3, 4] and the references therein). Most of them are, however, fairly complex (for instance, creating an optical breakdown or a radiofrequency or arc discharge inside a nozzle) and are only applicable to atoms, not molecules. They were largely developed for generating continuous beams.

Aerodynamic acceleration [5, 6] is used for this purpose most frequently. Molecules to be studied are then diluted with a lighter carrier gas (He, H_2, etc.). Another method involves heating molecules above the nozzle to high temperatures ($T_0 \approx 3000$ K); also possible is a combination of the two methods. Aerodynamic acceleration is not very effective when the ratio between the molecular weights of the gas to be studied and the carrier gas is small.

The energy of molecules in a beam generated by heating the nozzle is determined by gas temperature T_0 before expansion through the nozzle,

$$\frac{1}{2}mv^2 = \frac{\gamma}{\gamma-1}k(T_0-T), \qquad (1)$$

where v is the steady flow speed, m is the mass of the molecule, $\gamma = c_p/c_v$ is the ratio of the specific heat capacities, k is the Boltzmann constant, and T is the gas temperature in the flow.

In [7], argon atoms diluted with helium were accelerated to several electronvolts in a continuous beam by combining nozzle heating and aerodynamic acceleration. In [8], continuous beams of xenon atoms diluted

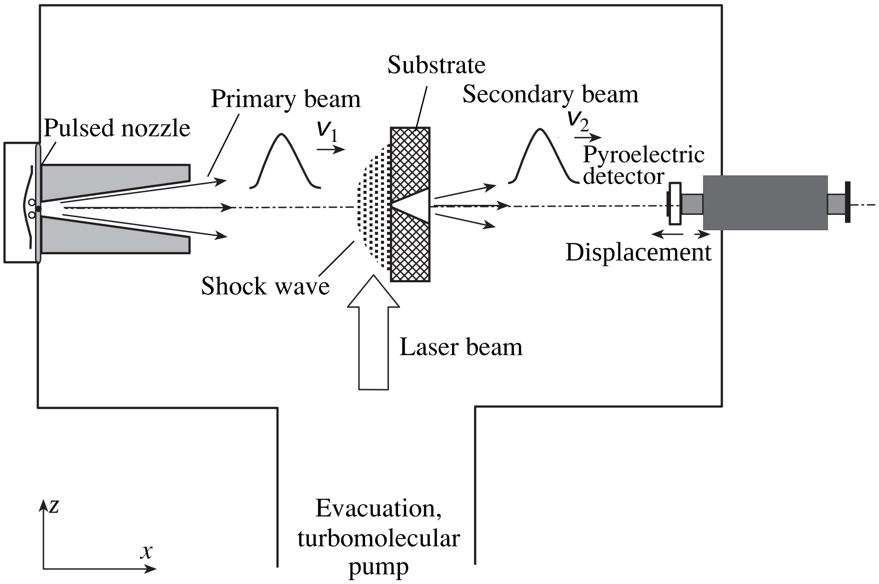


Fig. 1. Schematic of measurements. Secondary molecular beam formation with the use of a substrate with an orifice in the form of a divergent cone.

with hydrogen (Xe/H_2 = 0.23/100) with a kinetic energy up to $E_{kin} \lesssim 30$ eV were generated.

Heating pulse nozzles to high temperatures is a problematic approach, because the materials used in such nozzles (in particular, elastomers and plastics) decompose at high ($T \geq 200°C$ temperatures [2]. According to [9, 10], accelerated molecular beams can be generated using vibrational excitation of molecules by IR laser pulses in the hydrodynamic expansion region at the exit from the nozzle. Accelerated SF_6 and CF_3I molecular beams with kinetic energies approximately equal to 0.5 and 0.74 eV, respectively, were obtained [10, 11]. Using this approach, however, offers no possibility for obtaining high excitation energy densities because of the arising of optical breakdowns at nozzle beam-exit holes. In addition, these techniques can only be used to accelerate a small fraction of molecules in a beam.

It follows from the aforesaid that a promising technique for obtaining high-energy molecular beams would be excitation of molecules by high-power IR laser radiation inside a pulsed beam source itself, that is, before the gas flows through the nozzle. Precisely this possibility was used for the first time in the method for generating secondary pulsed molecular beams suggested in [12, 13]. Intense beams of accelerated SF_6 molecules with kinetic energies up to $E_{kin} \lesssim 1.5$ eV were obtained [13]. In this work, the suggested method is described in detail, and the results of a more complete study of acceleration of molecules in secondary pulsed molecular beams induced by high-power IR laser radiation are given.

2. EXPERIMENTAL SETUP AND PROCEDURE FOR MEASUREMENTS

The source of a secondary molecular beam in the suggested method is a compression shock (shock wave) [14–16] formed in interaction between a pulsed gas dynamically cooled molecular flow and a solid surface [17, 18]. A schematic of the procedure is given in Fig. 1. An intense [$\geq 10^{21}$ molecules/(sr s)] wide-aperture (divergence $\omega \approx 0.05$ sr) molecular beam (or flow) fell onto a solid surface, a substrate with an orifice in the center. A polished Dural substrate 7.5 mm thick was used; the substrate was mounted at a distance of $x \approx 60$ mm from the nozzle. The orifice had the form of a divergent cone with inlet and outlet diameters $d_{in} \approx 2$ mm and $d_{out} \approx 5$ mm. Orifice walls were polished.

When a primary beam fell on the substrate, a shock wave was formed in front of it; gas density, pressure, and temperature in the shock wave were substantially higher than in the incident beam [19, 20]. According to the estimates made in [19], the concentration of SF_6 molecules in the shock wave varied from approximately 10^{16} to 5×10^{17} cm^{-3} depending on the intensity of the primary beam. While the shock wave remained in front of the substrate, the gas flowed through the orifice into the high-vacuum part of the chamber. As a result, a new secondary pulsed molecular beam was formed, and its characteristics were different from those of the primary beam. Secondary beams were also generated with the use of hollow convergent truncated cones and convergent–divergent cones of the Laval nozzle type in place of a substrate with a cone orifice. The intensities of secondary molecular beams generated in cones were substantially (5–7 times) higher than those obtained using the substrate. When the primary beam interacted with a cone, a shock wave was formed inside the convergent part of the cone.

Molecules in a shock wave are very simple to excite (Fig. 1). When a cone was used to generate high-energy beams, the convergent part of the cone was replaced by a hollow truncated tetrahedral pyramid made of thin NaCl plates transparent to CO_2 laser radiation. This allowed us to excite molecules inside the pyramid directly prior to outflow from the secondary nozzle. Multiphoton absorption of molecules in a strong IR field [21] increased their internal (largely, vibrational) energy. Subsequent vibrational–translational (V–T) relaxation during gas expansion into a vacuum caused energy transfer from vibrational to translational degrees of freedom. As a result, molecules experienced acceleration. It follows that IR laser-induced acceleration of molecules can schematically be described by the equation

$$\mathrm{M}(v_0) + nh\nu \longrightarrow \mathrm{M}^*(v_0) \xrightarrow{V-T} \mathrm{M}(v_L), \qquad (2)$$

where M and M^* are molecules in the ground and vibrationally excited states, respectively; v_0 and v_L are the velocities of molecules in the absence of excitation and under pulsed laser excitation; and $nh\nu$ is the energy absorbed by one molecule from a laser pulse. Under the conditions of our experiments, the concentration of molecules in a shock wave in front of the surface was fairly high (the pressure was on the order of one to several Torr—see below). For this reason, both multiphoton excitation and V–T relaxation were comparatively effective and substantially increased the velocity of molecules in secondary beams. Gas heating in a shock wave caused by deceleration (see below) also increased the velocity of molecules in secondary beams.

Primary beams were generated by a pulsed nozzle of the current loop type [22]. The hole diameter was 0.75 mm. The time of opening was about 40 μs (at half-height). Gas pressure above the nozzle varied in the range of approximately 0.1 to 7 atm. The nozzle cut had the shape of a cone with an aperture angle of 15°. The cone length was 35 mm. The vacuum chamber in which molecular beams were formed was evacuated to about 1×10^{-6} Torr by a turbomolecular pump. The number of molecules flowing from the nozzle during one pulse depended on the gas pressure in front of the nozzle and varied approximately from 3×10^{15} to 1.1×10^{17} molecules/pulse in our experiments. A technique for measuring the number of molecules in a pulse was described in detail in [19, 20].

Vibrational excitation of molecules was created by a *TEA* CO_2 tunable laser. Pulse energy was up to 3 J. The generation pulse consisted of a peak about 100 ns wide at half-height and a tail about 0.6 μs wide containing about 50% of the total pulse energy. Molecules were excited in a shock wave directly before outflow from the secondary nozzle. Laser radiation was slightly focused onto the excitation zone by a NaCl lens with a 2-m focal length. The laser beam was directed parallel with the surface (normally to the primary beam). The laser spot size in the excitation zone was about 10×10 mm^2.

Molecular beams were detected by a pyroelectric detector with a 3–5 μs time resolution [23, 24]. The active element of the detector was 4×4 mm^2 in size. Molecules flowing within the body angle determined by the size of the detector active element and the distance from the beam source were recorded. The detector was described in detail in [25]. The pyroelectric detector could be moved along the beam axis. This allowed us to measure the time-of-flight spectra of molecules at different distances from primary and secondary beam sources [23]. The time-of-flight spectra were used to determine the velocities of the beams and the spread of the velocities of molecules within them. Our experiments aimed at determining the characteristics of secondary molecular beams and detecting high-energy molecules in them.

3. RESULTS AND DISCUSSION

3.1. The Formation of Secondary Molecular Beams and the Determination of Their Characteristics

As mentioned above, obtaining a secondary molecular beam requires that a shock wave be formed in front of the surface when the primary beam falls on it. The conditions of shock wave formation in interaction between a pulsed gas dynamically cooled supersonic molecular flow and the surface were studied in detail for the examples of SF_6 and CF_3I molecular flows in [19, 20] and [26], respectively. It was found that no shock wave was formed at low gas pressures above the nozzle ($P \le 0.2$ atm for SF_6), when the mean concentration of molecules in the incident flow was $N_1 \le 3 \times 10^{14}$ cm^{-3}.

In our experiments, we used a molecular beam source and a geometry of beam formation different from those described in [19, 20, 26]. For this reason, to create the conditions under which a shock wave was formed, we had to measure the number of molecules outflowing from the nozzle during one pulse (N_b) and estimate the intensity of the primary beam. The procedure for measuring N_b was described in detail in [19, 20]. The dependences of N_b on the pressure of SF_6 above the nozzle are shown in Fig. 2 for two values of the potential applied to the nozzle to open it U_n. The N_b value increases almost linearly at pressures ≤1 atm and reaches saturation at high pressures. At an SF_6 pressure above the nozzle of 3 atm and $U_n = 3.2$ kV, the number of molecules in the beam was $N_b \approx 9 \times 10^{16}$ cm^{-3}. The molecules flowed within a body angle of $\omega \approx 0.05$ sr determined by the aperture angle of the nozzle cone ($\alpha = 15°$), and the intensity of the primary SF_6 molecular beam per time $\Delta\tau_b \approx 100$ ns was $I_b \ge 10^{21}$ molecules/(sr s). Estimates of the mean concentration of SF_6 molecules in the primary beam are given in Subsection 3.3.

It was shown in [12, 13] that the primary molecular beam only flowed through the cone at a low gas pressure above the nozzle ($p_\Sigma \le 0.2$ atm), when the intensity of the primary beam was low and no shock wave was formed in front of the surface [19, 20]. The primary beam pulse narrowed as the pressure of the gas above the nozzle increased, when a shock wave was formed in the cone (in front of the substrate in Fig. 1). Simultaneously, a secondary beam pulse appeared and began to sharply increase in amplitude.

At a comparatively high primary molecular beam intensity [$\ge 10^{20}$ molecules/(sr s)], the intensity of the secondary beam is commensurate with the intensity of the unperturbed primary beam (in the absence of a substrate in its path). The dependences of the intensities of the unperturbed primary beam (curve *1*) and secondary beams (curves *2* and *3*) on the pressure of SF_6 above the nozzle are shown in Fig. 3. The secondary beams described by curves *2* and *3* were generated using a convergent cone (d_{in} = 11 mm, d_{out} = 2.8 mm, total length 32 mm) and a convergent–divergent cone of the Laval nozzle type (d_{in} = 14 mm, d_0 = 2 mm, d_{out} = 7 mm, total length 40 mm, convergent part length 30 mm), respectively. The distance between the nozzle and the detector was 143 mm, and the distance between the cone waist and the detector was 79 mm. At an SF_6 pressure above the nozzle of $p \ge 1.0$ atm, the pyroelectric signal from secondary molecular beam *3* was larger than that from the primary beam. This means that the intensity of secondary molecular beam *3* was comparable with the intensity of the primary beam, even considering an approximately quadratic dependence of the pyroelectric signal on the distance between the beam source and the detector.

The measured primary and secondary molecular beam parameters are listed in Table 1. It was found experimentally that the duration of the secondary molecular beam, its velocity, and the spread of molecular velocities in it did not differ substantially from the corresponding characteristics of the primary beam. As expected, the largest difference was between the spreads of molecular velocities. The spread of velocities was by 20–40% larger in the secondary beam. It follows from the data given in Table 1 that a fairly substantial gas cooling occurred in the secondary beam (the Mach number was $M_2 \approx v_2/\Delta v_2 \approx 5$).

3.2. Acceleration of Molecules in Secondary Beams

The dependence of the velocity of SF_6 secondary molecular beams on the energy density of CO_2 laser radiation is shown in Fig. 4. Molecules were excited in the beam source directly prior to outflow from the “secondary” nozzle (inside the hollow truncated tetragonal pyramid made of NaCl plates and attached to the front wall of the substrate with a conic orifice—see Fig. 1). The laser was tuned to a frequency of 938.7 cm^{-1} (the 10*P*(26) laser line), which was in resonance with

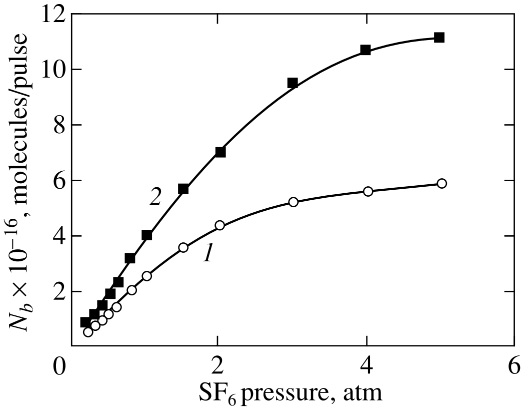


Fig. 2. Dependences of the total number of molecules flowing out of the nozzle in one pulse on SF_6 pressure above the nozzle. Voltage applied to the nozzle: U_n = (*1*) 3.0 and (*2*) 3.2 kV.

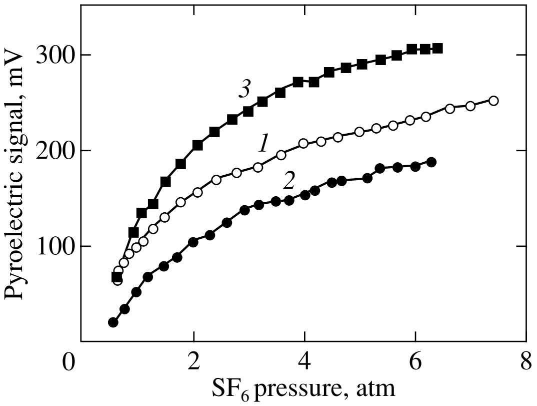


Fig. 3. Dependences of (*1*) unperturbed primary and (*2, 3*) secondary molecular beam intensities on SF_6 pressure above the nozzle. U_n = 3.2 kV. Secondary beams were generated in cones. See text for cone parameters and experimental conditions.

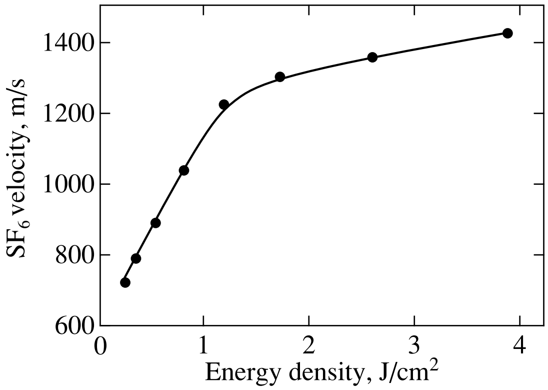


Fig. 4. Dependence of secondary SF_6 molecular beam velocity on laser radiation density. Molecules were excited at 938.7 cm^{-1} [10*P*(26) laser line]. SF_6 pressure above the nozzle 5.8 atm, U_n = 3.2 kV.

Table 1. Measured parameters of primary and secondary molecular beams

Gas composition	Pressure above nozzle, atm	Primary beam			Secondary beam		
		v_1, m/s	Δv_1, m/s	$v_1/\Delta v_1$	v_2, m/s	Δv_2, m/s	$v_2/\Delta v_2$
SF_6	6.2	480	77	6.2	447	87	5.1
SF_6	5.0	560	62	9.1	476	82	5.8
SF_6	3.0	545	88	6.2	517	133	3.9
SF_6/H_2 (1/10)	3.2	1130	97	11.6	1090	156	7.0
SF_6/H_2 (1/5)	3.0	1000	115	8.6	896	120	7.5
SF_6/He (1/10)	3.0	940	85	11.1	1000	230	4.4
SF_6/CH_4 (1/10)	3.1	870	99	8.8	835	128	6.5
CF_3I	4.6	417	53	7.9	405	57	7.1

the ν_3 mode of SF_6 molecules ($\nu \approx 948$ cm^{-1} [27]). The mean velocity of molecules in the absence of laser excitation was $v_0 = 460$ m/s; under laser excitation conditions, it increased to $v_L > 1400$ m/s.

The dependence of the kinetic energy of SF_6 molecules in the secondary molecular beam on the CO_2 laser exciting radiation frequency is shown in Fig. 5a. The secondary molecular beam was formed in a cone described above (the same as in Fig. 4). The SF_6 pressure above the nozzle was 5.8 atm. The laser radiation energy density was approximately 3.9 J/cm^2. The mean kinetic energy of molecules in the absence of laser excitation was $E^0_{\text{kin}} \approx 0.163$ eV. For comparison, the linear absorption spectrum of SF_6 in the region of ν_3 vibrations recorded in [28] at $T \approx 300$ K is shown in the bottom part of the figure. Our spectral dependence was fairly broad. This was likely to be caused by a comparatively high pressure and a fairly high vibrational and rotational temperature of SF_6 in the shock wave (see below). A maximum acceleration was observed when molecules were excited by laser lines 10*P*(22)–10*P*(26) (in the region of 940 cm^{-1}). A small maximum at 931 cm^{-1} might be related to excitation of $^{34}SF_6$ molecules ($\nu_3 \approx 930.5$ cm^{-1} [29]). The content of $^{34}SF_6$ in SF_6 with a natural isotopic composition, which was used in our experiments, was about 4.2%.

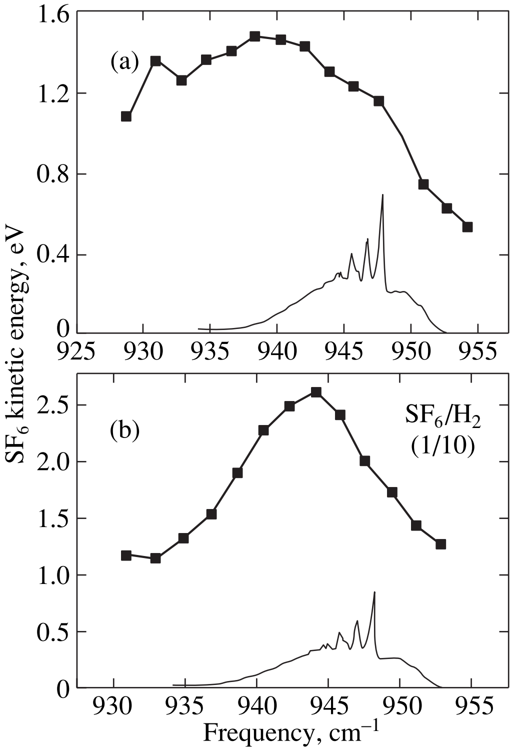


Fig. 5. Dependences of kinetic energy of SF_6 molecules in secondary molecular beams on exciting laser radiation frequency: (a) SF_6 molecular beam without a carrier gas and (b) $SF_6/H_2 = 1/10$ molecular beam.

A similar dependence obtained for an SF_6/H_2 (1/10) molecular beam is shown in Fig. 5b. This spectral dependence was much narrower, which was likely caused by fairly fast vibrational–translational relaxation of SF_6 in the H_2 carrier gas (see below, Subsection 3.5). The vibrational temperature of SF_6 in the shock wave was much lower than under excitation of SF_6 in the absence of a carrier gas. The IR multiphoton excitation spectrum was narrowed and displaced to the linear absorption spectrum region [18]. As a result, the spectral dependence of acceleration of molecules was also narrowed.

The data obtained in accelerating SF_6 and CF_3I molecules in secondary molecular beams are listed in Table 2. We generated SF_6 molecular beams with kinetic energies of $E^L_{\text{kin}} \approx 1.5$ eV ($v_L \approx 1400$ m/s) in the absence of a carrier gas and $E^L_{\text{kin}} \approx 2.5$, 2.6, and 2.7 eV in the presence of He, CH_4, and H_2 carrier gases, respectively. These values were substantially higher than those reported in [4, 10]. Note that the kinetic energies of SF_6 obtained in [12] were somewhat lower than in [13] and this work, because the cross-section size of the laser beam in the excitation zone (in the shock wave) was

Table 2. Experimental data on acceleration of SF_6 and CF_3I in secondary molecular beams

Gas composition	Pressure above nozzle, atm	CO_2 laser line	Energy density, J/cm^2	Mean velocity of molecules, m/s		Kinetic energy of molecules, eV	
				v_0	v_L	E^0_{kin}	E^L_{kin}
SF_6	6.6	10*P*(26)	3.7	460	1400	0.163	1.5
SF_6	3.0	10*P*(20)	3.7	474	1265	0.173	1.23
SF_6	1.0	10*P*(20)	3.7	470	1155	0.17	1.05
SF_6/H_2 (1/10)	3.2	10*P*(20)	3.5	1090	1870	0.91	2.7
SF_6/He (1/10)	6.0	10*P*(20)	3.7	1050	1800	0.85	2.5
SF_6/CH_4 (1/10)	5.0	10*P*(24)	3.7	1020	1835	0.8	2.6
CF_3I	4.6	9*R*(12)	1.2	417	1065	0.18	1.2

smaller in the earlier experiments ($\approx 8 \times 8$ mm^2, [12]) than in this work. A laser pulse therefore excited a smaller number of molecules. As a result, the kinetic energy of accelerated molecules was also lower.

3.3. Estimation of SF_6 Parameters in a Shock Wave

Briefly consider the parameters of SF_6 in a shock wave. Note that strongly nonuniform, nonequilibrium, and nonstationary shock wave conditions and a large number of processes that occur in shock waves considerably complicate estimating the parameters of a gas in a shock wave. Several techniques for estimating the concentration and temperature of SF_6 in a normal shock wave were described in [19, 20]. These estimates were based on using the parameters of a molecular flow (beam) incident on the surface and gas dynamic equations. We will estimate the concentration and temperature of SF_6 in a normal shock wave in front of the surface under the conditions of our experiments. The SF_6 concentration was probably higher in a shock wave formed inside the cone than in that formed in front of the surface.

The ratio between the concentrations of SF_6 in a shock wave (N_2) and in the incident (primary) beam (N_1) can roughly be assumed to equal the ratio between the incident beam extent (Δx_b) and the shock wave front width (Δx_{sh}),

$$\frac{N_2}{N_1} \approx \frac{\Delta x_b}{\Delta x_{sh}}.$$

In our experiments, $\Delta x_b \approx 4.8$ cm and $\Delta x_{sh} \approx 3$ mm [19, 20]. Therefore, $N_2 \approx 16N_1$. For instance, at a 3-atm pressure of SF_6 above the nozzle, the number of molecules outflowing from the nozzle per pulse equaled $N_b \approx 9 \times 10^{16}$ molecules/pulse (see Fig. 2). According to the estimates made in [19], the volume of the primary beam was $V_b \approx 15$ cm^3. Therefore, $N_1 \approx 6 \times 10^{15}$ cm^{-3} and $N_2 \approx 9.6 \times 10^{16}$ cm^{-3}.

The largest N_2 value can be found from pressure p_2 in the shock wave ($N_2 = p_2/kT_2$). Pressure p_2 can be estimated as

$$mv_1N_1 = p_2S_b\Delta\tau_b, \qquad (3)$$

where mv_1N_1 is the momentum of molecules in the beam incident on the surface, S_b is the cross section of the beam at the surface, and $\Delta\tau_b$ is the duration of the beam. Substituting the required characteristics of the SF_6 molecular beam [$m = 146 \times 1.67 \times 10^{-24}$ g, $v_1 \approx 4.8 \times 10^4$ cm/s, $N_b \approx 9 \times 10^{16}$, $S_b \approx 3$ cm^2, $\Delta\tau_b \approx 10^{-4}$ s, and $T_2 \approx 600$ K (see below)] yields $p_2 \approx 38$ mbar ($\approx$28.6 Torr) and $N_{2\max} \approx 5 \times 10^{17}$ cm^{-3}.

An increase in gas temperature (ΔT) caused by deceleration in a shock wave can be estimated from the law of the conservation of energy for SF_6 molecules in an incident beam and a shock wave [20]. We assume that the vibrational degrees of freedom of SF_6 remain unheated in a shock wave, which allows us to write

$$\frac{mv_1^2}{2} = 3k\Delta T.$$

Substituting the required characteristics of SF_6 beams yields $\Delta T \approx 635$ K. On the other hand, although vibrational–translational relaxation of SF_6 molecules occurs at a fairly low rate ($p\tau_{V-T} \approx 150$ μs Torr [30]), the vibrational degrees of freedom of the molecule also experience heating at gas pressures of several Torr in a shock wave, at least, in part. It is likely that, for this reason, ΔT was smaller than the estimate just obtained. The translational and rotational temperatures of SF_6 in incident beams were $T_{1,\text{tr}} \approx T_{1,\text{rot}} \le 40$ K [24]; it follows that, in the shock wave, these temperatures were $T_{2,\text{tr}} \approx T_{2,\text{rot}} \le 675$ K. The vibrational temperature of SF_6 in the incident beam was $T_{1,\text{vib}} \le 150$ K [24]. In the shock wave, it was likely to be higher.

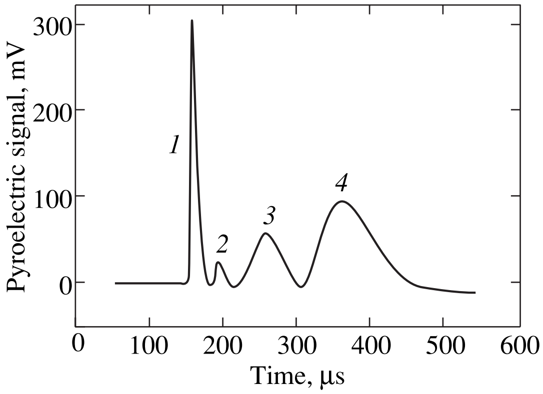


Fig. 6. Oscillogram showing time-of-flight spectra of CF_3 radicals (peak 2) and accelerated and unaccelerated CF_3I molecules (peaks 3 and 4, respectively) in a secondary molecular beam. The first peak corresponds to scattered exciting laser pulse radiation.

3.4. Estimation of the Fraction of Absorbed Energy Consumed in Acceleration

The literature data on the energy absorbed by SF_6 molecules from the IR laser pulse field and the data on acceleration of SF_6 obtained in this work can be used to estimate the fraction of the absorbed energy consumed to accelerate molecules. As mentioned above, the pressure of SF_6 in the shock wave during its formation is one to several Torr. At such pressures, all molecules in the gas volume subjected to laser radiation participate in absorption; that is, the fraction of excited molecules equals one. The number of absorbed quanta per molecule is $\langle n \rangle \approx 25$ when SF_6 is excited by a $10P(16)$ laser line and the energy density is $\Phi \approx 3.5$–4 J/cm^2 [31]. The mean energy absorbed by one SF_6 molecule therefore equals $E_{\mathrm{ab}} = \langle n \rangle h\nu \approx 2.94$ eV [$h\nu = 0.1176$ eV is the laser quantum energy at the $10P(16)$ line]. The kinetic energy of accelerated SF_6 molecule excited at the $10P(16)$ frequency is $E_{\mathrm{kin}}^{L} \approx 1.3$ eV (see Fig. 5a). It follows that about 45% of the absorbed energy is consumed to accelerate molecules. Note that some part of the absorbed energy becomes lost in collisions of molecules with nozzle walls, some part is consumed in heating the gas inside the shock wave itself, and some part is likely to remain in internal degrees of freedom, largely in the form of vibrational excitation. It follows from the results listed in Table 1 that there is time enough for the translational and rotational degrees of freedom of SF_6 to be cooled to $T \le 50$ K during gas expansion. A substantial absorbed energy part is likely to remain in vibrational degrees of freedom, because vibrational–translational relaxation in SF_6 (and CF_3I) molecules occurs at a fairly low rate,

$$p\tau_{V-T} \approx \begin{cases} 150\ \mu\text{s torr for } SF_6\ [30], \\ 350 \pm 100\ \mu\text{s torr for } CF_3I\ [32]. \end{cases}$$

Note that the rate of this process substantially increases in strong IR fields at high vibrational excitation levels [33]. It also increases in some gases [30], for instance,

$$p\tau_{V-T} \approx \begin{cases} 9.3\ \mu\text{s torr in } H_2, \\ 41\ \mu\text{s torr in He}, \\ 30\ \mu\text{s torr in } CH_4. \end{cases}$$

When these gases are used as carriers, vibrational–translational relaxation in SF_6 proceeds more effectively than in pure gaseous SF_6. As a result, molecules are more effectively accelerated in secondary beams (see Table 2).

3.5. Accelerated Radicals in Secondary Beams

We studied the possibility of obtaining accelerated radicals in secondary beams as a result of the dissociation of CF_3I molecules. The time-of-flight spectrum of CF_3 radicals and accelerated and unaccelerated CF_3I molecules is shown in Fig. 6. The secondary molecular beam was generated using a substrate with a conic orifice to which a hollow tetrahedral pyramid made of NaCl plates was attached. The pressure of CF_3I above the nozzle was 3 atm. When the primary beam fell on the pyramid, a shock wave formed inside it. Molecules in the shock wave were excited at a frequency of 1073.3 cm^{-1} [the $9R(12)$ laser line], which was in resonance with CF_3I ν_1 vibrations [34]. This frequency coincided with the maximum of the spectral dependence of the yield of IR multiphoton dissociation of CF_3I in a gas dynamically cooled flow [35]. The energy density of exciting radiation was $\Phi \approx 3$ J/cm^2. At such an energy density, the yield of CF_3I dissociation was ≥80% [36, 37].

The time-of-flight spectrum of CF_3 radicals was recorded by exciting molecules at the very beginning of shock wave formation. At long time lags between the initial molecular beam pulse and the laser pulse, the time-of-flight spectra of CF_3 radicals and accelerated CF_3I molecules overlapped each other, and CF_3 radicals could not be observed in the time-of-flight spectra recorded by the detection technique used in this work. Note that the IR multiphoton dissociation of CF_3I only yields CF_3 radicals and iodine atoms. Because of a fairly large difference between the masses of the CF_3 radical, the CF_3I molecule, and the iodine atom, we were able to record the time-of-flight spectrum of CF_3 radicals. We measured the velocity of CF_3 radicals and the spread of radical velocities in the secondary beam, which gave $v_2 \approx 800$ m/s and $\Delta v_2 \approx 120$ m/s (the Mach number equaled $M \approx v_2/\Delta v_2 \approx 6$). It follows that we obtained an intense beam of cooled CF_3 radicals with a kinetic energy of $E_{\mathrm{kin}} \approx 0.25$ eV in our experiments. CF_3 radicals were accelerated in the secondary beam because of an increase in gas temperature caused by deceleration. CF_3 radicals with a high kinetic energy can be obtained by IR or UV dissociation of CF_3I molecules accelerated in a secondary beam. For instance, the IR dissociation of CF_3I molecules accelerated to $E^L_{kin} \approx 1.2$ eV yields CF_3 radicals with $E_{kin} \geq 0.42$ eV. The dissociation of SF_6 molecules accelerated to ≈ 1.5 eV in the secondary beam yields accelerated SF_5 radicals with a kinetic energy of $E_{kin} \geq 1.23$ eV.

4. CONCLUSION

The method described in this work can be used to obtain intense molecular beams containing molecules whose kinetic energy can be controlled by vibrationally exciting them with the use of high-power laser pulses inside the beam source itself. The intensity of secondary molecular beams and the degree of gas cooling in them may be commensurate with similar characteristics of primary beams. Convergent–divergent cones of the Laval nozzle type provide optimal conditions for generating secondary molecular beams. Intense [$\geq 10^{20}$ molecules/(sr s)] SF_6 and CF_3I molecular beams with kinetic energies approximately equal to 1.5 and 1.2 eV, respectively, without carrier gases and SF_6 molecular beams with kinetic energies approximately equal to 2.5 and 2.7 eV with He (SF_6/He = 1/10) and H_2 ($SF_6/H_2 = 1/10$) as carrier gases, respectively, were obtained. It was shown that accelerated radicals could also be obtained in secondary beams by the dissociation of molecules under the action of high-power IR laser radiation in the secondary source or in the beam itself. Note in conclusion that the shortcoming of the suggested method for generating high-energy molecular beams is likely to be insufficient cooling of vibrational degrees of freedom of molecules.

ACKNOWLEDGMENTS

The author thanks V.M. Apatin and V.V. Nesterov for technical help. This work was financially supported by the Russian Foundation for Basic Research (project no. 00-03-33003-a) and the American Civil Research and Development Foundation for Independent Countries of the Former Soviet Union (grant no. RC1-2206).

REFERENCES

1. H. Pauly, in *Atomic and Molecular Beam Methods*, Ed. by G. Scoles (Oxford Univ. Press, New York, 1988), p. 124.
2. W. R. Gentry, in *Atomic and Molecular Beam Methods*, Ed. by G. Scoles (Oxford Univ. Press, New York, 1988), p. 54.
3. J. M. Girard, A. Lebehot, and R. Compargue, J. Phys. D **26**, 1382 (1993).
4. G. N. Makarov, Zh. Éksp. Teor. Fiz. **108**, 404 (1995) [JETP **81**, 218 (1995)].
5. E. W. Becker, K. Bier, and H. Burghoff, Z. Naturforsch. A **10**, 7 (1955).
6. E. Kolodney and A. Amirav, Chem. Phys. **82**, 269 (1983).
7. R. Comparque, A. Lebehot, J. C. Lemonnier, and D. Marette, in *Rarefied Gas Dynamics*, Ed. by S. S. Fisher (AIAA, New York, 1980), p. 823.
8. U. Buck, E. Lessner, and D. Pust, J. Phys. B **13**, L125 (1980).
9. V. N. Lokhman and G. N. Makarov, Pis'ma Zh. Éksp. Teor. Fiz. **61**, 172 (1995) [JETP Lett. **61**, 175 (1995)].
10. G. N. Makarov, Chem. Phys. Lett. **237**, 361 (1995).
11. V. N. Lokhman, G. N. Makarov, D. D. Ogurok, *et al.*, Khim. Fiz. **17**, 35 (1998).
12. V. M. Apatin, G. N. Makarov, and V. V. Nesterov, Pis'ma Zh. Éksp. Teor. Fiz. **73**, 735 (2001) [JETP Lett. **73**, 651 (2001)].
13. V. M. Apatin, G. N. Makarov, and V. V. Nesterov, Chem. Phys. Lett. **347**, 101 (2001).
14. Ya. B. Zel'dovich and Yu. P. Raizer, *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena* (Nauka, Moscow, 1966, 2nd ed.; Academic, New York, 1966).
15. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 6: *Fluid Mechanics* (Nauka, Moscow, 1986; Pergamon, New York, 1987).
16. G. N. Abramovich, *Applied Gas Dynamics* (Nauka, Moscow, 1991), Part 1.
17. G. N. Makarov and A. N. Petin, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 583 (2000) [JETP Lett. **71**, 399 (2000)].
18. G. N. Makarov and A. N. Petin, Chem. Phys. Lett. **323**, 345 (2000).
19. G. N. Makarov and A. N. Petin, Zh. Éksp. Teor. Fiz. **119**, 5 (2001) [JETP **92**, 1 (2001)].
20. G. N. Makarov and A. N. Petin, Chem. Phys. **266**, 125 (2001).
21. V. N. Bagratashvili, V. S. Letokhov, A. A. Makarov, and E. A. Ryabov, *Multiple Photon Infrared Laser Photophysics and Photochemistry* (Harwood Academic, New York, 1985).
22. W. R. Gentry and C. F. Giese, Rev. Sci. Instrum. **49**, 595 (1978).
23. V. M. Apatin, L. M. Dorozhkin, G. N. Makarov, and L. M. Pleshkov, Appl. Phys. B **B29**, 273 (1982).
24. V. M. Apatin and G. N. Makarov, Zh. Éksp. Teor. Fiz. **84**, 15 (1983) [Sov. Phys. JETP **57**, 8 (1983)].
25. R. V. Ambartzumian, L. M. Dorozhkin, G. N. Makarov, *et al.*, Appl. Phys. **22**, 409 (1980).
26. G. N. Makarov, S. A. Mochalov, and A. N. Petin, Kvantovaya Élektron. (Moscow) **31**, 263 (2001).
27. R. S. McDowell, J. B. Krohn, H. Flicker, and M. C. Vasquez, Spectrochim. Acta A **42**, 351 (1986).
28. J. L. Lyman, G. P. Quigley, and O. P. Judd, in *Multiple-Photon Excitation and Dissociation of Polyatomic Molecules*, Ed. by C. D. Cantrell (Springer-Verlag, Berlin, 1986), p. 34.

29. G. Baldacchini, S. Marchetti, and V. Montelatici, J. Mol. Spectrosc. **91**, 80 (1982).
30. J. I. Steifeld, I. Burak, D. G. Sutton, and A. V. Novak, J. Chem. Phys. **52**, 5421 (1970).
31. G. N. Makarov, Doctoral Dissertation in Physics and Mathematics (Inst. Spektrosk. Akad. Nauk SSSR, Troitsk, 1989).
32. J. M. Weulersse and R. Genier, Appl. Phys. **24**, 363 (1981).
33. S. A. Akhmanov, V. M. Gordienko, A. V. Mikheenko, and V. Ya. Panchenko, Pis'ma Zh. Éksp. Teor. Fiz. **26**, 603 (1977) [JETP Lett. **26**, 453 (1977)].
34. W. Fuss, Spectrochim. Acta A **38**, 829 (1982).
35. G. N. Makarov, V. N. Lokhman, D. E. Malinovskiĭ, and D. D. Ogurok, Kvantovaya Élektron. (Moscow) **25**, 545 (1998).
36. V. N. Bagratashvili, V. S. Dolzhikov, V. S. Letokhov, *et al.*, Zh. Éksp. Teor. Fiz. **77**, 2238 (1979) [Sov. Phys. JETP **50**, 1075 (1979)].
37. V. M. Apatin and G. N. Makarov, Kvantovaya Élektron. (Moscow) **10**, 1435 (1983).

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 93, No. 6, 2001, pp. 1231–1238.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 6, 2001, pp. 1421–1429.
Original Russian Text Copyright © 2001 by Smorodin.

The Onset of Convection of a Poorly Conducting Fluid in a Modulated Thermal Field

B. L. Smorodin
Perm State University, ul. Bukireva 15, Perm, 614600 Russia
e-mail: smorodin@psu.ru

Received July 20, 2001

Abstract—The stability of a poorly conducting fluid in a constant electric field of a horizontal capacitor is investigated under a variable temperature gradient. It is assumed that free charge in the fluid is generated only due to the nonhomogeneous conductivity of the fluid. The Floquet theory is used to determine the convection thresholds. The instability boundaries and the characteristics of critical perturbations are determined. In addition to the synchronous and subharmonic responses to an external action, the instability can be attributed to quasiperiodic perturbations. The low-frequency limit of modulation is considered by an asymptotic method. The critical electric Rayleigh number is represented as a function of inverse frequency and heating level. *© 2001 MAIK “Nauka/Interperiodica”.*

1. INTRODUCTION

The study of electrothermal convection is of interest in view of a possible control of convection in liquid dielectrics and a control of heat and mass transfer in high-voltage devices. On the other hand, the effect of electric fields on the motion of a fluid can be used for designing electrohydrodynamic energy transducers that directly transform the energy of the electric field into the kinetic energy of a moving fluid.

An electric field may strongly affect the convective motions of poorly conducting fluids due to the effect of specific electroconvective instability mechanisms [1]. The dielectrophoretic mechanism of instability is associated with the nonhomogeneity of the dielectric permittivity. The effect of nonhomogeneous polarization on the stability of a liquid dielectric in a constant electric field was investigated in the cases of a horizontal [2] and vertical [3] capacitor. The other two instability mechanisms are associated with free charge, accumulated in the fluid, that is generated either due to the charge injection or due to the gradient of conductivity inherent in a nonuniformly heated fluid. The interaction of the volume charge with the electric field results in the motion of the fluid. The convection of poorly conducting fluids in a constant electric field due to the electroconductive mechanism was experimentally investigated in [4, 5]. The injection of charge and the nonhomogeneous polarization of a medium were not considered in the theoretical analysis [5, 6]. This approach is justified because the threshold voltage starting from which the injection mechanism of charge generation in a fluid substantially influences the stability of equilibrium is greater than the potential difference, used in experiments, across the boundaries of a layer [7]. Moreover, the conductivity of the fluids used in experiments depends much more strongly on temperature as compared with the dielectric permittivity. Investigations revealed oscillatory modes of electroconvection of a dielectric in a horizontal capacitor. These modes were studied, and the effect of the electric-charge relaxation time on the dynamics of a convective system was analyzed.

On the other hand, the presence of a modulated parameter in a mechanical system strongly affects its stability [8, 9] and can be used for controlling convective motions in various technological situations. Vibrations, as well as variable electric fields or temperature gradients, are important examples of periodic action on mechanical systems and, in particular, on fluids. A classical example of changing the stability of equilibrium in a vibration field is given by the Kapitsa pendulum [10].

The effect of temperature modulation at the boundaries of a fluid layer on its convective stability was analyzed in [11, 12]. Two types of critical perturbations were revealed: the perturbations of the first type vary synchronously with the external action, while the perturbations of the second type oscillate at a half frequency. The parametric excitation of thermoelectric instability under a thermal wave in liquid semiconductors or ion melts was investigated in [13]. The electroconvective instability of a nonhomogeneously heated dielectric fluid under a variable electric field was investigated in the cases when the charge generation in the fluid is associated either with dielectrophoresis [14] or with electric conductivity [15].

The present work is devoted to the study of instability of a poorly conducting fluid in a constant electric field under a periodic temperature gradient. In addition to the synchronous and subharmonic responses to an external action, the instability may be attributed to quasiperiodic perturbations. We show that resonance instability regions appear at frequencies related to the frequency of neutral oscillatory perturbations of a poorly conducting fluid in constant fields.

2. STATEMENT OF THE PROBLEM

Consider a poorly conducting fluid that fills a parallel-plate capacitor. Assume that the plates with the coordinates $z = \pm h$ (where h is the half-width of the layer) are perfect heat and electric conductors. Suppose that these plates are held at different constant potentials

$$\varphi(\pm h) = \mp U$$

and different temperatures varying by the law

$$T(\pm h) = \mp\Theta(\eta_1 + \eta_2\cos\Omega t).$$

Here, Θ is a characteristic scale of temperature, and η_1 and η_2 are relative amplitudes of the constant and variable components of the temperature difference between the boundaries. In our case, η_2 varies continuously, while η_1 assumes two values: $\eta_1 = 0$ for a variable temperature difference and $\eta_1 = 1$ for a modulation against a constant background.

The electric force exerted on a unit volume of a fluid can be expressed as [16]

$$f_e = \rho_e\mathbf{E} - \frac{1}{2}E^2\nabla\varepsilon + \frac{1}{2}\nabla\left(\rho\frac{\partial\varepsilon}{\partial\rho}E^2\right). \tag{1}$$

Here, ε is the dielectric permittivity of the fluid, ρ_e is the free charge of the unit volume, and $\mathbf{E}$ is the electric field strength.

The last term in (1) has a gradient form and only leads to the redefinition of pressure. The second (dielectrophoretic) part of the electric field, which is associated with the nonhomogeneity of ε, is unessential. This approach is justified in the case when the nonhomogeneity of conductivity due to the temperature gradient is much greater than the nonhomogeneity of dielectric permittivity. Here, we apply an electrohydrodynamic approximation, where electric effects are large as compared with magnetic ones. For layers of thickness less than 10^{-1} m, this approximation is justified if the modulation frequency of the electric field is less than 10^9 rad/s and the conductivity of the fluid satisfies the inequality $\sigma \ll 10^{-1}$ $(\Omega\ \mathrm{m})^{-1}$ [17]. For fluids with very low conductivity, $\sigma \sim 10^{-11}$–10^{-9} $(\Omega\ \mathrm{m})^{-1}$ (transformer oil, capacitor oil, or corn oil), the role of induced magnetic fields proves to be really negligible in a wide range of frequencies. We assume that the conductivity of a fluid increases linearly with temperature,

$$\sigma = \sigma_0(1 + \beta_\sigma T),$$

where β_σ is the conductivity temperature coefficient, which is positive in general.

Below, we will use dimensionless variables based on the scales of length (h), time (h^2/ν), velocity (χ/h), temperature (Θ), pressure ($\rho_l\nu\chi/h^2$), potential (U), field (U/h), and charge density ($\varepsilon U/h^2$) (here, ρ_l is the fluid density and ν and χ are the kinematic viscosity and thermal diffusivity of the fluid, respectively).

Neglecting the electric charge injection, electric loss, and viscous dissipation [17], we obtain the following dimensionless system of equations and boundary conditions that describe the convection of a liquid dielectric in constant gravity and electric fields and a variable thermal field:

$$\begin{gathered}
\left(\frac{\partial\mathbf{v}}{\partial t} + \frac{1}{\mathrm{P}}(\mathbf{v}\nabla)\mathbf{v}\right)\\
= -\nabla p + \nabla^2\mathbf{v} + \mathrm{Ra}T\mathbf{e} + \mathrm{Ge}\rho\mathbf{E},\\
\mathrm{div}\mathbf{v} = 0,\\
\mathrm{P}\frac{\partial T}{\partial t} + (\mathbf{v}\nabla)T = \nabla^2 T,\\
\frac{\partial\rho}{\partial t} + \frac{1}{\mathrm{Pe}}\mathrm{div}(\sigma\mathbf{E}) + \frac{1}{\mathrm{P}}(\mathbf{v}\nabla)\rho = 0,\\
\mathrm{div}\mathbf{E} = \rho, \quad \mathbf{E} = -\nabla\varphi, \quad \mathbf{e} = (0, 0, 1),\\
\sigma = 1 + ST,\\
z = \pm 1\colon \mathbf{v} = 0,\\
T = \mp(\eta_1 + \eta_2\cos\omega t), \quad \varphi = \mp 1.
\end{gathered} \tag{2}$$

Here, $\mathbf{v}$ is velocity, p is pressure, T is temperature measured from a certain mean value, $\mathbf{E}$ and φ are the strength and potential of the electric field, and ρ is the free charge density.

System (2) involves the following dimensionless parameters:

$$\begin{gathered}
\mathrm{Ra} = g\beta\Theta h^3/\nu\chi, \quad \mathrm{Ge} = \varepsilon U^2/\nu\chi\rho_l, \quad P = \nu/\chi,\\
\mathrm{Pe} = \varepsilon\nu/h^2\sigma_0, \quad \omega = \Omega h^2/\nu, \quad S = \beta_\sigma\Theta.
\end{gathered} \tag{3}$$

Here, Ra is the Rayleigh number (β is the thermal-expansion coefficient of the fluid), Ge is an electric analogue of the Galilean number, P is the Prandtl number, Pe is the electric Prandtl number, and ω is the dimensionless modulation frequency; the parameter S characterizes the nonhomogeneity of conductivity. In poorly conducting fluids that exhibit electroconvective phenomena in experiments, $\beta_\sigma \approx 0.03$–$0.058$ deg^{-1} [5, 7] and the condition $S < 1$ holds if the temperature difference does not exceed 10 degrees. The results of theoretical investigations of electrothermal convection in constant electric fields, obtained under the stronger condition $S \ll 1$, show a good agreement with experimental results [5, 7].

The problem formulated admits a quasi-equilibrium solution where the fluid is at rest,

$$\mathbf{v}_0 = 0,$$

while its other characteristics vary in space and time:

$$T_0 = T_0(z, t), \quad p_0 = p_0(z, t),$$
$$\mathbf{E}_0 = (0, 0, E_0(z, t)), \quad \rho_0 = \rho_0(z, t).$$

We will not need an explicit expression for pressure p_0. The time-dependent temperature distribution $T_0(z, t)$ satisfies the one-dimensional heat equation and the corresponding boundary conditions:

$$P\frac{\partial T_0}{\partial t} = \frac{\partial^2 T_0}{\partial z^2}, \qquad (4)$$
$$z = \pm 1: T_0 = \mp(\eta_1 + \eta_2 \cos\omega t).$$

The temperature distribution in the quasi-equilibrium is determined by the superposition of a linear profile and two thermal waves that propagate from the boundaries toward the bulk of the fluid:

$$T_0 = -\eta_1 z - \eta_2[T_s \sin\omega t + T_c \cos\omega t], \quad \kappa = \sqrt{\frac{\mathrm{P}\omega}{2}},$$
$$T_s = \frac{\sin(\kappa(1+z))\sinh(\kappa(1-z)) - \sin(\kappa(1-z))\sinh(\kappa(1+z))}{\cos(2\kappa) - \cosh(2\kappa)}, \qquad (5)$$
$$T_c = \frac{\cos(\kappa(1+z))\cosh(\kappa(1-z)) - \cos(\kappa(1-z))\cosh(\kappa(1+z))}{\cos(2\kappa) - \cosh(2\kappa)}.$$

Using the temperature distribution (5) and the boundary condition for the potential φ, we can determine the field and free charge distributions in a fluid at rest:

$$E_0 = 1 + \eta_1 S z + \frac{\eta_2 S}{1 + \omega^2 \mathrm{Pe}^2}((T_s + \omega \mathrm{Pe} T_c)\sin\omega t$$
$$+ ((T_c - \omega \mathrm{Pe} T_s)\cos\omega t) + O(S^2)),$$
$$\rho_0 = \eta_1 S + \frac{\eta_2 S}{1 + \omega^2 \mathrm{Pe}^2}((T_{sz} + \omega \mathrm{Pe} T_{cz})\sin\omega t \qquad (6)$$
$$+ ((T_{cz} - \omega \mathrm{Pe} T_{sz})\cos\omega t) + O(S^2)),$$
$$T_{sz} = \frac{\partial T_s}{\partial z}, \quad T_{cz} = \frac{\partial T_c}{\partial z}.$$

To analyze the stability of the base state, let us consider its small perturbations. Upon eliminating the pressure and the horizontal components of velocity, we represent the perturbations of the vertical component of velocity v_z, temperature ϑ, and charge density ρ as follows:

$$v_z = w(z, t)\exp[ikx], \quad \vartheta = \theta(z, t)\exp[ikx], \qquad (7)$$
$$\rho = \rho(z, t)\exp[ikx].$$

Here, w, θ, and ρ are the amplitudes, and k is the wavenumber. The vectors $\mathbf{E}_0$ and ∇T_0 are directed normal to the boundaries; therefore, the problem is isotropic in the plane of the layer, and one can direct the x axis along the wavevector **k**.

The quasi-equilibrium field (6) represents a series

$$E_0 = E_{00} + S E_{01} + O(S^2).$$

The first term E_{00} is the external field; it is independent of S and characterizes the field between the conductive plates in the absence of a fluid. All the other terms of the series $SE_{01} + O(S^2)$ are associated with the redistribution of charge in the fluid,

$$\rho_0 = S\rho_{01} + O(S^2),$$

which is generated due to the nonhomogeneous conductivity of the medium. The associated perturbations of the Coulomb force can be represented as

$$\mathbf{F}'_k = \mathrm{Ge}(\mathbf{E}_{00}\rho' + S(\mathbf{E}_{01}\rho' + \mathbf{E}'\rho_{01})),$$

where $\mathbf{E}'$ and ρ' are the perturbations of the field and charge density. The smallness of the parameter S allows one to apply an induction-free electrohydrodynamic approximation in which the electric field associated with the redistribution of charge in the fluid is neglected as compared with the external field [5]:

$$\mathbf{F}'_k = \mathrm{Ge}\mathbf{E}_{00}\rho.$$

Substituting the perturbed fields into system (2) and linearizing with respect to the perturbations, we obtain the following amplitude problem:

$$\frac{\partial \Delta w}{\partial t} = \Delta^2 w - \mathrm{Ra}k^2\theta - Bk^2\rho,$$
$$\mathrm{P}\frac{\partial\theta}{\partial t} = (\eta_1 + \eta_2(T_{cz}\cos\omega t + T_{sz}\sin\omega t))w + \Delta\theta, \qquad (8)$$
$$\mathrm{Pe}\frac{\partial\rho}{\partial t} = -\frac{\partial\theta}{\partial z} - \rho,$$
$$z = \pm 1: w = 0, \quad w' = 0, \quad \theta = 0,$$

where

$$\Delta = \partial^2/\partial z^2 - k^2.$$

Note that, despite the smallness of the temperature-induced nonhomogeneity S of conductivity, the electric analogue of the Rayleigh number $B = \mathrm{Ge}S$ remains finite (the parameter Ge is large). If $\eta_2 \neq 0$, the system has periodic coefficients; then, a dynamic excitation or the suppression of electroconvection may occur.

The system of equations and boundary condition (8), together with the time-periodicity conditions for all variables, specify an eigenvalue problem for B or Ra as functions of the remaining parameters. The boundaries of the convective instability specified by the conditions for the existence of periodic solutions to system (8) can be determined by the classical Floquet method.

3. METHOD OF SOLUTION

Problem (8) for the modulation of arbitrary frequency was solved by the Galerkin method. To approximate the perturbations, we use the sets of spatial basis functions with time-dependent coefficients:

$$w = \sum_{m=0}^{M-1} a_m w_m, \quad \theta = \sum_{m=0}^{M-1} b_m \theta_m, \qquad (9)$$
$$\rho = \sum_{m=0}^{M-1} c_m \frac{\partial \theta_m}{\partial z}.$$

The space bases were constructed from normalized eigenfunctions of the amplitude problem for a fluid layer at rest [18]. The fourth- and second-order eigenfunctions were used as the bases for approximating the vertical component of velocity, the temperature, and the charge density:

$$\Delta^2 w_m = -\mu_m \Delta w_m, \quad w_m(\pm 1) = w'_m(\pm 1) = 0, \quad (10)$$

$$\Delta \theta_m = -P\nu_m \theta_m, \quad \theta_m(\pm 1) = 0. \quad (11)$$

Substituting expansions (9) into (8) and orthogonalizing by the Galerkin method, we obtain $K = 3M$ ordinary differential equations in a_r, b_s, and c_t:

$$\frac{\partial a_r}{\partial t} = -\mu_r a_r + k^2 \sum_{m=0}^{M-1} [\mathrm{Ra} E_{mr} b_m + B D_{rm} c_m],$$

$$\mathrm{P}\frac{\partial b_s}{\partial t} = \sum_{m=0}^{M-1} [\eta_1 E_{sm} + \eta_2 (S_{sm} \sin \omega t + C_{sm} \cos \omega t)]$$
$$\times a_m - \mathrm{P}\nu_s b_s,$$

$$\mathrm{Pe}\frac{\partial c_t}{\partial t} = -b_t - c_t, \quad r, s, t = 0, 1, \ldots, M-1, \qquad (12)$$

$$E_{mr} = \int_{-1}^{1} \theta_m w_r dz, \quad D_{rm} = \int_{-1}^{1} w_r \theta'_m dz,$$

$$S_{sm} = \int_{-1}^{1} \theta_s T_{sz}(z) w_m dz,$$

$$C_{sm} = \int_{-1}^{1} \theta_s T_{cz}(z) w_m dz.$$

Equations (12) can be rewritten as

$$\frac{\partial u_i}{\partial t} = L_{ij}(\omega t) u_j, \quad i, j = 1, \ldots, 3M,$$
$$u(t) = \begin{pmatrix} a_r \\ b_s \\ c_t \end{pmatrix}, \qquad (13)$$

where the matrix L composed of the coefficients of a_r, b_s, and c_t from (12) is periodic with period $2\pi/\omega$ and $u(t)$ is a K-dimensional vector function. The classical Floquet theory [19] allows one to express any solution to system (13) as

$$u(t) = \gamma u_0(t) = e^{\lambda t} u_0(t) \qquad (14)$$

with the time-periodic (with period $T = 2\pi/\omega$) vector $u_0(t)$ (where γ is the Floquet multiplier and $\lambda = \lambda_r + i\lambda_i$ are characteristic exponents, which are generally complex numbers). For various initial conditions

$$u_i^p(0) = \delta_{ip}, \quad p = 1, \ldots, K,$$

we obtain K linearly independent solutions $u_i^p(t)$. System (13) is integrated by the fourth-order Runge–Kutta method. The fundamental solutions taken at the end of the modulation period constitute K columns of a $K \times K$ monodromy matrix whose eigenvalues represent the Floquet multipliers. The values of the characteristic exponents determine the stability of the base state of quasi-equilibrium. If we arrange the exponents so that

$$\mathrm{Re}(\lambda_1) \geq \mathrm{Re}(\lambda_2) \geq \ldots \geq \mathrm{Re}(\lambda_K),$$

then the base state is stable when $\mathrm{Re}(\lambda_1) < 0$. The condition $\mathrm{Re}(\lambda_1) = 0$ determines the domain of periodic solutions in the space of parameters B, η_1, η_2, Ra, P, Pe, ω, and k. Here, the case

$$\mathrm{Re}(\lambda_1) = 0, \quad \mathrm{Im}(\lambda_1) = \omega/2$$

corresponds to subharmonic perturbations with the period twice that of the external action. If

$$\mathrm{Re}(\lambda_1) = 0, \quad \mathrm{Im}(\lambda_1) = \omega,$$

neutral perturbations vary synchronously with the external force and their periods coincide. For a pair of complex conjugate eigenvalues with unit modulus,

$$\mathrm{Re}(\lambda_1) = 0, \quad \mathrm{Im}(\lambda_1) \neq 0,$$

we have quasiperiodic neutral perturbations. For the majority of solutions obtained, 24 basis functions were used ($M = 8$). In the test computations carried out with 30 basis functions ($M = 10$), the variations in the convection threshold were less than 1%.

4. LOW-FREQUENCY MODULATIONS

For small modulation frequencies ω, the modulation period is greater than all characteristic times of the system:

$$T \gg \max\left[\frac{h^2}{\nu}, \frac{h^2}{\chi}, \frac{\varepsilon}{\sigma}\right], \tag{15}$$

or, in a dimensionless form,

$$\omega \ll \min[1, P^{-1}, \mathrm{Pe}^{-1}]. \tag{16}$$

Since the computation time increases proportionally to the modulation period *T*, the application of a numerical method is inexpedient when $\omega \longrightarrow 0$; here, one can apply the Wentzel–Kramers–Brillouin asymptotic method [20] with a small parameter ω.

After rescaling time, $\tau = \omega t$, we use the matrix form for the system of amplitude equations:

$$\omega\frac{\partial}{\partial\tau}Au = Nu \tag{17}$$

with the vector function

$$u = \begin{pmatrix} w \\ \theta \\ \rho \end{pmatrix}$$

and the constant (*A*) and 2π-periodic in τ(*N*) operators

$$A = \begin{pmatrix} \Delta & 0 & 0 \\ 0 & \mathrm{P} & 0 \\ 0 & 0 & \mathrm{Pe} \end{pmatrix},$$

$$N = \begin{pmatrix} \Delta^2 & -\mathrm{Ra}k^2 & -Bk^2 \\ f(t) & \Delta & 0 \\ 0 & -\dfrac{\partial}{\partial z} & -1 \end{pmatrix}, \tag{18}$$

where $f(\tau) = \eta_1 + \eta_2(T_{sz}\sin\tau + T_{cz}\cos\tau)$.

Using the smallness of ω, we expand *N*, *B*, and the solution *u* in series in ω:

$$N = N_0 + \omega N_1 + \omega^2 N_2 + \ldots,$$

$$B = B_0 + \omega B_1 + \omega^2 B_2 + \ldots,$$

$$u = \exp\left[\frac{1}{\omega}\int_0^\tau \lambda(\tau')d\tau'\right](u_0 + \omega u_1 + \omega^2 u_2 + \ldots), \tag{19}$$

$$\lambda = \lambda_0 + \omega\lambda_1 + \omega^2\lambda_2 + \ldots,$$

where λ is a characteristic increment of perturbations for the case of a slow quasistatic variation in the temperature gradient (a 2π-periodic function of τ). The Floquet stability boundary in the zeroth order in ω is determined from the integral condition

$$\int_0^{2\pi} \lambda_{0r}(B_0, \cos\tau)d\tau = 0, \tag{20}$$

where λ_{0r} is the real part of λ_0. To solve the eigenvalue problem in the zeroth order, we used the Galerkin method with basis (9).

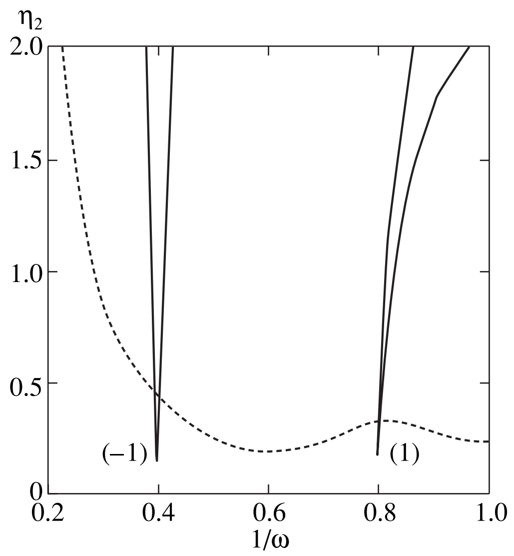


Fig. 1. Modulation amplitude η_2 as a function of inverse frequency 1/ω for Ra = 0. The wavenumber is fixed: k = 2.354. In the absence of modulation, the system is stable. The solid curves represent the stability boundaries for periodic perturbations, while the dashed curves represent the stability boundaries for quasiperiodic modes.

5. RESULTS

The poorly conducting fluids used in experiments have sufficiently high Prandtl numbers, P ≥ 100. The electric Prandtl number Pe depends not only on the electrophysical parameters of a fluid but also on the thickness of the layer. As an illustration, consider the behavior of the fluid with the parameters P = 100 and Pe = 0.04. First, let us discuss the onset of convection in the absence of gravity (Ra = 0). In constant electric and thermal fields, the instability is attributed to the electroconductive mechanism and is associated with the oscillatory mode. A convection threshold of $B_* = 2514.24$ corresponds to a critical wavenumber of $k = 2.354$ and a frequency of $\omega_0 = 1.256$. The effect of the modulation of the temperature gradient on the stability of the quasi-equilibrium of $\eta_2 = f(1/\omega)$ for fixed values of $k = 2.354$ and $B = 2510$ is shown in Fig. 1.

In the absence of modulation, $\eta_2 = 0$ (the abscissa), the quasi-equilibrium of the fluid is stable. (The electric Rayleigh number does not exceed its critical value, $B < B_*$.)

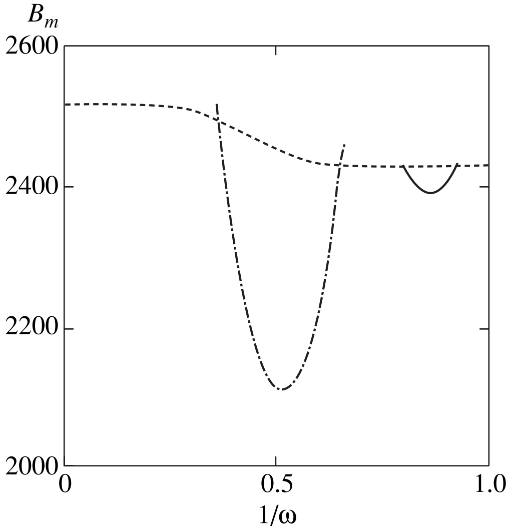


Fig. 2. The critical electric Rayleigh number B_m versus inverse frequency 1/ω for Ra = 0, $\eta_1 = 1$ and $\eta_2 = 1$. The dot-and-dash curve represents the stability boundary for subharmonic perturbations, the solid curve represents the stability boundary for synchronous perturbations, and the dashed curve is the same as in Fig. 1.

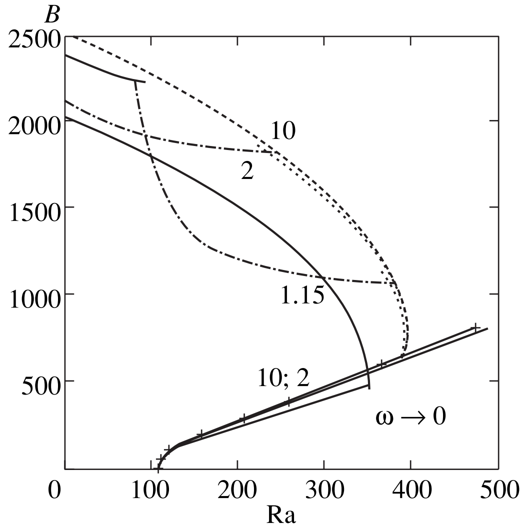


Fig. 3. Stability maps on the plane (Ra, *B*) in a modulated field for various frequencies of external action: ω = 10 (the stability boundary for the thermogravitational mode is marked by crosses), ω = 2, ω = 1.15, and ω ⟶ 0. The dot-and-dash, dashed, and solid curves are the same as in Fig. 2.

The growth of the modulation amplitude gives rise to increasing perturbations. The type of critical perturbations depends on frequency. The first resonance domain (–1) corresponds to the perturbations that are subharmonic with respect to the external action; its minimum is at $\omega = 2.5 \approx 2\omega_0$ and corresponds to a modulation amplitude of $\eta_2 = 0.151$. The domain (+1) of synchronous perturbations has a minimum of $\eta_2 = 0.170$ at a frequency of $\omega = 1.254 \approx \omega_0$. The effect of destabilization in this domain is weaker. For moderate modulation amplitudes, there exist domains of critical quasiperiodic perturbations between the domains of synchronous and subharmonic responses to the external field.

The critical electric Rayleigh number B_m versus inverse frequency 1/ω is presented in Fig. 2 (Ra = 0, $\eta_1 = 1$, and $\eta_2 = 1$; the minimization is carried out over the wavenumber *k*). The parametric excitation of convection manifests itself when the relation between the driving frequency and the frequency of the oscillatory mode in a constant field is expressed as $2\omega_0/m$ with integer *m*. Figure 2 represents the first two domains of instability.

Consider the electrothermoconvective instability in a static gravity field. Now, the behavior of perturbations should be considered from the viewpoint of amplitude equations (8) with Ra ≠ 0. In the absence of an electric field (*B* = 0), the critical thermal Rayleigh number determines the onset of convection under the heating from below, $\mathrm{Ra}_{m0} = 106.75$, and corresponds to the thermogravitational monotonic mode of instability [8]. The rise of an electric field and its increase stabilize the equilibrium, and the convection starts at larger values of the Rayleigh number. Moreover, due to the specific electroconductive mechanism of instability, the electric field gives rise to an oscillatory instability [5]. The modulation of the field changes the stability map. The critical electric Rayleigh number *B* versus Ra is shown in Fig. 3 for several values of frequency: ω = 10, 2, 1.15, and ω ⟶ 0 in the case of $\eta_1 = 1$ and $\eta_2 = 1$. The domain of stability lies between the curves and the coordinate axes. For the thermogravitational mode, critical perturbations are synchronous with the modulation of the temperature gradient. A decrease in the frequency of the external field results in a monotonic decrease in the boundary of the gravitational mode. The threshold of the electroconductive mode varies nonmonotonically as frequency decreases. Quasiperiodic perturbations remain critical for sufficiently large modulation frequencies (ω = 10).

At a frequency close to the doubled fundamental frequency of the oscillatory system, $\omega = 2 \approx 2\omega_0$, a competition between two types of oscillations at the stability boundary leads to a break on the stability boundary. The point of break corresponds to a change in the character of neutral oscillations. When 0 < Ra < 80, the fluid responds to the variation in the temperature gradient with synchronous oscillations, whereas, when Ra > 80, it responds with subharmonic oscillations. At a frequency of ω = 1.15, the stability boundary has points of break that correspond to the transitions between various types of perturbations, which differ both in the spatial period and in the time behavior. At a frequency of ω = 1.15, the parametric effect reduces the stability threshold by a factor of almost two when Ra ~ 150. The boundary ω ⟶ 0 is calculated in the low-frequency limit.

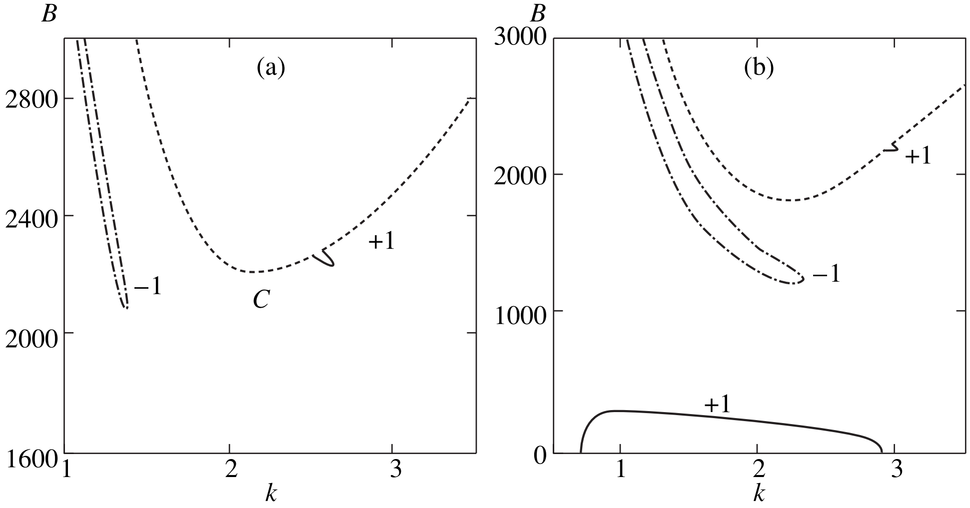


Fig. 4. Neutral curves of electrothermal convective instability $B(k)$; $\eta_1 = 1$, $\eta_2 = 1$, $\omega = 1.15$. (a) Subcritical heating, Ra = 85, and (b) supercritical heating, Ra = 200. The dot-and-dash, dashed, and solid curves are the same as in Fig. 2.

Figure 4a and 4b present the neutral curves $B(k)$ for a frequency of $\omega = 1.15$ for two values of the Rayleigh number. The boundary of instability consists of several parts that correspond to different types of neutral perturbations. A synchronous response is observed in domains +1; a subharmonic one, in domains –1; and a quasiperiodic response, in domain *C*. Under subcritical heating, $\mathrm{Ra} = 85 < \mathrm{Ra}_{m0}$ (Fig. 4a), the instability boundary consists of two parts. The first part is a neutral curve for subharmonic perturbations; the minimum of this curve determines the critical Rayleigh number. The second part of the neutral curve is generally quasiperiodic. However, a domain of synchronous perturbations appears near $k = 2.52$ as a result of bifurcation. Under supercritical heating, $\mathrm{Ra} = 200 > \mathrm{Ra}_{m0} = 106.7$, the topology of the neutral curves is changed (Fig. 4b). Now, $B(k)$ becomes a multivalued function and has a minimum and a maximum. The domain of synchronous perturbations moves down and expands. A new domain of synchronous perturbations that are associated with the thermogravitational mode appears near the abscissa.

6. CONCLUSION

The problem of electroconvective instability of a poorly conducting fluid in a constant electric field of a parallel-plate horizontal capacitor is considered within the framework of electrohydrodynamic equations under nonstationary heating of the fluid. It is assumed that the conductivity of the fluid depends linearly on temperature, while the physical properties of the fluid and electrodes are such that one can neglect the injection of charge and the nonhomogeneity of the polarization of the fluid. The boundaries of the instability domains are determined. For the weightlessness case, the amplitude of the critical modulation is obtained as a function of frequency. In the presence of gravity, there is an interaction between the thermogravitational and electroconductive mechanisms of instability. The analysis is carried out for a Prandtl number of P = 100. For greater Prandtl numbers, which are characteristic of insulating oils, the fundamental frequency ω_0 in the unmodulated case decreases; however, the main features of the parametric excitation of electroconductive instability are preserved: (1) there is a competition between subharmonic, synchronous, and quasiperiodic perturbations, and (2) the first domain of the subharmonic response corresponds to a frequency of $\omega \approx 2\omega_0$, while the first domain of the synchronous response corresponds to a frequency of $\omega \approx \omega_0$.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project nos. 98-01-00507 and 01-01-00515.

REFERENCES

1. A. Castellanos, P. Atten, and M. G. Velarde, Phys. Fluids **27**, 1607 (1984).
2. P. H. Roberts, J. Mech. Appl. Math. **22**, 211 (1969).
3. M. Takashima and H. Hamabata, J. Phys. Soc. Jpn. **53**, 1728 (1984).
4. R. J. Turnbull, Phys. Fluids **11**, 2597 (1968).
5. Ch. O. Lee, in *Proceedings of the 5th International Heat Transfer Conference, Tokyo, 1974*, Vol. 3, p. 173.
6. R. J. Turnbull, Phys. Fluids **11**, 2588 (1968).
7. S. A. Zhdanov, S. R. Kosvintsev, and I. Yu. Makarikhin, Zh. Éksp. Teor. Fiz. **117**, 398 (2000) [JETP **90**, 352 (2000)].

8. G. Z. Gershuni and E. M. Zhukhovitsky, *Convective Stability of Incompressible Liquid* (Nauka, Moscow, 1972).
9. G. Ahlers, P. C. Hohenberg, and M. Lücke, Phys. Rev. A **32**, 3493 (1985).
10. P. L. Kapitsa, Zh. Éksp. Teor. Fiz. **21**, 588 (1951).
11. G. Venezian, J. Fluid Mech. **35**, 243 (1969).
12. C. S. Yin and C. H. Li, J. Fluid Mech. **54**, 143 (1972).
13. B. L. Smorodin, G. Z. Gershuni, and M. G. Velarde, Int. J. Heat Mass Transf. **42**, 3159 (1999).
14. B. L. Smorodin and M. G. Velarde, J. Electrostat. **50** (3), 13 (2001).
15. B. L. Smorodin and M. G. Velarde, J. Electrostat. **48**, 261 (2000).
16. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media* (Nauka, Moscow, 1982; Pergamon, New York, 1984).
17. M. K. Bologa, F. P. Grosu, and I. A. Kozhukhar', *Electric Convection and Heat Exchange* (Shtiintsa, Kishinev, 1977).
18. G. I. Petrov, Prikl. Mat. Mekh. **4**, 3 (1940).
19. E. A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations* (McGraw-Hill, New York, 1955; Inostrannaya Literatura, Moscow, 1958).
20. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory* (Nauka, Moscow, 1989, 4th ed.; Pergamon, New York, 1977, 3rd ed.).

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 93, No. 6, 2001, pp. 1239–1244.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 6, 2001, pp. 1430–1437.
Original Russian Text Copyright © 2001 by Dilanyan, Khasanov, Bredikhin, Gurov, Kveder, Osip'yan, Shalynin.

SOLIDS
Structure

Phase Transitions and the Structure of the C_{60} Crystal Doped with Lithium by Electrodiffusion

R. A. Dilanyan*, S. S. Khasanov, S. I. Bredikhin, A. F. Gurov,
V. V. Kveder, Yu. A. Osip'yan, and A. I. Shalynin**
Institute of Solid-State Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia
**e-mail: ruben@issp.ac.ru*
***e-mail: kveder@issp.ac.ru*

Received March 28, 2001

Abstract—The structure and phase transitions of C_{60} crystals doped with lithium by injecting metal ions from the superionic crystal–C_{60} single crystal heterojunction under electrodiffusion conditions are reported. The sample experienced irreversible transitions resulting in the virtually complete disappearance of EPR signals and MW conduction in the temperature range 320–370 K. In this temperature interval, a new C_{60} phase was formed; the phase contained polymeric chains of C_{60} molecules along the crystallographic *c* axis and lithium clusters. The structure of this phase was determined. Annealing at 620 K restored the EPR signal and, according to the X-ray data, the initial cubic structure of pure C_{60}. The X-ray pattern, however, contained additional diffraction peaks, which was evidence of the presence of one more phase with a structure yet unknown. *© 2001 MAIK “Nauka/Interperiodica”.*

1. INTRODUCTION

Much attention has been given to studying the physical properties of fullerene intercalates containing alkali metals (fullerides), first and foremost, because some of them exhibit superconducting properties. Under normal conditions, the initial fullerene C_{60} phase has a face-centered cubic structure with an $a = 14.16$ Å lattice parameter. Because of the presence of octahedral voids, which form three-dimensional networks of channels extended along [111] crystallographic axes, a large number of elements can be intercalated into the structure of the initial C_{60} crystal. Fullerene C_{60} molecules can be added to atoms at their vinyl double bonds and can form polymeric chains under the action of radiation and pressure. Compounds of the general formula A_xC_{60}, where A = K, Rb, and Cs, have been studied in much detail [1]. The conduction properties of these compounds change depending on x. For instance, A_3C_{60} intercalates possess metallic conduction and experience a transition to the superconducting state as temperature lowers [2, 3]. Compounds with $x = 4$, however, are dielectric materials [4]. It is also well known that, below 400 K, AC_{60} intercalates polymerize to produce an orthorhombic structure [5–7] in which C_{60} molecules form parallel charged chains.

Lithium fullerides have been studied less thoroughly. The special features of liquid- and solid-state lithium intercalation are considered in [8]. The results of studying the physicochemical properties of lithium fullerides obtained by the liquid-phase technique are described in [9]. Li_xC_{60} with $0 < x < 12$ formed interstitial solid solutions. A further increase in the concentration of lithium caused the formation of lithium clusters; C_{60} molecules formed polymeric chains in these samples. An X-ray study of freshly prepared samples showed them to have a monoclinic structure. However samples stored for some time were orthorhombic and had lattice parameters $a = 9.11$ Å, $b = 9.87$ Å, and $c = 14.76$ Å.

Clearly, understanding processes that occur in A_xC_{60} systems requires studying the behavior of C_{60} single crystals at low degrees of doping them with alkali metals. Such samples are very difficult to obtain by usual diffusion techniques because of their decomposition into stable phases with integral x values. The corresponding systems still remain virtually unstudied because of the complexity of preparing them.

In [10], we developed and suggested a new approach to controlled doping of C_{60} single crystals with metal ions through injecting the ions from the superionic crystal–C_{60} single crystal heterojunction by electrodiffusion and studied the electronic properties of the Li_xC_{60} single crystals prepared in this way. We showed that the properties of the samples did not change in time below 250 K, whereas the EPR and IR reflectance spectra changed substantially as time passed at $T > 280$ K; in particular, the EPR and MW conduction signals virtually disappeared. At room temperature, the properties of Li_xC_{60} crystals irreversibly changed.

The purpose of this work was to determine the structure of C_{60} crystals doped with lithium and study the

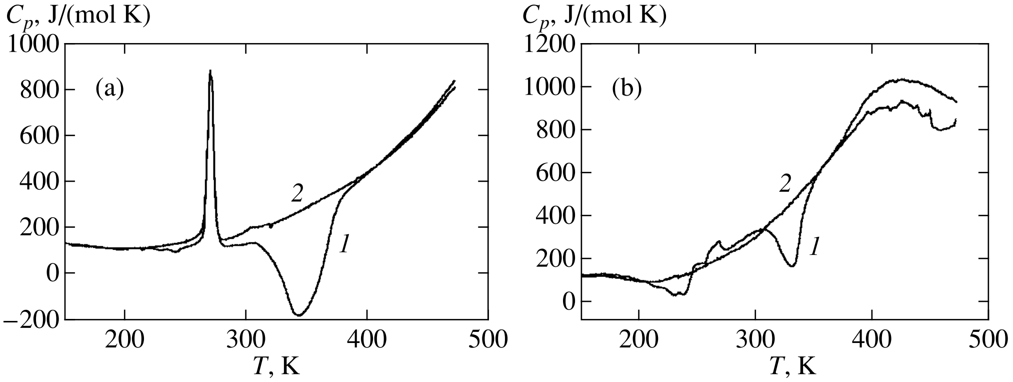


Fig. 1. Temperature dependences of heat capacity for (a) weakly doped ($x \ll 1$) and (b) strongly doped ($x \geq 1$) Li_xC_{60} samples recorded during the first (curves *1*) and second (curves *2*) heatings.

phase transitions that occurred in the Li_xC_{60} system by X-ray diffraction and calorimetrically.

2. EXPERIMENTAL

Fullerene C_{60} single crystals were grown at the Institute of Solid-State Physics, Russian Academy of Sciences, by physical vapor transport (sublimation) under controlled temperature gradient conditions in an evacuated ampule at temperatures of 870 to 910 K. The initial C_{60} sample was purified by chromatography. Traces of the solvent were removed by repeatedly resubliming the compound in a vacuum. The final product was no less than 99.98% pure. Measurements were performed using single crystals in the form of thin plane-parallel plates with natural faceting. Faceting planes corresponded to (111)-type planes. The plates were 0.7–2 mm thick; the other two dimensions were usually 2–4 mm.

Lithium ions were injected as follows. Electrodes were mechanically clamped to opposite plane-parallel (111) C_{60} single crystal faces. The cathode was a graphite electrode. Used as the anode were two-layer $Li_{0.2}WO_3$ | Li_7SiPO_8 electrodes that had a high ionic (lithium cations) conduction and blocked the electronic current component. As mentioned above, the use of cells of the type

$$(+)|Li_{0.2}WO_3|Li_7SiPO_8|C_{60}\ \text{crystal}|\text{Graphite}|(-)$$

allowed us to inject lithium cations into C_{60} single crystals and perform controlled doping of the samples. The cell for doping was mounted in a vacuum thermostat. The majority of experiments were performed at cell voltages of 10–200 V and temperatures of 400–520 K. Characteristic current flowing through the cell was of the order of 2–20 μA. The duration of experiments varied from 1 to 100 h. After electrodiffusion, the samples were rapidly transferred to a quartz ampule and stored at a liquid nitrogen temperature up to the time of measurements. Calorimetric studies were performed on a DSC-7 (Perkin Elmer) microcalorimeter in the temperature range 77–500 K at a 20 K/min heating rate. Freshly prepared samples were studied on an ENRAF NONIUS CAD-4 diffractometer. After storage at room temperature, the samples were ground into powder. The X-ray diffraction data were obtained on a SIEMENS D500 diffractometer. The structure of the samples was determined from their X-ray powder patterns by the Rietveld method using the RIETAN-98 program [11].

3. RESULTS AND DISCUSSION

As mentioned above, holding Li_xC_{60} at room temperature caused radical changes in the IR reflectance spectra of the samples and drastically decreased their EPR signal and MW conduction. In [10], this phenomenon was explained by the formation of new phases or compounds in Li_xC_{60} doped samples. To verify this assumption and elucidate the nature of the phase transitions, we performed a calorimetric study of C_{60} single crystals doped with lithium.

Studies of weakly doped Li_xC_{60} samples ($x \ll 1$) showed that the temperature dependence of C_p recorded when the sample was heated for the first time contained heat release peaks at 220–250 and 320–370 K (Fig. 1a, curve *1*). Repeat (curve *2*) and subsequent heatings gave a single peak at 220–250 K. This was evidence that irreversible changes occurred in the sample at 320–370 K. Heat release at 220–250 K corresponded to the well-known orientation phase transition in pure C_{60}. The presence of the corresponding peak was evidence that some part of the sample remained in the initial state after doping. Virtually no such peak was observed in the temperature dependences of C_p recorded for heavily doped Li_xC_{60} samples (Fig. 1b). This led us to suggest that the formation of new phases occurred in the major fraction of the sample. Calculations of the heat released during heating the sample for the first time (curve *1*) gave ΔQ = 15.1 kJ/mol, or 0.16 eV per molecule. Irreversible transitions, however, only occurred in some part of the sample, and the calculated heat effect value (0.16 eV per molecule) was obviously underestimated.

The experimental $\Delta Q > 0.16$ eV value per molecule was evidence that a new phase or compound was formed in doped Li_xC_{60} samples at 320–370 K. We deemed it interesting to determine the crystal structure of this phase.

A preliminary analysis of the diffraction data showed that the phase formed at 320–370 K (for convenience, this phase will be called phase II) was orthorhombic and had the parameters

$$a = 14.74\ \text{Å}, \quad b = 9.84\ \text{Å}, \quad c = 9.09\ \text{Å}.$$

The crystallographic axes of this orthorhombic cell (a, b, c) were related to the crystallographic axes of the cubic cell of pure C_{60} as $a \approx a_C$, $b \approx a_C/\sqrt{2}$, and $c \approx a_C/\sqrt{2}$, where a_C is the period of the undoped phase with distortions (1.04, 0.98, 0.91). Systematic reflection extinctions could be described by two space groups, *Immm* and *Pnnm*. The low (virtually, at the background level) intensity of the reflections permitted for *Pnnm* and forbidden for *Immm* prevented us from unambiguously determining the space group. We selected the *Pnnm* group because the C_{60} molecule could only then assume various orientations, whereas, if the space group were *Immm*, its orientation would be fixed. Note that the obtained orthorhombic cell parameters coincided with those reported in [9]. Note also that our data coincided with the results reported in [12], where the structure of the high-pressure phase of pure C_{60} was studied. The characteristic c parameter value was evidence that polymeric chains of C_{60} molecules might be formed in the c direction [12–14].

The initial model for analyzing the structure of phase II was based on the parameters reported in [12]. The positions of carbon atoms were somewhat changed to place them on an ideal spherical surface. The B_{iso} isotropic temperature factors were set equal to 1.0 for all carbon atoms. The parameters to be refined were the lattice constants, the background, and the diffraction line profiles. This gave the following results:

$$R_{wp} = 3.67\%, \quad R_B = 1.36\%,$$

$$\chi^2 = (R_{wp}/R_E)^2 = 7.3.$$

At this stage of analyzing the structure of phase II, an important problem was to determine the orientation of C_{60} molecules in the orthorhombic lattice. As C_{60} molecules formed chains in the c direction, the orientation of a molecule could be varied by rotating it as a whole about axis c through some angle. We refined the structure in calculations for a discrete set of C_{60} rotation angles in the angle range $0 \le \alpha \le 180°$; the angle step size was 5°. The initial orientation ($\alpha = 0$) corresponded to the arrangement of polymeric bonds in a plane parallel to ab. The dependence of χ^2 on α is shown in Fig. 2. According to this figure, there was a global minimum in the α range 20°–30°. This minimum corresponded to the most probable orientation of C_{60} molecules.

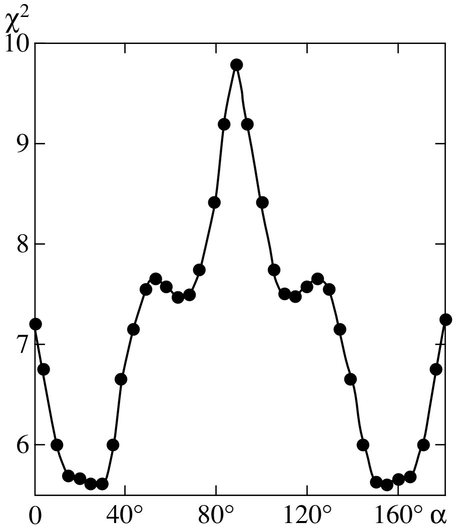


Fig. 2. Dependence of the χ^2 parameter on the angle of rotation of C_{60} molecules about the c axis.

Another criterion for analyzing the orientation of C_{60} molecules was the results reported in [15], where the molecular form factor of C_{60} was studied. It was, in particular, shown that diffraction lines with a large scattering vector ($SD > 2$, where S is the scattering vector and D is the diameter of the molecule) were most sensitive to fullerene molecule orientations. For this reason, checking refinement at large diffraction angles could also be used to determine the most probable orientation of C_{60} molecules.

In the further analysis, we used the $\alpha = 30°$ orientation. The B_{iso} isotropic temperature factors (equal for all carbon atoms) and the atomic positions were refined. The x and y coordinates were fully determined by the rotation angle of C_{60} molecules about the c axis, and therefore only the z coordinates were varied. The final diffraction picture is shown in Fig. 3. The final refined parameters and the coordinates of atoms are listed in Tables 1 and 2, respectively. The mean lengths of C–C single and C=C double bonds were found to be 1.431 and 1.392 Å, respectively. The shortest intermolecular distances were 9.09 Å along the [001] direction, 9.84 Å along [010], and 9.96 Å along [111]. The value of 9.09 Å for two neighboring C_{60} molecules (the [001] direction) shows that molecules are linked by polymeric bonds in this direction. The volume per molecule obtained in this

Table 1. Lattice parameters and final refinement parameters (space group *Pnnm*)

a, Å	14.745(2)	R_{wp}, %	2.47
b, Å	9.843(2)	R_B, %	0.79
c, Å	9.092(1)	R_E, %	1.36
V, Å^3	1319.5(4)	χ^2	3.30

Table 2. Structural parameters of the orthorhombic phase

Atom	g	x	y	z	B_{iso}
C1	1.0	0.2276	−0.1153	0.0	2.42(5)
C2	1.0	0.1178	0.3056	0.085(1)	2.42(5)
C3	1.0	0.0408	−0.0353	0.409(1)	2.42(5)
C4	1.0	0.2234	0.1103	0.078(1)	2.42(5)
C5	1.0	0.1208	0.1664	0.288(1)	2.42(5)
C6	1.0	0.1765	−0.6782	0.254(1)	2.42(5)
C7	1.0	0.2260	−0.0293	0.132(1)	2.42(5)
C8	1.0	0.1713	0.2073	0.157(1)	2.42(5)
C9	1.0	0.1228	0.0307	0.329(1)	2.42(5)
C10	1.0	−0.1805	0.2376	0.0	2.42(5)
C11	1.0	−0.0479	0.3449	0.081(1)	2.42(5)
C12	1.0	0.0358	0.2399	0.283(1)	2.42(5)
C13	1.0	−0.1275	0.1950	0.265(1)	2.42(5)
C14	1.0	−0.1299	0.2785	0.133(1)	2.42(5)
C15	1.0	0.0342	0.3258	0.164(1)	2.42(5)
C16	1.0	−0.0436	0.1746	0.336(1)	2.42(5)

Table 3. Volume per C_{60} molecule

Symmetry	*Pnnm*	*Pnnm*	*Pm*3*m*
V_0, Å^3	659.7	663.6*	711.6**

* High-pressure orthorhombic polymeric phase [12].

** Initial C_{60} cubic phase, $a = 14.17$ Å.

work is compared with the corresponding characteristics of the initial cubic crystal and the high-pressure phase in Table 3. The volume of the unit cell decreases, and this decrease is comparable with a similar effect reported for the high-pressure phase (see [12]).

The results of our calorimetric and X-ray studies correlate well with the earlier data on the evolution of the EPR spectra of C_{60} doped with lithium [10]. It was found [10] that, unlike the initial samples, which had empty EPR spectra, C_{60} crystals doped with lithium gave an intense EPR signal comprising several lines with g factors close to 2.0. The principal lines always present in the EPR spectra of doped crystals were divided into three groups, X_1, X_2, and X_3. Below 250 K, the intensities of the EPR lines did not change as time passed. At $T > 280$ K, the intensities of lines X_1 and X_2, however, changed dramatically. This is qualitatively shown in Fig. 4a, where spectrum *1* was measured immediately after rapidly (for several minutes) cooling the sample from 520 K (the sample was saturated with lithium at 520 K) to room temperature. The spectra *2–4* were measured sequentialy with an interval of several hours. The sample always occurred at room temperature. The evolution of the EPR spectra was explained on the assumption of the occurrence of two simultaneous independent processes:

(1) a decrease in the concentration of X_1 centers (a broad EPR line) caused by their transformation into X_2 (a narrow EPR line);

(2) a decrease in the concentration of X_2 centers because of their transformation into EPR-inactive complexes.

It can be suggested that centers X_1 (a broad EPR line) correspond to electrons in the conduction band of a weakly doped region in $\mathrm{Li}_x\mathrm{C}_{60}$ with $x \ll 1$, whereas centers X_2 correspond to electrons on C_{60} molecules in the newly formed phase. Process (1) of transforming centers X_1 into X_2 is then likely to correspond to the formation of the new phase at 220–250 K. Process (2) leading to the disappearance of the X_2 EPR signal is, we believe, more complex. This process is likely to be related to the formation of the new phase and to correspond to the heat release peak of the heat capacity curve at 320–370 K. This suggestion finds support in the

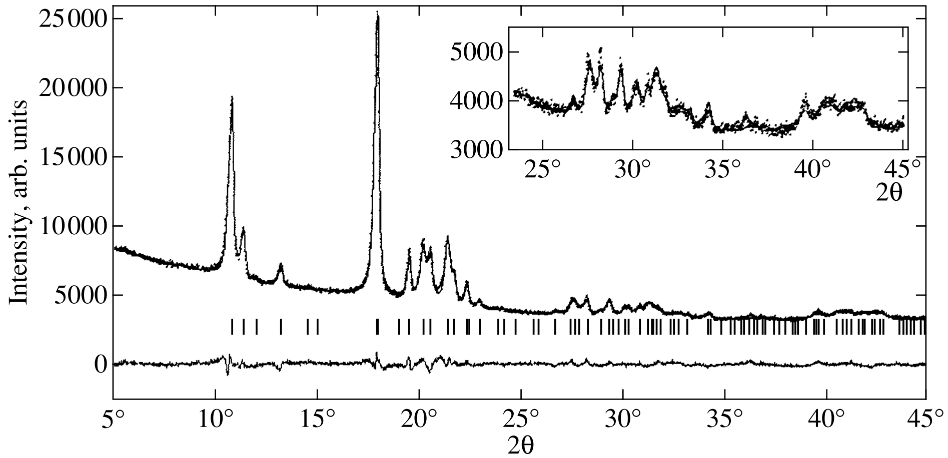


Fig. 3. Experimental, calculated, and difference diffraction patterns of the orthorhombic C_{60} phase.

observation that the EPR signal of Li_xC_{60} samples disappears after heat release is completed.

To summarize, we found that a polymeric orthorhombic structure is formed in the temperature range 320–370 K. Polymeric chains oriented in the [0, 0, z] and [1/2, 1/2, z] directions are rotated through $\alpha \approx 30°$ counterclockwise and clockwise, respectively. The question arises whether or not lithium is present in this structure. Clearly, the contribution of lithium to the intensity of diffraction lines is insignificant (small scattering cross section). For this reason, we did not take lithium into account in structure refinement. A small amount of lithium can be present in the structure, for instance, in tetrahedral voids. The possibilities of such a position order were considered in [6]. In addition, lithium can be concentrated in regions of C_{60} polymer chain beaks and can close chain termini. We, however, believe that the most probable process in the specified temperature range is the formation of an orthorhombic polymeric C_{60} phase and lithium clusters (this also follows from the IR data reported in [9]). Such a suggestion explains a strong decrease in the intensity of the EPR signal.

Studies of polymerized phases of C_{60} crystals formed in polymerization under high pressure showed that polymeric phases were unstable at high temperatures and their high-temperature annealing (620 K) restored the initial monomeric C_{60} crystal phases. To answer the question whether or not the polymeric fulleride phase described above was stable at $T = 520–620$ K, we studied the influence of high-temperature annealing on the properties of single crystals doped with lithium.

First note that annealing at 620 K restored the EPR signal. By way of example, Fig. 4b contains the EPR spectra of a Li_xC_{60} sample after prolonged storage at room temperature (curve *1*) and of the same sample after annealing at 620°C for one hour. Clearly, annealing at $T = 620$ K restored the X_1 and X_2 EPR lines; the appearance of these lines was evidence of polymeric chain rupture in the orthorhombic structure of phase II. However, note that the sample formed in annealing differed from freshly doped Li_xC_{60} samples: the concentrations of its EPR-active X_1 and X_2 centers changed insignificantly as time passed.

According to the X-ray data, annealing at 620 K restores the initial cubic structure with the $a = 14.177(2)$ Å lattice parameter. The presence of additional peaks in the diffraction pattern is, however, evidence of the formation of one more phase, whose structure is yet unknown. The positions of its diffraction lines cannot be described using the parameters of the orthorhombic lattice specified above (otherwise, the presence of this phase might be explained by annealing duration being insufficient for complete transition of the polymeric orthorhombic phase into the cubic structure). It can be suggested that one more phase transition occurs during heating (phase II') with the formation of a new structure and that traces of this phase are observed in the diffraction pattern. It can also be suggested that polymeric chain rupture that occurs during annealing yields not only isolated C_{60} molecules but also C_{60}–C_{60} dimers (the II–II' transition). In addition, lithium clusters can also participate in reactions with C_{60} molecules or C_{60}–C_{60} dimers at the temperature of annealing. It follows that high-temperature annealing results in the formation of Li_xC_{60} phases and restores the initial cubic C_{60} phase. However, these are but suggestions, and elucidating the nature of the unknown phase requires more thorough experiments *in situ*.

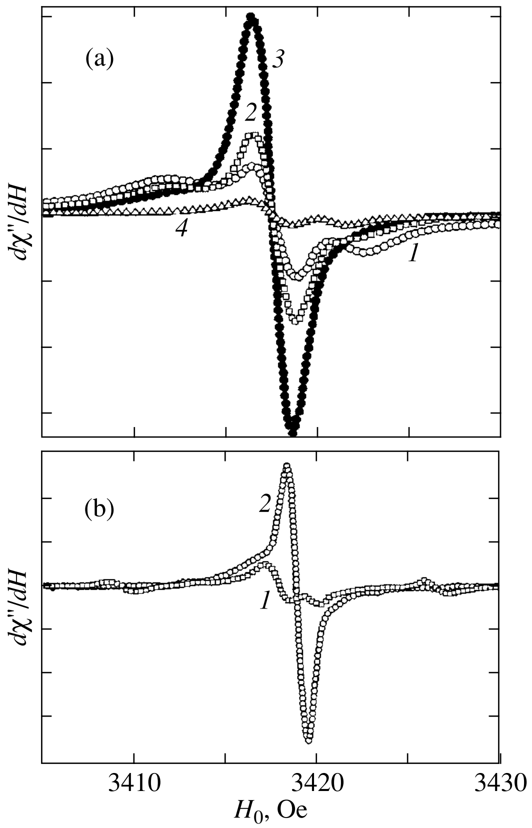


Fig. 4. Transformation of the EPR spectra of (a) C_{60} crystal doped with lithium by electrodiffusion and (b) Li_xC_{60} sample annealed at $T = 620$ K.

In conclusion, several points should be mentioned concerning the low-temperature phase (phase I). The Laue pattern and the rotating-crystal X-ray photograph of a C_{60} single crystal doped with lithium are shown in Figs. 5a and 5b, respectively; the measurements were taken at $T = 170$ K. The splitting of some cubic cell reflections, for instance, of the $(220)_C$, $(311)_C$, and $(422)_C$ reflections, is evidence that the introduction of lithium causes a decrease in symmetry. Several unsplit reflections were assigned to the cubic system with an $a = 14.11$ Å parameter. The presence of cubic phase peaks lends support to the data given above according to which lithiun intercalation only occurs in some part

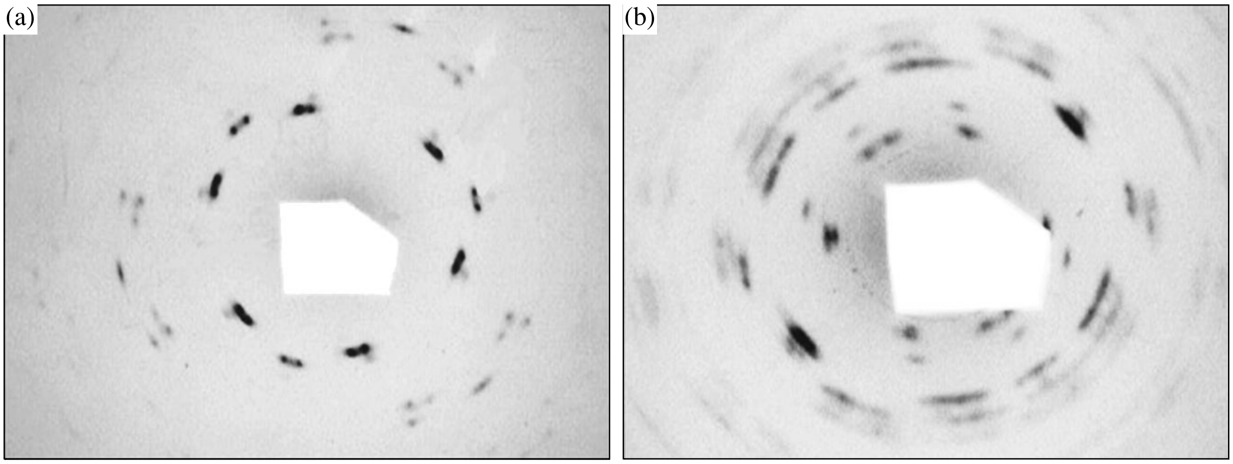


Fig. 5. (a) Laue pattern and (b) rotating-crystal X-ray photograph of a C_{60} single crystal doped with lithium by electrodiffusion (T = 170 K, MoK_α radiation).

of a sample. One of the problems that complicate the determination of the structure of phase I is that of the actual number of phases or structural states in this temperature interval. The point is that several phase transitions can occur during cooling samples from the temperature of their synthesis; one of these transitions occurs at 320–370 K. Currently, no information about the behavior of the polymeric orthorhombic phase described above at low temperatures is available. It can be suggested that a very complex state including several phases arises at $T < 250$ K. We believe that rapidly cooling samples from synthesis temperatures to $T < 250$ K would allow this problem to be solved.

4. CONCLUSIONS

To summarize, we

(1) studied the structure and phase transitions of C_{60} crystals doped with lithium by electrodiffusion;

(2) found that irreversible transformations occurred in the samples at 320–370 K;

(3) found that these transformations only occurred in some part of each sample and the size of this part depended on the degree of doping;

(4) determined the structure of the new phase (orthorhombic; space group *Pnnm*; lattice parameters a = 14.745 Å, b = 9.843 Å, and c = 9.092 Å; C_{60} molecules form polymeric chains along the crystallographic c axis);

(5) found that annealing at 620 K restored the EPR signal.

REFERENCES

1. D. M. Poirier, D. W. Owens, and J. H. Weaver, Phys. Rev. B **51**, 1830 (1995).
2. A. F. Hebard, M. J. Rosseinsky, R. C. Haddon, *et al.*, Nature **350**, 600 (1991).
3. W. Kratschmer, L. D. Lamb, K. Fostiropulos, and D. R. Huffan, Nature **347**, 354 (1990).
4. R. Kerkoud, P. Auban-Senzier, D. Jerome, *et al.*, Synth. Met. **77**, 205 (1996).
5. S. Pekker, L. Forro, L. Mihaly, and A. Janossy, Solid State Commun. **90**, 349 (1994).
6. P. W. Stephens, G. Bortel, G. Faigel, *et al.*, Nature **370**, 636 (1994).
7. S. Pekker, A. Janossy, L. Mihaly, *et al.*, Science **265**, 1077 (1994).
8. G. Yu. Mitronova, A. P. Savchenkova, A. F. Maĭorova, *et al.*, Dokl. Ross. Akad. Nauk **348**, 491 (1996).
9. A. P. Savchenkova, V. V. Avdeev, N. A. Zubareva, *et al.*, Zh. Fiz. Khim. **72**, 1337 (1998).
10. A. V. Bazhenov, S. I. Bredikhin, V. V. Kveder, *et al.*, Zh. Éksp. Teor. Fiz. **116**, 1706 (1999) [JETP **89**, 923 (1999)].
11. F. Izumi, International Union of Crystallography Newsletter, No. 20, 22 (1998).
12. V. Agafonov, V. A. Davydov, L. S. Kasherova, *et al.*, Chem. Phys. Lett. **267**, 193 (1997).
13. Y. Iwasa, T. Arima, R. M. Fleming, *et al.*, Science **264**, 1570 (1994).
14. M. Nucez-Requeiro, L. Marques, J.-L. Hodeau, *et al.*, Phys. Rev. Lett. **74**, 278 (1995).
15. E. V. Shulakov, R. A. Dilanian, O. G. Rybchenko, and V. Sh. Shekhtman, Kristallografiya **41**, 39 (1996) [Crystallogr. Rep. **41**, 33 (1996)].
16. V. L. Aksenov, Yu. A. Osip'yan, and V. S. Shakhmatov, Pis'ma Zh. Éksp. Teor. Fiz. **62**, 417 (1995) [JETP Lett. **62**, 428 (1995)].

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 93, No. 6, 2001, pp. 1245–1249.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 6, 2001, pp. 1438–1444.
Original Russian Text Copyright © 2001 by Tsiok, Khvostantsev.

SOLIDS
Structure

Phase Transitions in Cerium at High Pressures (up to 15 GPa) and High Temperatures

O. B. Tsiok* and L. G. Khvostantsev

Institute for High Pressure Physics, Russian Academy of Sciences, Troitsk, Moscow oblast, 142190 Russia
**e-mail: tsiok@ns.hppi.troitsk.ru*

Received June 7, 2001

Abstract—Phase transitions in cerium have been studied by the electrical resistance method in the 15-GPa pressure range at high temperatures. At pressures above 10 GPa, cerium represents a mixture of stable and metastable phases, the composition of this mixture being dependent on the trajectory in the *P*–*T* plane that leads to a given point. Transformations in both stable and metastable components of the mixture proceeding rather independently display a complicated picture of phase transitions. It was assumed that only the α (fcc) and α' (α-*U*) phases are stable at pressures above the well-known γ–α transition, the other phases being metastable. The proposed bow-shaped equilibrium phase diagram includes an extremely wide hysteresis region, where stable and metastable phases can coexist. The fcc α phase alone survives upon heating above 500°C at 15 GPa. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

For a long time, cerium (Ce) has been a very attractive object for investigations in the field of high pressure physics. In addition to the well-known, carefully studied isostructural γ–α transition, Ce exhibits numerous low-symmetry structures within a 12-GPa pressure interval [1, 2]. In addition, the bowlike shape of transition lines on the phase diagram discovered many years ago [3] and then verified [2] still looks very unusual. One should also mention a long discussion regarding the structures proposed for the so-called "underbow" phases. This question was studied only recently [4, 5]. In addition, all the low-symmetry structures, including the tetragonal (bct) one, are distortions of the fcc structure and are very close in energy to each other [6], which may result in a confused picture of phase transitions at pressures above 10 GPa. Thus, despite the availability of modern powerful X-ray methods, the phase diagram of Ce at pressures above 12–14 GPa at room temperature and 9–10 GPa at elevated temperature still remained unknown.

At the beginning of our work, we only planned to verify the phase diagram reported in [3] and investigate the phase boundary between the α (fcc) and ε (bct) phases above the hypothetical triple point in the region of 8.5 GPa, but the results were really surprising. While our results below 8–9 GPa were in good agreement with the previously published data [2, 3], they was not so above 9 GPa. Probably, this was a result of limited experimental facilities employed in [1–3]. In any case, the transition between the α (fcc) and ε (bct) phases has not been studied at all. Some speculations about the corresponding line position in [2] were based on the analysis of the *c*/*a* ratio for the tetragonal phase; however, no experimental evidence for this transition was presented. Unless data on the α–ε transition are available, there is no ground to discuss the ε phase stability. We studied this transition and came to a conclusion that the ε phase is metastable just like the α'' phase. The ε phase disappears upon heating above 500°C at pressures above 12 GPa, converting into the α phase.

2. EXPERIMENTAL

The experiments under high pressure in the 15-GPa range were performed in a specially designed apparatus with a total high-pressure zone volume of 0.3 cm^3. The other experiments were carried out using the toroidal devices [7] with a central recess of 15 and 25 mm in diameter. The larger apparatus was used for experiments in the 10-GPa range under hydrostatic or close to hydrostatic conditions with precise pressure control. A detailed description of the toroidal high pressure apparatus, the design of a hydrostatic liquid-filled ampule, and the experimental methods were presented in [8]. The only difference was in the pressure-transmitting media used. Traditionally utilized petroleum ether decomposes at high temperatures releasing hydrogen. This medium can be used at temperatures up to 600°C but only if the heating continues for several minutes and the sample is not sensitive to hydrogen. Our experiments involved long heating times during the pressure cycles carried out at high temperatures. It was found that even at 400°C and 3–8 GPa the samples absorbed a large amount of hydrogen. As a result, the *P*–*T* parameters of the transitions were strongly changed, and even some new transitions appeared. In the course of experiments, it was established that polysiloxane (PES5) is the most suitable liquid for our purposes. This medium can serve up to 500°C without any chemical interactions with Ce. Moreover, polysiloxane prevents the oxi-

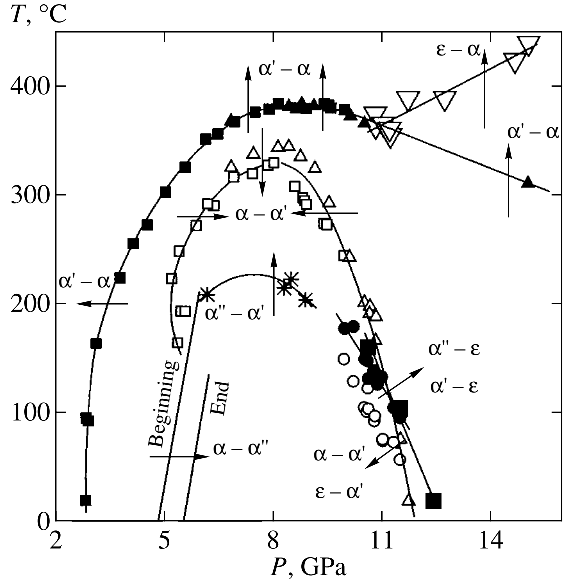


Fig. 1. The scheme of phase transitions in Ce. Black and open small squares asterisks, and lines without symbols correspond to the transitions under hydrostatic conditions; other denote quasi-hydrostatic experiments, arrows show the path in the *P–T* plane.

dation of Ce during storage, mechanical treatment, preparation of the sample assembly, and even during the spot welding of contacts. Thus, the ampules with two isolated compartments filled with different liquids were used. Polysiloxane was used for the "hot" compartment, and a 4 : 1 methanol–ethanol mixture, for the "cold" one. The accuracy of pressure determination was better than 0.05 GPa.

The experiments at higher pressures were carried out using two devices with the central recess 15 mm in diameter. One of these was the classical device [8] tolerating a 12-GPa limiting pressure and usually employed up to 10 GPa. The other device was specially designed for this work by the authors. The point is that the 12-GPa pressure range was insufficient for refining the confused picture of phase transitions in Ce. The new device reproducibly operates with a pressure–load characteristic being linear up to 15 GPa. The details will be published later. Both devices were calibrated against the well-known transitions in Bi, Sn, Ba, Pb, and ZnS. In addition, the calibration was checked against the α–α'' transition in Ce [9] in each experiment. The estimated accuracy of pressure determination in this pressure was about 0.2–0.3 GPa. This value depends mainly on the pressure variations caused by heating. In order to avoid errors associated with possible chemical reactions, MgO, pyrophyllite, and Teflon were used as pressure media. The temperature was measured using a chromel–alumel thermocouple calibrated at zero pressure without any corrections for the pressure effect on the thermal emf. A Ce ingot with a 99.99% purity refined through the zone melting was placed at our disposal by Prof. I.A. Smirnov (Ioffe Physicotechnical Institute, St. Petersburg). The samples were cut from the ingot and polished under droplets of polysiloxane in order to obtain a rectangular shape and remove the oxide traces. The typical size of the samples was $2 \times 0.8 \times 0.5$ mm^3. The electrical contacts were made from Ni or alumel wires 0.05 and 0.2 mm in diameter. The wires were spot-welded to the sample for hydrostatic experiments or simply pressed to it in the case of quasi-hydrostatics. The electrical resistance was measured by the four-point-probe ac technique using a lock-in amplifier. During the experiments, the resistance was measured as a function of continuously varying parameters (dT/dt = 0.25 K/s, dP/dt = 0.03–0.1 GPa/min).

3. RESULTS AND DISCUSSION

More than 20 experiments with multiple pressure and temperature cycles were carried out in the course of this study. The results are presented in Fig. 1. The identification of phases and establishment of the correspondence between resistance anomalies and transitions of particular phases were carried out as follows:

(1) Below 9 GPa, the phase diagram has been well studied by X-ray diffraction [2, 4, 5], and we simply use this information for establishing the correspondence between resistance anomalies and particular phase transitions.

(2) Above 9 GPa, the situation is more complicated. It is known that the ε (bct) phase appears at room temperature and a pressure of about 12 GPa [1]. The ε–α transition was identified by the sign of the resistance anomalies, which is opposite to that of the α'–α transition. This will be discussed in detail below.

(3) Different transitions have different kinetics, and this fact also was used for the identification of phases.

Below 9 GPa, our results concerning the α–α' transition are in good agreement with the results obtained in [3]. At room temperature and on heating up to 200°C, the α phase transforms into the α'' phase [9, 4, 5]. The latter phase seems to be metastable and appears due to kinetics reasons. This conclusion is based on the following facts: (i) the α'' phase irreversibly transforms into α' upon heating, and (ii) the α–α'' transition exhibits a martensitic character and takes place in the hysteresis region of the α–α' transition. From the experimental standpoint, the term "martensitic" means that this transition is temperature-independent and time-independent, i.e., the sample resistance *r* in the transition region does not vary with time. The shape of the *r*(*P*) dependences and the transition hysteresis are the same in the entire temperature range where the martensitic transition takes place. Thus, the phase diagram of Ce in this pressure range is well established, and our results are in good agreement with the previously published data [2–5]. However, it is not the case in the region above 9 GPa.

It is well known that cerium transforms into the ε (bct) phase above 12 GPa [1]. Let as consider this *P–T* field more carefully. In order to obtain an understandable picture of phase transitions in this *P–T* region, a number of sophisticated experiments have

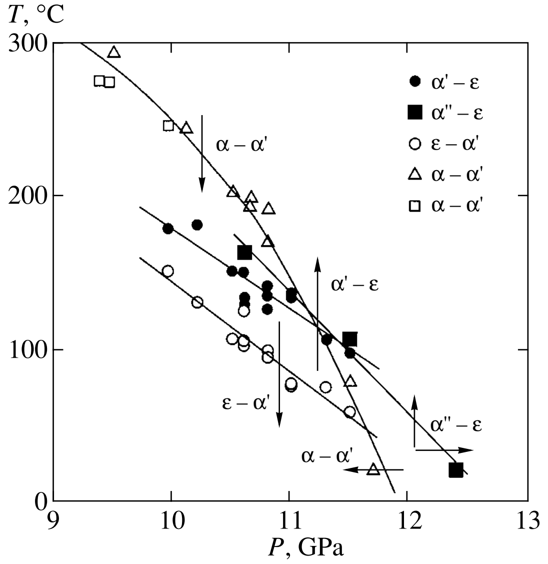


Fig. 2. A fragment of Fig. 1 demonstrating the complicated bundle of transition lines in the regime of phase mixture. One squares and triangles refer to the α–α' transition under hydrostatic and quasihydrostatic conditions.

been performed. A fragment of the scheme of these phase transitions is shown in Fig. 2. In the small *P–T* region, we have two lines of the ε–α' transitions (direct and reverse), the line of the direct α''–ε transition, and the line of the reverse α–α' one. This means that the sample resistance displays anomalies corresponding to each transition, with the anomaly amplitudes depending on the previous paths of the sample on the *P–T* plane in the course of experiment. The only way to understand the intricate picture of phase transitions in Ce above 9 GPa is to assume that Ce is a multiphase object in this pressure range. The phase transitions in each component of the mixture proceed to some extent independently in the sense that the resistance anomalies are observed on each line. The anomaly amplitudes behave in the following way: when one is large the other is small and vice versa.

The examples of transitions between "pure" phases are shown in Fig. 3, where part (a) corresponds to the above discussed α–α' transition below 9 GPa, and part (b) illustrates transitions to the ε (bct) phase. Starting from the α'' phase, the first temperature run shows the α''–ε transition proceeding through the nucleation and growth mechanism. The sample resistance in this transition region is time-dependent, the *r*(*T*) curve shape depending on the temperature growth rate. However, the reverse transition leads to the α' phase, which was proved in a separate experiment. When releasing the pressure at room temperature, the transitions from the α' and α'' phases to the α phase take place at different pressures, roughly at 3 and 5 GPa, respectively. In our case, only the α'–α transition in the region of 3 GPa was observed. The mechanism of the α'–ε transition differs from that of the α''–ε transition. The resistance in the α'–ε transition region is time-independent. Therefore, the system exhibits one more martensitic transition.

The transitions between "pure" phases in Ce take place only under certain special conditions. The pressure must not exceed 9 GPa for the α–α' transition, and the temperature must be lower than 300°C at a pressure higher than 10.5 GPa for the α'–ε transition. Some examples of such transitions in a mixture of Ce phases are presented in Fig. 4. At pressures exceeding 8.5 GPa, the hysteresis of the α–α' transition starts to increase. Then, additional resistance anomalies appear, and the higher the pressure, the larger the amount of the ε phase involved in the phase transitions.

An example of the "relatively pure" ε–α transition is shown in Fig. 5. During this transition, the resistance grows in contrast to the α'–α case, providing the opportunity to separate these transitions. The temperature cycling before and after the transition showed different *r*(*T*) behavior for the ε and α phases. As the temperature decreased, the reverse ε–α' and α–α' transitions took place. One more example illustrating transitions in the regime of phase mixture is shown in Fig. 6. In this experiment, the sample was pressurized to 15 GPa at

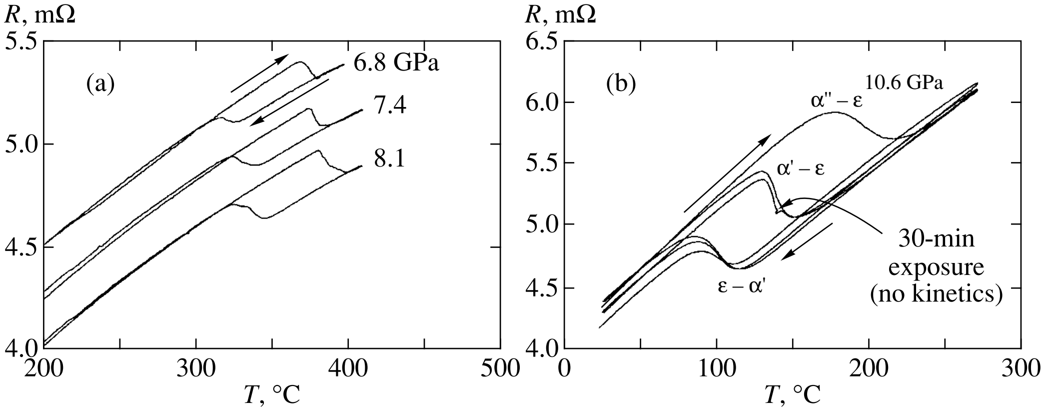


Fig. 3. The examples of "pure" transitions: (a) transitions α''–α and α–α' below 8.5 GPa (the curves for 7.4 and 8.1 GPa are shifted downward); (b) the transition to a tetragonal phase under some special conditions discussed in the text.

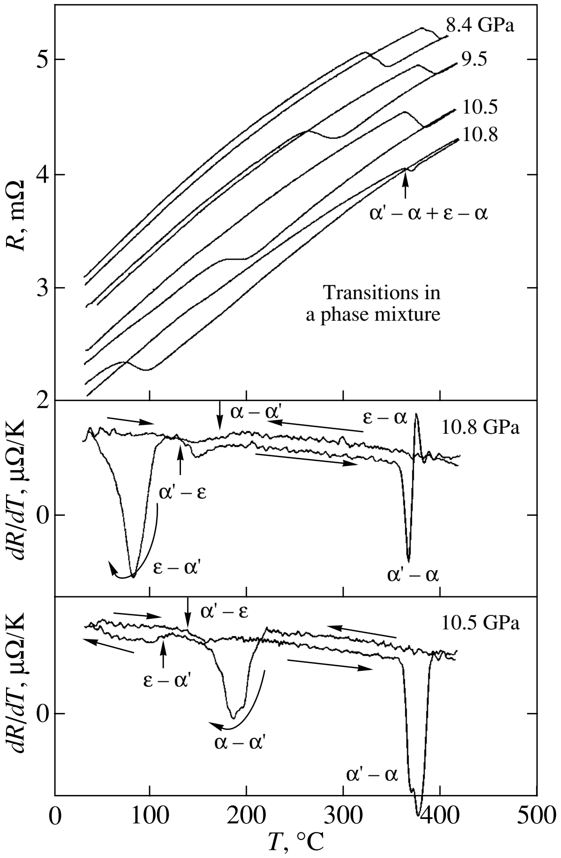


Fig. 4. The typical case of transitions in the mixture of phases (only the mathematical treatment is capable of shedding some light upon this confused picture). The top part shows the plots of resistance versus temperature (the curves for 9.5, 10.5, and 10.8 GPa are shifted downward); the two bottom parts show the numerically calculated derivatives plotted on an arbitrary scale.

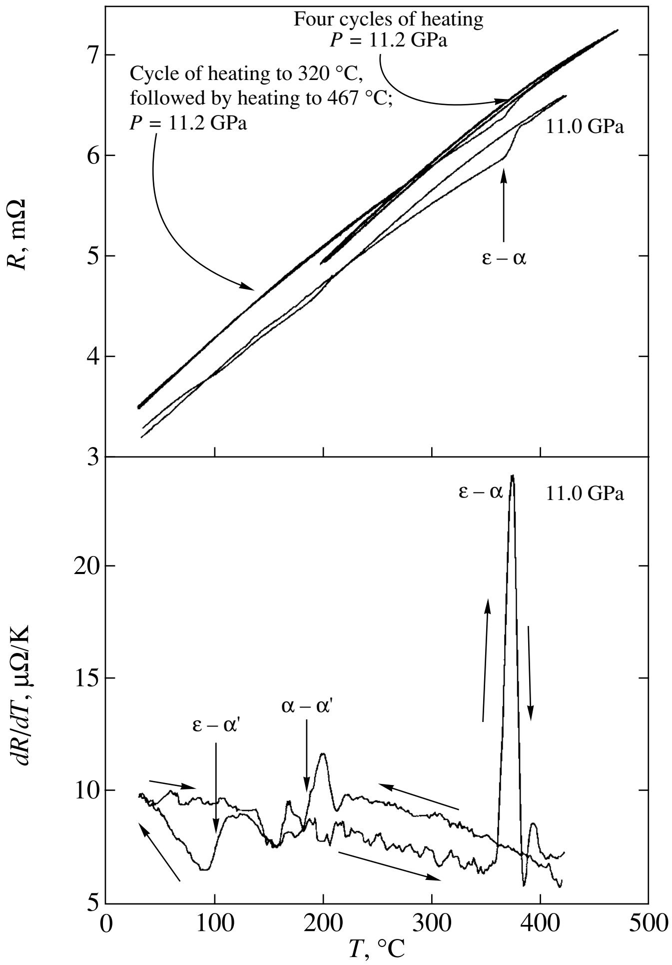


Fig. 5. The examples of (a) a transition from the ε (bct) phase (specially prepared through some *P*–*T* treatment) to the α (fcc) phase (the curve for 11.0 GPa is shifted downward) and (b) a numerical derivative of the last run at 11 GPa (the "additional" transitions are clearly seen).

room temperature, then heated to 200°C for removing traces of the α'' phase, and then the first temperature run was made. As can be seen, the α'–α and ε–α transitions are well separated. The resistance anomaly amplitudes are nearly equal, showing a large admixture of the α' phase in *P*–*T* region well above that of the martensitic α'–ε transition. The subsequent heating cycle displays the absence of transitions. The α phase is retained even at room temperature and transforms into the α' phase only when the pressure decreases to 11.7 GPa.

Summarizing, one should say that such a sophisticated scheme of phase transition with a very wide hysteresis region is a consequence of the fact that the energies of all structures under consideration are very close to each other [6]. Of course, our work leaves a lot of questions unsolved. Some of these are related to stability of the ε (bct) phase and the lines of the α'–ε and ε–α' transitions. The resistance anomalies corresponding to these transitions become weaker and weaker at lowering pressure and then disappear at all. Is this an indication of the ε phase metastability? Is it possible to bypass this transition at high temperature, retaining the α' phase? A large amount of work should be done to clarify the questions. We plan to do this in future.

Let us try to construct the equilibrium phase diagram of Ce on the basis of available data. The α'' and ε phases seem to be metastable. The α'' phase has already been discussed. The case of the ε phase is more difficult. The α'–ε transition is clearly martensitic, and it seems to disappear somewhere in the region of 10 GPa and 170°C, but it takes place far below the middle line between the α'–α and α–α' transitions. In addition, there is a transition from the ε to α phase but there is no reverse transition (perhaps, it takes place only below room temperature?). With lowering pressure and temperature, the α phase transforms directly to the α' phase bypassing the ε phase, although with an extremely large field of hysteresis. In our opinion, only the α (fcc) and α' (α–*U*) phases are thermodynamically stable in the *P*–*T* region under consideration.

The proposed equilibrium phase diagram of Ce is presented in Fig. 7. At present, the bow-shaped line on

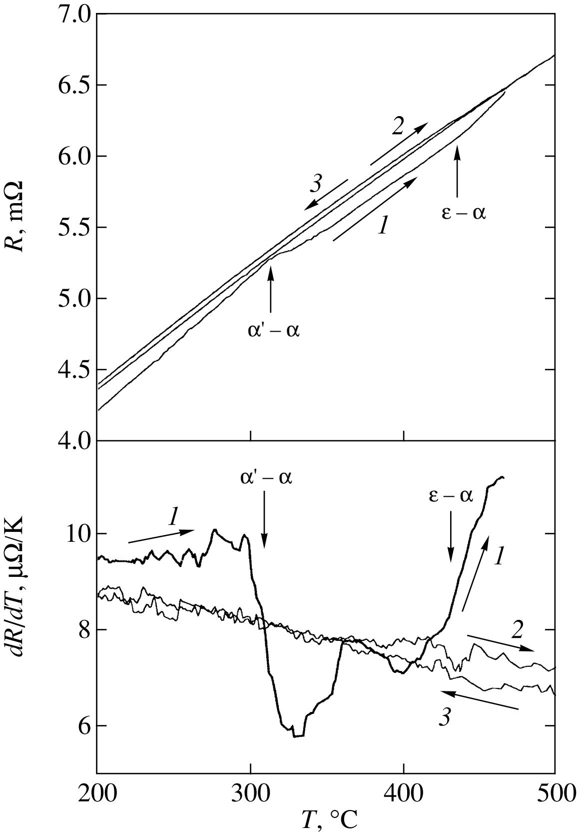


Fig. 6. The special experiment at pressure up to 15 GPa showing the presence of different phases far beyond the transition to the ε (bct) phase (the transitions are well separated). Only the α (fcc) phase is retained after heating above 500°C; the other phases are metastable. Arrow *1* indicates heating in the first cycle; arrows *2* and *3* indicate heating and cooling in the second cycle. The bottom part shows the numerical derivatives of resistance.

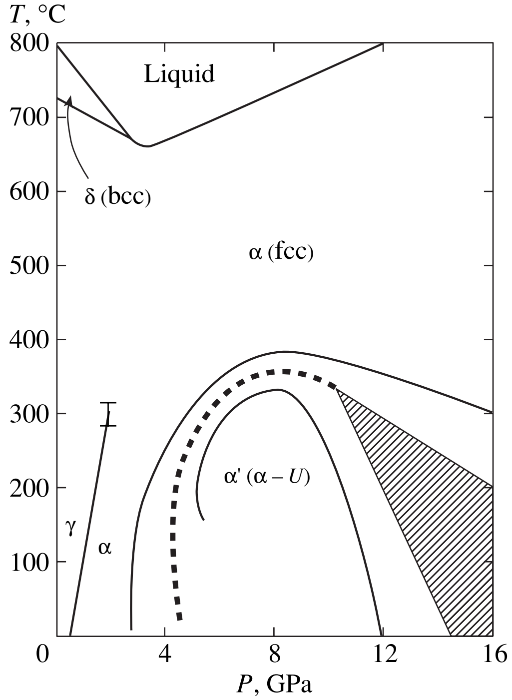


Fig. 7. The proposed equilibrium phase diagram of Ce. Note the extremely large “hysteresis” region at relatively high temperature.

the phase diagram is not so surprising as it was two decades ago [3]. For example, the phase diagram of La shows the same feature [10]. A rather unusual feature is a very large hysteresis region at relatively high temperatures. It looks like a “field of equilibrium” (of course, this is impossible in thermodynamics) where the α, α', and ε (and α" at low temperature) phases of Ce can coexist, a relative composition of the mixture being dependent on the trajectory in the *P*–*T* plane leading to a given state (i.e., on the method by which this state was obtained).

4. CONCLUSION

The phase transitions in Ce at pressures up to 15 GPa and at temperatures above 500°C have been studied. The part of the *P*–*T* diagram below 9 GPa is in good agreement with the previously reported data [2–5]. At higher pressures, the ε–α transition has been studied. The reverse transition was not observed. Once obtained, the α phase remains stable at room temperature and transforms into α' only when the pressure decreases to 11.7 GPa. Thus, the ε phase seems to be metastable. In addition, the extremely large field of hysteresis has been detected at pressures above 12 GPa. On the basis of the experimental data obtained, a new phase diagram of Ce has been constructed.

REFERENCES

1. S. Endo and N. Fujioka, Phys. Lett. A **70**, 475 (1979).
2. Y. C. Zhao and W. B. Holzapfel, J. Alloys Compd. **246**, 216 (1997).
3. T. E. Antonova, I. T. Belash, and E. G. Ponyatovskiĭ, Fiz. Met. Metalloved. **51**, 131 (1981); Phys. Met. Metallogr. **51**, 110 (1981).
4. M. I. McMahon and R. J. Nelmes, Phys. Rev. Lett. **78**, 3884 (1997).
5. M. I. McMahon and R. J. Nelmes, Rev. High Pressure Sci. Technol. **7**, 313 (1998).
6. P. Ravindran, L. Nordstrom, R. Ahuja, *et al.*, Phys. Rev. B **57**, 2091 (1998).
7. L. G. Khvostantsev, L. F. Vereshchagin, and A. P. Novikov, High Temp.–High Press. **9**, 637 (1977).
8. L. G. Khvostantsev, V. A. Sidorov, and O. B. Tsiok, in *Properties of Earth and Planetary Materials at High Pressures and Temperatures*, Ed. by M. H. Manghnani and T. Yagi (American Geophysical Union, Washington, 1998), p. 89.
9. L. G. Khvostantsev and N. A. Nikolaev, Phys. Status Solidi B **114**, K135 (1982).
10. F. Porsch and W. B. Holzapfel, Phys. Rev. Lett. **70**, 4087 (1993).

Translated by O. Tsiok

Journal of Experimental and Theoretical Physics, Vol. 93, No. 6, 2001, pp. 1250–1260.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 6, 2001, pp. 1445–1456.
Original Russian Text Copyright © 2001 by Kazei, Kolmakova, Shishkina.

SOLIDS
Structure

Crystal Field and Quadrupole Interaction Effects in $YbXO_4$ Zircons (X = V, P)

Z. A. Kazei[a, *], **N. P. Kolmakova**[b], **and O. A. Shishkina**[b]
[a]*Moscow State University, Moscow, 119899 Russia*
**e-mail: kazei@plms.phys.msu.su*
[b]*Bryansk State Technical University, Bryansk, 241035 Russia*

Received June 21, 2001

Abstract—The magnetic and magnetoelastic properties of $YbPO_4$ and $YbVO_4$ crystals are investigated experimentally and theoretically; the crystal field parameters are determined, as well as the magnetoelastic coefficients B^μ and total quadrupole coupling constants G^μ for all symmetry modes. It is found that, for $H \parallel [100]$, γ-symmetric quadrupole interactions predominate and are responsible for a significant contribution to the third-order susceptibility, magnetization, magnetostriction, and elastic constant. It is demonstrated that, in the absence of an external field, these interactions do not lead to quadrupole ordering, because the respective deformation susceptibility χ_γ is several times less than the critical value of $1/G^\gamma$. The influence of an external magnetic field along different symmetry axes on the quadrupole effects and quadrupole interactions in Yb zircons is investigated. It is demonstrated that, for $H \parallel [110]$, the susceptibility χ_γ increases with the field, so that in a fairly strong field in the investigated crystals one can expect a γ-symmetric stimulated phase transition. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Rare-earth paramagnets with a tetragonal structure of RXO_4 zircon (X = P, V) are ideal for use in the investigation of the effects due to crystal field and quadrupole interactions and for validation of theoretical models. Systematic investigations reveal that the behavior of the entire family of rare-earth zircons is described within a unified approach using the crystal field parameters and the parameters of single-ion and pairwise interactions, which vary systematically over the series of rare-earth zircons [1, 2]. At present, it is safe to assume that the presence or absence of quadrupole ordering (Jahn–Teller cooperative effect) or, in a more general case, of observed quadrupole effects for different compounds of this family is due to the ratio between the crystal field, which forms the spectrum and wave functions of a rare-earth ion, on the one hand, and the values of quadrupole coupling constants, on the other hand. It is the splitting in the crystal field that defines the presence in the spectrum of degenerate or quasi-degenerate levels with high quadrupole moments, which are essential for quadrupole ordering. As to the values of total quadrupole coupling constants G^μ ($\mu = \alpha, \gamma, \delta, \varepsilon$), the main contribution to which for the family of rare-earth zircons is made by single-ion magnetoelastic interaction, these values vary systematically over the series of rare-earth zircons in accordance with the second-order Stevens parameter α_J [2]. Because the quadrupole interaction constants G^μ of different symmetries vary over the series of rare-earth zircons as α_J^2, the strongest quadrupole effects are to be expected for rare-earth ions with high values of α_J, such as Tb, Tm, and Dy, for which spontaneous quadrupole ordering is observed most frequently.

In the case of a Yb ion, the value of the parameter α_J is also fairly high; therefore, for these compounds one can expect, if not a phase transition, at least (provided the necessary conditions in the spectrum of ions in a crystal field are met) the quadrupole effects observed. However, no systematic investigation of the magnetic and magnetoelastic properties of Yb zircons and no reliable determination of microscopic interaction constants on the basis of these properties have been performed up to now. One can only note the investigation of γ-symmetric elastic constants in $YbPO_4$, which reveals a softening of $\Delta C^\gamma(T)/C^\gamma(300\text{ K}) \sim 20\%$ as the temperature decreases to 4.2 K [3]. Studies of Raman spectra in $YbPO_4$ resulted in revealing the temperature-dependent electron–phonon interaction which was an order of magnitude higher than that for other rare-earth systems [4]. The results of investigation of the Mössbauer effect demonstrate that magnetic ordering occurs in $YbPO_4$ and $YbVO_4$ at temperatures of approximately 0.1 K [5]; the antiferromagnetic state in $YbVO_4$ was investigated by the NMR method [6].

This paper deals with the experimental and theoretical investigation of the magnetic and magnetoelastic properties of $YbPO_4$ and $YbVO_4$ zircons in a wide range of temperatures and magnetic fields for the purpose of determining the microscopic parameters of interactions and developing a complete pattern for the entire family of rare-earth zircons.

2. SAMPLES AND MEASUREMENT PROCEDURE

We measured the magnetization and magnetostriction curves for $YbPO_4$ and $YbVO_4$ crystals along symmetric [100], [110], and [001] directions in the temperature range from 1.5 to 300 K and with steady magnetic fields of up to 140 kOe. Crystals of Yb zircons were grown by spontaneous crystallization from a solution in a ($PbO–PbF_2$)-based melt and had average dimensions of $2 \times 2 \times 3$ mm^3. Very informative in the case of investigation of systems with strong magnetoelastic and quadrupole interactions is the nonlinear susceptibility, i.e., the term at H^3 in the field expansion of the magnetic moment $M = \chi_M H + \chi_M^{(3)} H^3$ [7]. In order to determine the first- and third-order magnetic susceptibility, the magnetic data were represented in $H/M = \chi_M^{-1} + bM^2$ coordinates as a function of M^2. The foregoing correlations in fairly weak fields are linear; in this case, the value of H/M for $M^2 = 0$ gives the ordinary magnetic susceptibility χ_M^{-1}, and the slope b of the straight line is related to the nonlinear susceptibility as $\chi_M^{(3)} = -b/(\chi_M^{-1})^4$.

The positive value of b for the linearized dependences $H/M(M^2)$ corresponds to the negative value of the coefficient at H^3, i.e., to the most commonly encountered case of a magnetization curve with saturation. The negative value of b is observed less frequently and corresponds to a magnetization curve with a kink. The positive nonlinear susceptibility may be due either to Jahn–Teller correlations or to the effects of level crossing in strong magnetic fields.

The magnetostriction was measured by miniature cross-shaped sensors which enabled one to simultaneously investigate two deformation components. The variation of the strain gage resistance $\Delta R/R$ during its deformation was measured by a bridge circuit into which a second strain gage (glued onto a diamagnetic analog of the sample being investigated) was connected to compensate for the variation of the strain gage resistance due to temperature and magnetic field.

In what follows, the following contracted notation is used for the variation of the length $\lambda_{ij} = \lambda^{\beta_1\beta_2\beta_3}_{\alpha_1\alpha_2\alpha_3}$. The first subscript i corresponds to the measurement direction of the strain gage, and the second subscript j corresponds to the field direction ($i, j = a \equiv [100]$, $a' \equiv [010]$, $b \equiv [110]$, $b' \equiv [\bar{1}10]$, $c \equiv [001]$). The data were represented in the form of functions of the square of the field, and different linear combinations of λ_{ij} corresponded to different normal deformation modes ε^μ. The range of linearity of the curves $\varepsilon^\mu(H^2)$ depends both on the temperature and on the normal mode being investigated. For measuring the fully symmetric $\alpha 1$ and $\alpha 2$ modes, the strain gage was glued in the (100) plane such that its directions of strain sensitivity were parallel with the [001] and [010] axes. When γ- and δ-symmetric orthorhombic modes were investigated, the strain gage was glued in the (001) plane; in this case, the directions of strain sensitivity were parallel with the [100] and [010] or [110] and $[\bar{1}10]$ axes.

3. THEORETICAL TREATMENT

The complete Hamiltonian for a single 4f ion includes the crystal field Hamiltonian H_{CF} and the Zeeman term H_Z describing the interaction between the magnetic moment and external field $\mathbf{H}$, as well as the quadrupole interaction Hamiltonian H_{QT}:

$$H = H_{CF} + H_Z + H_{QT}. \tag{1}$$

One can use the method of equivalent operators and the molecular field approximation for pairwise quadrupole interactions to write these terms in the form (for more detail, see, for example, [8])

$$\begin{aligned} H_{CF} = \alpha_J B_2^0 O_2^0 + \beta_J (B_4^0 O_4^0 + B_4^4 O_4^4) \\ + \gamma_J (B_6^0 O_6^0 + B_6^4 O_6^4), \end{aligned} \tag{2}$$

$$H_Z = -g_J \mu_B \mathbf{H} \cdot \mathbf{J}, \tag{3}$$

$$\begin{aligned} H_{QT} = -G^\alpha \langle O_2^0 \rangle O_2^0 - G^\gamma \langle O_2^2 \rangle O_2^2 - G^\delta \langle P_{xy} \rangle P_{xy}; \\ P_{xy} = \frac{1}{2}(J_x J_y + J_y J_x). \end{aligned} \tag{4}$$

In these expressions, O_n^m and B_n^m ($n = 2, 4, 6$; $m = 0$, 4; $m \le n$) are the crystal field operators and parameters; α_J, β_J, and γ_J are the Stevens parameters; g_J and μ_B are the Lande g-factor and Bohr magneton, respectively; and

$$\langle O_n^m \rangle = \sum_i \langle i | O_n^m | i \rangle \frac{1}{Z} \exp\left(-\frac{E_i}{k_B T}\right)$$
$$(O_n^m = O_2^0, O_2^2, P_{xy})$$

are quadrupole moments. The quadrupole interaction Hamiltonian includes both pairwise quadrupole interaction and single-ion magnetoelastic interaction; after formal transformations, they reduce to the generalized quadrupole interaction Hamiltonian H_{QT}, in which the quadrupole coupling constants

$$G^\mu = G^\mu_{ME} + K^\mu = \frac{(B^\mu)^2}{C_0^\mu} + K^\mu \quad (\mu = \gamma, \delta)$$

include contributions both from single-ion magnetoelastic interaction B^μ and from pairwise quadrupole

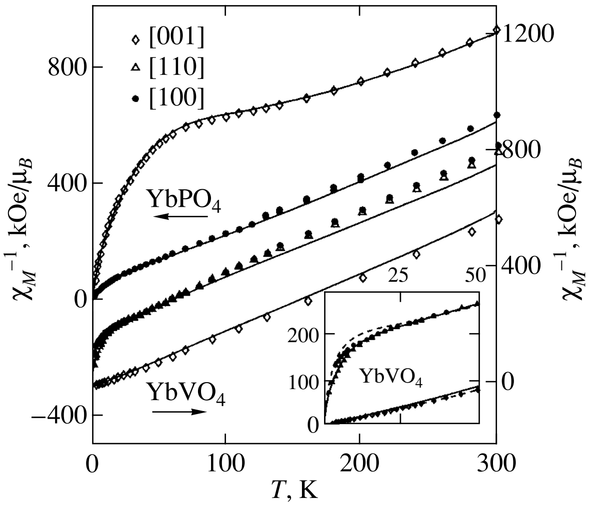


Fig. 1. The temperature dependence of the inverse magnetic susceptibility of YbPO$_4$ and YbVO$_4$ crystals along the [100], [110], and [001] axes. The lines indicate the dependences calculated disregarding bilinear interactions for sets of crystal field parameters M and N (dashed curves in the inset) from the table.

interaction K^μ (C_0^μ is the background elastic constant in the absence of interactions). In the Hamiltonian H_{QT}, the terms are omitted which correspond to ε symmetry and produce no contribution for the magnetic field orientation in the basal plane or along the tetragonal axis.

In the presence of minor external action, analytical expressions were derived in the perturbation theory approximation for free energy and for different generalized susceptibilities with the magnetic field orientation along symmetric directions in crystal [8]. For example, the third-order magnetic susceptibility, i.e., the initial curvature of the magnetization curve,

$$\chi_M^{(3)} = \frac{1}{(1-n\chi_0)^4} \times \left[\chi_0^{(3)} + \frac{2G^\alpha(\chi_\alpha^{(2)})^2}{1-G^\alpha\chi_\alpha} + \frac{2G^\mu(\chi_\mu^{(2)})^2}{1-G^\mu\chi_\mu}\right], \tag{5}$$

depends on four single-ion susceptibilities ($n = \theta^*/C$, $C = N\mu_{\text{eff}}^2/3k_B$ is the Curie constant). The single-ion susceptibilities χ_0 and $\chi_0^{(3)}$ describe the slope and curvature of the magnetization curve in the absence of any interactions. The deformation, $\chi_\mu = \partial\langle O_2^\mu\rangle/\partial\varepsilon^\mu$, and quadrupole-field, $\chi_\mu^{(2)} = \partial\langle O_2^\mu\rangle/\partial H^2$, susceptibilities characterize the response of the quadrupole moments $\langle O_2^\mu\rangle$ ($\langle O_2^\mu\rangle = \langle O_2^0\rangle$, $\langle O_2^2\rangle$, P_{xy} for μ = α, γ, δ, respectively) to deformation and to the magnetic field and are responsible for the softening of the respective elastic constant and for the behavior of magnetostriction in weak fields,

$$C^\mu = C_0^\mu - \frac{(B^\mu)^2\chi_\mu}{1-K^\mu\chi_\mu}, \tag{6}$$

$$\varepsilon^\mu = \frac{B^\mu}{C_0^\mu}\frac{\chi_\mu^{(2)}}{1-G^\mu\chi_\mu}H^2 + \dots. \tag{7}$$

The anisotropic susceptibilities $\chi_0^{(3)}$, χ_μ, and $\chi_\mu^{(2)}$ may be calculated given the rare-earth ion eigenfunctions and energy levels, preassigned by the crystal field. The description of the first-order magnetic susceptibility along the [001] axis and in the basal plane defines the bilinear interaction factor θ*. The third-order magnetic susceptibility, magnetostriction, and elastic constants produce different combinations of the coefficients B^μ and K^μ. Different experimental configurations enable one to separately study the α, γ, and δ modes and determine the respective coefficients of single-ion and pairwise interactions. For example, one can use a comparison of the calculated curve $1/(\chi_\mu^{(2)})^{1/2}$ and the linearized temperature dependence $H/|\varepsilon^\mu|^{1/2}$ for magnetostriction to derive the factors $A^\mu = B^\mu/C_0^\mu$ and G^μ (see Eq. (7)).

The interaction parameters for all symmetry modes, determined as a result of magnetic measurements in relatively weak fields, were then used to describe the properties in strong magnetic fields. The eigenvalues and eigenfunctions required for the calculation of magnetization and magnetostriction in strong fields were determined by way of numerical diagonalization of a complete Hamiltonian in which α-, γ-, and δ-symmetric quadrupole interactions dependent on the electronic configuration were taken into account in a self-consistent manner. The symmetrized magnetoelastic deformations ε^μ (μ = α1, α2, γ, δ) are linearly related to the quadrupole moments $\langle O_2^\mu\rangle$,

$$\varepsilon^\mu = A^\mu\langle O_2^\mu\rangle = \frac{B^\mu}{C_0^\mu}\langle O_2^\mu\rangle, \tag{8}$$

where the factors A^μ were determined during measurements in relatively weak fields.

4. EXPERIMENTAL RESULTS AND DISCUSSION

4.1. First-Order Magnetic Susceptibility and Refinement of Crystal Field Parameters

Experimental data on the initial magnetic susceptibility $\chi_M(T)$, which is defined only by the state of rare-earth ions in the crystal field and bilinear interactions, are of interest because they may be used in determining the crystal field. The magnetic susceptibility of investigated YbPO$_4$ and YbVO$_4$ zircons is anisotropic along and perpendicular to the tetragonal axis (Fig. 1); how-

The crystal field parameters of rare-earth phosphate $YbPO_4$ and vanadate $YbVO_4$, obtained using the results of optical (*L*), neutron (*N*), and magnetic (*M*) measurements, and Stark splitting of the ground multiplet Yb^{3+} in $YbXO_4$ for the given crystal field parameters

RXO_4	Method	Ref.	B_2^0, K	B_4^0, K	B_4^4, K	B_6^0, K	B_6^4, K	E_1, cm^{-1}	E_2, cm^{-1}	E_3, cm^{-1}	E_4, cm^{-1}
$Yb : YPO_4$	*L*(1)	[9]	244	13	975	–62	–4	0	92	307	315
$Yb : LuPO_4$	*L*(2)	[9]	184	3	915	–63	–16	0	101	281	288
$YbPO_4$	*M*		196	1	980	–62	–39	0	99	290	305
$YbVO_4$	*N*	[10]	–154	77	1062	–84	–156	0	58	280	353
$YbVO_4$	*M*		–106	92	950	–103	–318	0	58	277	353

ever, it remains isotropic in the basal plane, as one would expect for tetragonal symmetry. For the entire family of rare-earth zircons, the direction of the axis of easy magnetization at high temperatures is defined by the signs of the second-order crystal field parameter B_2^0 and Stevens parameter α_J. In accordance with the variation of α_J over the series of rare-earth elements for phosphates with ions from Tb to Ho, the easy axis is the tetragonal axis which changes for the basal plane for phosphates with Tm and Yb. Exactly the opposite is the situation in the case of rare-earth vanadates in which the parameter B_2^0 changes sign. The anisotropy of magnetic susceptibility $\Delta\chi/\bar{\chi} = \Delta\chi^{-1}/\bar{\chi}^{-1}$ at high temperature in phosphate is twice that in vanadate, in accordance with the value of the parameter B_2^0.

One can see in Fig. 1 that, for both zircons at a temperature above 150 K, the correlation $\chi_M^{-1}(T)$ obeys the Curie–Weiss law; however, the slopes of linear portions for the *c* and *a* axes differ somewhat from one another. This means that the crystal field effects at temperatures of about 300 K still play an important part, and the effective magnetic moment does not reach its value $\mu_{eff} = 4.54\mu_B$ for a free ion of Yb^{3+}. As the temperature decreases, the susceptibility along these two axes increases monotonically, so that the anisotropy χ_M remains moderate. The magnetic susceptibility along the hard magnetization axis exhibits characteristic kinks at 70 and 20 K, which are associated with the variation of the population of excited levels when the temperature decreases.

It is well known that, in the case of tetragonal symmetry, it is difficult to unambiguously determine a complete set of parameters of the crystal field on the basis of the magnetic properties alone. As a rule, more than one set of parameters may be identified providing a plausible description of magnetic properties within the experimental error. Therefore, in determining the crystal field in the case of tetragonal symmetry, it is absolutely necessary to simultaneously employ spectroscopic data, the use of which often results in rendering the problem uniquely determined. In addition, the single-ion deformation susceptibility and single-ion quadrupole-field susceptibility, which define the behavior of elastic constants and magnetostriction, enable one in a number of cases to confirm or define more exactly the crystal field.

The crystal field parameters B_n^m for a Yb^{3+} ion in the phosphate matrix are known only for alloyed compounds of $Yb{:}LuPO_4$ and $Yb{:}YPO_4$ [9]; they differ considerably by the parameter B_2^0 and by the position of the first excited doublet (see table). The wave functions and Stark splitting of the ground multiplet, obtained in calculations involving these parameters, differ appreciably from spectroscopic data for a concentrated compound of $YbPO_4$ (0, 99, 258 ± 12, and 347 ± 16 cm^{-1}) [3]. A detailed investigation of the electronic structure of a rare-earth ion by the method of inelastic neutron scattering [10] was performed for $YbVO_4$, and the crystal field parameters were determined, as well as the wave functions and values of the components of the *g* tensor, which proved to be comparable in magnitude with those obtained previously for a dilute compound [9]. These parameters, however, fail to provide a good description of the experimentally obtained correlation $\chi_M^{-1}(T)$ along the hard magnetization axis (see the inset in Fig. 1).

The crystal field parameters for Yb zircons were refined on the basis of our data on the initial magnetic susceptibility χ_M and all of the available spectroscopic data using the FLEX optimization procedure [11]. The space of the parameters B_n^m was used to search for all solutions of the inverse problem for which the residual sum on the array of experimental data was minimal and did not exceed some value defined by the accuracy of the employed data. Used as the input set of parameters in the case of Yb phosphate was a set of crystal field parameters for $Yb{:}LuPO_4$ (see table), for which the calculated temperature dependence $\chi_M(T)$ along the [001] and [100] axes is closer to that obtained experimentally. Because bilinear interactions with the parameter $\theta^* \approx 0.1$ K obtained by estimation using De Gennes' formula do not lead to any appreciable modification of the magnetic properties, the parameter θ^* was then

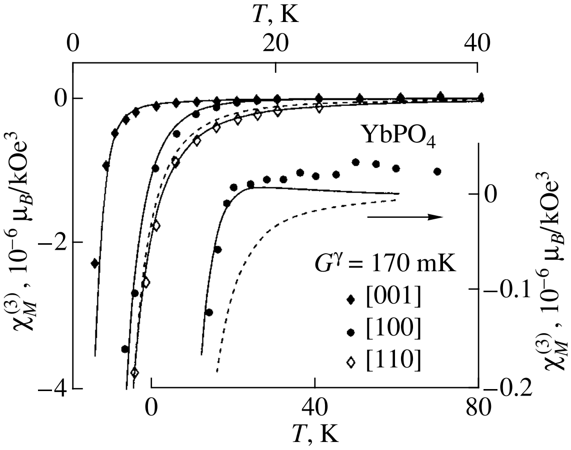


Fig. 2. The temperature dependence of the third-order magnetic susceptibility of YbPO$_4$ crystal along the [100], [110], and [001] axes. The solid (dashed) curves indicate the dependences calculated in view of (disregarding) quadrupole interactions for a set of crystal field parameters M (see table) and G^γ = 170 mK. The inset gives the $\chi_M^{(3)}(T)$ dependences for $H \parallel [100]$ on a larger scale.

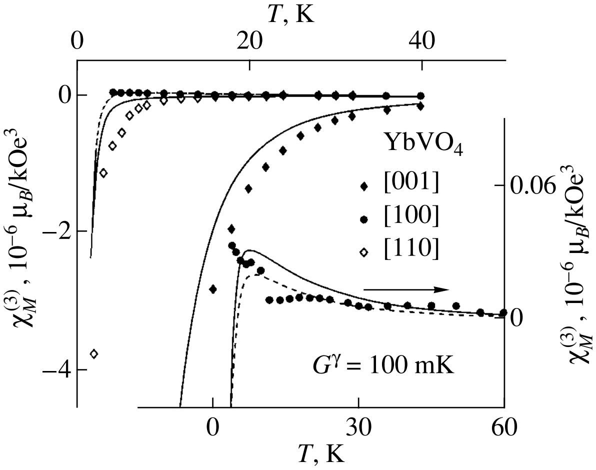


Fig. 3. The temperature dependence of the third-order magnetic susceptibility of YbVO$_4$ crystal along the [100], [110], and [001] axes. The solid (dashed) curves indicate the dependences calculated in view of (disregarding) quadrupole interactions for a set of crystal field parameters M (see table) and G^γ = 100 mK. The inset gives the $\chi_M^{(3)}(T)$ dependences for $H \parallel [100]$ on a larger scale.

assumed to be zero. An optimal description of the experimental data was obtained using the parameters M given in the table. The set of crystal field parameters determined by us is in good agreement with the parameters obtained in an analogous array of experimental data for other rare-earth phosphates with ions from Tb to Tm [2].

For YbVO$_4$, the parameters obtained by us (the set of M, see table) provide a description of the magnetic susceptibility that is closer to experiment (compare the solid and broken lines in the inset in Fig. 1), and the multiplet splitting agrees well with the data of neutron measurements [10]. Note that, for both Yb compounds, the procedure stably produces a single minimum for different initial conditions of search from the space region ($\bar{B}_n^m \pm \Delta B_n^m$), where $\bar{B}_n^m$ denotes the average (for a series of rare-earth phosphates or vanadates) values of crystal field parameters, and the $\Delta B_n^m / \bar{B}_n^m$ ratio does not exceed 30%.

One can see that the most significant difference in the crystal fields for two Yb zircons consists in that the parameter B_2^0 for vanadate is negative and almost twice less in magnitude than the analogous parameter in phosphate. Note further the somewhat underestimated value of the fourth-order parameter B_4^0 for Yb phosphate and some difference of the parameters B_4^0 and B_6^0 for Yb vanadate from the average values over the series of rare-earth zircons.

4.2. Third-Order Magnetic Susceptibility

The third-order magnetic susceptibility $\chi_M^{(3)}$ along the principal symmetric directions is given in Figs. 2 and 3 for Yb vanadate and Yb phosphate, respectively. The behavior of $\chi_M^{(3)}$ along the [100] axis is shown in more detail in the insets in these figures. One can see that, along the [001] and [110] axes, the susceptibility $\chi_M^{(3)}$ for both crystals is negative and well described by the curves calculated using the crystal field parameters determined above without any fitting parameters. Note the somewhat poorer description of susceptibility along the [110] and [001] axes for vanadate. The greater magnitude of $\chi_M^{(3)}$ along the [110] axis in experiment may be attributed, for example, to the effects of misorientation (the $\chi_M^{(3)}$ contribution from the [001] axis), because the anisotropy of $\chi_M^{(3)}$ at low temperatures in vanadate is appreciably higher than in phosphate.

The experimentally obtained dependences $\chi_M^{(3)}(T)$ along the [100] axis are nonmonotonic and of positive sign in some temperature range. For their description, one must take into account the contribution made by γ-symmetric quadrupole interactions, which give a positive addition to the negative susceptibility $\chi_M^{(3)}$ defined by the crystal field (compare the solid and broken lines in the figures and insets). This addition is comparable with $\chi_0^{(3)}$ because of the high values of the quadrupole-field, $\chi_\gamma^{(2)}$, and deformation, χ_γ, susceptibilities corresponding to γ symmetry (see Eq. (5)). For Yb vanadate at $H \parallel [100]$, the susceptibility $\chi_0^{(3)}$ defined by the crystal field is positive even at low temperatures, and the contribution to $\chi_M^{(3)}$ by quadrupole interactions is much lower than in the case of Yb phosphate.

4.3. Magnetostriction

As was already noted, in order to investigate the normal deformation modes $\varepsilon^{\alpha 1}$, $\varepsilon^{\alpha 2}$, ε^{γ}, and ε^{δ}, the magnetostrictive deformation was simultaneously measured in parallel with and perpendicular to the external magnetic field applied along one of the symmetry axes [001], [100], and [110]. Experiment reveals that the sample length variations λ_{ij} are quadratic in H up to 140 kOe at $T > 30$ K; at lower temperatures, this field range is reduced.

Fully symmetric α *mode.* For a magnetic field oriented along the [001] axis, the length variations λ_{cc} and λ_{ac} define the volume, $\varepsilon^{\alpha 1} = (\lambda_{cc} + 2\lambda_{ac})/\sqrt{3}$, and tetragonal, $\varepsilon^{\alpha 2} = \sqrt{2/3}\,(\lambda_{cc} - \lambda_{ac})$, modes. The temperature dependences of the initial slopes of magnetostriction curves, $(\lambda_{cc} + 2\lambda_{ac})/H^2$ and $(\lambda_{cc} - \lambda_{ac})/H^2$, for Yb phosphate and Yb vanadate are given in Fig. 4. Because the value of magnetostrictive deformation remains low even at low temperatures, random errors of measurement lead to some scatter of experimental data. In addition, the compensation of the strain gage magnetoresistance starts acquiring a considerable importance. The optimal compensation is attained when measurements are performed using a standard strain gage glued onto a diamagnetic analog of the sample being investigated, for example, YPO_4 for $YbPO_4$. In other cases, the incompletely compensated magnetoresistance of the strain gage at $T < 40$ K leads to a systematic error of measurement, as is seen in Fig. 4 for $YbVO_4$. We believe that this is why the values of the factors $A^{\alpha 1}$ and $A^{\alpha 2}$ for Yb vanadate are somewhat underestimated (see below).

For $YbPO_4$, both modes exhibit similar temperature dependences, for which the deformation changes its sign in the neighborhood of 50 K. In the case of Yb vanadate, the volume and tetragonal modes retain their sign in the entire investigated temperature range. In the absence of bilinear interactions, the calculated dependences of initial slopes are defined by the expression

$$\sqrt{3}\frac{A^{\alpha 1}H^2}{|\varepsilon^{\alpha 1}|} = \sqrt{3/2}\frac{A^{\alpha 2}H^2}{|\varepsilon^{\alpha 2}|} = \frac{1}{\chi_\alpha^{(2)}}(1 - G^\alpha \chi_\alpha), \qquad (9)$$

in which the constants $A^{\alpha 1}$ and $A^{\alpha 2}$ depend on the background elastic constants $C_0^{\alpha 1}$, $C_0^{\alpha 2}$, and $C_0^{\alpha 12}$ and on the magnetoelastic coefficients $B^{\alpha 1}$ and $B^{\alpha 2}$. The calculation results demonstrate that the inclusion of G^α in Eq. (9) in the case of Yb zircons is of no importance. A comparison with experiment gives values of the factors $A^{\alpha 1} = -1.1 \times 10^{-4}$ and $A^{\alpha 2} = 1.6 \times 10^{-4}$ for phosphate and $A^{\alpha 1} = -0.4 \times 10^{-4}$ and $A^{\alpha 2} = 0.9 \times 10^{-4}$ for vanadate. Moreover, in accordance with experiment, the susceptibility $\chi_\alpha^{(2)}$ is positive at high temperatures and changes sign at a temperature below 50 K for phosphate and remains positive for vanadate.

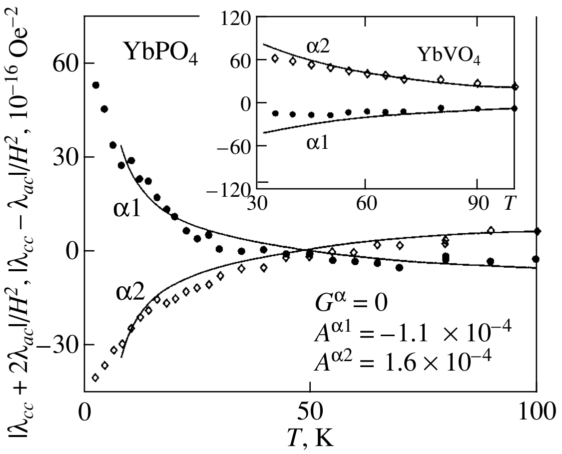


Fig. 4. The temperature dependence of α-symmetric magnetostriction of $YbPO_4$ and $YbVO_4$ crystals (inset) for a magnetic field along the [001] axis experimentally obtained (points) and calculated using the given parameters (curves).

In order to determine the magnetoelastic coefficients $B^{\alpha 1}$ and $B^{\alpha 2}$ for $YbPO_4$, use was made of values of symmetrized elastic constants $C_0^{\alpha 1} = 17.3$, $C_0^{\alpha 2} = 14.4$, and $C_0^{\alpha 12} = 0.55$ (in 10^5 K), which were recalculated proceeding from the experimental data of Nipko *et al.* [3] (because of the absence of experimental data for c_{13}, it was assumed that $c_{13} \approx c_{12}$). In this manner, the values of $B^{\alpha 1}/\alpha_J = -5.7 \times 10^3$ K and $B^{\alpha 2}/\alpha_J = 7.1 \times 10^3$ K and the contribution $G_{ME}^\alpha = 56$ mK of magnetoelastic interaction were found. Because the quadrupole component $\langle O_2^0 \rangle$ is other than zero even in the absence of an external field, in the calculation of the initial susceptibility, the α-symmetric magnetoelastic interactions must be included during diagonalization of the Hamiltonian $H_{CF} + H_{QT}$. However, the calculation results have demonstrated that the value of $G^\alpha \sim 50$ mK does not cause a variation of either single-ion susceptibilities proper or their renormalization; therefore, in what followed, the α-symmetric quadrupole interactions were not taken into account ($G^\alpha = 0$ K). The effect of the parameter $G^\alpha \sim 50$ mK on $\chi_M^{(3)}$ along the [100] axis is likewise negligible (Fig. 2). In view of the indeterminacy of the coefficients $A^{\alpha 1}$ and $A^{\alpha 2}$ and elastic con-

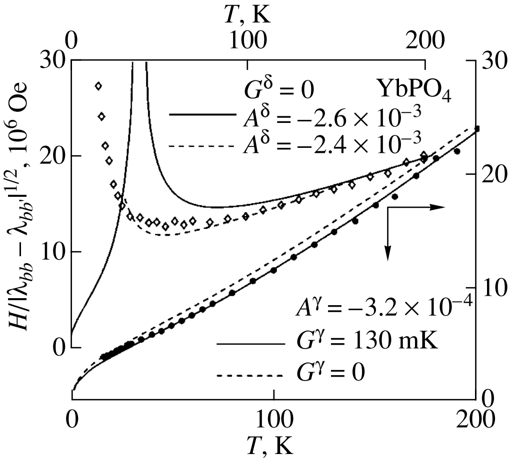


Fig. 5. The temperature dependence of inverse γ and δ symmetric magnetostriction of YbPO$_4$ crystal for a magnetic field along the [100] and [110] axes experimentally obtained (points) and calculated using the given parameters (curves). The dashed curves for the δ mode indicate the dependences calculated using the alternative parameters of the crystal field.

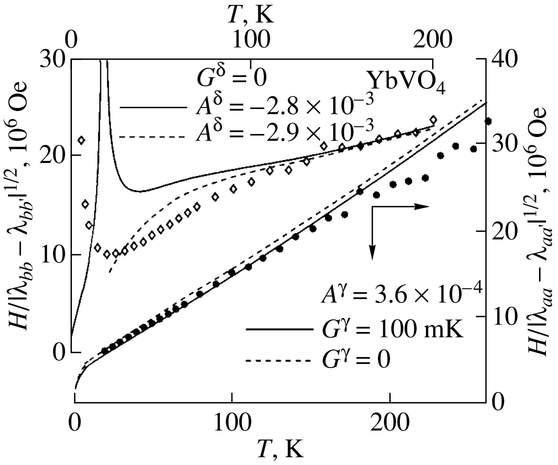


Fig. 6. The temperature dependence of inverse magnetostriction, which correspond to γ and δ symmetry of YbVO$_4$ crystal for a magnetic field along the [100] and [110] axes experimentally obtained (points) and calculated using the given parameters (curves). The dashed curves for the δ mode indicate the dependences calculated using the alternative parameters of the crystal field.

stants $C^{\alpha i}$, no estimates of the magnetoelastic coefficients $B^{\alpha 1}$ and $B^{\alpha 2}$ are given here.

Low-symmetry γ and δ modes. The modes of orthorhombic symmetry for two Yb zircons exhibit common regularities at low temperatures. In the entire investigated temperature range, the quadrupole-field susceptibility $\chi_\gamma^{(2)}$ is positive both for YbPO$_4$ and for YbVO$_4$; this provides for the positive values of magnetostriction $\varepsilon^\gamma = (\lambda_{aa} - \lambda_{aa'})/\sqrt{2}$. The calculation results demonstrate that, at temperatures below ~15 K, the range of fields in which the magnetostriction is quadratic in the field becomes narrow (less than 10 kOe); as a result, it is difficult to experimentally determine the initial derivative ε^γ/H^2 in this range. A comparison of the experimentally obtained and calculated temperature dependences of inverse magnetostriction $H/|\varepsilon^\gamma|^{1/2}$ (Figs. 5 and 6) gives values of the factors $A^\gamma = 3.2 \times 10^{-4}$ and $A^\gamma = 3.6 \times 10^{-4}$ in phosphate and vanadate, respectively. This gives the magnetoelastic coefficients $B^\gamma/\alpha_J = 13.9 \times 10^3$ K and $B^\gamma/\alpha_J = 14.2 \times 10^3$ for the background elastic constants $C_0^\gamma = 1.36 \times 10^6$ K [3] and $C_0^\gamma = 1.24 \times 10^6$ K (the data for TmVO$_4$ were borrowed from Melcher [12]). In this case, contributions of the magnetoelastic coupling to the quadrupole constant $G_{ME}^\gamma \sim 142$ mK and $G_{ME}^\gamma \sim 163$ mK are obtained for Yb phosphate and Yb vanadate, respectively. The inclusion of pairwise quadrupole interactions with total constants $G^\gamma = 130$ mK and $G^\gamma = 100$ mK leads to an increase in magnetostriction and to an almost parallel shift of the curve $H/(\varepsilon^\gamma)^{1/2}(T)$ in accordance with experiment. The value of $G^\gamma \approx 130$ mK for YbPO$_4$ is close to that derived from the third-order susceptibility.

A characteristic feature of the magnetostriction $\varepsilon^\delta = (\lambda_{bb} - \lambda_{bb'})/\sqrt{2}$ for both Yb compounds is the change of sign of deformation at low temperatures (Figs. 5 and 6). The δ-symmetric magnetostriction, as well as the third-order susceptibility, is not highly dependent on quadrupole interactions, because the susceptibilities $\chi_\delta^{(2)}$ and χ_δ are not high. The calculations involving the crystal field parameters which describe well all of the properties discussed heretofore produce an appreciably higher characteristic temperature of change of sign than that observed experimentally. This difference cannot be explained by the possible contribution by the γ mode due to misorientation of the field and strain gage relative to the [110] axis, experimental systematic errors of measurement (which are small in this temperature range), and other factors.

An adequate description of δ-symmetric magnetostriction for both Yb zircons can only be obtained using another spectrum for Yb^{3+} ion, i.e., another set of crystal field parameters that would differ considerably from the set used to describe the neutron experiments and initial magnetic susceptibility. For both Yb compounds, the anomaly of δ-symmetric magnetostriction proved to be most sensitive to the values of diagonal parameters of the fourth, B_4^0, and sixth, B_6^0, orders. The search for an alternative set, which allows a variation of the parameters B_4^0 and B_6^0 within ±30% of their average values, leads to the parameters $\bar{B}_4^0 = B_4^0 + \Delta B_4^0$ and

$\bar{B}_6^0 = B_6^0 + \Delta B_6^0$, which differ from the parameters of sets M (see table) by the corrections ΔB_4^0 and ΔB_6^0. The corrections amount to $|\Delta B_4^0| \approx |\Delta B_6^0| \approx 20$ K for both Yb compounds. The examples of calculation of the curves $H/(\varepsilon^\delta)^{1/2}(T)$ using alternative sets of parameters are given in the figures by dashed lines. These calculations produce values of the factors $A^\delta = -2.4 \times 10^{-3}$ and $A^\delta = -2.9 \times 10^{-3}$ and of the magnetoelastic coefficients $B^\delta/\alpha_J = -16.2 \times 10^3$ K and $B^\delta/\alpha_J = -19.2 \times 10^3$ K (for the background elastic constant $C_0^\delta = 2.1 \times 10^5$ K) in $YbPO_4$ and $YbVO_4$, which do not differ strongly from the average values for the series of rare-earth zircons.

The experimental data for the δ mode indicate that the second doublet must be located somewhat lower than it follows from the results of neutron experiments, which could be the case if the crystal field parameters were temperature-dependent. One of the mechanisms responsible for the temperature dependence of the crystal field parameters may be the α-symmetric magnetoelastic interaction. For example, in the case of $DyPO_4$, a variation of the quadrupole moment $\langle O_2^0 \rangle$ with temperature brings about a variation of the effective second-order parameter of the crystal field $B_2^0 = \bar{B}_2^0 - G^\alpha \langle O_2^0 \rangle / \alpha_J$ and a shift of the second doublet by approximately 20 K [13]. For both Yb zircons, the multipole moments $\langle O_4^0 \rangle$ and $\langle O_6^0 \rangle$ are not low and have clearly defined singularities in the desired temperature range; these singularities may, in principle, bring about a desired variation of the parameters (positive additions for phosphate and negative additions for vanadate).

4.4. Elastic Constants

The results of investigation of magnetic susceptibility and magnetostriction were supplemented with the analysis of the experimental results of Nipko *et al.* [3], who used Brillouin light scattering to measure the temperature dependences of all elastic constants (except for c_{13}) for $YbPO_4$ and compared the results with the data of Armbruster *et al.* [14] for $LuPO_4$. In the temperature range from 300 to 15 K, all elastic constants for $YbPO_4$ (with the exception of C^γ) differ by not more than 2–3%, and their behavior differs little from the behavior of elastic constants for a nonmagnetic analog of $LuPO_4$. The constant C^γ exhibits a softening by approximately 20% (Fig. 7). Nipko *et al.* [3] describe the $C^\gamma(T)$ dependence within pseudospin formalism on the basis of two lower doublets. Note, however, that Nipko *et al.* [3] give inaccurate values of the parameters.

A more systematic description may be obtained within the formalism of deformation susceptibilities $\chi_\mu(T)$. The calculation results demonstrate that, for $YbPO_4$, the variation of χ_γ in the temperature range from 0 to 300 K is one and a half to two orders of magnitude greater than the variations of χ_δ and χ_ε. The variation of χ_α has an intermediate value and is approximately five times less than that in the case of χ_γ. This correlates fully with the experimental results of Nipko *et al.* [3], which are indicative of the presence of the effect of substantial softening only for the γ-symmetric elastic constant.

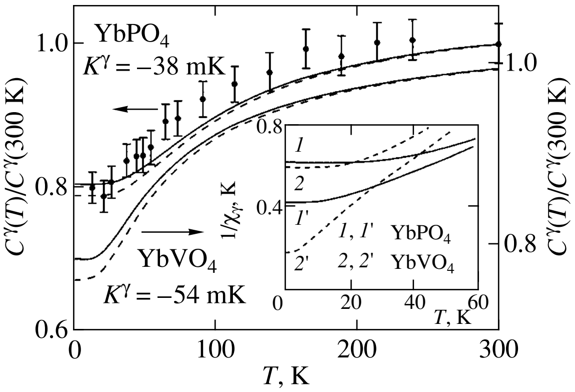


Fig. 7. The temperature dependence of the elastic interaction constant C^γ of $YbPO_4$ and $YbVO_4$ crystals experimentally obtained (points, data borrowed from [3]) and calculated using the given parameters (curves). The inset shows the dependences $1/\chi_\gamma(T)$ at $H = 0$ (curves *1* and *2*) and $H = 300$ kOe along the [110] axis (curves *1'* and *2'*).

A plausible description of the curve $C^\gamma(T)/C^\gamma(300\text{ K})$ within the experimental error on the basis of the calculated dependence $\chi_\gamma(T)$ may be obtained using different sets of the parameters B^γ and K^γ (see Eq. (6)). The choice of the optimal set of coefficients B^γ and K^γ was performed employing additional data on the γ-symmetric magnetostriction (which is quite sensitive to the coefficient B^γ. It was assumed in the calculations that the value of C_0^γ defined by phonons increased by 2% from 300 to 50 K and did not vary thereafter. As a result, the coefficients $B^\gamma/\alpha_J = 14.4 \times 10^3$ K and $K^\gamma = -38$ mK were obtained, satisfying the ratio $K^\gamma/G_{ME}^\mu \sim -0.25$ and simultaneously describing the magnetostriction and the temperature dependence of the elastic constant within the experimental error (solid curve in Fig. 7).

The optimal agreement between the experimentally obtained and calculated dependences $C^\gamma(T)$ within the crystal field parameters M (see table) may be attained for values of the coefficient $B^\gamma/\alpha_J = 9.77 \times 10^3$ K and of the pairwise quadrupole coupling constant $K^\gamma = 347$ mK, which clearly fail to satisfy the ratio $K^\gamma/G_{ME}^\mu = -1/3$. Note that the alternative set of crystal field parameters, which describes well the δ-symmetric magnetostriction in phosphate, also provides a better description of the contribution by the magnetoelastic interaction to the elastic constant $C^\gamma(T)$.

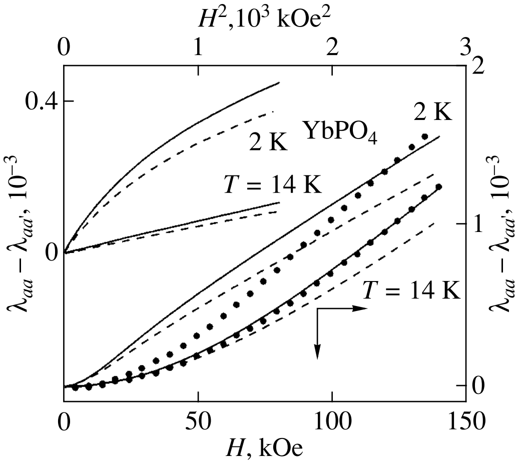


Fig. 8. The field dependence of the magnetostriction, corresponding to γ symmetry $\sqrt{2}\,\varepsilon^{\gamma} = (\lambda_{aa} - \lambda_{aa'})$ of YbPO$_4$ crystal for a magnetic field along the [100] axis at $T = 2$ K and $T = 14$ K. The inset gives the calculated dependences $\sqrt{2}\,\varepsilon^{\gamma}(H^2)$.

No experimental data on elastic constants are available for YbVO$_4$; however, the results of deformation susceptibility calculations indicate that for vanadate as well one must expect a significant softening of the constant C^{γ}. Figure 7 gives the calculated temperature dependence of the elastic constant $C^{\gamma}(T)/C^{\gamma}$ (300 K) with the parameters $B^{\gamma}/\alpha_J = 14.2 \times 10^3$ K and $K^{\gamma} = -54$ mK obtained during analysis of the magnetostriction data.

4.5. Magnetic and Magnetoelastic Properties in Strong Fields

Investigations of the magnetic and magnetoelastic properties of Yb zircons in weak fields reveal the observed quadrupole effects for $H \parallel [100]$, namely, significant contributions by quadrupole interactions to the third-order susceptibility $\chi_M^{(3)}$, magnetostriction, and the γ-symmetric elastic constant. These interactions, however, are not strong enough to lead to quadrupole ordering which takes place if the condition $1/\chi_{\gamma} \le G^{\gamma}$ is valid, i.e., if the respective deformation susceptibility reaches a critical value. One can see in the inset of Fig. 7 that, for YbPO$_4$ and YbVO$_4$, the inverse susceptibility $1/\chi_{\gamma}$ at low temperatures is approximately 0.6 K, i.e., $1/\chi_{\gamma}$ is several times the value of the total quadrupole coupling constant $G^{\gamma} \sim 150$ mK for these zircons. Therefore, YbPO$_4$ and YbVO$_4$ remain in the above-critical state in the absence of external magnetic fields.

The external magnetic field changes the spectrum and wave functions of a Yb^{3+} ion and brings the system closer to the critical state; in so doing, the situation depends strongly on the field orientation. The magnetic field $H \parallel [100]$, which induces the deformation of the same symmetry as the expected spontaneous one, causes the emergence in the Hamiltonian H_{QT} of γ-symmetric quadrupole terms. These terms bring about a modification of the electronic structure of rare-earth ions and additional contributions to the magnetic and magnetoelastic properties. Figure 8 gives, by way of example, the experimentally obtained and calculated isotherms of γ-symmetric magnetostriction at $T = 2$ K and 14 K for YbPO$_4$. A comparison of the curves calculated in view of (solid curves) and disregarding (broken curves) quadrupole interactions reveals that the contribution by quadrupole interactions with the experimentally obtained constant $G^{\gamma} = 130$ mK at low temperatures in the field of 140 kOe reaches 15–20%. The experimental data are described fairly well at temperatures above 10–15 K; the reasons for divergence at lower temperatures (2 K) still remain to be understood. Also given in the inset are the calculated dependences $\sqrt{2}\,\varepsilon^{\gamma}(H^2)$, which demonstrate that the range of quadraticity of magnetostriction is reduced considerably at helium temperatures. Similarly, for the magnetization curves $M(H)$, quadrupole interactions at helium temperatures in a field of 140 kOe produce an additional contribution of approximately 10–15% in both Yb compounds.

The situation is more interesting as regards the field orientation $H \parallel [110]$, because the induced and expected spontaneous deformations are characterized by different symmetries. The electronic configuration of a Yb^{3+} ion in the field varies in such a manner that χ_{γ} increases and the system approaches the critical state. For example, in a field of 300 kOe, as is seen in the inset of Fig. 7 (curves *1*' and *2*'), the susceptibility χ_{γ} at helium temperatures increases; it increases especially strongly in the case of YbVO$_4$. With such a tendency, χ_{γ} in a fairly strong field will reach the critical value of $1/G^{\gamma}$, and a γ-symmetric stimulated phase transition will be realized in the system. However, the results of calculations involving real parameters of the crystal field and quadrupole interactions demonstrate that the value of the critical field H_c in which the condition $1/\chi_{\gamma}(H_c) \le G^{\gamma}$ is valid for the first time is fairly high ($H_c > 1000$ kOe for YbPO$_4$ and $H_c > 500$ kOe for YbVO$_4$) and hardly accessible for experimental investigation. The situation described is similar to that observed and experimentally investigated by different methods in a virtual Jahn–Teller compound TmPO$_4$ with respect to δ-symmetric stimulated quadrupole ordering [15].

5. SINGLE-ION MAGNETOELASTIC AND PAIRWISE QUADRUPOLE PARAMETERS IN RARE-EARTH ZIRCONS

The investigation results enable one to determine all magnetoelastic parameters for Yb zircons and, thereby, obtain a full pattern for the entire family of rare-earth vanadates and phosphates. Experiments in magnetostriction directly produce the factors $A^{\mu} = B^{\mu}/C_0^{\mu}$; in order to determine the magnetoelastic coefficients B^{μ}, one needs to know the background elastic constants C_0^{μ}. It is only for some zircons that experimental data on these constants are available in the literature, obtained by different methods and revealing irregular variations and/or considerable scatter over the series of rare-earth zircons. Therefore, it makes sense to perform a comparison, for different rare-earth zircons, of the factors A^{μ} determined directly from experiment and to recalculate the factors B^{μ} for background elastic constants averaged over the series of rare-earth zircons.

For the fully symmetric α1 and α2 modes in $YbPO_4$, values of the coefficients $A^{\alpha 1} = -1.1 \times 10^{-4}$ and $A^{\alpha 2} = 1.6 \times 10^{-4}$ were obtained, comparable in magnitude and exhibiting opposite signs. For $YbVO_4$, the coefficients $A^{\alpha 1}$ and $A^{\alpha 2}$ are of the same signs and comparable in magnitude. The difference between the coefficients $A^{\alpha i}$ for $YbPO_4$ and neighboring $TmPO_4$ ($A^{\alpha 1} = -3.2 \times 10^{-5}$ and $A^{\alpha 2} = 4.7 \times 10^{-5}$ [16]) is due to the difference between the Stevens parameters $\alpha_J = 2/63$ for a Yb ion and $\alpha_J = 1/99$ for a Tm ion. Therefore, in what follows, the comparison is performed, over the family of rare-earth phosphates, of the coefficients $\bar{A}^{\mu} = A^{\mu}/\alpha_J$ normalized to the Stevens parameter α_J. The fairly high values of fully symmetric coefficients $A^{\alpha 1}$ and $A^{\alpha 2}$ represent a characteristic feature of rare-earth zircons and may cause unusual effects, for example, a variation in the temperature dependence of the spectrum in the magnetically disordered or quadrupole-disordered phases.

For the orthorhombic γ mode, the constants $\bar{A}^{\gamma} = B^{\gamma}/\alpha_J C_0^{\gamma} = 10.2 \times 10^{-3}$ and $\bar{A}^{\gamma} = 11.4 \times 10^{-3}$ were obtained for Yb phosphate and Yb vanadate, respectively. These values are comparable with the values of $\bar{A}^{\gamma} \sim 10 \times 10^{-3}$ for a series of rare-earth phosphates [2] and $\bar{A}^{\gamma} \sim 12 \times 10^{-3}$ for a series of rare-earth vanadates [17, 18], obtained as a result of analysis of both magnetostriction and elastic constants. For the rhombic δ mode, the factors $\bar{A}^{\delta} = B^{\delta}/\alpha_J C_0^{\delta} = -7.5 \times 10^{-2}$ and $\bar{A}^{\delta} = -9.2 \times 10^{-2}$ were obtained for $YbPO_4$ and $YbVO_4$, respectively. Note, however, that, because of problems involved in describing the δ-symmetric magnetostriction, the indeterminacy of values of the coefficients $\bar{A}^{\delta}$ for Yb compounds may exceed somewhat that for other zircons. The contribution by the magnetoelastic coupling to the total quadrupole constant G_{ME}^{μ} for both orthorhombic modes in ytterbium zircons proves to be an order of magnitude higher than, for example, that in the $TmPO_4$ compound [2] because of the difference between the Stevens coefficients for Yb and Tm ions.

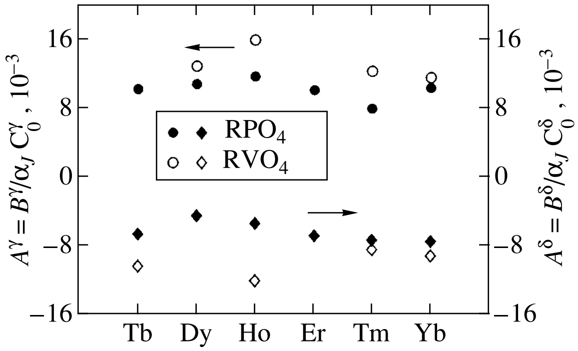


Fig. 9. The variation of the coefficients $\bar{A}^{\gamma} = B^{\gamma}/\alpha_J C_0^{\gamma}$ (circles) and $\bar{A}^{\delta} = B^{\delta}/\alpha_J C_0^{\delta}$ (rhombs) for weakly symmetric γ and δ modes over the series of rare-earth phosphates RPO_4 and vanadates RVO_4.

Figure 9 gives the constants $\bar{A}^{\delta}$ and $\bar{A}^{\gamma}$ obtained previously for a series of rare-earth phosphates [2] and rare-earth vanadates [18] supplemented by new data for Yb zircons. One can see in the figure that all of the obtained values are comparable in magnitude for rare-earth zircons, and those for the δ mode are almost an order of magnitude higher because of the much lower value of the background constant $C_0^{\delta} \approx C_0^{\gamma}/5$. Moreover, values of the magnetoelastic coefficients B^{γ}/α_J and B^{δ}/α_J, recalculated with average elastic constants C_0^{γ} and C_0^{δ}, also prove to be comparable over the series of rare-earth zircons and with one another.

Of interest is the ratio K^{μ}/G_{ME}^{μ} for the family of rare-earth zircons, because this ratio contains information about the importance of acoustic and optical phonons in the mechanism of quadrupole interaction between rare-earth ions. In the absence of contribution by optical phonons, the value of $K^{\mu}/G_{ME}^{\mu} = -1/3$ was obtained for the simplest model of a two-level system. This ratio is valid with sufficient accuracy for the soft δ mode in $TmVO_4$, $TbVO_4$, and $TbPO_4$ compounds exhibiting structural phase transitions and in $HoVO_4$ in the absence of structural transition. In the case of $DyVO_4$ which experienced a structural phase transition, for the soft γ mode this ratio ranges, according to different data, from –1.2 to +0.40. For $TmPO_4$, a simultaneous analysis of experimental data on magnetostriction and elastic constant leads to the value of $K^{\delta}/G_{ME}^{\delta} \approx -1/6$. A similar analysis for neighboring (in the series of rare-earth phosphates) $YbPO_4$ gives the ratio $K^{\gamma}/G_{ME}^{\gamma} = -0.25$; for $YbVO_4$, the experimental data are described fairly well at $K^{\gamma}/G_{ME}^{\gamma} \approx -1/3$.

The magnetic and magnetoelastic properties of Yb zircons at $H \parallel [100]$ exhibit a number of observed quadrupole effects both in weak and in strong magnetic fields. Quadrupole interactions are responsible for a significant contribution to the third-order susceptibility, magnetization, magnetostriction, and γ-symmetric elastic constant. These interactions, however, are not strong enough to bring about quadrupole ordering in either $YbPO_4$ or $YbVO_4$, because the respective deformation susceptibility χ_γ is several times less than the critical value $1/G^\gamma$. Experiment and theory reveal that, in both Yb compounds, the contribution by quadrupole interactions with the constant $G^\gamma \approx 150$ mK to the magnetostriction and magnetization curves for $H \parallel [100]$ at low temperatures in a field of 140 kOe reaches 15–20%. For the field orientation of $H \parallel [110]$, the deformation susceptibility χ_γ increases, so that in a fairly strong field in $YbPO_4$ and $YbVO_4$ crystals one can expect a γ-symmetric stimulated phase transition.

ACKNOWLEDGMENTS

This study received partial financial support from the Russian Foundation for Basic Research (project 00-02-17756).

REFERENCES

1. G. A. Gehring and K. A. Gehring, Rep. Prog. Phys. **38**, 1 (1975).
2. P. Morin and Z. Kazei, J. Phys.: Condens. Matter **11**, 1289 (1999).
3. J. Nipko, M. Grimsditch, C.-K. Loong, *et al.*, Phys. Rev. B **53**, 2286 (1996).
4. P. C. Becker, G. M. Williams, N. Edelstein, *et al.*, Phys. Rev. B **45**, 5027 (1992).
5. J. A. Hodges, J. Phys. (Paris) **44**, 833 (1983).
6. M. J. Prandolini, W. D. Hutchison, D. H. Chaplin, *et al.*, J. Magn. Magn. Mater. **177–181**, 1054 (1998).
7. P. Morin and D. Schmitt, Phys. Rev. B **23**, 5936 (1981).
8. P. Morin, J. Rouchy, and D. Schmitt, Phys. Rev. B **37**, 5401 (1988).
9. P. C. Becker, T. Hayhurst, G. Shalimoff, *et al.*, J. Chem. Phys. **81**, 2872 (1984).
10. J. Nipko, C.-K. Loong, S. Kern, *et al.*, J. Alloys Compd. **250**, 569 (1997).
11. E. B. Dovidovich, N. P. Kolmakova, and A. A. Mukhin, Preprint No. 228, IOFAN (General Physics Inst., Academy of Sciences of USSR, 1984).
12. R. L. Melcher, in *Physical Acoustics: Principles and Methods*, Ed. by W. P. Mason and R. N. Thurston (Academic, New York, 1976), Vol. 12.
13. Z. A. Kazei and V. V. Snegirev, Pis'ma Zh. Éksp. Teor. Fiz. **73**, 95 (2001) [JETP Lett. **73**, 90 (2001)].
14. A. Armbruster, R. Thoma, and H. Wehrle, Phys. Status Solidi A **24**, K71 (1974).
15. P. Morin and Z. Kazei, Phys. Rev. B **55**, 8887 (1997).
16. P. Morin, J. Rouchy, and Z. Kazei, J. Phys.: Condens. Matter **8**, 7967 (1996).
17. P. Morin, J. Rouchy, and Z. Kazei, Phys. Rev. B **51**, 15103 (1995).
18. Z. A. Kazei, N. P. Kolmakova, and O. A. Shishkina, Vestn. Mosk. Univ., Ser. 3: Fiz., Astron., No. 6, 40 (2000).

Translated by H. Bronstein

Journal of Experimental and Theoretical Physics, Vol. 93, No. 6, 2001, pp. 1261–1269.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 6, 2001, pp. 1457–1467.
Original English Text Copyright © 2001 by Andreev, Melnikovsky.

SOLIDS
Structure

Two-Velocity Elasticity Theory and Facet Growth[¶]

A. F. Andreev* and L. A. Melnikovsky**
Kapitza Institute for Physical Problems, Russian Academy of Sciences, Moscow, 117334 Russia
** e-mail: andreev@kapitza.ras.ru*
*** e-mail: leva@kapitza.ras.ru*

Received July 18, 2001

Abstract—We explain the linear growth of smooth solid helium facets by the presence of lattice point defects. To implement this task, the framework of very general two-velocity elasticity theory equations is developed. Boundary conditions for these equations for various surface types are derived. We also suggest additional experiments the concept to justify. © *2001 MAIK “Nauka/Interperiodica”.*

1. INTRODUCTION

The existence of two distinct states of a crystal surface is well known: it may be either smooth or rough (for a review, see [1]). A smooth surface is characterized by a long-range order and small fluctuations. On the contrary, a rough surface behaves differently: it does not exhibit a long-range order and its displacement fluctuates heavily. These equilibrium properties lead to different kinetic properties. While the rough surface is usually supposed to grow easily (as described by the growth coefficient), the smooth one is characterized by the zero growth coefficient and grows with nuclei of a new atomic layer (if the crystal has no dislocations). In accordance with this mechanism, one should not observe the linear growth rate if low overpressure is applied. The reality is different: experiments [2] demonstrate that a smooth helium surface free of screw dislocations grows linearly. This work is an attempt to explain this behavior by the presence of lattice point defects (vacancies). The idea is similar to that suggested by Herring [3] and by Lifshitz [4] as an explanation of the flow of polycrystals. It is quite simple: the mass flux in bulk helium is attributed to the motion of vacancies. This flux is the mass transfer through the lattice. Therefore, if vacancies are allowed to be created at the bottom edge of the sample (the boundary between the crystal and the wall, see Fig. 1) and to be annihilated at the top of it (at the smooth crystal–liquid interface), then the crystal grows.

The suggested crystal growth mechanism can be explained as follows. Because the smooth crystal facet (the top one in Fig. 1) cannot move with respect to the lattice, it moves upward adhering to the lattice. Vacuities appearing due to this at the bottom edge of the crystal transform into lattice defects (ordinary vacancies) and go up through (and faster than) the bulk helium. They finally vanish in the liquid on the top smooth surface of the sample. In other words, the crystal grows on the boundary between helium and the wall, rather than on the smooth solid–liquid interface (which nevertheless provides mass supply for the growth). It is important to emphasize that this scenario can occur if and only if the vacancies are allowed to emerge at the bottom edge of the crystal. One can say that this boundary is in some sense “atomically rough”—it can grow new atomic layers. For this condition to be satisfied, the wall surface must have a disordered shape or be slightly tilted with respect to the basal planes of the crystal (thereby forming a vicinal interface). This ensures that the surface can play the role of a source or a sink of vacancies. An atomically flat wall parallel to the basal plane should, in contrast, behave like a normal smooth surface—it is fixed to the lattice. This is because new atomic layer nuclei must be created for the surface to move.

This paper is organized as follows. In Sections 2 and 3, we derive very general two-velocity elasticity theory equations. They consist of conditions for the variables of the conventional elasticity theory (including lattice velocity) and equations for a macroscopic description of the quasiparticle gas (including the quasiparticle gas velocity).

Equations to be derived are similar to those of the two-velocity superfluid hydrodynamics. In our equations, the lattice and excitation gas velocities replace the superfluid and normal component velocities of the two-fluid hydrodynamics. Similarly to the regular linear phonon hydrodynamics (see [5, § 71]), the probabil-

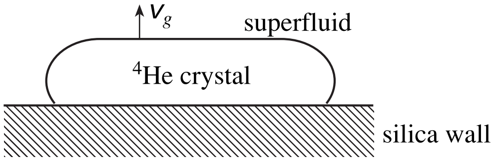


Fig. 1. Typical experimental layout.

[¶] This article was submitted by the authors in English.

ity of Umklapp processes (which result in the nonconservation of the total quasimomentum in quasiparticle collisions) is supposed to be low. In the low-temperature region considered here, this assumption is quite reasonable. We also neglect dissipation here. This means that our analysis is limited to the terms of the first order in gradients.

The exact (nonlinear) hydrodynamics equations for a superfluid can be derived (see [6]) from phenomenological considerations, using conservation laws. A constitutive argument for this derivation is the statement that the superfluid flow is potential. This is an intrinsic property of the order parameter in a superfluid. This condition is unavailable for a crystal (moreover, there is no quasimomentum conservation relation in the nonlinear description, see Eq. (15) below).

We deduce the two-velocity elasticity theory equations using a more general approach (see the paper by Pushkarov and one of the authors [7], as well as [8] and [9]). It is based on the kinetic equation description of the quasiparticle dynamics. The realization of this technique *per se*, in a nonlinear situation in particular, is a matter of considerable interest not only for a solid but also for a superfluid (that this procedure is possible is mentioned in [5, § 77]). With this technique, we find exact expressions for all hydrodynamic variables and their dependence (in terms of the quasiparticle energy spectrum) on the relative velocity of components. It is trivial to extend the equations obtained for the solid dynamics to the simpler case of superfluid hydrodynamics.

Boundary conditions for our equations depend on the surface type; in Section 4, we thoroughly consider three possibilities: rough (Section 4.1.1) and smooth (Section 4.1.2) interfaces between solid and liquid helium, and the rough boundary between solid helium and the normal hard wall (Section 4.2.1). Finally, in Section 5, we calculate the growth rate for the crystal.

2. DEFINITIONS

Following the principles in [7–9], we employ the Euler approach to the lattice description. We thus introduce three “node numbers” N^α ($\alpha = 1, 2, 3$). They are functions of space coordinates $\mathbf{r}$ and time t, $N^\alpha = N^\alpha(\mathbf{r}, t)$. From now on, Greek indices (like α here) are used for the “lattice space” components and Latin indices (e.g., i in x_i for the components of $\mathbf{r}$) for the real space components. Defining the reciprocal lattice vectors as

$$\mathbf{a}^\alpha = \partial N^\alpha / \partial \mathbf{r},$$

we obtain the elementary lattice translation vectors a_β as

$$\mathbf{a}^\alpha \mathbf{a}_\beta = \delta^\alpha_\beta.$$

Taking the time derivative, we obtain the lattice velocity as

$$\mathbf{w} = -\mathbf{a}_\alpha \dot{N}^\alpha.$$

The elastic energy E_l of the lattice is a function of the deformation. Moreover, because it depends not on the spatial orientation of the infinitesimal sample (the space is isotropic), but on the relative position of the a_i^α vectors, we can write $E_l = E_l(g^{\alpha\beta})$, where $g^{\alpha\beta} = \mathbf{a}^\alpha \mathbf{a}^\beta$ is a symmetric “metric tensor” of the lattice space.

We are now ready to describe quasiparticle degrees of freedom. We do not specify the quasiparticle nature at the moment (the quasiparticles can be phonons, vacancies as in [9], or electrons as in [7]). All the equations written below imply the summation over all branches of excitations; we do not explicitly write the sum for brevity. Any quasiparticle should be characterized by its mass m (which is zero for phonons, positive for electrons, and negative for vacancies), coordinate, and momentum. Because quasiparticles exist in the lattice background, the quasimomentum should be used. The quasiparticle energy in the frame of reference of the lattice $\epsilon = \epsilon(\mathbf{a}_\alpha(\mathbf{p} - m\mathbf{w}), g^{\alpha\beta})$ is a periodic function of the quasimomentum $\mathbf{p}$ (with the periods $2\pi\hbar\mathbf{a}^\beta$). In laboratory frame of reference, we have the quasiparticle energy (see [7])

$$\tilde{\epsilon} = \epsilon + m\mathbf{w}\frac{\partial \epsilon}{\partial \mathbf{p}} + m\frac{w^2}{2}.$$

We also use the variables $\mathbf{k} = \mathbf{p} - m\mathbf{w}$ and $k_\alpha = \mathbf{a}_\alpha \mathbf{k}$. Quasiparticle dynamics is determined by the Hamilton function

$$H = \epsilon + \mathbf{p}\mathbf{w} - mw^2/2.$$

We now introduce the distribution function $f(\mathbf{r}, \mathbf{p})$ (it is also a periodic function of the quasimomentum $\mathbf{p}$). Its kinetics is governed by the Boltzmann equation

$$\frac{\partial f}{\partial t} + \frac{\partial f}{\partial \mathbf{r}}\frac{\partial H}{\partial \mathbf{p}} - \frac{\partial f}{\partial \mathbf{p}}\frac{\partial H}{\partial \mathbf{r}} = \mathrm{St}\, f. \tag{1}$$

Using this distribution function, we can obtain macroscopic quantities, such as the mass density,

$$\rho = \rho_l + \langle mf \rangle = M\det(g^{\alpha\beta})^{1/2} + mn,$$

where the angle brackets denote the integration over quasimomentum space,

$$\langle \; \rangle = \int \frac{d\mathbf{p}}{(2\pi\hbar)^3},$$

ρ_l is the lattice density, M is the mass of an elementary cell, and $n = \langle f \rangle$.

We consider a quasi-equilibrium distribution function. The complete set of quantities conserved in quasiparticle collisions consists of their mass (which is proportional to their quantity for “real” particles like electrons and vacancies and is zero for phonons), energy, and quasimomentum (in the low-temperature region, Umklapp processes can be neglected). Consequently, the most general quasi-equilibrium distribution is a function of

$$z = \frac{\epsilon - \mathbf{k}\mathbf{v} - m\mu_0}{T} = \frac{\epsilon - (\mathbf{p} - m\mathbf{w})\mathbf{v} - m\mu_0}{T}$$

$$= \frac{\epsilon - \mathbf{p}\mathbf{v} - m\mu_0 + m\mathbf{w}\mathbf{v}}{T} = \frac{\epsilon - \mathbf{p}\mathbf{v} - m\phi}{T},$$

where

$$\mu_0 - \mathbf{w}\mathbf{v} = \phi.$$

The respective Lagrange coefficients T, v, and μ_0 denote the temperature, the velocity relative to the lattice, and the chemical potential of the quasiparticle gas. For definiteness, we assume that the excitations are Bose particles. The distribution function is then given by

$$f = \frac{1}{e^z - 1} = \left(\exp\frac{\epsilon - \mathbf{p}\cdot\mathbf{v} - m\phi}{T} - 1\right)^{-1} \tag{2}$$

and[1]

$$\ln\frac{f+1}{f} = z.$$

We can now calculate other macroscopic parameters with this distribution function. For the mass flux, we have

$$\mathbf{J} = \rho_l\mathbf{w} + \mathbf{w}mn + \mathbf{j} = \rho\mathbf{w} + \mathbf{j} = \rho\mathbf{w} + mn\mathbf{v}, \tag{3}$$

where the mass flux with reference to the lattice is

$$\mathbf{j} = m\left\langle f\frac{\partial\epsilon}{\partial\mathbf{p}}\right\rangle = mn\mathbf{v}.$$

Using $\mathbf{J}$, we can write the mass conservation as

$$\dot{\rho} + J_{i,\,i} = 0.$$

The number of real (massive) particles is also conserved in the bulk, and we therefore have one additional conservation law

$$\dot{\rho}_l + (\rho_l w_i)_{,i} = 0.$$

Similarly, the energy density is given by

$$\begin{aligned} E &= \rho_l\frac{w^2}{2} + E_l(g^{\alpha\beta}) + \langle\tilde{\epsilon} f\rangle \\ &= \rho\frac{w^2}{2} + E_l(g^{\alpha\beta}) + \mathbf{w}\mathbf{j} + \langle\epsilon f\rangle. \end{aligned} \tag{4}$$

This equation allows us to prove (and find) exact macroscopic equivalents of the microscopic quantities introduced above. The total energy density of the crystal can be obtained via a Galilean transformation,

$$E = E_0 + \frac{\rho w^2}{2} + \mathbf{j}\mathbf{w}, \tag{5}$$

where $E_0 = E_0(\mathbf{a}^\alpha, S, \rho, \mathbf{K})$ is the energy in the frame of reference of the lattice, with $\mathbf{K} = \langle\mathbf{k}f\rangle$ characterizing the quasimomentum density. A reasonable expression for the E_0 differential

$$dE_0 = \lambda_{ij}a_{\nu j}da_i^\nu + TdS + \mu d\rho + \mathbf{v}d\mathbf{K} \tag{6}$$

can be obtained with the conventional definition of the entropy density for the Bose gas,

$$S = \langle(f+1)\ln(f+1) - f\ln f\rangle = \langle fx + \ln(f+1)\rangle.$$

Its differential is

$$dS = \langle(\ln(f+1) + 1 - \ln f - 1)df\rangle = \langle zdf\rangle.$$

Subtracting the differentials of (4) and (5), we obtain

$$\begin{aligned} 0 &= \lambda_{ij}a_{\nu j}da_i^\nu + TdS + \mu d\rho + \mathbf{v}d\mathbf{K} - dE_l(g^{\alpha\beta}) \\ &\quad - d\langle\epsilon f\rangle = \lambda_{ij}a_{\nu j}da_i^\nu + T\langle zdf\rangle + \mu d\rho + \mathbf{v}d\mathbf{K} \\ &\quad - dE_l(g^{\alpha\beta}) - \langle\epsilon df\rangle - \langle fd\epsilon\rangle \qquad (7) \\ &= \lambda_{ij}a_{\nu j}da_i^\nu + \langle(Tz - \epsilon + \mathbf{v}\mathbf{k})df\rangle + \mu d\rho \\ &\quad - dE_l(g^{\alpha\beta}) + \mathbf{v}\langle fd\mathbf{k}\rangle - \langle fd\epsilon\rangle. \end{aligned}$$

We now transform the part of this equation related to the lattice deformation:

$$\begin{aligned} &\frac{\mu}{2}\rho_l g_{\alpha\beta}dg^{\alpha\beta} - dE_l(g^{\alpha\beta}) + \mathbf{v}\langle fd\mathbf{k}\rangle - \langle fd\epsilon\rangle \\ &= \frac{\mu}{2}\rho_l g_{\alpha\beta}dg^{\alpha\beta} - \frac{\partial E_l}{\partial g^{\alpha\beta}}dg^{\alpha\beta} - m\mathbf{v}\langle fd\mathbf{w}\rangle \\ &- \left\langle f\left(\frac{\partial\epsilon}{\partial k_\alpha}((\mathbf{p} - m\mathbf{w})d\mathbf{a}_\alpha - m\mathbf{a}_\alpha d\mathbf{w}) + \left(\frac{\partial\epsilon}{\partial g^{\alpha\beta}}\right)_{k_\alpha}dg^{\alpha\beta}\right)\right\rangle \\ &= -\left(\frac{\partial E_l}{\partial g^{\alpha\beta}} + \left\langle f\left(\frac{\partial\epsilon}{\partial g^{\alpha\beta}}\right)_{k_\alpha}\right\rangle\right)dg^{\alpha\beta} \\ &+ \left(\mu\rho_l\delta_{ij} - \left\langle f\frac{\partial(Tx + \mathbf{p}\mathbf{v})}{\partial p_j}(p_i - mw_i)\right\rangle a_j^\alpha\right)da_{\alpha i} \\ &= -\left(\frac{\partial E_l}{\partial g^{\alpha\beta}} + \left\langle f\left(\frac{\partial\epsilon}{\partial g^{\alpha\beta}}\right)_{k_\alpha}\right\rangle\right)2a_i^\alpha a_j^\beta a_{\nu j}da_i^\nu \\ &+ (\delta_{ij}(T\langle\ln(f+1)\rangle + \mu\rho_l) + v_iP_j - mnv_iw_j)a_{\nu j}da_i^\nu \\ &= \left(\delta_{ij}\left(TS + \rho\frac{w^2}{2} + E_l - E + \mathbf{P}\mathbf{v} + \mu\rho_l + mn\mu_0\right)\right. \end{aligned}$$

---

[1] For further convenience, we also provide here the result of the distribution function integration:

$$\int fdz = \ln\frac{e^z - 1}{e^z} = -\ln(f+1).$$

$$-\Lambda_{ij} + v_i P_j - mnv_i w_j\Big)a_{\nu j}da_i^{\nu}.$$

From (7), we finally obtain

$$0 = (\mu - \mu_0)mdn$$

$$+\left\{\lambda_{ij} - \Lambda_{ij} + \delta_{ij}\left(TS + E_l - E + \mu\rho + \mathbf{P}\mathbf{v} + \rho\frac{w^2}{2}\right)\right. \tag{8}$$

$$\left. + v_i P_j - mnv_i w_j\right\}a_{\nu j}da_i^{\nu},$$

where we introduced $P_i = \langle p_i f\rangle$ and

$$\Lambda_{ij} = 2a_i^{\alpha}a_j^{\beta}\left(\frac{\partial E_l}{\partial g^{\alpha\beta}} + \left\langle\left(\frac{\partial \epsilon}{\partial g^{\alpha\beta}}\right)_{k_\alpha}\right\rangle\right). \tag{9}$$

The terms in (8) are independent, and each of them must therefore be equal to zero. That is,

$$\mu = \mu_0,$$

$$\lambda_{ij} = \Lambda_{ij} - \delta_{ij}\left(TS + E_l - E + \mu\rho + \mathbf{P}\mathbf{v} + \rho\frac{w^2}{2}\right) \tag{10}$$

$$-v_i P_j + mnv_i w_j.$$

3. EQUATIONS AND FLUXES

Here, we derive dynamics equations and thermodynamic fluxes for the system. Neglecting dissipation at this point, we assume that the entropy conservation law is valid,

$$\dot{S} + F_{i,i} = 0,$$

where the entropy flux F_i is determined by

$$\mathbf{F} = S(\mathbf{v} + \mathbf{w}).$$

We continue with the equation for the momentum flux found in [7],

$$\dot{j}_i + \Pi_{ik,k} = 0, \tag{11}$$

where

$$\Pi_{ik} = \rho w_i w_k - E_l\delta_{ik} + 2a_i^{\alpha}a_k^{\beta}\left(\frac{\partial E_l}{\partial g^{\alpha\beta}} + \left\langle f\left(\frac{\partial \epsilon}{\partial g^{\alpha\beta}}\right)_{k^\alpha}\right\rangle\right)$$

$$+ w_i j_k + w_k j_i = \rho w_i w_k - E_l\delta_{ik} + mn(w_i v_k + w_k v_i)$$

$$+ 2a_i^{\alpha}a_k^{\beta}\left(\frac{\partial E_l}{\partial g^{\alpha\beta}} + \left\langle f\left(\frac{\partial \epsilon}{\partial g^{\alpha\beta}}\right)_{k^\alpha}\right\rangle\right) \tag{12}$$

$$= \rho w_i w_k - E_l\delta_{ik} + mn(w_i v_k + w_k v_i) + \Lambda_{ik},$$

where we used definition (9) for Λ_{ik}.

Taking the appropriate equation for the energy flux from [7], we have

$$\dot{E} + Q_{i,i} = 0, \tag{13}$$

where

$$Q_i = w_i E_l + \left\langle \epsilon\frac{\partial H}{\partial p_i}f\right\rangle - \frac{w^2}{2}J_i + w_k\Pi_{ik}.$$

To find the second term, we again use the distribution function from (2),

$$\left\langle \epsilon\frac{\partial H}{\partial p_i}f\right\rangle = \left\langle \epsilon\left(\frac{\partial \epsilon}{\partial p_i} + w\right)f\right\rangle = \left\langle \epsilon\left(T\frac{\partial z}{\partial p_i} + v_i + w_i\right)f\right\rangle$$

$$= (v_i + w_i)\langle\epsilon f\rangle + T\left\langle (Tx + \mathbf{p}\mathbf{v} + \phi)\frac{\partial z}{\partial p_i}f\right\rangle$$

$$= (v_i + w_i)\langle\epsilon f\rangle - T\left\langle (\mathbf{p}\mathbf{v})\frac{\partial \ln(1+f)}{\partial p_i}\right\rangle$$

$$= (v_i + w_i)\langle\epsilon f\rangle + Tv_i\langle\ln(1+f)\rangle = (v_i + w_i)\langle\epsilon f\rangle$$

$$+ Tv_i\left\langle\ln(1+f) + f\ln\left(\frac{f+1}{f}\right)\right\rangle - Tv_i\left\langle f\frac{\epsilon - \mathbf{p}\mathbf{v} - m\phi}{T}\right\rangle$$

$$= (v_i + w_i)\langle\epsilon f\rangle + Tv_i S - v_i\langle f\epsilon\rangle + v_i(\mathbf{v}\mathbf{P})$$

$$+ mn\phi = w_i\langle\epsilon f\rangle + v_i(TS + \mathbf{v}\mathbf{P} + mn\phi).$$

For the energy flux, we finally have

$$Q_i = w_i\langle\epsilon f\rangle + v_i(TS + \mathbf{v}\mathbf{P} + mn\phi)$$

$$-\frac{w^2}{2}(\rho w_i + mnv_i) + w_k(\rho w_i w_k + mn(w_i v_k + w_k v_i)$$

$$+ \Lambda_{ik}) = w_i\langle\epsilon f\rangle + v_i(TS + \mathbf{v}\mathbf{P} + mn\phi)$$

$$+ \frac{w^2}{2}(\rho w_i + mnv_i) + w_k(mnw_i v_k + \Lambda_{ik}) \tag{14}$$

$$= w_i\left(\langle\epsilon f\rangle + \mathbf{w}\mathbf{j} + \rho\frac{w^2}{2}\right)$$

$$+ v_i\left(TS + \mathbf{v}\mathbf{P} + mn\left(\phi + \frac{w^2}{2}\right)\right) + w_k\Lambda_{ik}$$

$$= v_i\left(TS + \mathbf{v}\mathbf{P} + mn\left(\phi + \frac{w^2}{2}\right)\right) + w_i(E - E_l) + w_k\Lambda_{ik}.$$

This formula completes the list of the conventional elasticity theory equations. An additional equation is required to govern the quasiparticle degrees of freedom. We now find the time derivative of P_i. We multiply Boltzmann equation (1) by p_i and integrate over the momentum space. We temporarily neglect Umklapp processes, which are supposedly rare. If needed, dissipation can be explicitly introduced into the final result. In other words, the quasimomentum **p** is conserved in (normal) collisions, and the term involving the collision integral is therefore zero. The left-hand side of the Boltzmann equation gives

$$\dot{P}_i = \left\langle p_i\left(\frac{\partial f}{\partial \mathbf{p}}\frac{\partial H}{\partial \mathbf{r}} - \frac{\partial f}{\partial \mathbf{r}}\frac{\partial H}{\partial \mathbf{p}}\right)\right\rangle$$

$$= \left\langle p_i\left(\frac{\partial f}{\partial \mathbf{p}}\left(\frac{\partial \epsilon}{\partial \mathbf{r}} + (p_j - mw_j)\frac{\partial w_j}{\partial \mathbf{r}}\right) - \frac{\partial f}{\partial \mathbf{r}}\left(\frac{\partial \epsilon}{\partial \mathbf{p}} + \mathbf{w}\right)\right)\right\rangle$$

$$= \left\langle -f\left(\frac{\partial \epsilon}{\partial x_i} + (\mathbf{p} - m\mathbf{w})\frac{\partial \mathbf{w}}{\partial x_i}\right)\right.$$

$$\left. - f p_i\left(\frac{\partial^2 \epsilon}{\partial \mathbf{r}\partial \mathbf{p}} + \frac{\partial \mathbf{w}}{\partial \mathbf{r}}\right) - p_i\frac{\partial f}{\partial \mathbf{r}}\frac{\partial \epsilon}{\partial \mathbf{p}} - p_i\frac{\partial f}{\partial \mathbf{r}}\mathbf{w}\right\rangle$$

$$= -\left\langle f\frac{\partial \epsilon}{\partial x_i} + f p_i\frac{\partial^2 \epsilon}{\partial \mathbf{r}\partial \mathbf{p}} + p_i\frac{\partial f}{\partial \mathbf{r}}\frac{\partial \epsilon}{\partial \mathbf{p}}\right\rangle$$

$$+ \langle f\rangle m\mathbf{w}\frac{\partial \mathbf{w}}{\partial x_i} - \left\langle \mathbf{p}\frac{\partial \mathbf{w}}{\partial x_i}f + p_i\frac{\partial \mathbf{w}}{\partial \mathbf{r}}f + p_i\frac{\partial f}{\partial \mathbf{r}}\mathbf{w}\right\rangle.$$

The first term can be transformed as

$$\left\langle f\frac{\partial \epsilon}{\partial x_i} + f p_i\frac{\partial^2 \epsilon}{\partial \mathbf{r}\partial \mathbf{p}} + p_i\frac{\partial f}{\partial \mathbf{r}}\frac{\partial \epsilon}{\partial \mathbf{p}}\right\rangle$$

$$= \left\langle f p_i\frac{\partial^2 \epsilon}{\partial \mathbf{r}\partial \mathbf{p}} + f\frac{\partial \epsilon}{\partial x_i} + p_i\frac{\partial f}{\partial \mathbf{r}}\left(T\frac{\partial z}{\partial \mathbf{p}} + \mathbf{v}\right)\right\rangle$$

$$= \left\langle f p_i\frac{\partial^2 \epsilon}{\partial \mathbf{r}\partial \mathbf{p}} + f\frac{\partial \epsilon}{\partial x_i} + p_i\frac{\partial f}{\partial \mathbf{r}}\mathbf{v}\right.$$

$$\left. + p_i\frac{\partial f}{\partial z}\frac{\partial z}{\partial \mathbf{p}}\left(\frac{\partial \epsilon}{\partial \mathbf{r}} - p_k\frac{\partial v_k}{\partial \mathbf{r}} - m\frac{\partial \phi}{\partial \mathbf{r}} - z\frac{\partial T}{\partial \mathbf{r}}\right)\right\rangle$$

$$= \left\langle p_i\frac{\partial f}{\partial \mathbf{r}}\mathbf{v} + mf\frac{\partial \phi}{\partial x_i} + f\mathbf{p}\frac{\partial \mathbf{v}}{\partial x_i}\right.$$

$$\left. + fz\frac{\partial T}{\partial x_i} + p_i f\frac{\partial \mathbf{v}}{\partial \mathbf{r}} + p_i f\frac{\partial z}{\partial \mathbf{p}}\frac{\partial T}{\partial \mathbf{r}}\right\rangle$$

$$= nm\frac{\partial \phi}{\partial x_i} + \mathbf{P}\frac{\partial \mathbf{v}}{\partial x_i} + \frac{\partial}{\partial \mathbf{r}}(P_i\mathbf{v}) + \left\langle \delta_{ik}fz + p_i f\frac{\partial z}{\partial p_k}\right\rangle\frac{\partial T}{\partial x_k}$$

$$= nm\frac{\partial \phi}{\partial x_i} + \mathbf{P}\frac{\partial \mathbf{v}}{\partial x_i} + \frac{\partial}{\partial \mathbf{r}}(P_i\mathbf{v})$$

$$+ \left\langle \delta_{ik}f\ln\frac{1+f}{f} - p_i\frac{\partial z}{\partial p_k}\frac{\partial}{\partial z}\ln(f+1)\right\rangle\frac{\partial T}{\partial x_k}$$

$$= nm\frac{\partial \phi}{\partial x_i} + \mathbf{P}\frac{\partial \mathbf{v}}{\partial x_i} + \frac{\partial}{\partial \mathbf{r}}(P_i\mathbf{v}) + \left\langle f\ln\frac{1+f}{f} + \ln(f+1)\right\rangle$$

$$\times\frac{\partial T}{\partial x_i} = nm\frac{\partial \phi}{\partial x_i} + \mathbf{P}\frac{\partial \mathbf{v}}{\partial x_i} + \frac{\partial}{\partial \mathbf{r}}(P_i\mathbf{v}) + S\frac{\partial T}{\partial x_i}.$$

Consequently,

$$\begin{aligned} \dot{P}_i &= nm\mathbf{w}\frac{\partial \mathbf{w}}{\partial x_i} - \mathbf{P}\frac{\partial \mathbf{w}}{\partial x_i} - \frac{\partial}{\partial \mathbf{r}}P_i\mathbf{w} \\ &\quad - nm\frac{\partial \phi}{\partial x_i} - \mathbf{P}\frac{\partial \mathbf{v}}{\partial x_i} - \frac{\partial}{\partial \mathbf{r}}(P_i\mathbf{v}) - S\frac{\partial T}{\partial x_i} \\ &= nm\frac{\partial}{\partial x_i}\left(\frac{w^2}{2} - \phi\right) - S\frac{\partial T}{\partial x_i} \\ &\quad - \mathbf{P}\frac{\partial}{\partial x_i}(\mathbf{w} + \mathbf{v}) - \frac{\partial}{\partial \mathbf{r}}(P_i(\mathbf{w} + \mathbf{v})). \end{aligned} \tag{15}$$

The sought complete set of the two-velocity elasticity theory equations consists of Eqs. (15), (13) (with Q defined by (14)), and (11) (with Π_{ij} defined by (12)).

4. BOUNDARY CONDITIONS

We now turn to boundary conditions. They immediately follow from the conservation relations satisfied at the interface. It is much easier to perform all transformations in the frame of reference of the interface itself. All the velocities are therefore taken relative to the boundary. Moreover, we simplify the problem by restricting it to the one-dimensional case: all fluxes are supposed to be perpendicular to the flat surface; we let the z axis run along this direction. Because no curvature is ascribed to the surface, we ignore capillary effects. All calculations performed here are valid within the linear approximation. Naturally, boundary conditions must depend on the type of the boundary and on the type of the media on the other side of the interface. We begin with the situation extensively discussed in the literature, the solid-liquid interface [1]. Because the possibility of the mass flux through the lattice is taken into account, the results are different, however.

4.1. Solid-Liquid Interface

The liquid on the other side of the interface (being superfluid) is characterized by the chemical potential μ^L, normal and superfluid densities ρ_n^L and ρ_s^L, normal and superfluid velocities v_n^L and v_s^L, temperature T^L, pressure p^L, and the entropy density S^L (see Fig. 2):

$$\begin{aligned} R + S^S(v^S + w^S) &= S^L v_n^L, \\ w^S(E^S - E_l^S + \Lambda_{zz}^S) + v^S(T^S S^S + m^S n^S \phi^S) & \\ = \mu^L(\rho_s^L v_s^L + \rho_n^L v_n^L) &+ S^L T^L v_n^L, \\ v^S m^S n^S + \rho^S w^S &= \rho_s^L v_s^L + \rho_n^L v_n^L, \\ \Lambda_{zz}^S - E_l^S &= p^L. \end{aligned} \tag{16}$$

The superscript S indicates that the appropriate quantities refer to the solid. The first equation is the entropy

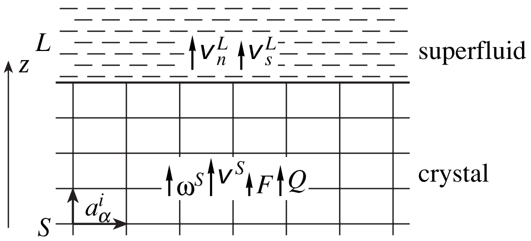


Fig. 2. Solid–liquid boundary: fluxes in one dimension.

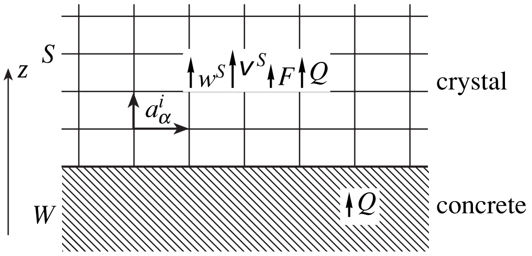


Fig. 3. Solid–wall boundary: fluxes in one dimension.

growth condition and R is the surface dissipative function. The last three equations in (16) are simply the requirements for the energy, mass, and momentum conservation for the surface, respectively. The surface dissipative function must be a positive square form. Using (16) and (10), it can be expressed as

$$\begin{aligned} RT^L &= v^S(S^S(T^S - T^L) + m^S n^S(\phi^S - \mu^L)) \\ &\quad + w^S(E^S - E_l^S + \Lambda_{zz}^S - \rho^S\mu^L - T^L S^S) \\ &= v^S(m^S n^S(\phi^S - \mu^L) + S^S(T^S - T^L)) \\ &\quad + w^S(\lambda_{zz}^S + \rho^S(\phi^S - \mu^L) + S^S(T^S - T^L)). \end{aligned}$$

We now recall that the solid-liquid boundary can be either atomically rough or atomically smooth, depending on the temperature. The nature of the surface may (or may not) impose certain restrictions on the dynamics. For both types of the surface, the equation

$$\Lambda_{zz}^S - E_l^S = p^L \tag{17}$$

is satisfied.

4.1.1. Rough surface. Employing the Onsager principle, we obtain

$$\begin{aligned} v^S &= \alpha(m^S n^S(\phi^S - \mu^L) + S^S(T^S - T^L)) \\ &\quad + \eta(\lambda_{zz}^S + \rho^S(\phi^S - \mu^L) + S^S(T^S - T^L)), \\ w^S &= \eta(m^S n^S(\phi^S - \mu^L) + S^S(T^S - T^L)) \\ &\quad + \nu(\lambda_{zz}^S + \rho^S(\phi^S - \mu^L) + S^S(T^S - T^L)). \end{aligned} \tag{18}$$

The kinetic matrix $\begin{pmatrix} \alpha & \eta \\ \eta & \nu \end{pmatrix}$ is positively definite.

4.1.2. Smooth surface. A smooth surface implies immobility of the interface relative to the lattice. That is,

$$w^S = 0.$$

For the quasiparticle gas velocity, we then obtain a restricted version of (18),

$$v^S = \alpha(m^S n^S(\phi^S - \mu^L) + S^S(T^S - T^L)), \tag{19}$$

with the kinetic coefficient $\alpha > 0$.

4.2. Solid-Wall Boundary

By a wall, we imply a macroscopically flat structure-less medium, in short, “concrete.” The “solid-wall” boundary occurs between solid helium and some normal rigid solid (silica in experiment [2]). A “concrete” wall is characterized by no mass flux in it (i.e., through the interface). The wall can supply an arbitrary energy flux; we let Q denote the flux and T^W the wall temperature (see Fig. 3). A concrete wall is characterized by fewer variables than liquid, and the appropriate equations are therefore somewhat simpler.

As for the solid-liquid interface, the actual boundary conditions must depend on the microscopic pattern of the surface. We can imagine a smooth basal plane of the crystal adjacent to an atomically flat concrete wall. This plane must stay at rest with respect to the wall because its motion requires the creation of new atomic layer nuclei. The plane is similar to the smooth solid–liquid interface, and we can naturally say that such an interface is smooth. The boundary condition is then given by

$$w^S = v^S = 0.$$

Another, much more interesting scenario is realized if the interface is slightly tilted with respect to the basal plane. Such planes may move by growing additional nodes at the edge. This means that no restrictions are imposed on the lattice velocity near the interface. In other words, vacancies are allowed to freely appear and vanish on the surface (in this sense, the surface is similar to a grid of dislocations arranged at the boundary of the crystal that serve as sources or sinks for vacancies; similar speculations can be found in [4] in explaining polycrystal plasticity). We call this type of interface “rough.” In this sense, the solid-wall boundary can be either smooth or rough. The suggested growth mechanism can be applied only to the rough boundary.

4.2.1. Rough boundary. Assuming the boundary to be rough and using the same approach as for the liquid, we write the conservation laws

$$\begin{aligned} S^S(v^S + w^S) &= R + Q/T^W, \\ w^S(E^S - E_l^S + \Lambda_{zz}^S) + v^S(T^S S^S + m^S n^S \phi^S) &= Q, \\ v^S m^S n^S = \rho^S w^S &= 0. \end{aligned}$$

Again, R is the surface dissipative function,

$$RT^W = v^S\left(\frac{m^S n^S}{\rho^S}(E^S - E_l^S + \Lambda_{zz}^S - \rho^S\phi^S - T^W S^S) + S^S(T^W - T^S)\right)$$

$$= v^S\left(\frac{m^S n^S}{\rho^S}\lambda_{zz}^S + \left(1 - \frac{m^S n^S}{\rho^S}\right)S^S(T^W - T^S)\right).$$

It must be positive, and for the quasiparticle velocity on the surface we therefore have

$$v^S = \beta\left(\frac{m^S n^S}{\rho^S}\lambda_{zz}^S + \left(1 - \frac{m^S n^S}{\rho^S}\right)S^S(T^W - T^S)\right), \quad (20)$$

where $\beta > 0$ is the surface kinetic coefficient.

5. THE GROWTH RATE

We now use the equations and boundary conditions obtained above. The physical system discussed in what follows is solid helium with elementary excitations represented by phonons and vacancies. We first introduce a certain amount of friction between the quasiparticle gas and the lattice. To obtain a physically sound result, we again restrict our analysis to one dimension. Furthermore, for simplicity, all our calculations are performed within the linear approximation. We can write the quasimomentum density (with the superscript S omitted) as $K = \rho_K v$, where

$$\rho_K \sim \frac{\hbar}{a^4}\frac{T^4}{\Theta_D^4 c}. \quad (21)$$

Here, Θ_D is the Debye temperature, c is the velocity of sound in the crystal, and a is the lattice period. The last equation is quite obvious. It follows from the fact that in the low-temperature region, the quasimomentum is mainly associated with phonons (the number of vacancies is exponentially small). The result therefore coincides with the one for the mass density (and the momentum density) of the normal component of the superfluid,

$$\rho_K \sim \rho_n \sim T^4/\hbar^3 c^5$$

(see [10]).

To describe (rare) Umklapp events, we introduce the appropriate relaxation time parameter τ_U. It is a "between-Umklapp-collision time." From (15), we then have

$$0 = \dot{K} = -mn\nabla\phi - S\nabla T - \frac{K}{\tau_U}$$

$$= -mn\nabla\phi - S\nabla T - \rho_K\frac{v}{\tau_U}.$$

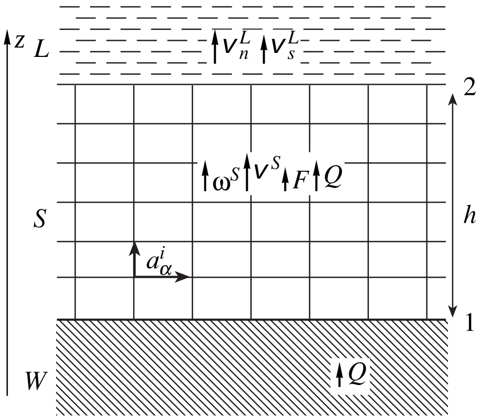


Fig. 4. Crystal growth in one dimension.

It is worth mentioning that τ_U may well depend on both phonons and vacancies, despite the fact that the population of vacancies is much lower than that of phonons. For instance, if the vacancy energy band is sufficiently narrow, the probability of Umklapp processes is significantly higher for vacancies than for phonons. This might overcome the low concentration of vacancies.

Interestingly enough, these formulas allow us to obtain the growth rate for a smooth surface. The quasiparticles playing the crucial role here (that of mass carriers) are vacancies, with their mass given by $m = -m_{\mathrm{He}}$.

To estimate the rate, we write the temperature gradient as

$$\nabla T = \frac{T_2 - T_1}{h},$$

where the subscripts 1 and 2 stand for the solid-wall and the solid–liquid interfaces respectively (see Fig. 4). Similarly, we write $\nabla\phi = (\phi_2 - \phi_1)/h$ for the chemical potential. We now use boundary conditions (19) and (20),

$$(\alpha + \beta)v = \alpha\beta\left(mn(\phi_2 - \mu^L) + S(T_2 - T^L) + \frac{mn}{\rho}\left(\lambda_{zz1} + \left(1 - \frac{mn}{\rho}\right)S(T^W - T_1)\right)\right)$$

$$= \alpha\beta\left\{mn(\phi_2 - \phi_1) + S(T_2 - T_1) + S(T^W - T^L) + mn\left(\phi_1 - \mu^L + \frac{1}{\rho}(\lambda_{zz1} - S(T^W - T_1))\right)\right\}$$

$$= \alpha\beta\left\{-\rho_K h\frac{v}{\tau_U} + S(T^W - T^L) + mn\left(\frac{1}{\rho}(\lambda_{zz1} - S(T^W - T_1)) + \phi_1 - \mu^L\right)\right\}.$$

In other words,

$$v\left(\frac{1}{\alpha}+\frac{1}{\beta}+\frac{\rho_K h}{\tau_U}\right) = S(T^W - T^L)$$

$$+ mn\left(\frac{1}{\rho}(\lambda_{zz1} - S(T^W - T_1)) + \phi_1 - \mu^L\right),$$

where (using Eqs. (10) and (17))

$$\begin{aligned}\lambda_{zz1} &= \Lambda_{zz1} - T_1S_1 - E_{l1} + E_1 - \phi_1\rho_1\\ &= p^L - T_1S_1 + E_1 - \phi_1\rho_1.\end{aligned}$$

In equilibrium,

$$\lambda_{zz1} = 0, \quad \phi_1 = \mu^L,$$

and

$$T^W = T^L = T_1.$$

If the liquid temperature and pressure change by ΔT and Δp, we can write an equation for the growth rate v_g,

$$-v_g\frac{\rho}{mn}\left(\frac{1}{\alpha}+\frac{1}{\beta}+\frac{\rho_K h}{\tau_U}\right) = -S\Delta T$$

$$+ mn\left(\frac{1}{\rho}(\Delta p - S\Delta T_1 - T\Delta S_1 + \Delta E_1\right.$$

$$\left. - \rho\Delta\phi_1 - \phi\Delta\rho_1 + S\Delta T_1) + \Delta\phi_1 + \frac{S^L\Delta T - \Delta p}{\rho^L}\right) \quad (22)$$

$$= -S\Delta T + mn\left(\frac{\Delta p}{\rho} + \frac{S^L\Delta T - \Delta p}{\rho^L}\right)$$

$$= -\left(S - S^L\frac{mn}{\rho^L}\right)\Delta T + mn\left(\frac{1}{\rho} - \frac{1}{\rho^L}\right)\Delta p,$$

where we used Eq. (6), the thermodynamic equality

$$d\mu = \frac{dp - SdT}{\rho}$$

for the liquid, and the obvious relation

$$v_g = -\frac{vmn}{\rho}.$$

We now consider this equation where the second term in the right-hand side is equal to zero. This is a usual scenario for heat conductivity measurements. The heat flux $Q = vTS$ can then be expressed as

$$Q = -\frac{\Delta T}{R_{K1} + R_{K2} + h/\kappa},$$

where R_{K1} and R_{K2} are the Kapitza thermal resistances on the solid–wall and solid–liquid boundaries and κ is the heat conductivity of the crystal. Taking the inequalities

$$R_{K2} \ll R_{K1}, \quad mn \ll \rho, \quad \rho - \rho^L \ll \rho$$

into account, we immediately obtain

$$R_{K1} = \frac{1}{\beta TS^2}, \quad \kappa = \frac{\tau_U TS^2}{\rho_K}.$$

As a result, the growth rate is given by

$$v_g = \frac{mn}{\rho(TS^2R_K + TS^2h/\kappa)}\left(S\Delta T + mn\frac{\rho - \rho^L}{\rho^2}\Delta p\right).$$

Strictly speaking, the last equality implies that thermodynamic properties of the crystal mainly depend on phonons, while the contribution of vacancies to the effect under consideration is limited to the mass transfer. The growth rate here depends on the overpressure as well as on the temperature difference between the liquid and the wall.

In the real experiment [2], the temperature is lower than 100 mK. In this region, the phonon free path is much longer than the experimental cell size and the impact of phonons on the vacancy behavior is proportional to a high power of the small ratio T/Θ_D. Consequently, as $\Theta_D \longrightarrow \infty$, we can consider the vacancy gas as an independent component and neglect the influence of phonons on it. The crystal growth is accounted for by the presence of vacancies; hence, to estimate the growth rate in the experimental situation, we can simply substitute the vacancy-only quantities for all variables in Eq. (22). Because it seems that there were essentially no temperature gradients in the experiment, we consider the isothermal case $\Delta T = 0$.

The kinetic coefficients α and β are determined by the vacancy annihilation probabilities in vacancy–surface collisions. The vacancy gas velocity v at the interface should be expressed in terms of the accommodation coefficient W (which is the ratio of the number of annihilated vacancies to the total number of incident vacancies) as

$$v \sim \frac{\Delta f}{f}V_T W, \quad (23)$$

where

$$\Delta f \sim fm\Delta\mu/T$$

is the difference between the incident and reflected distribution functions and

$$V_T \sim \sqrt{T/m^*}$$

is the thermal velocity. Here, m^* is the effective mass near the bottom of the vacancy energy band. This mass can be estimated from the energy band width Δ as

$$m^* \sim \frac{\hbar^2}{a^2\Delta}.$$

The accommodation coefficient W, as any other inelastic process probability in quasiparticle–surface interactions [11], is approximately the squared ratio of the lattice period to de Broglie wavelength,

$$W \sim \left(\frac{a}{\lambda}\right)^2 \sim \frac{T}{\Delta}.$$

We can now compare Eqs. (23), (19), and (20). For the coefficients, this yields

$$\alpha \sim \beta \sim \frac{a}{n\hbar}\sqrt{\frac{T}{\Delta}}.$$

An estimate of the relaxation time τ_N characterizing the normal (non-Umklapp) vacancy collisions can be obtained from

$$\tau_N \sim \frac{1}{n\sigma V_T},$$

where $\sigma \sim a^2$ the vacancy-vacancy scattering cross section. The Umklapp relaxation time is exponentially longer, $\tau_U \sim \tau_N \exp(\Delta_U/T)$, where $\Delta_U < \Delta$ is a certain energy specific to the vacancy Umklapp processes.

Using the obvious relation $\rho_K = m^*n$, we proceed to the growth rate. From Eq. (22), it follows that

$$v_g \sim \frac{m^2n^2(\rho-\rho^L)}{(1/\alpha + 1/\beta + m^*n/\tau_U)\rho^3}\Delta p$$

$$\sim \frac{m^2n^2(\rho-\rho^L)}{(\sqrt{\Delta/T}n\hbar/a + hm^*n^2a^2V_T\exp(-\Delta_U/T))\rho^3}\Delta p$$

$$\sim \frac{a^4(\rho-\rho^L)}{\hbar\rho}\left(\exp\left(\frac{\epsilon_0}{T}\right)\sqrt{\frac{\Delta}{T}} + \exp\left(-\frac{\Delta_U}{T}\right)\frac{h}{a}\sqrt{\frac{T}{\Delta}}\right)^{-1}\Delta p.$$

Here, as an estimate, we set $\rho a^3 \sim m$ and $na^3 \sim \exp(-\varepsilon_0/T)$, where ε_0 is the bottom of the vacancy energy band. For the facet mobility $\mu_f^* = v_g/\Delta p$ introduced in [2], we have

$$\mu_f^* \sim \frac{a^4}{\hbar}\frac{\rho-\rho^L}{\rho} \times\left(\exp\left(\frac{\epsilon_0}{T}\right)\sqrt{\frac{\Delta}{T}} + \exp\left(-\frac{\Delta_U}{T}\right)\frac{h}{a}\sqrt{\frac{T}{\Delta}}\right)^{-1}. \quad (24)$$

6. CONCLUSION

Formula (24) provides a reasonable correspondence between the theory proposed here and the experiment [2]. It suggests three main predictions to be verified in further experiments.

1. The facet mobility has a maximum at some finite temperature. If the temperature decreases below the point of the maximum, the growth rate must also decrease. Otherwise, if the mobility does not tend to zero as the temperature tends to zero, this should be considered as an indication of the presence of zero-point vacancies (see [12]).

2. The observed growth rate depends on the height of the sample.

3. The crystal grows at the boundary between the solid and the wall. This fact can potentially be observed experimentally using some small foreign object frozen into the crystal in its upper part.

ACKNOWLEDGMENTS

We thank K.O. Keshishev and A.Ya. Parshin for fruitful discussions. This study was performed under partial support from the Russian Foundation for Basic Research, INTAS, and NWO.

REFERENCES

1. P. Noziéres, in *Solids Far From Equilibrium*, Ed. by C. Godréche (Cambridge Univ. Press, Cambridge, 1991), p. 1.
2. J. P. Ruutu, P. J. Hakonen, A. V. Babkin, *et al.*, Phys. Rev. Lett. **76**, 4187 (1996).
3. C. Herring, J. Appl. Phys. **21**, 437 (1950).
4. I. M. Lifshitz, Zh. Eksp. Teor. Fiz. **44**, 1349 (1963) [Sov. Phys. JETP **17**, 909 (1963)].
5. E. M. Lifshitz and L. P. Pitaevskii, *Physical Kinetics* (Nauka, Moscow, 1979; Pergamon, Oxford, 1981).
6. I. M. Khalatnikov, *The Theory of Superfluidity* (Nauka, Moscow, 1971; Mir, Moscow, 1971).
7. A. F. Andreev and D. I. Pushkarov, Zh. Éksp. Teor. Fiz. **89**, 1883 (1985) [Sov. Phys. JETP **62**, 1087 (1985)].
8. A. F. Andreev and Ya. B. Bazaliy, Zh. Éksp. Teor. Fiz. **98**, 1480 (1990) [Sov. Phys. JETP **71**, 827 (1990)].
9. A. F. Andreev, Ya. B. Bazaliy, and A. D. Savishchev, J. Low Temp. Phys. **88**, 101 (1992).
10. E. M. Lifshitz and L. P. Pitaevskiĭ, *Course of Theoretical Physics*, Vol. 5: *Statistical Physics* (Nauka, Moscow, 1978; Pergamon, New York, 1980), Part 2.
11. A. F. Andreev, Usp. Fiz. Nauk **105**, 113 (1972) [Sov. Phys. Usp. **14**, 609 (1972)].
12. G. A. Lengua and J. M. Goodkind, J. Low Temp. Phys. **79**, 251 (1990).

Journal of Experimental and Theoretical Physics, Vol. 93, No. 6, 2001, pp. 1270–1279.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 6, 2001, pp. 1468–1477.
Original Russian Text Copyright © 2001 by Vas'ko.

SOLIDS
Electronic Properties

Intersubband Absorption Peak Broadening Due to Unscreened Inhomogeneity of Levels

F. T. Vas'ko
Institute of Semiconductor Physics, National Academy of Sciences of Ukraine, Kiev, 03650 Ukraine
e-mail: ftv@hotmail.com

Received January 30, 2001

Abstract—The shape of the interband absorption peak in quantum wells with uneven heteroboundaries is studied theoretically. Although the large-scale variations of the ground level in strongly doped structures are screened, the energy of intersubband transitions remains inhomogeneous in the 2D plane due to unscreened changes of the energy of an excited level. The equations for intersubband polarization are derived taking into account the Coulomb contributions proportional to e^2 and leading to a depolarization shift and the exchange renormalization of the spectrum. The shape of the intersubband absorption peak is analyzed both in the local approximation and taking into account the nonlocality of the response in the 2D plane. In the case of single-layer irregularities of heteroboundaries, this mechanism makes the main contribution to the intersubband absorption peak broadening for the far and intermediate IR range. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In spite of the fact that resonant intersubband transitions have been studied actively during the last decade [1–3], the shape of the absorption peak has not been investigated in detail and the relative contributions of various mechanisms of broadening remain unclear. This is due to the fact that spectral studies in the IR and submillimeter spectral ranges are rather complicated; besides, the broadening is usually determined by the competition between several mechanisms [4–7]. Moreover, the electron-electron interaction leading to a depolarization resonance shift and to the exchange renormalization of the spectrum noticeably affects intersubband transitions (see the literature cited in [8, 9]). The considerable influence of Coulomb effects on the shape of the absorption peak in nonideal quantum wells (QW) is also associated with the fact that large-scale variations of the interband transition energy caused by the variations of the width of nonideal QWs are not screened even in heavily doped structures [10]. In an analysis of the transport of 2D electrons filling the ground state of 2D-inhomogeneous QWs, energy variations in this state are completely screened (we consider the case when such variations are smaller than the Fermi energy) so that large-scale inhomogeneities of heteroboundaries become immaterial. However, in an analysis of intersubband transitions, inhomogeneities lead to a considerable absorption peak broadening since the variations of the excited level energy (which are stronger than in the ground state; see the diagram in Fig. 1) remain unscreened. For a simple model of a QW of width $d_{\mathbf{x}} = d + \delta d_{\mathbf{x}}$ with high barriers, the interband energy $\varepsilon_{2\mathbf{x}} - \varepsilon_{1\mathbf{x}} = \varepsilon_{21} + \delta\varepsilon_{\mathbf{x}}$ is defined by the relations

$$\begin{gathered}\varepsilon_{s\mathbf{x}} \approx \frac{(s\pi\hbar/d_{\mathbf{x}})^2}{2m} \approx \varepsilon_s\left(1 - 2\frac{\delta d_{\mathbf{x}}}{d}\right),\\ \delta\varepsilon_{\mathbf{x}} \approx 6\varepsilon_1\frac{\delta d_{\mathbf{x}}}{d}.\end{gathered} \tag{1}$$

Here, ε_{21} is the energy of an interband transition in an ideal QW, $\delta\varepsilon_{\mathbf{x}}$ is the variation of the transition energy ($\mathbf{x}$ is the coordinate on the 2D plane), m is the effective mass, d is the average width of the QW, and $\delta d_{\mathbf{x}}$ describes its variation due to irregularities of heteroboundaries.

In this work, we consider the effect of unscreened inhomogeneities of a QW on the shape of the interband absorption peak. We will derive the equations for linear corrections to nondiagonal (in the band number) components of the density matrix, $\delta f_{12}(\mathbf{p}, \mathbf{x})$ and $\delta f_{21}(\mathbf{p}, \mathbf{x})$, taking into account the Coulomb contributions proportional to e^2. Such contributions describe the high-frequency correction to the self-consistent field as well as the influence of exchange effects on intersubband tran-

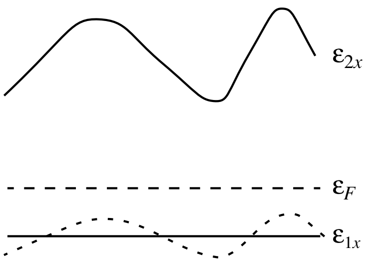


Fig. 1. Variations of energy $\varepsilon_{21}(\mathbf{x})$ of the intersubband transition along the 2D plane in a strongly doped QW.

sitions. The inhomogeneity of the problem in the 2D plane is taken into account in the semiclassical approximation so that equations for $\delta f_{12}(\mathbf{p}, \mathbf{x})$ and $\delta f_{21}(\mathbf{p}, \mathbf{x})$ contain the frequency $(\varepsilon_{21} + \delta\varepsilon_{\mathbf{x}})/\hbar$ of intersubband transitions and conventional drift contributions proportional to $\nabla\delta f_{ss'}(\mathbf{p}, \mathbf{x})$, while Coulomb contributions are determined by integrals over the **p** plane (Section 2). The summation of such a system over 2D momenta gives a chain of balance equations for the moments of the response (Section 3). In the case of extreme long-wave inhomogeneities (the local mode of the response, for which we can disregard the drift contributions and consider only the equation for intersubband polarization), we arrive at a simple expression describing the nonuniform broadening of the absorption peak (Section 4). The nonlocality of the response is taken into account using power approximations for the kernels of the integral terms describing exchange contributions, which enables us to express the intersubband polarization in terms of a one-particle Green function. This function is defined as a continual integral and describes (after averaging over inhomogeneities) the transformation from the Gaussian absorption peak to the Lorentzian shape of the line with a smaller characteristic scale of inhomogeneities (Section 5). The last section is devoted to approximations and concluding remarks.

2. SELF-CONSISTENT DESCRIPTION OF RESPONSE

The linear response of a QW inhomogeneous in the **x** plane, which is excited by a transverse electric field $E_{\perp}\exp(-i\omega t)$ (polarized along the z axis) below $\mathbf{r} = (\mathbf{x}, z)$ is determined by the current density $J_{\omega\mathbf{x}}\exp(-i\omega t)$:

$$J_{\omega\mathbf{x}} = \frac{2e}{L^2}\sum_{\mathbf{p}} v_{\perp}[\delta f_{12}(\mathbf{p}, \mathbf{x}) - \delta f_{21}(\mathbf{p}, \mathbf{x})]. \qquad (2)$$

Here, $v_{\perp} = \langle 2|\hat{v}_z|1\rangle$ is the matrix element of the velocity, which is calculated for the $|s\rangle$ states described by orbitals φ_{sz} (here and below, we take into account only the pair of levels $s = 1$ and $s = 2$ between which transitions take place), L^2 is the normalization area, and the factor 2 appears on account of summation over spins. The linearized density matrix, which is nondiagonal in the band numbers, is written here using the Wigner representation in the variables describing the motion in the 2D plane:

$$\delta f_{ss'}(\mathbf{p}, \mathbf{x}) = \langle s|\widehat{\delta\rho}|s'\rangle_{\mathbf{p},\mathbf{x}}.$$

Here, $\widehat{\delta\rho}$ is the correction to the density matrix, which is proportional to $\exp(-i\omega t)$ and satisfies the operator equation (see, for example, [11])

$$-i\omega\widehat{\delta\rho} + \frac{i}{\hbar}[\tilde{h}, \widehat{\delta\rho}] + \frac{1}{\hbar}[\widehat{\delta h}, \hat{\rho}] = J(\widehat{\delta\rho}), \qquad (3)$$

where $\hat{\rho}$ is the equilibrium density matrix and $J(\widehat{\delta\rho})$ is the linearized collision integral. The steady-state Hamiltonian $\tilde{h}$ appearing in relation (3) includes the one-electron Hamiltonian $\hat{h}$ of a nonideal QW as well as Coulomb corrections proportional to e^2 (self-consistent field and exchange contribution):

$$\tilde{h} = \hat{h} + \sum_{\mathbf{Q}} v_Q[e^{-i\mathbf{Q}\cdot\mathbf{r}}\mathrm{Sp}(e^{i\mathbf{Q}\cdot\mathbf{r}}\hat{\rho}) - e^{-i\mathbf{Q}\cdot\mathbf{r}}\hat{\rho}e^{i\mathbf{Q}\cdot\mathbf{r}}]. \qquad (4)$$

Here, summation is carried out over the 3D wave vector $\mathbf{Q} = (\mathbf{q}, q_{\perp})$, $v_Q = 4\pi e^2/(\epsilon Q^2L^3)$ is the Coulomb matrix element with permittivity ϵ homogeneous across the layer, and the trace also includes the summation over the levels $s = 1, 2$. The contribution to the perturbation operator $\widehat{\delta h}$ in Eq. (3), which is proportional to $\exp(-i\omega t)$, can be written in the form

$$\begin{aligned}\widehat{\delta h} &= \frac{ie}{\omega}E_{\perp}\hat{v}_{\perp} \\ &+ \sum_{\mathbf{Q}} v_Q[\delta n_{\mathbf{Q}}e^{-i\mathbf{Q}\cdot\mathbf{r}} - e^{-i\mathbf{Q}\cdot\mathbf{r}}\widehat{\delta\rho}e^{i\mathbf{Q}\cdot\mathbf{r}}],\end{aligned} \qquad (5)$$

where the contribution from the self-consistent field is expressed in terms of the induced density $\delta n_{\mathbf{Q}} = \mathrm{Sp}[\exp(i\mathbf{Q}\cdot\mathbf{r})\widehat{\delta\rho}]$.

Approximating the collision integral by the relaxation frequency ν, we obtain the following system of equations for $\delta f_{ss'}$,

$$\begin{aligned}&[\omega + i\nu + \omega_{21}(\mathbf{x}, p) + i\mathbf{v}\cdot\nabla_x] + \delta f_{12}(\mathbf{p}, \mathbf{x}) \\ &\qquad + \delta h_{12}(\mathbf{p}, \mathbf{x})f_{Fp}/\hbar = 0, \\ &[\omega + i\nu - \omega_{21}(\mathbf{x}, p) + i\mathbf{v}\cdot\nabla_x] + \delta f_{21}(\mathbf{p}, \mathbf{x}) \\ &\qquad + \delta h_{21}(\mathbf{p}, \mathbf{x})f_{Fp}/\hbar = 0,\end{aligned} \qquad (6)$$

in which $\mathbf{v} = \mathbf{p}/m$ is the velocity of 2D electrons, f_{Fp} is the spatially homogeneous Fermi distribution on the ground level $s = 1$ (we consider the case of an unfilled excited level), and the intersubband transition frequency $\omega_{21}(\mathbf{x}, p) = \varepsilon_{21}(\mathbf{x}, p)/\hbar$ takes into account the random correction to the interleave energy described by relations (1) as well as the exchange contributions:

$$\begin{aligned}\hbar\omega_{21}(\mathbf{x}, p) &= \varepsilon_{21} + \delta\varepsilon_{\mathbf{x}} \\ -\sum_{\mathbf{Q}} v_Q[\langle 2|&e^{-i\mathbf{Q}\cdot\mathbf{r}}\hat{\rho}e^{i\mathbf{Q}\cdot\mathbf{r}}|2\rangle - \langle 1|e^{-i\mathbf{Q}\cdot\mathbf{r}}\hat{\rho}e^{i\mathbf{Q}\cdot\mathbf{r}}|1\rangle].\end{aligned} \qquad (7)$$

The matrix elements of perturbation (5) can be written in the Wigner representation in the form

$$\delta h_{ss'}(\mathbf{p},\mathbf{x}) - \frac{ie}{\omega}E_\perp\begin{bmatrix} v_\perp, (ss') = (21) \\ -v_\perp, (ss') = (12)\end{bmatrix}$$
$$= \int\frac{d\mathbf{p}_1}{2\pi m}\sum_{ab}\delta f_{ab}(\mathbf{p}_1,\mathbf{x}) \tag{8}$$
$$\times\left[2M_{ss'ba}(0) - M_{sabs'}\left(\frac{|\mathbf{p}-\mathbf{p}_1|}{\hbar}\right)\right],$$

where the Coulomb contributions are expressed in terms of the factor

$$M_{abcd}(q) = \frac{e^2 m}{\pi\epsilon\hbar^2}\int_{-\infty}^{\infty}\left(dq_\perp\frac{\langle a|e^{-iq_\perp z}|b\rangle\langle c|e^{iq_\perp z}|d\rangle}{q_\perp^2 + q^2}\right)$$
$$= \frac{1}{a_B q}\int dz\varphi_{az}\varphi_{bz}\int dz'\varphi_{cz'}\varphi_{dz'}e^{-q|z-z'|}. \tag{9}$$

In the equality on the right-hand side, we carried out integration with respect to $q_\perp$. Using such a notation, we can write the frequency of intersubband transitions introduced in Eq. (7) in the form

$$\omega_{21}(\mathbf{x}, p) = \omega_{21}(\mathbf{x}) - \int\frac{d\mathbf{p}_1}{2\pi\hbar m}$$
$$\times f_{Fp_1}\left[M_{2112}\left(\frac{|\mathbf{p}-\mathbf{p}_1|}{\hbar}\right) - M_{1111}\left(\frac{|\mathbf{p}-\mathbf{p}_1|}{\hbar}\right)\right], \tag{10}$$

where $\omega_{21}(\mathbf{x}) = (\varepsilon_{21} + \delta\varepsilon_\mathbf{x})/\hbar = \omega_{21} + \delta\omega_\mathbf{x}$ takes into account the inhomogeneity of the problem in the 2D plane. The last term in relation (10) describes the exchange renormalization of the intersubband transition frequency, which is a function of momentum p.

Further simplification is possible in the limiting 2D case, when $p_F \ll \hbar/d$ and the kernels $M_{abcd}(q)$ of the system of equations under investigation can be presented in the form

$$M_{abcd}(q) \approx \frac{\delta_{ab}\delta_{cd}}{a_B q} - \frac{L_{abcd}}{a_B},$$
$$L_{abcd} = \int dz\int dz'\varphi_{az}\varphi_{bz}\varphi_{cz'}\varphi_{dz'}|z - z'|, \tag{11}$$

where a_B is the Bohr radius and the characteristic length L_{abcd} has been introduced. It is convenient to introduce, instead of $\delta f_{ss'}$, new unknown functions $\psi^{\pm}_{\mathbf{p},\mathbf{x}}$ defined as

$$\delta f_{12}(\mathbf{p},\mathbf{x}) \pm \delta f_{21}(\mathbf{p},\mathbf{x}) = \frac{2ie}{\hbar\omega}E_\perp v_\perp f_{Fp}\psi^{\pm}_{\mathbf{p},\mathbf{x}}. \tag{12}$$

As a result, the system of equations (6) can be transformed to

$$[\omega + i\nu + i\mathbf{v}\cdot\nabla_x]\psi^{+}_{\mathbf{p},\mathbf{x}}$$
$$+ \omega_{21}(\mathbf{x})\psi^{-}_{\mathbf{p},\mathbf{x}} + \Omega(\psi^{-}|\mathbf{p},\mathbf{x}) = 1,$$
$$[\omega + i\nu + i\mathbf{v}\cdot\nabla_x]\psi^{-}_{\mathbf{p},\mathbf{x}} \tag{13}$$
$$+ \omega_{21}(\mathbf{x})\psi^{+}_{\mathbf{p},\mathbf{x}} + \Omega(\psi^{+}|\mathbf{p},\mathbf{x}) = 0.$$

This system should be considered in the region $|\mathbf{p}| < p_F$. The Coulomb contributions $\Omega(\psi^{\pm}|\mathbf{p},\mathbf{x})$ in Eqs. (13) are defined by integrals,

$$\begin{bmatrix}\Omega(\psi^{+}|\mathbf{p},\mathbf{x}) \\ \Omega(\psi^{-}|\mathbf{p},\mathbf{x})\end{bmatrix} = \int_{|\mathbf{p}|<p_F}\frac{d\mathbf{p}_1}{2\pi\hbar m}(\psi^{\pm}_{\mathbf{p},\mathbf{x}} - \psi^{\pm}_{\mathbf{p}_1,\mathbf{x}})$$
$$\times\left(\frac{\hbar}{a_B|\mathbf{p}-\mathbf{p}_1|} + \frac{L_{1212}}{a_B}\right) - \int_{|\mathbf{p}|<p_F}\frac{d\mathbf{p}_1}{2\pi\hbar m} \tag{14}$$
$$\times\left(\psi^{\pm}_{\mathbf{p},\mathbf{x}}\frac{L_{1111}}{a_B} - \psi^{\pm}_{\mathbf{p}_1,\mathbf{x}}\begin{bmatrix}(L_{1122} - 2L_{1212})/a_B \\ L_{1122}/a_B\end{bmatrix}\right),$$

so that Eqs. (13) form a system of differential equations in **x** (in first-order partial derivatives) and integral equations in **p**.

3. BALANCE EQUATIONS

In accordance with Eqs. (2) and (12), the induced current density can be written in terms of $\psi^{-}_{\mathbf{p},\mathbf{x}}$ as

$$J_{\omega\mathbf{x}} = i\frac{(2e)^2|v_\perp|^2}{\hbar\omega L^2}E_\perp{\sum_{\mathbf{p}}}'\psi^{-}_{\mathbf{p},\mathbf{x}} = \sigma_{\omega\mathbf{x}}E_\perp, \tag{15}$$

where the sum with the prime corresponds to summation inside the Fermi circle $|\mathbf{p}| < p_F$. The local conductivity $\sigma_{\omega\mathbf{x}}$ in Eq. (15) contains the factor $\delta n^{-}_{\mathbf{x}} = (2/L^2){\sum_{\mathbf{p}}}'\psi^{-}_{\mathbf{p},\mathbf{x}}$, whose dimension is equal to concentration divided by frequency. Relative absorption $\xi(\omega)$ of a 2D layer (defined as the ratio of energy fluxes absorbed in the QW and transmitted through the 2D layer; see [2]) is introduced through the relation

$$\xi(\omega) = \frac{4\pi}{c\sqrt{\varepsilon}}\mathrm{Re}\langle\sigma_{\omega\mathbf{x}}\rangle = \frac{e^2}{\hbar c}\frac{8\pi|v_\perp|^2}{\sqrt{\varepsilon}\omega}\mathrm{Im}\langle\delta n^{-}_{\mathbf{x}}\rangle, \tag{16}$$

where $\langle\ldots\rangle$ denotes averaging over inhomogeneities in the QW thickness. Carrying out the summation of Eq. (13) over the **p** plane, we obtain the system of balance equations for $\delta n^{\pm}_{\mathbf{x}}$,

$$\begin{gathered}(\omega+i\nu)\delta n_{\mathbf{x}}^{+}+i\nabla_x\cdot\mathbf{i}_{\mathbf{x}}^{+}+\omega_{21}(\mathbf{x})\delta n_{\mathbf{x}}^{-}\\ +\frac{2}{L^2}\sum_{\mathbf{p}}{}'\Omega(\psi^{-}|\mathbf{p},\mathbf{x}) = \frac{2}{L^2}\sum_{\mathbf{p}}{}'1,\\ (\omega+i\nu)\delta n_{\mathbf{x}}^{-}+i\nabla_x\cdot\mathbf{i}_{\mathbf{x}}^{-}+\omega_{21}(\mathbf{x})\delta n_{\mathbf{x}}^{+}\\ +\frac{2}{L^2}\sum_{\mathbf{p}}{}'\Omega(\psi^{+}|\mathbf{p},\mathbf{x}) = 0,\end{gathered}\tag{17}$$

in which $\sum_{\mathbf{p}}'1$ denotes the summation over filled states and the flux contributions

$$\mathbf{i}_{\mathbf{x}}^{\pm} = \frac{2}{L^2}\sum_{\mathbf{p}}{}'\mathbf{v}\psi_{\mathbf{p},\mathbf{x}}^{\pm}$$

appear. The system of equations for $\mathbf{i}_{\mathbf{x}}^{\pm}$ is obtained as a result of summation of Eqs. (13) with weight $\mathbf{v}$,

$$\begin{gathered}(\omega+i\nu)\mathbf{i}_{\mathbf{x}}^{+}+\frac{i}{2}\nabla_x w_{\mathbf{x}}^{+}+\omega_{21}(\mathbf{x})\mathbf{i}_{\mathbf{x}}^{-}\\ +\frac{2}{L^2}\sum_{\mathbf{p}}{}'\mathbf{v}\Omega(\psi^{-}|\mathbf{p},\mathbf{x}) = 0,\\ (\omega+i\nu)\mathbf{i}_{\mathbf{x}}^{-}+\frac{i}{2}\nabla_x w_{\mathbf{x}}^{-}+\omega_{21}(\mathbf{x})\mathbf{i}_{\mathbf{x}}^{+}\\ +\frac{2}{L^2}\sum_{\mathbf{p}}{}'\mathbf{v}\Omega(\psi^{+}|\mathbf{p},\mathbf{x}) = 0,\end{gathered}\tag{18}$$

and acquires the following moments:

$$\frac{2}{L^2}\sum_{\mathbf{p}}{}'v_\alpha v_\beta\psi_{\mathbf{p},\mathbf{x}}^{\pm}\approx\frac{\delta_{\alpha\beta}w_{\mathbf{x}}^{\pm}}{2}.$$

Here, we have used the smallness of anisotropy $\psi_{\mathbf{p},\mathbf{x}}^{\pm}$ in the case of smooth inhomogeneities and retained only the diagonal contribution in α, β while writing the next mean $w_{\mathbf{x}}^{\pm}$. After summation of system (13) over the **p** plane with weight v^2, we obtain the following equations for this contribution:

$$\begin{gathered}(\omega+i\nu)w_{\mathbf{x}}^{+}+\omega_{21}(\mathbf{x})w_{\mathbf{x}}^{-}\\ +\frac{2}{L^2}\sum_{\mathbf{p}}{}'v^2\Omega(\psi^{-}|\mathbf{p},\mathbf{x}) = \frac{2}{L^2}\sum_{\mathbf{p}}{}'v^2,\\ (\omega+i\nu)w_{\mathbf{x}}^{-}+\omega_{21}(\mathbf{x})w_{\mathbf{x}}^{+}\\ +\frac{2}{L^2}\sum_{\mathbf{p}}{}'v^2\Omega(\psi^{+}|\mathbf{p},\mathbf{x}) = 0,\end{gathered}\tag{19}$$

where we have neglected the gradients $(2/L^2)\sum_{\mathbf{p}}'v^2\mathbf{v}\psi_{\mathbf{p},\mathbf{x}}^{\pm}$ of the next moments.

Let us now consider the Coulomb contributions to the chain of equations (17)–(19). In the summation of the integral term in Eq. (14) over **p**, the contribution proportional to $|\mathbf{p}-\mathbf{p}_1|^{-1}$ is left out and we arrive at the simple relation

$$\frac{2}{L^2}\sum_{\mathbf{p}}{}'\Omega(\psi^{\pm}|\mathbf{p},\mathbf{x}) = \delta\omega_{\pm}\delta n_{\mathbf{x}}^{\pm},$$

in which the Coulomb corrections to frequencies have the form

$$\begin{gathered}\delta\omega_{\pm} = \frac{\varepsilon_F}{\hbar a_B}\\ \times\begin{cases}L_{1122}-L_{1111}-2L_{1212}\\ L_{1122}-L_{1111}\end{cases} = \frac{\varepsilon_F}{\hbar}\frac{d}{a_B}A_{\pm}.\end{gathered}\tag{20}$$

The last equality is written here after calculating L_{abcd} by formula (11) for the model of QW of width d with high barriers, when $L_{1111}\approx 0.2d$, $L_{1122}\approx 0.27d$, and $L_{1212}\approx -0.11d$ so that coefficients $A_{\pm}$ are equal to $A_{+}\approx 0.29$ and $A_{-}\approx 0.063$. In the system of equations (18), there appear the Coulomb contributions

$$\begin{gathered}\frac{2}{L^2}\sum_{\mathbf{p}}{}'\mathbf{v}\Omega(\psi^{\pm}|\mathbf{p},\mathbf{x}) = \frac{\varepsilon_F}{\hbar}\left[\frac{L_{1212}-L_{1111}}{a_B}\mathbf{i}_{\mathbf{x}}^{\pm}\right.\\ \left.+\frac{\hbar}{a_B p_F}\frac{2}{L^2}\sum_{\mathbf{p}}{}'\psi_{\mathbf{p},\mathbf{x}}^{\pm}v_F\frac{\mathbf{p}}{p}S\left(\frac{p}{p_F}\right)\right],\end{gathered}\tag{21}$$

while system (19) contains the contributions

$$\begin{gathered}\frac{2}{L^2}\sum_{\mathbf{p}}{}'v^2\Omega(\psi^{\pm}|\mathbf{p},\mathbf{x})\\ = \frac{\varepsilon_F}{\hbar}\left[\frac{\hbar}{a_B p_F}\frac{2}{L^2}\sum_{\mathbf{p}}{}'\psi_{\mathbf{p},\mathbf{x}}^{\pm}v_F^2R\left(\frac{p}{p_F}\right)+\frac{L_{1212}-L_{1111}}{a_B}w_{\mathbf{x}}^{\pm}\right.\\ \left.-\begin{Bmatrix}3L_{1212}-L_{1122}\\ L_{1212}-L_{1122}\end{Bmatrix}\delta n_{\mathbf{x}}^{\pm}\frac{v_F^2}{2a_B}\right].\end{gathered}\tag{22}$$

In these expressions, the integrals proportional to $|\mathbf{p}-\mathbf{p}_1|^{-1}$ do not vanish and are expressed in terms of the dimensionless functions $S(p/p_F)$ and $R(p/p_F)$.

The functions $S(z)$ and $R(z)$ in (21) and (22) are defined by the integrals

$$\begin{gathered}S(z) = \int\limits_{x<1}\frac{d\mathbf{x}}{\pi z^2}\frac{z^2-\mathbf{x}\cdot\mathbf{z}}{|\mathbf{z}-\mathbf{x}|}\approx S,\\ R(z) = \int\limits_{x<1}\frac{d\mathbf{x}}{\pi}\frac{z^{2-x^2}}{|\mathbf{z}-\mathbf{x}|}\approx -R+rz^2\end{gathered}\tag{23}$$

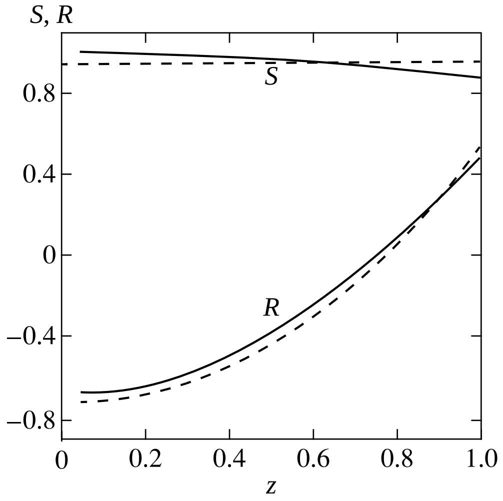


Fig. 2. Functions $S(z)$ and $R(z)$ defined by formulas (23); the dashed curves correspond to the approximations used for closing the balance equations.

and are quite smooth (see the results of numerical calculations presented in Fig. 2). These dependences can be approximated by a constant and a parabola written in the right equalities of Eqs. (23). The following numerical values of these parameters can be chosen with an accuracy of about 5%: $S \approx 0.95$, $R \approx 0.7$, and $r \approx 1.35$. When such an approximation is used, the integral contribution on the right-hand side of Eq. (21) can be expressed in terms of $\delta n_{\mathbf{x}}^{\pm}$ and $\mathbf{i}_{\mathbf{x}}^{\pm}$ and the right-hand side of Eq. (22) acquires the terms $\delta n_{\mathbf{x}}^{\pm}$ and $w_{\mathbf{x}}^{\pm}$. Thus, Eqs. (17)–(19) form an approximate closed system of differential equations (17), (18) and algebraic equations (19). These equations contain the random correction $\delta\omega_{\mathbf{x}}$ to the transition frequency, which is responsible for the absorption peak broadening.

4. LOCAL APPROXIMATION

Let us first consider the case of extremely smooth inhomogeneities, when the flux contributions $\nabla_x \cdot \mathbf{i}_{\mathbf{x}}^{\pm}$ in Eqs. (17) can be neglected. This leads to the simple algebraic system for $\delta n_{\mathbf{x}}^{\pm}$,

$$\begin{vmatrix} \omega + i\nu & \tilde{\omega}_{21} + \delta\omega_{\mathbf{x}} \\ \bar{\omega}_{21} + \delta\omega_{\mathbf{x}} & \omega + i\nu \end{vmatrix} \begin{vmatrix} \delta n_{\mathbf{x}}^{+} \\ \delta n_{\mathbf{x}}^{-} \end{vmatrix} = \begin{vmatrix} n_{2D} \\ 0 \end{vmatrix}, \qquad (24)$$

where n_{2D} is the concentration of 2D electrons, and the frequencies $\tilde{\omega}_{21} = \omega_{21} + \delta\omega_{-}$ and $\bar{\omega}_{21} = \omega_{21} + \delta\omega_{+}$ have been introduced. The solution of system (24) can easily be determined, and it remains for us to calculate the average $\langle \delta n_{\mathbf{x}}^{-} \rangle$. However, such an averaging in the general case is quite cumbersome and we will confine our analysis to the shape of the curves near resonance. In this approximation, it is convenient to diagonalize first the non-Hermitian matrix

$$\begin{vmatrix} 0 & \tilde{\omega}_{21} \\ \bar{\omega}_{21} & 0 \end{vmatrix}$$

using the nonunitary transformation

$$\hat{\Lambda} = \begin{vmatrix} \bar{\eta} & -\tilde{\eta} \\ \bar{\eta} & \tilde{\eta} \end{vmatrix}, \quad \hat{\Lambda}^{-1} = \frac{1}{2}\begin{vmatrix} 1/\bar{\eta} & 1/\bar{\eta} \\ -\tilde{\eta} & \tilde{\eta} \end{vmatrix}, \qquad (25)$$

in which the parameters

$$\bar{\eta} = \sqrt{\bar{\omega}_{21}/(\tilde{\omega}_{21} + \bar{\omega}_{21})}$$

and

$$\tilde{\eta} = \sqrt{\tilde{\omega}_{21}/(\tilde{\omega}_{21} + \bar{\omega}_{21})}$$

have been introduced. The result of diagonalization has the form

$$\hat{\Lambda}\begin{vmatrix} 0 & \tilde{\omega}_{21} \\ \bar{\omega}_{21} & 0 \end{vmatrix}\hat{\Lambda}^{-1} = \begin{vmatrix} -\Omega_{21} & 0 \\ 0 & \Omega_{21} \end{vmatrix}, \qquad (26)$$

where $\Omega_{21} = \sqrt{\bar{\omega}_{21}\tilde{\omega}_{21}}$ is the resonance transition frequency. For a small displacement of the peak, we have

$$\Omega_{21} \approx \omega_{21} + A\frac{\varepsilon_F}{\hbar}\frac{d}{a_B}$$

with coefficient $A \approx 0.18$ calculated for the model of a QW with high barriers, which was used for deriving Eq. (20). We can now carry out a similar transformation for matrix

$$\begin{vmatrix} 0 & \delta\omega_{\mathbf{x}} \\ \delta\omega_{\mathbf{x}} & 0 \end{vmatrix}$$

and for the inhomogeneous part of Eqs. (24) by writing the system of equations for

$$\hat{\Lambda}\begin{vmatrix} \delta n_{\mathbf{x}}^{+} \\ \delta n_{\mathbf{x}}^{-} \end{vmatrix}.$$

For frequencies close to Ω_{21}, only the resonance component $P_{\mathbf{x}}$ is large in such a column, which satisfies the equation

$$[\delta\varepsilon + i\gamma - u_{\mathbf{x}}]P_{\mathbf{x}} = n_{2D}\bar{\eta}. \qquad (27)$$

Here, we have introduced the energy detuning $\delta\varepsilon = \hbar(\omega - \Omega_{21})$ and the energy of uniform broadening $\gamma = \hbar\nu$ as well as the random potential $u_{\mathbf{x}} = K\hbar\delta\omega_{\mathbf{x}}$ (parameter $K = (\sqrt{\tilde{\omega}_{21}/\bar{\omega}_{21}} + \sqrt{\bar{\omega}_{21}/\tilde{\omega}_{21}})/2$ is close to unity for the small Coulomb renormalization of the intersubband transition energy, $\delta\omega_{\pm} \ll \omega_{21}$). The function $\delta n_{\mathbf{x}}^{-}$

appearing in Eq. (16) can be expressed in terms of $P_\mathbf{x}$ in accordance with relation $\delta n_\mathbf{x}^- \approx -\hbar P_\mathbf{x}/2\tilde{\eta}$.

Substitution of the solution of Eq. (27) into relation (16) gives the relative absorption

$$\xi(\delta\varepsilon) = -\frac{e^2}{\hbar c}\frac{4\pi|v_\perp|^2\hbar n_{2D}}{\sqrt{\varepsilon}\omega} \times \sqrt{\frac{\bar{\omega}_{21}}{\tilde{\omega}_{21}}}\mathrm{Im}\langle[\delta\varepsilon + i\gamma - u_\mathbf{x}]^{-1}\rangle. \tag{28}$$

In order to average this expression over the random potential $u_\mathbf{x}$, it is convenient to use the relation

$$(A + i\gamma)^{-1} = -i\int_{-\infty}^{0} d\tau \exp[(-iA + \gamma)\tau].$$

In this case, $u_\mathbf{x}$ appears in the exponent and, hence, the average over the variations of the QW thickness can be obtained using the relation

$$\langle\exp(-iu_\mathbf{x}\tau)\rangle = \exp[-(K\overline{\delta\varepsilon}\tau)^2/2],$$

where $\overline{\delta\varepsilon}$ is the mean square variation of the interband energy (1). As a result, the relative absorption is given by the formula (which can also be written in terms of the probability integral of a complex argument)

$$\xi(\delta\varepsilon) = \frac{e^2}{\hbar c}\frac{4\pi|v_\perp|^2 n_{2D}}{\sqrt{\varepsilon}\omega}\sqrt{\frac{\bar{\omega}_{21}}{\tilde{\omega}_{21}}}\mathrm{Re}\int_{-\infty}^{0} dt \exp\left[\frac{(\gamma - i\delta\varepsilon)t}{\hbar}\right] \times \exp\left[-\frac{1}{\sqrt{2}}\left(\frac{\Gamma t}{\hbar}\right)^2\right] = \frac{e^2}{\hbar c}\frac{4\pi|v_\perp|^2\hbar n_{2D}}{\sqrt{\varepsilon}\omega}\sqrt{\frac{\bar{\omega}_{21}}{\tilde{\omega}_{21}}} \times \begin{cases} 2\gamma/[(\delta\varepsilon)^2 + \gamma^2], & \gamma \gg \Gamma, \\ \sqrt{2\pi}\Gamma\exp[-(\delta\varepsilon/\Gamma)^2/2], & \gamma \ll \Gamma, \end{cases} \tag{29}$$

where we have introduced the characteristic energy of nonuniform broadening, $\Gamma = 6K\varepsilon_1\overline{\delta d}/d$, corresponding to the mean square variation of the QW width equal to $\overline{\delta d}$. The last equality gives the asymptotic forms for dominating collision or nonuniform mechanisms of broadening, for which a Lorentzian of a Gaussian peak is realized. Figure 3 illustrates the transformation of the line shape between these limiting cases upon an increase in the ratio Γ/γ. Here, the maximum value $\xi_{\max} = \xi(\delta\varepsilon = 0)$ of relative absorption has also been introduced:

$$\xi_{\max} = \frac{e^2}{\hbar c}\frac{(2\pi)^{3/2}|v_\perp|^2\hbar n_{2D}}{\sqrt{\varepsilon}\Omega_{21}\Gamma}\sqrt{\frac{\bar{\omega}_{21}}{\tilde{\omega}_{21}}}G\left(\frac{\gamma}{\sqrt{2}\Gamma}\right), \quad G(z) = e^{z^2/2}[1 + \Phi(z)], \tag{30}$$

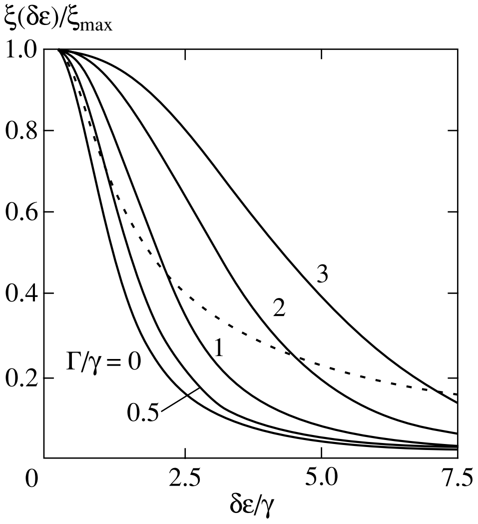


Fig. 3. Transformation of the shape of the absorption peak upon an increase in the contribution of nonuniform broadening (the values of Γ/γ are given on the curves); the function $G(z)$ from Eq. (30), describing the decrease in the peak height upon its broadening, is shown by the dashed curve.

where the function $G(z)$ describing the decrease in the peak height upon an increase in its width is expressed through the probability integral $\Phi(z)$.

The numerical estimate of the value of Γ for the parameters of GaAs QW corresponding to the spectral region of 10 μm in the case of monolayer variations of heteroboundaries ($\overline{\delta d} \approx 2.5$ Å) gives a line width of approximately 10 meV, which is in accord with a number of measurements (see Chapter 8 in [2] and [12]). Recent measurements [5, 7] of smaller peak widths (of the order of a few millielectronvolts) became possible owing to a higher quality of heteroboundaries (for which $\overline{\delta d} \approx 1$ Å). These measurements also demonstrate a noticeable decrease in the value of Γ upon an increase in the QW width, which agrees with the results of our analysis. A similar estimation for an InAs QW with the parameters corresponding to the spectral region of 5 μm gives a value of Γ of the order of 30 meV, which corresponds to the results obtained in [6]. The measurements of emission in the submillimeter spectral region [13–15] give line widths of the order of several millielectronvolts, which also conforms to the expression for Γ presented by us here (in these numerical estimates, it was assumed that the energy $3\varepsilon_1$ is close to the energy of the transition between energy levels). Although a detailed analysis of the shape of the peak requires additional self-consistent calculations and special measurements, a satisfactory agreement of the simple expression for Γ with the experimental results demonstrates a significant role of nonuniform broadening in the experiments [5–7] and [12–15].

5. NONLOCAL BROADENING MODE

Let us now consider the case of finite values of correlation length l_c, for which the flux contributions to the formation of the intersubband absorption peak must be taken into account and the equations for $\delta n_{\mathbf{x}}^{\pm}$, $\mathbf{i}_{\mathbf{x}}^{\pm}$, and $w_{\mathbf{x}}^{\pm}$ must be analyzed. Substituting approximations (23) into the right-hand sides of (21) and (22), we can write Eqs. (18) and (19) in the form

$$\begin{vmatrix} \omega + i\nu & \tilde{\omega} + \delta\omega_{\mathbf{x}} \\ \tilde{\omega} + \delta\omega_{\mathbf{x}} & \omega + i\nu \end{vmatrix} \begin{vmatrix} \mathbf{i}_{\mathbf{x}}^{+} \\ \mathbf{i}_{\mathbf{x}}^{-} \end{vmatrix} = -\frac{i}{2} \begin{vmatrix} \nabla_x w_{\mathbf{x}}^{+} \\ \nabla_x w_{\mathbf{x}}^{-} \end{vmatrix}, \tag{31}$$

$$\begin{aligned} &\begin{vmatrix} \omega + i\nu & \bar{\omega} + \delta\omega_{\mathbf{x}} \\ \bar{\omega} + \delta\omega_{\mathbf{x}} & \omega + i\nu \end{vmatrix} \begin{vmatrix} w_{\mathbf{x}}^{+} \\ w_{\mathbf{x}}^{-} \end{vmatrix} \\ &= \frac{\varepsilon_F}{\hbar} \begin{vmatrix} (\varepsilon_F/\hbar)/\pi + \alpha v_F^2 \delta n_{\mathbf{x}}^{-} \\ \beta v_F^2 \delta n_{\mathbf{x}}^{+} \end{vmatrix}, \end{aligned} \tag{32}$$

where we have introduced the parameters

$$\alpha = \frac{3L_{1212} - L_{1122}}{2a_B} + \frac{R\hbar}{a_B p_F},$$

$$\beta = \frac{L_{1212} - L_{1122}}{2a_B} + \frac{R\hbar}{a_B p_F},$$

as well as the renormalization frequencies

$$\begin{aligned} \tilde{\omega} &= \omega_{21} + \Delta\omega + R\frac{\varepsilon_F}{a_B p_F}, \\ \bar{\omega} &= \omega_{21} + \Delta\omega + r\frac{\varepsilon_F}{a_B p_F}, \end{aligned} \tag{33}$$

and $\hbar\Delta\omega = \varepsilon_F(L_{1212} - L_{1111})/2a_B$. The substitution of the solution of Eq. (32) into Eq. (31) and the subsequent substitution of the solution of Eq. (31) into Eq. (17) give a closed system of equations for $\delta n_{\mathbf{x}}^{\pm}$.

Confining our analysis to the resonance approximation, we will henceforth disregard small random corrections in nonresonance contributions while writing the solutions of Eqs. (31) and (32) and assume that the value of ω is close to Ω_{21}. As a result, we obtain the following system of equations for the functions $\delta n_{\mathbf{x}}^{\pm}$ determining the intersubband polarization:

$$\begin{aligned} &\left\{ \hat{\mathcal{K}} \nabla_x^2 + \begin{vmatrix} \omega + i\nu & \tilde{\omega}_{21} + \delta\omega_{\mathbf{x}} \\ \tilde{\omega}_{21} + \delta\omega_{\mathbf{x}} & \omega + i\nu \end{vmatrix} \right\} \\ &\times \begin{vmatrix} \delta n_{\mathbf{x}}^{+} \\ \delta n_{\mathbf{x}}^{-} \end{vmatrix} = \begin{vmatrix} n_{2D} \\ 0 \end{vmatrix}, \end{aligned} \tag{34}$$

which differs from system (24) only in the emergence of a nonlocal (proportional to ∇_x^2) contribution determined by the tensor

$$\begin{aligned} \hat{\mathcal{K}} &\approx \frac{\varepsilon_F v_F^2/2\hbar}{(\Omega_{21}^2 - \tilde{\omega}^2)(\Omega_{21}^2 - \bar{\omega}^2)} \\ &\times \begin{vmatrix} -\alpha\Omega_{21}(\tilde{\omega} + \bar{\omega}) & \beta(\Omega_{21}^2 + \tilde{\omega}\bar{\omega}) \\ \alpha(\Omega_{21}^2 + \tilde{\omega}\bar{\omega}) & -\beta\Omega_{21}(\tilde{\omega} + \bar{\omega}) \end{vmatrix}. \end{aligned} \tag{35}$$

Further, we carry out the transformation of Eq. (25) and separate the resonance component of system (34), which leads to the equation

$$\left[\frac{\hbar^2}{2M}\nabla_x^2 - u_{\mathbf{x}} + \delta\varepsilon + i\gamma\right] P_{\mathbf{x}} = n_{2D}\bar{\eta}, \tag{36}$$

generalizing Eq. (27) by taking into account the nonlocal contribution with the effective mass

$$M = -m\frac{\hbar^2(\Omega_{21} - \bar{\omega})(\Omega_{21} - \tilde{\omega})}{(\alpha + \beta)\varepsilon_F^2} \tag{37}$$

(here $M = \hbar/2\mathcal{K}_{11}$) along with the random potential $u_{\mathbf{x}} \propto \delta\omega_{\mathbf{x}}$.

It is convenient to write the solution of Eq. (36) by introducing the Green function of this equation in accordance with

$$\left[\frac{\hat{\mathbf{p}}^2}{2M} + u_{\mathbf{x}} - \delta\varepsilon + i\gamma\right] g_{\delta\varepsilon}(\mathbf{x}, \mathbf{x}') = \delta(\mathbf{x} - \mathbf{x}') \tag{38}$$

($\hat{\mathbf{p}} \equiv -i\hbar\nabla_x$). Substituting now $\delta n_{\mathbf{x}}^{-} = -\hbar P_{\mathbf{x}}/2\tilde{\eta}$ into relation (16), we obtain the following expression for relative absorption:

$$\xi(\delta\varepsilon) = \frac{e^2}{\hbar c}\frac{4\pi|v_\perp|^2 n_{2D}\hbar}{\sqrt{\varepsilon}\Omega_{21}} \mathrm{Im}\left\langle \int d\mathbf{x}' g_{\delta\varepsilon}(\mathbf{x}, \mathbf{x}') \right\rangle. \tag{39}$$

The Green function introduced into Eq. (38) in the standard manner can be written in terms of the continual integral. In this integral, we carry out averaging over the random potential and separate a rectilinear trajectory along the interval ($\mathbf{x}$, $\mathbf{x}'$) (see [16]). As a result, $\xi(\delta\varepsilon)$ can be expressed in terms of the continual contour integral,

$$\begin{aligned} \xi(\delta\varepsilon) &= \frac{e^2}{\hbar c}\frac{4\pi|v_\perp|^2 n_{2D}\hbar}{\sqrt{\varepsilon}\Omega_{21}} \mathrm{Im}\frac{i}{\hbar}\int d\mathbf{u} \\ &\times \int_{-\infty}^{0} dt \exp\left[\frac{i(\delta\varepsilon - i\gamma)t}{\hbar}\right] \exp\left(\frac{iMu^2}{2\hbar t}\right) \\ &\times \oint \mathcal{D}\{\mathbf{x}_t\} \exp\left[\frac{iM}{2\hbar}\int_0^t d\tau \dot{x}_\tau^2\right. \\ &\left. - \frac{1}{\hbar^2}\int_0^t d\tau \int_0^i d\tau' W\left(\left|\mathbf{u}\frac{\tau - \tau'}{t} + \mathrm{x}_\tau - \mathrm{x}_{\tau'}\right|\right)\right], \end{aligned} \tag{40}$$

which is independent of $\mathbf{x}$. Here, $\mathbf{u} = \mathbf{x} - \mathbf{x}'$ and we have used the Gaussian correlation function

$$W(|\mathbf{x} - \mathbf{x}'|) = \langle \delta\varepsilon_{\mathbf{x}} \delta\varepsilon_{\mathbf{x}'} \rangle = (\overline{\delta\varepsilon})^2 \exp[-|\mathbf{x} - \mathbf{x}'|^2/l_c^2]$$

with the correlation length l_c and the characteristic energy $\overline{\delta\varepsilon} = \Gamma/K$ (the broadening Γ was introduced into Eq. (29); henceforth, K will be replaced by unity).

In order to estimate the contributions of deviations from the rectilinear trajectory to the continual integral (40), we use the method of the optimal trajectory. According to [16], such a trajectory corresponds to the minimal variation of action in the exponent in Eq. (40) and can be determined from the Euler–Lagrange equation

$$M\ddot{\mathbf{x}}_\tau - \frac{2i}{\hbar l_c^2} \int_0^t d\tau' \left(\mathbf{u}\frac{\tau - \tau'}{t} + \mathbf{x}_\tau - \mathbf{x}_{\tau'} \right) \times w\left(\left| \mathbf{u}\frac{\tau - \tau'}{t} + \mathrm{x}_\tau - \mathrm{x}_{\tau'} \right| \right) = 0. \tag{41}$$

This equation must be solved with the initial conditions $\mathbf{x}_{\tau = 0,\, t} = 0$. The right-hand side of Eq. (41) can be estimated by using the inequality $|\mathbf{x}W(x)| \le l_c \overline{\delta\varepsilon}^2$, which gives a simple upper estimate for $\mathbf{x}_\tau$ (see analogous calculations in Appendix B in [17]). The maximum deviation of the optimal trajectory from the rectilinear trajectory, $\mathbf{x}_\tau^{\max}$, takes place for $\tau = t/2$. Thus, we can disregard the deviations from the rectilinear trajectory while writing the correlation function in Eq. (40) if the following inequality is satisfied:

$$|\mathbf{u}| \frac{\overline{\delta\varepsilon} t^2}{2M\hbar l_c} \ll l_c. \tag{42}$$

After substituting the continual integral into Eq. (40), we obtain the expression for the relative absorption for free motion in the form of a multiple integral:

$$\xi(\delta\varepsilon) = \frac{e^2}{\hbar c} \frac{4\pi |v_\perp|^2 n_{2D} \hbar}{\sqrt{\varepsilon}\Omega_{21}} \times \mathrm{Im} \int d\mathbf{u} \int_{-\infty}^{0} \frac{dtM}{2\pi\hbar t} \exp\left[\frac{i(\delta\varepsilon - i\gamma)t}{\hbar} \right] \times \exp\left[i\frac{Mu^2}{2\hbar t} - \frac{t^2}{2\hbar^2} \int_0^1 dx \int_0^1 dx' W(|\mathbf{u}|(x - x')) \right]. \tag{43}$$

It should be noted that time intervals of the order of $\hbar/\gamma_{\mathrm{eff}}$ are significant for calculating Eq. (43); here γ_{eff} determines the resulting peak broadening.

For large-scale variations of the QW width, for which

$$\varepsilon_c \equiv \frac{2}{M}\left(\frac{\hbar}{l_c}\right)^2 \ll \gamma_{\mathrm{eff}},$$

the term in Eq. (43) describing the nonuniform broadening assumes the form $\exp[-(\overline{\delta\varepsilon}\, t/\hbar)^2/2]$; here, the energy ε_c has been introduced for estimating the drift contributions to the peak broadening. After elementary integration with respect to $\mathbf{u}$, relative absorption is given by the simple integral (29) with respect to time. Thus, the inequality $\varepsilon_c \ll \gamma_{\mathrm{eff}}$ defines the limits of applicability of the local approximation discussed in Section 4. Such a dependence is also valid in the tails of the absorption peak under the condition $\varepsilon_c \le |\delta\varepsilon|$ so that the Lorentzian spectrum is always realized for high detuning energies.

In the opposite limiting case $\varepsilon_c \gg \gamma_{\mathrm{eff}}$ of small-scale inhomogeneities, a more detailed analysis of the shape of the central part of the peak is required, where $|\delta\varepsilon| \le \varepsilon_c$. Since only small quantities $|x - x'|$ make a contribution to the term in Eq. (40) proportional to t^2, the relative absorption is transformed to

$$\xi(\delta\varepsilon) \approx \frac{e^2}{\hbar c} \frac{4|v_\perp|^2 n_{2D}}{\sqrt{\varepsilon}\Omega_{21}} \mathrm{Im} \int_0^\infty dt \exp\left[\frac{i(\delta\varepsilon + i\gamma)t}{\hbar} \right] \times \int d\mathbf{s} \exp(iv^2) \exp\left[-\frac{\sqrt{\pi}\,\overline{\delta\varepsilon}^2 (t/\hbar)^{3/2}}{2v\sqrt{\varepsilon_c}} \right], \tag{44}$$

where the direction of the time axis is changed and the dimensionless variable $\mathbf{s}$ is introduced instead of displacement $\mathbf{u}$. The term in the exponent proportional to $\overline{\delta\varepsilon}$ can be replaced by unity under the condition

$$\Gamma_{\mathrm{sr}}^{3/2} \ll \gamma_{\mathrm{eff}}^{3/2}, \quad \Gamma_{\mathrm{sr}} \equiv \sqrt[3]{\frac{\pi}{4}} \frac{\overline{\delta\varepsilon}^{4/3}}{\varepsilon_c^{1/3}} = \sqrt[3]{\pi} \frac{\Gamma^{4/3}}{\varepsilon_c^{1/3}}, \tag{45}$$

where Γ_{sr} is the characteristic energy of broadening in the case of small-scale inhomogeneities. In the limiting case (45), the absorption peak is of the Lorentzian type with a uniform collision broadening γ. If, however, the inequality opposite to (45) is realized, the integration in Eq. (44) should be carried out numerically.

Figure 4 shows the dependences of relative absorption on $\delta\varepsilon/\gamma$ for various values of the ratio $\gamma/\Gamma_{\mathrm{sr}}$. For large values of this parameter, a conventional Lorentzian peak of half-width γ is realized. As the value of $\gamma/\Gamma_{\mathrm{sr}}$ decreases, the peak becomes slightly asymmetric and its half-width is of the order of Γ_{sr}; the value of $\xi(\delta\varepsilon)$ also varies slowly in the vicinity of the peak. Such a shape of the peak is better approximated by the sum of two Lorentzian contributions with noticeably different half-widths than by a Lorentzian or Gaussian curve with any parameters. A more complex approximation for the absorption peak determined by formula (44) can

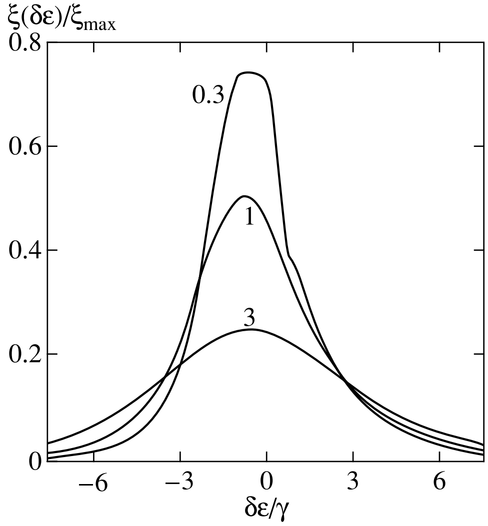


Fig. 4. Relative absorption (in the units of $\bar{\xi}_{max} = (e^2/\hbar c)(4\pi v_\perp n_{2D}/\sqrt{\kappa}\,\Omega_{21})(\hbar/\gamma)))$ for various values of γ/Γ_{sr} indicated on the curves.

be obtained by presenting $\int d\mathbf{s}$ as the sum of the asymptotic forms for $t \longrightarrow 0$ and for large times (in this region, the integral is evaluated by the steepest descent technique). As a result, $\xi(\delta\varepsilon)$ is given by the sum of the Lorentzian peak of half-width γ and a correction with a half-width of the order of $\gamma + \Gamma_{sr}$, which decreases in proportion to $|\delta\varepsilon|^{-3/2}$.

The characteristic energy ε_c determining the conditions of applicability of small-scale inhomogeneities as well as energy Γ_{sr} can be expressed in terms of the mass introduced in Eq. (37). For a GaAs QW of width 100 Å and with an electron concentration 5.6×10^{11} cm^{-2}, we obtain $M \approx 1.7m$ so that we have $\varepsilon_c \approx 3$meV for values of l_c approximately equal to 200 Å (for which the conditions $l_c > d$, a_B are still observed). Using the nonuniform broadening $\Gamma = 3$ meV corresponding to $\overline{\delta d}$, we obtain from Eq. (45) $\Gamma_{sr} \approx 4.4$ meV. For such parameters, ε_c only becomes equal to γ_{eff}; i.e., the conditions of the applicability of the short-wave limiting case are observed for decreasing $\overline{\delta d}$ and l_c. In [5, 12], peaks exhibiting a weak dependence on temperature and having a shape close to the Lorentzian were observed; the broadening of such peaks is associated with heterostructure irregularities. Since the mobility in the samples investigated in [5, 12] did not correlate with the line width, the contribution from irregularities to uniform broadening (which is significant for $l_c \le \hbar/p_F$) can be regarded as small so that the broadening mechanism described above (and corresponding to the condition $l_c \ge \hbar/p_F$) is realized.

6. CONCLUSIONS

The results obtained in this work demonstrate a significant contribution of large-scale unscreened variations of intersubband transition energy to the absorption peak broadening. Such a broadening mechanism may dominate not only in the far, but also in the intermediate IR region. Upon a decrease in the correlation length of inhomogeneities, the nonlocal response mode is realized and the Gaussian shape of the line is transformed into the Lorentzian shape.

Let us briefly discuss the approximations used in the present work. The interband transition energy and its variations due to irregularities of heteroboundaries were estimated by using the simplest model of a QW with high barriers, and the transition broadening due to collisions were described by the phenomenological relaxation frequency ν. Since the nonideality of the heterostructure changes the surface permittivity insignificantly, the exciting field $E_\perp$ is assumed to be uniform in the 2D plane. These approximations are generally accepted and the main approximation used here is a simple approximation of functions (23), which makes it possible to close the chain of balance equations and to carry out the description of the peak shape taking into account nonlocal effects with the help of the standard averaging of the continual integral. The applicability of this approximation is based on the matching of the results of numerical calculations presented in Fig. 2 and approximation of functions (23). We also assume that the inhomogeneities of the QW are smooth: the correlation length l_c is regarded as noticeably larger than a_B so that the variations of the ground state energy are screened completely. The variations of the intersubband energy are assumed to be small as compared to the Fermi energy, which allows us to confine our analysis to the contribution of rectilinear trajectories while calculating the integral (40). A numerical analysis of the opposite limiting case of electron localization in a weakly doped QW was carried out recently in [18]. An analysis of the transition between small and large inhomogeneities requires special attention (a similar problem appears in the description of the shape of the exciton absorption peak in a nonideal QW [19]).

In spite of the simplifications presented in Sections 4 and 5, our results describe a transition between two modes of nonuniform and uniform broadening of the interband absorption peak and give the line width depending on the QW parameters and the nature of inhomogeneity of the heteroboundaries. The numerical estimates demonstrate a considerable contribution of the proposed mechanism to the absorption peak broadening. Such a broadening mechanism must be taken into account not only in the description of the spectral parameters of various devices (detectors and modulators) for the far and intermediate IR range, but also in the analysis of the resonance amplification of monopolar lasers. This mechanism of broadening must also be manifested in inelastic scattering of light involving intersubband transitions [20].

REFERENCES

1. *Intersubband Transitions in Quantum Wells*: *Physics and Devices*, Ed. by S. S. Li and Y.-K. Su (Kluwer, Boston, 1998).
2. F. T. Vasko and A. Kuznetsov, *Electronic States and Optical Transitions in Semiconductor Heterostructures* (Springer-Verlag, Berlin, 1998).
3. M. Helm, Semicond. Semimet. **62**, 1 (2000).
4. B. Gelmont, V. Gorfinkel, and S. Luryi, Appl. Phys. Lett. **68**, 2171 (1996).
5. K. L. Campman, H. Schmidt, A. Imamoglu, and A. C. Gossard, Appl. Phys. Lett. **69**, 2554 (1996).
6. G. Beadie, W. S. Rabinovich, D. S. Katzer, and M. Goldenberg, Phys. Rev. B **55**, 9731 (1997).
7. S. Tsujino, M. Rufenacht, H. Nakajima, *et al.*, Phys. Rev. B **62**, 1560 (2000).
8. D. E. Nikonov, A. Imamoglu, and M. O. Scully, Phys. Rev. B **59**, 12212 (1999).
9. O. E. Raichev and F. T. Vasko, Phys. Rev. B **60**, 7776 (1999).
10. F. T. Vasko, J. P. Sun, G. I. Haddad, and V. V. Mitin, J. Appl. Phys. **87**, 3582 (2000).
11. F. T. Vas'ko and V. M. Rozenbaum, Fiz. Tverd. Tela (Leningrad) **21**, 648 (1979) [Sov. Phys. Solid State **21**, 383 (1979)].
12. R. J. Warburton, K. Weilhammer, C. Jabs, *et al.*, Physica E (Amsterdam) **7**, 191 (2000).
13. M. Helm, P. England, E. Colas, *et al.*, Phys. Rev. Lett. **63**, 74 (1989).
14. M. Rochat, J. Faist, M. Beck, *et al.*, Appl. Phys. Lett. **73**, 3724 (1998).
15. J. Ulrich, R. Zobl, K. Unterrainer, *et al.*, Appl. Phys. Lett. **74**, 3158 (1999).
16. V. L. Bonch-Bruevich, I. P. Zvyagin, R. Kapper, *et al.*, *Electron Theory of Disordered Semiconductors* (Nauka, Moscow, 1981).
17. F. T. Vasko, O. G. Balev, and N. Studart, Phys. Rev. B **62**, 12940 (2000).
18. C. Metzner and G. H. Dohler, Phys. Rev. B **60**, 11005 (1999).
19. Zh. S. Gevorkyan and Yu. E. Lozovik, Fiz. Tverd. Tela (Leningrad) **27**, 1800 (1985) [Sov. Phys. Solid State **27**, 1079 (1985)].
20. A. Pinczuk, S. Schmitt-Rink, G. Danan, *et al.*, Phys. Rev. Lett. **63**, 1633 (1989); S. Ernst, A. R. Goni, K. Syassen, and K. Eberl, Phys. Rev. Lett. **72**, 4029 (1994).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 93, No. 6, 2001, pp. 1280–1287.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 6, 2001, pp. 1478–1485.
Original Russian Text Copyright © 2001 by Buchstaber, Karpov, Tertychniy.

SOLIDS
Electronic Properties

Electrodynamic Properties of a Josephson Junction Biased with a Sequence of Delta-Function Pulses

V. M. Buchstaber, O. V. Karpov* , and S. I. Tertychniy

*National Scientific and Research Institute for Physical–Technical and Radiotechnical Measurements,
Mendeleevo, Moscow oblast, 141570 Russia*

**e-mail: mera@vniiftri.ru*

Received May 8, 2001

Abstract—An analytic method is developed for studying the electrodynamic properties of Josephson junctions biased with a sequence of delta-function pulses. Such investigations are of interest in connection with developing, on the basis of the Josephson effect, analog-to-digital converters and synthesizers producing signals of a preset form with a fundamental precision. Analytic expressions are obtained that determine phase lock conditions, the mean voltage across the junction, and the boundaries of Shapiro's steps. The results obtained are presented in a graphical form for the cases of unipolar and bipolar bias of a junction. The results are discussed and compared with those known from previous investigations. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Presently, Volt standard based on the Josephson effect and intended for reproducing, with a fundamental precision, the unit of dc electric voltage at levels of 1 and 10 V have been created and are successfully operating in a number of metrological laboratories in leading countries worldwide [1–4], including Russia [5–7]. A array of Josephson junctions that contains 15 to 20 or more thousands of junctions, depending on the frequency of microwave bias and the nominal voltage to be reproduced, form the hub of such devices.

Originally [8–10], such arrays were implemented with the aid of superconductor–insulator–superconductor (SIS) Josephson structures, which are characterized by a considerable junction capacitance and, as a consequence, by a hysteresis in the current–voltage characteristic [11, 12]. The electrodynamic properties of SIS structures are described by the first-order nonlinear differential equation {cf. Eq. (2.48) in [11] or with Eq. (6.3.5) in [12]}

$$\alpha\ddot{\varphi} + \beta(\varphi)\dot{\varphi} + \sin\varphi = q(t),$$

where the factor

$$\beta(\varphi) = (1 + \varepsilon\cos\varphi), \quad |\varepsilon| \le 1,$$

takes into account dissipation and the interference between the current of Cooper pairs and the quasiparticle current in tunnel structures and α is a hysteresis parameter. The parameter ε is small in general, so that the contribution of the $\cos\varphi$ term can be disregarded in many cases of practical importance.

The success in creating the standards of dc voltage gave impetus to undertaking efforts to develop devices generating (synthesizing), with a fundamental precision, an ac voltage [13–19]. For this, it is necessary to have so-called overdamped ($\alpha \ll 1$) Josephson structures, whose current–voltage characteristic does not possess a hysteresis. The electrodynamic properties of such a Josephson junction can be described in terms of a first-order differential equation [17–19].

Work on developing, on the basis of the Josephson effect, programmable analog-to-digital converters and devices synthesizing signals of a preset shape are being performed in a number of centers that possess a considerable technological potential for creating Josephson structures, including those that employ superconductor–normal metal–superconductor (SNS) junctions (at the National Institute of Standards and Technologies, USA) [13] and those that employ superconductor–insulator–normal metal–insulator–superconductor (SINIS) junctions (at the Physikalische-Techninikalische Bundesanstalt, Germany) [14].

It should be noted that, in contrast to SNS structures, two-barrier structures of the SINIS type possess a finite capacitance. In describing the electrodynamic properties of Josephson junctions manufactured on the basis of SINIS technologies, it is therefore necessary, in general, to take into account the term that involves the second derivative. Nevertheless, experimental investigations of SINIS arrays operating in a frequency region around 10 GHz that were manufactured at the Physikalische-Techninikalische Bundesanstalt and which are intended for application in devices providing the standard of ac voltage showed the absence of a hysteresis in their current–voltage characteristics (see, for example, [6, 7]).

One possible method for generating an ac voltage of a preset shape on the basis of the Josephson effect consists in varying the frequency of bias according to a specific law. If, however, sinusoidal microwave radiation is used for bias, the admissible frequency range within which the current–voltage characteristic of a junction contains Shapiro's steps is bounded, its width does not exceed, as a rule, 10–15% of the central value.

In [15], it is stated that, if a sinusoidal microwave voltage for bias a junction is replaced by a pulsed voltage, the width of Shapiro's steps in the current–voltage characteristic of a Josephson junction becomes virtually independent of the pulse-repetition frequency over a wide range for all frequencies below some critical value.

Devices that are intended for generating (synthesizing) an ac voltage of a preset shape and which are being designed at present are based on forming the output ac voltage by controlling, according to a specific digital code, the frequency of repetition of electric pulses bias the Josephson junction used [16].

In view of the importance of this approach, attempts were made in [17–19] to study theoretically the electrodynamic properties of a Josephson junction biased with a sequence of pulses having a specific shape. In the present study, we propose a somewhat different approach that makes it possible to describe a number of radically new features in the behavior of a overdamped Josephson junction biased with a sequence of delta-function pulses. For the mathematical model of bias pulses, we have chosen a delta function because a comparatively simple experimental realization of this model is possible—this can be achieved, for example, by differentiating an electric bias signal that has the shape of a meander.

2. FORMULATION OF THE PROBLEM

As was indicated above and in [11, 12, 17–19], the response of a overdamped Josephson junction of the SNS or the SINIS type can be described in terms of the first-order differential equation

$$\dot{\varphi}(t) + \sin\varphi(t) = q(t), \tag{1}$$

where $\varphi(t)$ is the phase of the order parameter and $t = \omega_c\tau$ is a dimensionless time, with $\omega_c = 2\pi K_J R_N I_c$ being a frequency characteristic of a specific junction. Here, $K_J = 483597.9$ GHz/V is the Josephson constant, R_N is the normal resistance of the junction, and I_c is its critical current. The quantity $q(t)$ on the right-hand side of Eq. (1) is a periodic function having a period T and describing the signal bias the junction. The period T is defined here as the dimensionless quantity $T = \omega_c/f$, with f being the dimensional pulse-repetition frequency (the inverse of the dimensional pulse-repetition period).

If a Josephson junction is biased with a periodic sequence of delta-function pulses, the function $q(t)$ can be represented in the form

$$q(t) = i_0 + 2\pi i_1 \sum_{n \geq 1} \delta(t - t_0 - nT), \tag{2}$$

where t_0 is an initial instant of time ($t > t_0$), $2\pi i_1 = 2\pi I_1/I_c$ is the integrated amplitude of the pulses, and $i_0 = I_0/I_c$ is the constant component of the junction bias current minus the component induced by the delta-function pulses. It is worth noting that, upon averaging the right-hand side of Eq. (2), there arises the expression $i_0 + 2\pi i_1/T$, which involves the term $2\pi i_1/T$. This term represents the mean current induced in the junction by the bias pulses. The constant bias current i is preset by an external power-supply unit. Since the microwave-bias circuit contains, as a rule, a separating electric capacitor, we will henceforth set $i_0 = i - 2\pi i_1/T$.

The mean voltage U across a Josephson junction is given [11, 12] by

$$U = \frac{1}{2\pi K_j \tau_a}[\varphi(t_0 + \omega_c\tau_a) - \varphi(t_0)], \tag{3}$$

where τ_a is the time of averaging; in practice, this time is usually much longer than the pulse-repetition period f^{-1}. In the presence of the phase lock effect [11, 12, 17–19], which leads to the appearance of Shapiro's steps in the current–voltage characteristic of a junction, the condition

$$\varphi(t + \omega_c\tau_a) - \varphi(t) = 2\pi k\left(\frac{\omega_c\tau_a}{T}\right)[1 + \mu(\tau_a)] \tag{4}$$

holds at $t \longrightarrow \infty$ for $\tau_a \longrightarrow \infty$ with $\mu(\tau_a) \longrightarrow 0$. For sufficiently long averaging times τ_a, Eq. (3) reduces to the well-known Josephson relation $K_J U = kf$, where $k = 0, \pm1, \pm2\ldots$ is the order of Shapiro's step.

In order to construct a solution to Eq. (1) with the right-hand side specified by Eq. (2), we make use of the method proposed in [20]. Following this method, we seek a general solution $\varphi(t)$ to Eq. (1) in the form of the sum

$$\varphi(t) = \sum_{n \geq 0} \varphi_n(t), \tag{5}$$

where $\varphi_n(t)$ satisfies the relation

$$\exp[j\varphi_n(t)] = \exp[j\varphi_*(t - t_n)]\frac{1 + C_n\overline{F(t - t_n)}}{1 + C_n F(t - t_n)} \tag{6}$$

in the interval $t_n \leq t < t_n + T$, $t_n = t_0 + nT$ and vanishes for other values of $t > t_0$. In Eq. (6), j stands for an imaginary unit; $\varphi_*(t)$ is a solution to the equation

$$\dot{\varphi}_* + \sin\varphi_* = i_0, \tag{7}$$

where the right-hand side i_0 is a constant; C_n is a real-valued constant; and $F(t)$ is specified by the expressions

$$\begin{gathered} F(t) = Q(t) + jP(t), \\ P(t) = \exp\left[-\int_0^t \cos\varphi_*(t)dt\right], \\ Q(t) = \int_0^t P(t)\sin\varphi_*(t)dt \end{gathered} \tag{8}$$

(note that the definition of $P(t)$ differs from that used in [20]). In addition, the functions $\varphi_n(t)$ must satisfy the matching conditions

$$\lim_{t \to t_{n+1} - 0} \varphi_n(t) + 2\pi i_1 = \varphi_{n+1}(t_{n+1}), \tag{9}$$

which, together with Eq. (6), determine recursively the entire sequence of the constants C_n and the required solution $\varphi(t)$, apart from an additive constant term (an integral multiple of 2π).

For the phase lock effect, which leads to the appearance of Shapiro's steps in the current–voltage characteristic, to occur, it is necessary and sufficient that the condition

$$\varphi_{n+1}(t+T) - \varphi_n(t) \longrightarrow 2\pi k \tag{10}$$

for $n \longrightarrow \infty$ be satisfied uniformly in the interval $t_n \le t < t_{n+1}$. The existence of the limit in (10) is equivalent to the condition of convergence of the sequence C_n, the limit $C = \lim C_n$ satisfying a quadratic equation that will be explicitly obtained below [Eq. (16)]. Accordingly, the phase lock condition proves to be equivalent to a condition that ensures the existence of a real-valued solution to the aforementioned quadratic equation (more precisely, there are generally two such solutions, but only one of these corresponds to the limit of the sequence C_n).

Thus, the problem specified by Eq. (1) with the right-hand side in the form (2) has been reduced to making use of the well-known [11, 12] analytic solution to Eq. (7) with a constant right-hand side on the interval $[0, T]$, whereupon it only remains to derive conditions under which relation (10) is satisfied.

3. ANALYTIC SOLUTION

In order to analyze the behavior of the phase function within the framework of the above approach, we must choose a solution φ_* to Eq. (7). It is necessary to consider the cases of $|i_0| > 1$ and $|i_0| < 1$ separately. First, we consider the case of $|i_0| < 1$, where it is convenient to employ the obvious time-independent solution $\varphi_*(t) = \varphi_{(*)}$, where the constant $\varphi_{(*)}$ satisfies the equation

$$\sin\varphi_{(*)} = i_0,$$

which corresponds to the time-independent Josephson effect [11, 12]. Let us note that, in the regions

$$-\pi/2 + 2m\pi \le \varphi_{(*)} \le \pi/2 + 2m\pi, \quad m = 0, \pm 1, \pm 2, \ldots,$$

where the time-independent solution is stable, the following relation holds:

$$\cos\varphi_{(*)} = \sqrt{1 - i_0^2} > 0.$$

Substituting the expressions for $\sin\varphi_{(*)}$ and $\cos\varphi_{(*)}$ into (8), we can derive, after a simple algebra, the following formula for the functional F:

$$\begin{aligned} F(t) = {} & \tan\varphi_{(*)}[1 - \exp(-t\cos\varphi_{(*)})] \\ & + j\exp(-t\cos\varphi_{(*)}). \end{aligned} \tag{11}$$

In the case of $|i_0| > 1$, Eq. (7) does not have time-independent solutions. Suppose that $i_0 > 1$ (the case of $i_0 < 1$ can be treated in a similar way). In the interval $0 \le t < T$, that time-dependent solution which reduces to a time-independent one for $i_0 \longrightarrow 1 + 0$ can be represented in the form

$$\begin{aligned} \exp(j\varphi_*(t)) = {} & j\left[\sqrt{i_0+1}\cos\left(\frac{t}{2}\sqrt{i_0^2-1}\right)\right. \\ & \left. + j\sqrt{i_0-1}\sin\left(\frac{t}{2}\sqrt{i_0^2-1}\right)\right] \\ & \times\left[\sqrt{i_0+1}\cos\left(\frac{t}{2}\sqrt{i_0^2-1}\right) - j\sqrt{i_0-1}\sin\left(\frac{t}{2}\sqrt{i_0^2-1}\right)\right]^{-1}, \end{aligned} \tag{12}$$

which can be shown equivalent to expression (6.2.6) from [12].

It is worth noting at this point that, if $\sin\varphi_* \ne i_0$ and if the function φ_* satisfies Eq. (7), there is an explicit formula that directly expresses $F(t)$ in terms of $\varphi_*(t)$, which reads

$$F(t) = \frac{\cos(\varphi_*(0)) + ji_0 - \exp(j\varphi_*(t))}{i_0 - \sin\varphi_*(0)}. \tag{13}$$

Together with formula (12), this yields

$$\begin{aligned} F(t) = {} & j\frac{\sqrt{i_0+1}}{\sqrt{i_0-1}} \\ & \times\frac{i_0 + \sqrt{i_0^2-1} - \exp(jt\sqrt{i_0^2-1})}{i_0 + \sqrt{i_0^2-1} + \exp(jt\sqrt{i_0^2-1})}. \end{aligned} \tag{14}$$

Taking into account relation (6), we can recast the phase lock condition into a form that ensures the existence of a real-valued solution C to the equation

$$\exp[j\varphi_*(0) - j\varphi_*(T)] \times \frac{1 - jC}{1 + jC} \cdot \frac{1 + CF(T)}{1 + C\overline{F(T)}} = \exp(2\pi j i_1). \tag{15}$$

For $|i_0| < 1$, Eq. (15) reduces to a quadratic equation for the constant C; that is,

$$aC^2 + 2bC + c = 0, \tag{16}$$

where

$$a = 2\cos(\pi i_1)\tan\varphi_{(*)}\sinh\left[\frac{1}{2}\cos(\varphi_{(*)})T\right] + \sin(\pi i_1)\exp\left[-\frac{1}{2}\cos(\varphi_{(*)})T\right],$$

$$b = \sinh\left[\frac{1}{2}\cos(\varphi_{(*)})T\right] \times [\cos(\pi i_1) + \sin(\pi i_1)\tan(\varphi_{(*)})],$$

$$c = \sin(\pi i_1)\exp\left[\frac{1}{2}\cos(\varphi_{(*)})T\right].$$

The relevant discriminant has the form

$$D = \left\{\cosh\left[\frac{1}{2}\cos(\varphi_{(*)})T\right]\cos(\pi i_1) - \tan(\varphi_{(*)})\sinh\left[\frac{1}{2}\cos(\varphi_{(*)})T\right]\sin(\pi i_1)\right\}^2 - 1. \tag{17}$$

For Eq. (16) to have a real-valued solution, it is necessary and sufficient that the condition $D \geq 0$ be satisfied. It follows that the phase lock condition for $|i_0| < 1$ can eventually be reduced to the inequality

$$\left|\cosh\left(\frac{T}{2}\sqrt{1 - i_0^2}\right)\cos(\pi i_1) - \frac{i_0}{\sqrt{1 - i_0^2}}\sinh\left(\frac{T}{2}\sqrt{1 - i_0^2}\right)\sin(\pi i_1)\right| > 1. \tag{18}$$

In the case of $|i_0| > 1$, the relevant calculations are similar to those presented above, but they are somewhat more lengthy, since φ_* is a function in that case rather than a constant. For the discriminant, we obtain

$$D = \left\{\cos\left[\frac{T}{2}\sqrt{i_0^2 - 1}\right]\cos(\pi i_1) - \frac{i_0}{\sqrt{i_0^2 - 1}}\sin\left[\frac{T}{2}\sqrt{i_0^2 - 1}\right]\sin(\pi i_1)\right\}^2 - 1. \tag{19}$$

This implies that, for $|i_0| > 1$, the phase lock condition assumes the form

$$\left|\cos\left(\frac{T}{2}\sqrt{i_0^2 - 1}\right)\cos(\pi i_1) - \frac{i_0}{\sqrt{i_0^2 - 1}}\sin\left(\frac{T}{2}\sqrt{i_0^2 - 1}\right)\sin(\pi i_1)\right| > 1. \tag{20}$$

Equations (18) and (20) are convenient for analyzing the regions in the parameter space that are characterized by the presence or the absence of the phase lock effect. Such an analysis will be performed in the concluding section of this article.

Let us now consider analytic expressions for the discriminant D in the case where the Josephson junction is biased by bipolar delta-function pulses of equal magnitude. In this case, the function $q(t)$ has the form

$$q(t) = i_0 + 2\pi i_1 \times \sum_{n > 1}\left[\delta(t - t_0 - nT) - \delta\left(t - t_0 - \left(n + \frac{1}{2}\right)T\right)\right]. \tag{21}$$

The analog of Eq. (15), which actually describes the jump in the phase at the instant of the pulse arrival, gives two such jumps in the interval of duration T: one at the instant $t = 0$, which is equivalent to the instant $t = T$ by virtue of periodicity, and the other at $t = T/2$. Specifically, we have

$$\exp[j\varphi_*(0) - \varphi_*(T)]\frac{1 - jC_1}{1 + jC_1} \cdot \frac{1 + C_2F(T)}{1 + C_2\overline{F(T)}} = \exp(2\pi j i_1), \tag{22}$$

$$\frac{1 + C_1\overline{F(T/2)}}{1 + C_1F(T/2)} \cdot \frac{1 + C_2F(T/2)}{1 + C_2\overline{F(T/2)}} = \exp(-2\pi j i_1), \tag{23}$$

where the constant C_1 describes the limiting phase function in the intervals $[t_0 + 2\pi m, t_0 + 2\pi(m + 1/2)]$, while the constant C_2 corresponds to the intervals $[t_0 + 2\pi(m + 1/2), t_0 + 2\pi(m + 1)]$.

For the above bias signal, the discriminant can be represented as

$$D = 4\frac{\cos^2(\pi i_1) - i_0^2}{1 - i_0^2}\sinh^2\left(\frac{T}{4}\sqrt{1 - i_0^2}\right) \times \left[\cosh^2\left(\frac{T}{4}\sqrt{1 - i_0^2}\right) - \frac{\sin^2(\pi i_1)}{1 - i_0^2}\sinh^2\left(\frac{T}{4}\sqrt{1 - i_0^2}\right)\right] \tag{24}$$

for $|i_0| < 1$ and as

$$D = -4\frac{i_0^2 - \cos^2(\pi i_1)}{i_0^2 - 1}\sin^2\left(\frac{T}{4}\sqrt{i_0^2 - 1}\right)$$
$$\times\left[\cos^2\left(\frac{T}{4}\sqrt{i_0^2-1}\right) - \frac{\sin^2(\pi i_1)}{i_0^2 - 1}\sin^2\left(\frac{T}{4}\sqrt{i_0^2-1}\right)\right] \tag{25}$$

for $|i_0| > 1$.

4. RESULTS AND DISCUSSION

The condition in (20), which determines, in the space of the parameters T, i_0, and i_1, the boundaries of the region where the phase lock effect occurs and, in particular, the width of Shapiro's steps in the current–voltage characteristic of a Josephson junction in the case of unipolar bias, can be reduced to a quadratic equation in $\tan((T/2)\sqrt{i_0^2 - 1})$ and, further, to the form

$$\frac{i_0\cos(\pi i_1) \pm 1}{\sin(\pi i_1)}\frac{\tan\left(\frac{T}{2}\sqrt{i_0^2-1}\right)}{\sqrt{i_0^2-1}} = -1, \tag{26}$$

which is convenient for the subsequent analysis and for a comparison with expression (23) from [17]—the latter, albeit being close to (26) in form, differs from it.

It should be noted that, in the absence of the phase lock effect, the voltage across the junction is given by

$$U = \frac{2\omega_c}{2\pi T K_J}\left\{\arccos\left[\cos\left(\frac{T}{2}\sqrt{i_0^2-1}\right)\cos(\pi i_1)\right.\right.$$
$$\left.\left. - \frac{i_0}{\sqrt{i_0^2-1}}\sin\left(\frac{T}{2}\sqrt{i_0^2-1}\right)\sin(\pi i_1)\right] + \pi k\right\}, \tag{27}$$

with the integer k being determined by the order of Shapiro's steps on the right and on the left of the segment considered in the current–voltage characteristic. Expression (27), which is obtained from Eq. (3) by varying the phase function at large values of the time parameter, can be of use in analyzing the electrodynamic properties of some types of SNS and SINIS Josephson junctions and in determining their current–voltage characteristics.

Let us now consider in greater detail some properties of a Josephson junction in the case of the bias current given by Eq. (2). As the simplest implication of the above relations, we will first examine the well-known case [11, 12] where there is no pulsed component in the bias signal (2)—that is, where $i_1 = 0$. The discriminant then takes the form

$$D = \begin{cases} \sinh^2\left(\frac{T}{2}\sqrt{1 - i_0^2}\right), & |i_0| < 1, \\ -\sin^2\left(\frac{T}{2}\sqrt{i_0^2 - 1}\right), & |i_0| > 1, \end{cases} \tag{28}$$

from which it follows that, for $|i_0| < 1$, there occurs the phase lock effect reducing to the function $\varphi_{(*)}$ asymptotically approaching a constant value. For $|i_0| > 1$, the phase φ changes irregularly, but the limit of the mean variation of φ over a large time interval exists and is equal to $\sqrt{i_0^2 - 1}$. Thus, the current–voltage characteristic of a Josephson junction consists of a horizontal segment $U = 0$ for $|i_0| < 1$ and hyperbolas

$$U = \pm\frac{\omega_c}{2\pi K_J}\sqrt{i_0^2 - 1}, \quad |i_0| > 1,$$

all of these being matched together. It is worth mentioning that the discriminant D vanishes at $i_0^{(k)} = \sqrt{1 + (2\pi k T^{-1})^2}$. These are the points near which Shapiro's steps arise when a small pulsed component with pulse-repetition period $f^{-1} = T/\omega_c$ is added to the bias current. By way of example, we indicate that, in the case of delta-function pulses of small amplitude i_1 (more precisely, for $0 < 2\pi i_1/T \ll 1$), Shapiro's steps occupy the i_0 intervals

$$\left(\sqrt{1 + \left(\frac{2\pi k}{T}\right)^2} - \frac{4\pi i_1}{T} + o(i_1), \sqrt{1 + \left(\frac{2\pi k}{T}\right)^2} + o(i_1)\right). \tag{29}$$

This illustrates the following well-known fact characteristic of the current–voltage characteristics of overdamped Josephson junctions [11, 12]: at a fixed bias frequency, a higher bias current corresponds to a greater number of Shapiro's steps; with increasing bias frequency, the Shapiro step characterized by a fixed number is shifted toward greater values of $|i_0|$.

In accordance with expressions (18) and (20), the regions where the phase lock effect is present or absent in case of unipolar and bipolar bias signals are shown in Figs. 1 and 2 (the regions where there is no this effect are shaded). In Fig. 1a, one can see shrinking bent shaded regions of funnel-like shape, which first become discontinuous and then disappear. These regions represent gaps between neighboring Shapiro's steps. In fact, they persist to exist still further, but the code used to plot them cannot depict this because their width becomes too small.

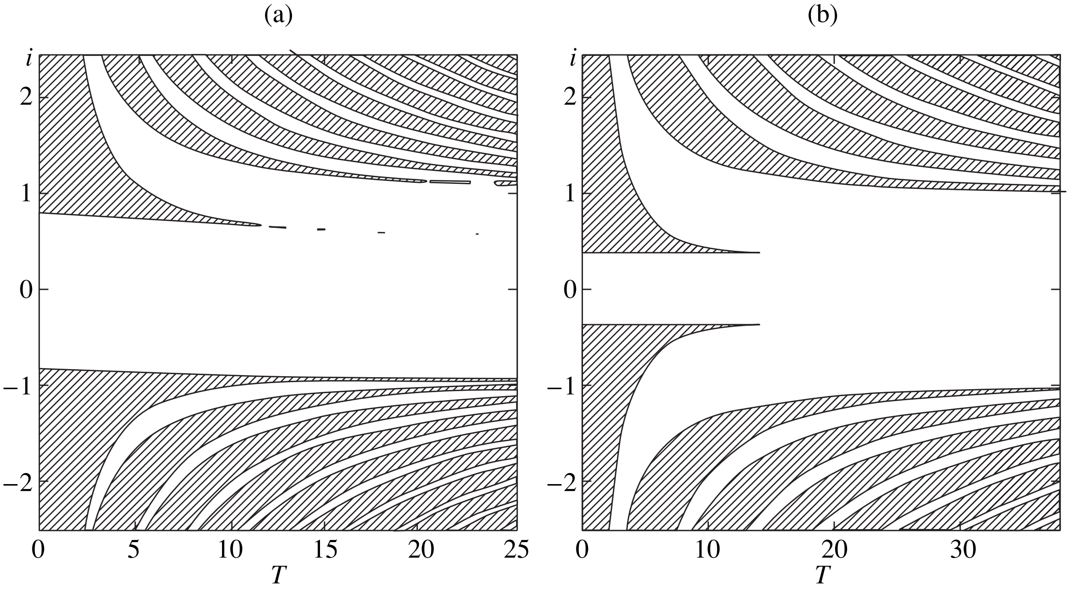


Fig. 1. Regions where the phase lock effect occurs and regions where this effect does not occur (shaded) in the problem specified by Eqs. (1) and (2). The results are presented in the plane spanned by the pulse-repetition period T and the constant component of the bias current i, where (a) $i = i_0 + 2\pi i_1/T$ for unipolar and (b) $i = i_0$ for symmetric bipolar pulses. The amplitude of the bias pulses is $2\pi i_1$ = (a) 1/3 and (b) 3/8.

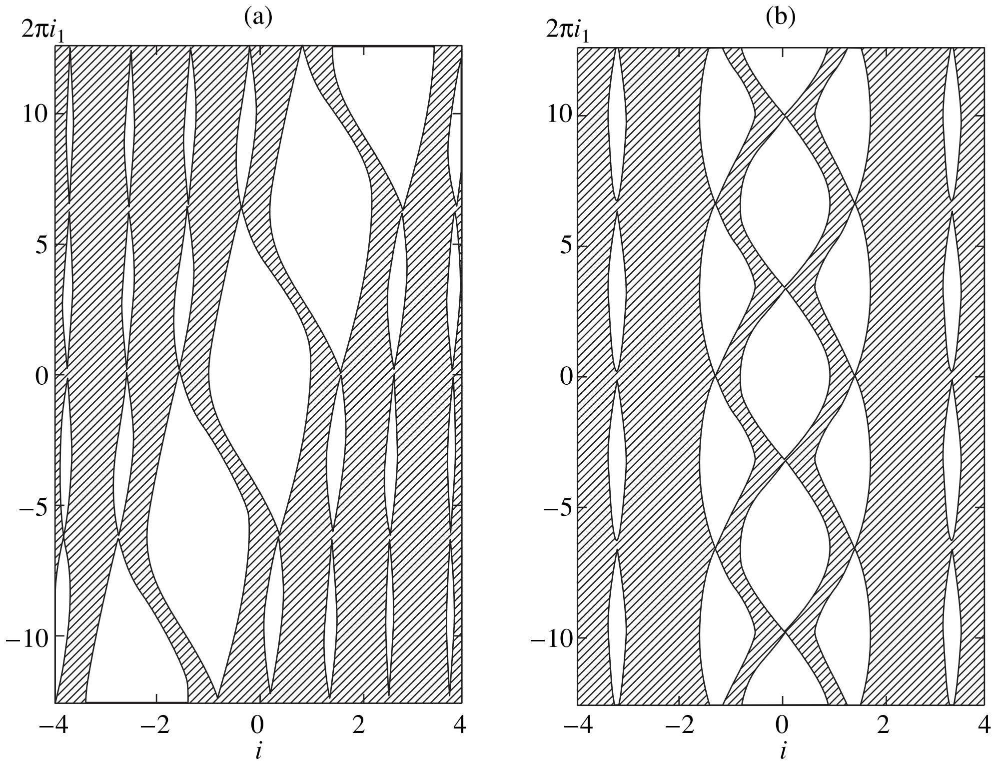


Fig. 2. Regions where the phase lock effect occurs and regions where this effect does not occur (shaded) in the problem specified by Eqs. (1) and (2). The results are presented in the plane spanned by the constant component of the bias current i and the pulse amplitude $2\pi i_1$ for (a) unipolar and (b) symmetric bipolar pulses. The pulse-repetition period is T = (a) 25 and (b) 5.3.

From Fig. 1a, it can be seen that, from some value of T, the widths of Shapiro's steps of order zero and one (minus one) are nearly independent of the pulse-repetition period. This relations was indicated in a number of studies (see, for example, [18]). However, a pronounced asymmetry with respect to the $i = 0$ axis is observed in the case of unipolar bias.

At the same time, it can be seen from Figs. 1a and 2 that, at a fixed bias current, the number of Shapiro's step appears to be a multi-valued function of the bias pulse repetition frequency—at $i = \text{const}$, there exists an entire set of Shapiro's steps characterized by different values of the number k. This relations is also observed in the case where a Josephson junction is biased with symmetric bipolar pulses (see Fig. 1b).

In order to explain this relations, we examine Fig. 2a (the case of unipolar bias with dimensionless period $T = \omega_c/f = 5.3$), from which one can see that Shapiro's steps of numbers 0, ±1, and $\pm k$ correspond to bias pulse amplitudes in the intervals $-2\pi \le 2\pi i_1 \le 2\pi$, $-2\pi \mp 2\pi \le 2\pi i_1 \le 2\pi \mp 2\pi$, and $-2\pi \mp 2\pi k \le 2\pi i_1 \le 2\pi \mp 2\pi k$, respectively.

According to Eqs. (18) and (20), the boundaries of the phase lock regions in Figs. 1a and 2a are determined by the quantity $i_0 = i - 2\pi i_1/T$. Moreover, these regions (possibly, with the exception of the zeroth region) undergo deformations and partly overlap or are replaced in response to variations in T. This case of interest calls for a dedicated consideration.

In order to perform a qualitative analysis of the current–voltage characteristic of a junction biased with a sequence of delta-function pulses, it is sufficient to take two sections in Fig. 2a (for example, for unipolar bias) with respect to the i and the i_1 axis. In the U–i plane, it is then necessary to depict regions corresponding to Shapiro's steps and regions where there is no phase lock. The scale along the i axis is specified by the figure itself, while the scale along the U axis is determined, for each specific value of T, according to Eq. (3) or Eq. (27).

In order to perform a full analysis of the current–voltage characteristic of a Josephson junction subjected to pulsed bias, we consider, in Figs. 1a and 2a, the regions where there is no phase lock. Equations (18) and (20) make it possible to draw some conclusions on the positions and widths of the zeroth and the first Shapiro step: if the bias current i is constant and if $-1 < i_0 < +1$, there are finite segments (of width not less than some minimal value) of the zeroth and the first Shapiro step, the interval that separates them and which has the boundaries

$$i_0 \simeq \cos(\pi i_1) \mp \sin^2(\pi i_1)\exp\left[-\frac{T}{2}\sin(\pi i_1)\right] \qquad (30)$$

determining, in the first order, the upper boundary of the zeroth and the lower boundary of the first Shapiro step. Accordingly, the gap between these steps is estimated as

$$2\sin^2(\pi i_1)\exp\left[-\frac{T}{2}\sin(\pi i_1)\right].$$

This means that it is possible to make arbitrarily small the region of transition from the zeroth Shapiro's step to the first one in the current–voltage characteristic of a Josephson junction. Thus, the effect of transient processes in devices whose operation is based on switching over the number k of Shapiro's steps can be eliminated completely or minimized, which opens radically new possibilities in telecommunication technologies.

5. CONCLUSION

The method proposed in the present article makes it possible to analyze the electrodynamic properties of a Josephson junction biased with a periodic sequence of delta-function pulses and to determine the corresponding current–voltage characteristics. The results obtained along these lines pave the way for new experimental investigations—for example, investigations aimed at developing fast analog-to-digital converters and signal synthesizers based on the Josephson effect [16].

ACKNOWLEDGMENTS

We are grateful to J. Niemeyer (Physikalische-Technikalische, Bundesanstalt, 38116, Braunschweig, Germany) for attention to this research.

REFERENCES

1. Y. Sakamoto, H. Yoshida, T. Sakuraba, *et al.*, IEEE Trans. Instrum. Meas. **40**, 312 (1991).
2. D. Reymann, IEEE Trans. Instrum. Meas. **40**, 309 (1991).
3. C. A. Hamilton, C. J. Burroughs, and K. Cheh, J. Res. Natl. Inst. Stand. Technol. **95**, 219 (1990).
4. R. Popel, J. Nimeyer, R. Fromknecht, *et al.*, J. Appl. Phys. **68**, 4294 (1990).
5. A. S. Katkov and V. N. Krutikov, Kontr.-Izmer. Tekh., No. 3, 25 (2000).
6. O. V. Karpov, V. O. Kutovoĭ, S. V. Sherstobitov, and J. Niemeyer, Prib. Tekh. Éksp., No. 5, 91 (2001) [Instr. Exp. Techn. **44**, 655 (2001)].
7. O. V. Karpov, V. D. Koutovoi, S. V. Sherstobitov, and J. Niemeyer, Metrologia **38** (5) (2001) (in press).
8. R. L. Kautz and G. Costabil, IEEE Trans. Magn. **MAG-17**, 780 (1981).
9. J. Niemeyer, L. Grinn, and W. Mier, Appl. Phys. Lett. **46**, 1222 (1985).
10. J. Niemeyer, J. H. Hinken, E. Vollner, *et al.*, Metrologia **22**, 213 (1986).

11. K. K. Likharev and B. T. Ul'rikh, *Systems with Josephson Contacts* (Mosk. Gos. Univ., Moscow, 1978).
12. A. Barone and G. Paterno, *Physics and Applications of the Josephson Effect* (Wiley, New York, 1982; Mir, Moscow, 1984).
13. C. A. Hamilton, S. P. Benz, C. J. Burroughs, *et al.*, IEEE Trans. Instrum. Meas. **46**, 224 (1997).
14. J. Nimeyer, Supercond. Sci. Technol. **13**, 546 (2000).
15. S. P. Benz and C. A. Hamilton, Appl. Phys. Lett. **68**, 3171 (1996).
16. J. X. Przybysz, A. H. Worsham, S. P. Benz, *et al.*, US Patent No. 5.812.078 (1998).
17. B. Ya. Shapiro, T. Dayan, M. Gitterman, *et al.*, Phys. Rev. B **46**, 8349 (1992).
18. J. Kim, A. Sosso, and A. F. Clark, J. Appl. Phys. **83**, 3225 (1998).
19. R. Monaco, J. Appl. Phys. **68**, 679 (1990).
20. S. I. Tertychniy, Usp. Mat. Nauk **55**, 195 (2000).

Translated by A. Isaakyan

Journal of Experimental and Theoretical Physics, Vol. 93, No. 6, 2001, pp. 1288–1295.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 6, 2001, pp. 1486–1494.
Original Russian Text Copyright © 2001 by Vasil'ev, Kan, Ohta, Hiruma.

SOLIDS
Electronic Properties

Condensation of Electron–Hole Pairs in Bulk GaAs at Room Temperature under Conditions of Femtosecond Cooperative Radiation

P. P. Vasil'ev[a, *], H. Kan[b], H. Ohta[b], and T. Hiruma[b]
[a] *Lebedev Institute of Physics, Russian Academy of Sciences, Moscow, 117924 Russia*
**e-mail: peter@mail1.lebedev.ru*
[b] *Hamamatsu Photonics K.K., Central Research Laboratory, 5000 Hirakuchi, Hamakita City, 434 Japan*

Received March 11, 2001

Abstract—The results are given of an experimental investigation of the spectral characteristics of cooperative recombination of a high-density electron–hole plasma in GaAs. It is demonstrated that, under conditions of generation of high-power femtosecond pulses of superradiation, the properties of electrons and holes differ considerably from their properties under conditions of lasing or regular spontaneous recombination. The peak of the cooperative radiation line ($\hbar\omega = 1.405$ to 1.407 eV) is shifted inward into a (nonrenormalized) forbidden band. It is located 20 meV lower on the energy scale than the lasing peak and more than 40 meV below the center of the line of spontaneous recombination at the same pumping level. This corresponds to electron–hole condensation to the bottom of the bands. The properties of cooperative recombination may be defined by the pairing of electrons and holes and by the formation of a short-lived coherent electron–hole BCS state. The estimated value of the order parameter Δ is approximately 2 meV. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The revealed and previously investigated mode of superradiation (or cooperative spontaneous recombination) in semiconductor laser structures [1–3] enables one to study high-density electron–hole (*e*–*h*) plasma placed under entirely unusual conditions. Indeed, thanks to the specific geometry of laser structure and to the very high pumping rate, it is possible to attain, during periods of time of several picoseconds, the level of concentration of electrons and holes in GaAs, which is several times higher than the lasing threshold of (1.5–2.0) × 10^{18} cm^{-3}. Under these conditions, the unsaturated gain may exceed 10^3 cm^{-1}.

It is well known [4–6] that the mode of superradiation in a system of quantum oscillators consists of two time phases of markedly different durations. First, a relatively slow relative phasing of oscillators by electromagnetic field occurs, along with the formation of a macroscopic dipole (macroscopic polarization of the medium). The second phase is characterized by a fast radiative recombination and collective radiation of a high-power and short electromagnetic pulse. In the case of a semiconductor medium, the duration of superradiation pulses is in the femtosecond range [2].

It is obvious that the properties of a coherent ensemble of *e*–*h* pairs may differ from the properties of a plasma of electrons and holes unbound to one another. In particular, their statistical parameters in an ordered coherent system must be different from the parameters of *e*–*h* plasma, say, in the case of spontaneous recombination or in the lasing mode. Previous experiments [1–3] have revealed that a coherent interaction between the optical field and semiconductor medium under conditions of superradiation, which showed up in the generation of a high-power femtosecond pulse, in coherent beats of optical field with a terahertz frequency, in the emergence of a doublet in the optical spectrum, and so on, indeed takes place. The coherence of interaction persisted for periods of time several times longer than the transverse relaxation time T_2. The mode of cooperative radiation differs greatly from all dynamic modes of semiconductor lasers (Q modulation, mode locking, etc.) both qualitatively and quantitatively.

On the other hand, for quite a number of years different authors have been developing an approach to the problem of interaction between an electromagnetic field and an *e*–*h* system, which involves the pairing of electrons and holes and the formation of quasiparticles [7–10]. Given a low density of electrons and holes and a fairly low temperature, their pairing apparently brings about the formation of excitons. Under certain conditions, the Bose–Einstein condensation of exitons is possible [11]. The paired electron and hole in the case of a high density may be regarded as a cooperative *e*–*h* state similar to the BCS state of Cooper pairs in a superconductor [9, 10]. It is the purpose of this study to find out whether any variations of statistical distributions of electrons and holes under conditions of superradiation are observed. The question also arises whether the unusual properties of the *e*–*h* system observed under conditions of superradiation and listed above may be attributed to the pairing of electrons and holes and to the formation of a coherent BCS-like *e–h* state. In addition, we will try to use the experimental data to evaluate the order parameter Δ. Preliminary results pertaining to the condensation of *e–h* pairs and to the formation of a coherent BCS *e–h* state were published previously [12].

2. SAMPLES AND MEASUREMENT PROCEDURE

In the experiments, use was made of semiconductor laser structures based on a GaAs/AlGaAs heterostructure, similar to those described previously [1–3]. The heterostructures were grown on an *n*-GaAs substrate by gas-phase deposition from organometallic compounds (MOCVD). In the top, highly alloyed layer of *p*-GaAs, a mesostructure 5 μm wide was etched to restrict the spreading of the injection current in the lateral direction. Three sections were formed by photolithography along the resonator axis. Two of these sections, arranged at the crystal edges, had a common electric contact and were pumped by current pulses with an amplitude up to 0.8–1.0 A. Figure 1 is a photograph of one of the structures. The lasing thresholds in analogous structures with a single top contact did not exceed 100 mA. The duration of current pulses varied from 1 to 10 ns, and the pulse repetition rate ranged from 10 Hz to 14 MHz. The middle portion of the semiconductor structure, which served as a saturable absorber, was connected to a d.c. source. A blocking voltage in the range from zero to –10 V was applied to this portion. The total length of the structures was 250, 350, and 450 μm. Laser crystals were placed on a special massive copper coolant line, which provided for good dissipation of heat from the active region. In spite of the high amplitude of current pulses, the average current through the structures did not exceed 10–15 mA. Owing to this, the heating of the active region relative to the coolant line was insignificant. All measurements were performed at room temperature.

The spectral characteristics of radiation were investigated using an MDR-23 monochromator with a resolution of less than 0.1 nm. The radiation dynamics were studied using an IMACON-500 image converter camera with a time resolution of 1.5 ps in the mode of frame-by-frame scan. The limiting resolution of the camera was measured with the aid of femtosecond laser pulses. The procedure of recording second-order autocorrelation functions during the second harmonic generation was used for more accurate measurement of the optical pulse duration [2, 3]. The other details of the experimental setup are given in [1, 2].

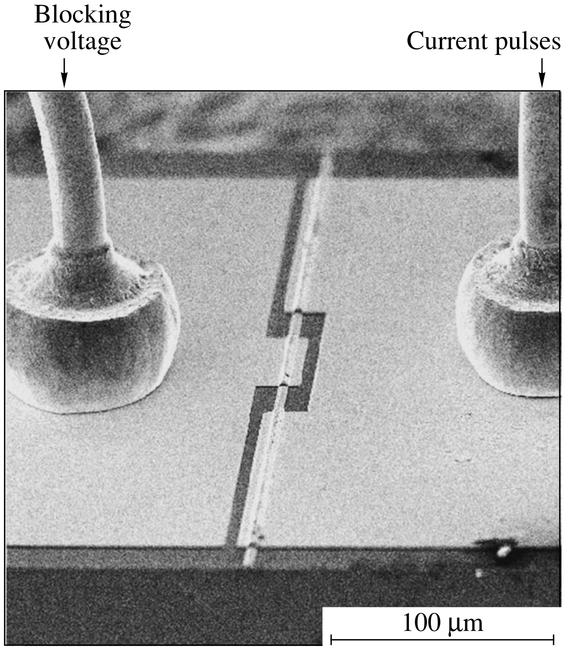


Fig. 1. A photograph of one of the investigated semiconductor structures.

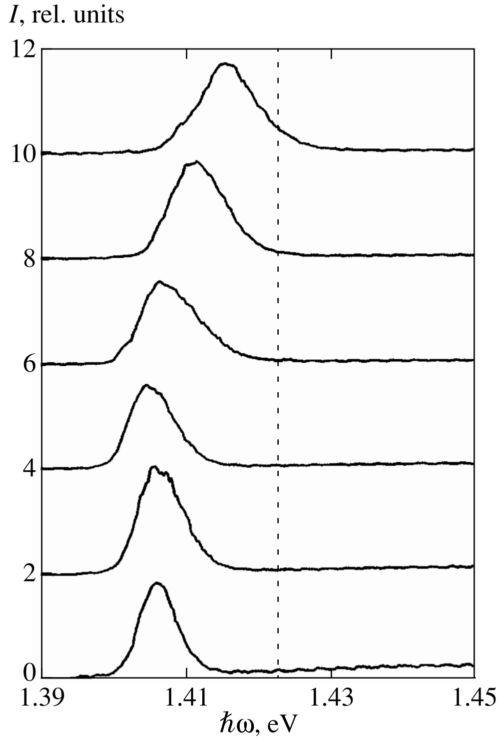


Fig. 2. Optical spectra of cooperative radiation for different values of blocking voltage on the absorber.

3. EXPERIMENTAL RESULTS

Figure 2 gives typical optical spectra of superradiation pulses, obtained for different pumping levels in a laser structure. The absolute magnitude of blocking voltage in the absorbing region of the structure increases in the figure, from top to bottom, from 4.2 V (top curve) to 7.1 V (bottom curve). Note that, at voltages below 4.2 V, the superradiation mode is less pronounced and accompanied by lasing; as the voltage approaches zero, this mode disappears altogether. In this case, only lasing is observed. One can clearly see in Fig. 2 how the spectrum center shifts to the long-wave region and its width decreases. The minimal value of the photon energy at the spectrum center is 1.406 meV, and that at the long-wave edge, 1.395 meV. The dotted line in the figure indicates the edge of the forbidden band in pure GaAs at $T = 300$ K, disregarding the narrowing of the band with an increase in the carrier concentration. The total width of the spectra at the base decreases from 20 meV to approximately 12 meV. The minimal attained spectral width was less than 8 meV (about 4 meV on half the amplitude). The shape of the spectra is most often asymmetric, with a steeper long-wave edge and a smoother short-wave edge, although, in a number of cases with an especially small width of the spectra (4–5 meV), their shape was almost symmetric. When the voltage was fairly high, no superradiation was observed. In this case, the electrons and holes recombined spontaneously, the radiation intensity was insignificant, and the optical pulse duration corresponded to the pumping pulse duration (several nanoseconds).

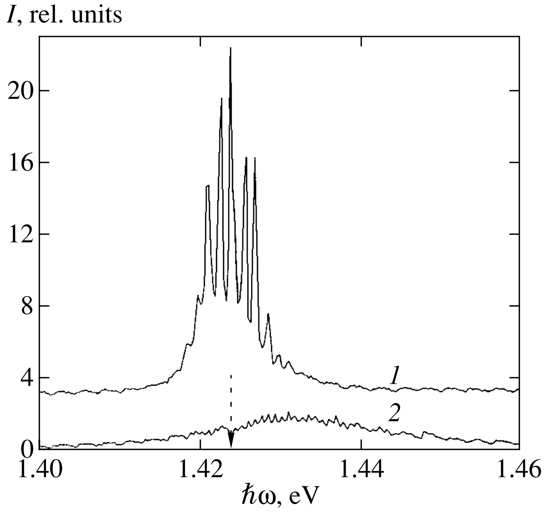


Fig. 3. Spectra of (*1*) lasing and (*2*) amplified spontaneous radiation at a low excitation level. The arrow indicates the position of the edge of nonrenormalized forbidden band.

Given in Fig. 3 for comparison are typical spectra of lasing and of regular spontaneous recombination of electrons and holes in the structures being investigated at a low pumping level. As usual, a lasing spectrum consists of several longitudinal resonator modes, with its center corresponding to the photon energy of 1.424 eV, which is almost 20 meV higher than in the case of cooperative radiation. The center of spontaneous radiation spectrum is located even higher on the energy scale (1.432 eV); it is very wide, with its width on half the amplitude being approximately *kT*.

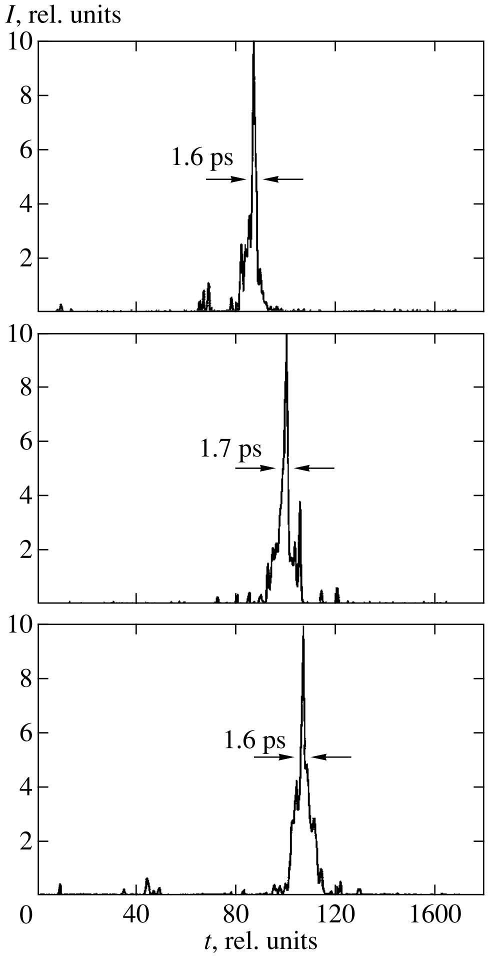


Fig. 4. Cooperative radiation pulses recorded by an image converter camera.

In the time domain, the femtosecond optical pulses given in Fig. 4 correspond to the optical spectra of Fig. 2. Shown in Fig. 4 are three envelopes of cooperative radiation pulses. The pulse duration on the screen slightly exceeds the time resolution of the camera (1.5 ps), which corresponds to the true pulse duration of less than 1 ps. For precise measurements of the pulse duration, use was made of the standard procedure of obtaining autocorrelation functions of intensity using the second harmonic generation. This procedure, characterized by femtosecond time resolution, was used to measure the true duration of superradiation pulses, which was found to be in the range from approximately 300 to 600 fs depending on the geometry of the laser structure and the conditions of excitation. Note that the time of light travel from edge to edge of the crystal varied from 4 to 6 ps depending on the length.

One can clearly see in Fig. 4 that the femtosecond duration of cooperative radiation pulses is obtained right away, as distinct from the conditions of mode locking, when tens and hundreds of passes in the laser resonator are required to obtain femtosecond pulses. The product of the duration of superradiation pulses into the width of their spectrum amounted to 0.6–0.9;

i.e., these pulses, as well as the pulses of lasers with locked modes, are in fact spectrally restricted. This means that the spectrum width is restricted from below by a value defined by the lifetime of cooperative state.

4. DISCUSSION

According to the experimental data given in the preceding section and obtained previously [1–3], we deal with a situation in which a semiconductor has two subsystems (two ensembles) of electrons and holes. One of these subsystems is involved in regular spontaneous recombination with a low power and nanosecond duration, and the other subsystem recombines collectively to form a high-power and ultrashort superradiation pulse. The phase (intraband) relaxation transfers the electrons and holes from the second to the first ensemble. The radiative recombination of these two ensembles is markedly spaced over the wavelength. Figure 5 gives spectra of cooperative radiation and regular spontaneous recombination for almost the same pumping level (carrier density). The photon energy of the maximum of cooperative radiation spectrum is 1.406 meV, while for spontaneous radiation it is 1.451 meV. An analysis of the above-identified optical spectra reveals that, prior to radiative collective recombination, the electrons and holes in the second ensemble are condensed to the bottom of bands in a narrow energy range (5–10 meV in the vicinity of the bottom of bands) and possess a minimal energy.

The question may arise whether the observed spectra are the result of cooperative recombination or of stimulated emission of a high-density *e*–*h* plasma. This is of considerable importance from the standpoint of correct interpretation of the experimental results. In the former case, the form and shape of spectra will be directly related to the energy distribution of particles in the condensate, and the approximation of spectra by the formulas given below will have a physical meaning [14–16]. In the latter case, the spectrum being recorded will be the result of amplification of the initial spectrum and will be distorted (modulated) by the spectral distribution of the optical gain. In this case, as in the laser mode, the spectral shape is not directly associated with the level distribution of electrons and holes, and the estimates of parameters of the *e*–*h* system (concentration, Fermi energy, etc.) lose their meaning. Much like in [17], we have investigated the effect of the amplification and reabsorption processes on the shape and position of the maximum of *e*–*h* recombination spectrum. It has turned out that, if we assume that we are dealing with radiative recombination of an *e*–*h* plasma and its subsequent amplification during propagation throughout the sample, the spectral distribution of the gain indeed strongly distorts the initial recombination spectrum. However, this distortion primarily affects the spectral width: the spectrum narrows considerably, with the position of the maximum remaining virtually unchanged; i.e., no long-wave (as well as short-wave) shift occurs. The same effect was detected by Dneprovskiĭ *et al.* [18]. Therefore, the sizable long-wave shift of recombination line observed cannot be attributed to amplification and reabsorption of radiation in the bulk of the semiconductor. Furthermore, in this case, the process of recombination would take up a much longer time interval in the time domain (tens and hundreds of picoseconds), and it would be impossible to explain the emergence of femtosecond monopulses given in Fig. 4.

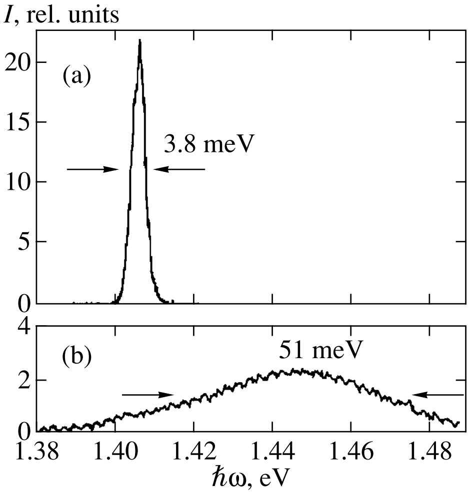


Fig. 5. Optical spectra of (a) cooperative radiation and (b) amplified spontaneous radiation for the same excitation level and somewhat differing levels of blocking voltage on the absorber.

We will now demonstrate that the radiation observed cannot be attributed to the laser effect. Two most important distinctions exist between the lasing and the cooperative radiation observed by us. These are, first, differences in the spectrum and, second, qualitative differences in the radiation dynamics. Indeed, an absorbing region is located at the center of the semiconductor structure; the absorption coefficient of this region depends on frequency. When reverse bias is applied to the absorber, the absorption edge shifts to the long-wave region (Franz–Keldysh effect). Therefore, the lasing with reverse biased absorption should have occurred at the very edge of spectral distribution of gain from the long-wave side. It is obvious, however, that, in this region at the band bottom, the density of states (and, accordingly, the gain) is many times less than at the center of the laser line where the lasing usually arises. Thus, resonator modes should have been observed in the lasing spectrum, and the spectrum in this case would have resembled the spectrum shown in Fig. 3. The cooperative radiation spectra are continuous, as one can readily see in Fig. 2.

Up to now, laser radiation dynamics have been very well studied both theoretically and experimentally [13]. The initial period of the emergence of lasing is many times longer than the time of the light passage past the laser resonator (equal to 3–12 ps for our semiconductor structures). In addition, it is always accompanied by relaxation oscillation. Dynamic modes of a semiconductor laser structure with three sections were studied theoretically in [3]. It has been demonstrated that a decrease in the initial (unsaturated) gain in the structure by a factor of only three or four brings about a qualitative variation in the radiation dynamics. In the case of insufficiently high gain, laser structures emit regular Q-switched pulses of characteristic duration of 20 to 30 ps. The dynamics and spectral behavior of Q-switched lasers have been studied in detail (see, for example, [13]). Therefore, the development of lasing at the long-wave tail of the gain line would either bring about the mode of Q modulation in the laser structure or relaxation oscillation. This would be readily recorded in the experiment, because both of these modes strongly differ qualitatively and quantitatively from the cooperative radiation mode.

Therefore, it remains to assume that we are dealing with the recombination of cooperative state as an integral whole. Its space dimension is approximately equal to the product of the velocity of light in the medium by the time of transverse relaxation (dephasing) and in our case amounts to 20–40 μm. Further attributing to this is the fact that the cooperative state is located in the vicinity of the structure ends (see below) and its radiation immediately leaves the sample [3]. Furthermore, it is only the concentration of electrons and holes in a narrow energy range that may provide for a high value of the gain necessary for the criterion of observation of superradiation to be valid [3]. Such a concentration is only possible under conditions of condensation of *e*–*h* pairs.

We will estimate the critical density of *e*–*h* pairs necessary for their condensation. Intensive studies of the condensation of electrons and holes (excitons) in semiconductors have been under way for quite a number of years [7, 11]. A system of electrons and holes of concentration n goes over to the condensed state when the so-called criterion of quantum degeneracy is valid [11],

$$n\lambda_D^3 > 1, \tag{1}$$

where $\lambda_D = \sqrt{2\pi\hbar^2/MkT}$ is the de Broglie wavelength and M is the mass of an *e*–*h* pair. In GaAs, we have $m_e = 0.07m_0$ for electrons, $m_{hh} = 0.5m_0$ for heavy holes, and $m_{lh} = 0.08m_0$ for light holes (m_0 is the electron mass in vacuum). In accordance with Eq. (1), a value of about 5×10^{18} cm^{-3} is obtained for the critical density at room temperature if the pairs are formed by heavy holes. For light holes and for a combination of light and heavy holes, this value is several times smaller, and the condensation starts with pairs formed by electrons and light holes. The experimentally obtained values of the density of *e*–*h* pairs in the coherent cooperative state are close to this value. The radiation of *e*–*h* condensate during recombination, as well as the radiation of *e*–*h* liquid, must be shifted to the long-wave region.

In our case of high density and room temperature, we are not dealing with excitons (the binding energy of an exciton in bulk GaAs is about 4 meV, with the exciton radius $a_0 \approx 140$ Å). With the density of *e*–*h* pairs in the above-identified range being $(2–5) \times 10^{18}$ cm^{-3}, the mean distance between particles $r_s = (3/4\pi n)^{1/3}$ ranges from 49 to 34 Å. Therefore, the coherent state of paired electrons in this case must resemble the state of an ensemble of electrons in a superconductor when the characteristic dimension of Cooper pairs is many times the mean distance between electrons (the pairs are strongly overlapped in space).

As was already noted in the Introduction, some authors [9, 10, 19] have indicated that under conditions of high density, $r_s/a_0 \le 1$, the cooperative state of electrons and holes may resemble the BCS state of Cooper pairs in superconductors. Multiparticle interactions in an *e*–*h* system result in the tendency for the electrons and holes to pairing in the entire density range. In a first approximation, pairs made up of two fermions behave as bosons and are capable of condensation. In our case, a decisive part in the pairing and in the setting up of coherence in an ensemble of electrons and holes is played by the electromagnetic field which is permanently present in the semiconductor volume. Note that the phasing of electrons and holes is preceded by the bleaching of the absorber of semiconductor structure by radiation propagating between the crystal faces. Because of the smallness of the photon pulse (compared with that of electrons and holes), the field produces and annihilates *e*–*h* pairs with zero total momentum. Therefore, when correlations are induced by the electromagnetic field, a coherent ensemble of particles (cooperative state) with zero total momentum arises in an *e*–*h* system at the initial stage of development of superradiation from all electrons and holes. In so doing, as was noted previously, in view of the presence of a reverse biased absorber in the semiconductor, the correlations (coherence) in the system are induced by long-wave photons, because the optical absorption is lower in the vicinity of the band edge. Because of the very high gain and exponential rise of the phasing coherent field towards the ends of the structure, the cooperative state is highly inhomogeneous in space. It is located mainly at both faces of the crystal. It is the collective radiative decay of this coherent ensemble that is observed experimentally in the form of a high-power femtosecond pulse.

All particles in the collective state are paired, with the pair size being many times the mean distance between particles. Because the total momentum of each pair is zero, as in the case of Cooper pairs, such a collective state is referred to as a BCS-like *e*–*h* state [9, 10, 19]. By analogy, the *e*–*h*-BCS state must have a minimal energy and the minimal possible quantum energy must be observed during its recombination. This may explain the observed strong long-wave shift of cooperative radiation compared both with lasing and with regular spontaneous recombination.

In order to obtain information about the parameters of an *e–h* system, we used an approximation of cooperative recombination spectra, analogous to that used to estimate the concentration and temperature when studying the luminescence of an *e–h* plasma and an *e–h* liquid in semiconductors [14–16]. As is demonstrated by the theoretical results of [9, 10, 19], the collective state of *e–h* pairs is described by the density of states of quasiparticles and by their level distribution, with the exact form of the levels being determined by numerical calculation. The energy spectrum of quasiparticles has a gap Δ whose size is restricted from above by the binding energy of an *e–h* pair (exciton). We will use the results of these studies to approximate the observed spectra of collective recombination.

It has been found that the spectra are well described by simple convolution of the density of states of quasiparticles, ρ, and of the occupancy functions of quasiparticles, ν^2, according to the formula

$$I(\hbar\omega) = I_0 \int\limits_{E'_g}^{\hbar\omega} \rho(E)\rho(\hbar\omega - E'_g - E) \times \nu^2(E)\nu^2(\hbar\omega - E'_g - E)\,dE, \tag{2}$$

where E'_g is the renormalized width of forbidden band and I_0 is a constant. The validity of this formula in our case is defined by the following.

As is known, Fermi distributions in an *e–h* system assume a steady state very rapidly, in times which are only several times longer than the time between electron–hole collisions. In our case of a very high density, the characteristic time between collisions is known to be less than 10 fs. As is demonstrated by the theoretical [3] and experimental [1, 2] results, the characteristic time of cooperative radiation has values of hundreds of femtoseconds, which considerably exceeds the relaxation time of Fermi distributions. In addition, under conditions of collective decay of the system, the moments of recombination of individual *e–h* pairs are correlated, and all particles recombine almost simultaneously. This means that the dependence $I(\hbar\omega)$ will be defined by the energy distribution of electrons and holes in accordance with Eq. (2) and will not be distorted by reabsorption and amplification in other regions of the structure. Finally, the ultrashort lifetime of the cooperative state implies the smearing-out of energy levels and, accordingly, integration with respect to energy. Indeed, in accordance with the uncertainty principle, the energy level width is the greater, the longer the time during which an electron or hole is at this level. Because the cooperative state exists for only several hundred femtoseconds, the width of individual levels in bands may be quite appreciable (over 1 meV), which must be taken into account when approximating the spectra by formula (2).

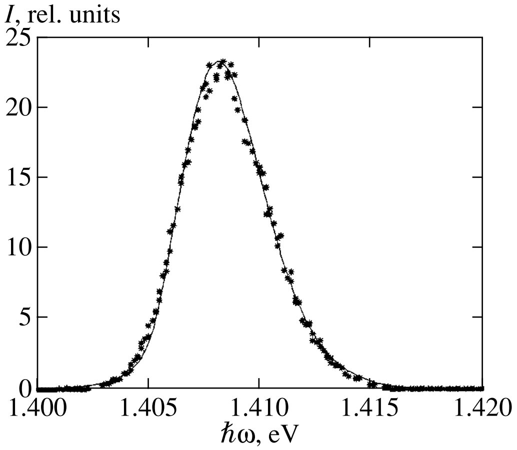


Fig. 6. A typical approximation of a spectrum of cooperative radiation by formulas (2)–(5) at E'_g = 1.4008 eV, μ = 7.3 meV, Δ_0 = 2.53 meV, Γ = 1.7 meV.

The density of states of quasiparticles, ρ(*E*), has a singularity at point *E* = Δ [20],

$$\rho \propto \frac{E}{\sqrt{E - \Delta}}. \tag{3}$$

In order to eliminate the divergence, one must take into account the width of energy levels. This can be done by introducing the effective density of states. For a simple Lorentz line shape, the effective density of states may be written as [17]

$$\rho_{\text{eff}}(E) = \frac{1}{\pi}\int \frac{\rho(E)dE_1}{\pi(E - E_1)^2 + \Gamma^2}, \tag{4}$$

where Γ is the effective level width, which is independent of energy in the simplest case.

The quasiparticle distribution function $\nu^2(E)$ is, as in the case of Cooper pairs [21], very close to the Fermi distribution and may be found by numerical methods [10]. We approximated it by the following function:

$$\nu^2(E) = \frac{1}{2}\left[1 - \frac{E - \mu}{\sqrt{(E - \mu)^2 + \Delta^2(E)}}\right], \tag{5}$$

where μ is the Fermi energy and the magnitude of the gap Δ, generally speaking, depends on the quasimomentum (energy) and must be calculated self-consistently [9, 10, 19]. This calculation falls outside the scope of this paper. We used the properties of the function Δ(*k*), namely, that the maximum of the function Δ_0 is located at $k = k_F$ and $\Delta(0) = \Delta(\infty) \longrightarrow 0$, and approximated it by the Lorentz function, with the Fermi quasimomentum corresponding to the chemical potential μ.

We used Eqs. (2)–(5) to approximate the experimentally obtained spectra, as is shown in Fig. 6. The figure gives one of the spectra; the experimental data are indicated by the asterisks, and the approximation by formulas (2)–(5) is indicated by the continuous line. The parameters of the curve include the gap Δ_0, the Fermi energy μ, the renormalized width of the forbidden gap E_g', and the level width Γ. One can see that the prediction curve describes the experimental data very well. Good agreement was obtained for all spectra of cooperative radiation without exception. The predicted values of the parameter Δ_0 were in the range from 2.2 to 3.6 meV; the values of the parameter E_g', from 1.400 to 1.405 meV; the values of the Fermi energy μ, from 3.5 to 7.4 meV; and the effective level width Γ for different spectra varied from 0.7 to 2.3 meV. The estimated values of the concentration of electrons and holes agree with the values experimentally obtained by measuring the number of *e*–*h* pairs being pumped and the energy of superradiation pulses. Note further that the values of the parameter Δ_0 decrease with an increase in the concentration of electrons and holes.

It is known that the scattering of Cooper pairs by one another does not lead to the loss of coherence in the ensemble [21]. In our case of cooperative *e*–*h* state, we have a similar situation. It was experimentally found that the coherent properties of interaction of electrons and holes with an optical field were retained for anomalously long times exceeding many times the transverse relaxation time T_2 (which is less than 100 fs) [1, 2]. Moreover, the classical case of superradiation should have been observed during times of obviously less than 100 fs [6]. The experimentally observed coherent beats with a frequency of over 1 THz lasting several picoseconds [2] count in favor of the fact that, in the collective state, the coherence persists for several hundred femtoseconds, although the time between collisions of particles in the ensemble is less than 10 fs. This may be attributed to the fact that no loss of coherence occurs during the scattering of electrons and holes within the ensemble. In the case of collision with external electrons and holes not included in the cooperative state, the ensemble coherence also persists, because the collective state apparently takes up the impact as an integral whole, as a giant molecule of sorts. Because the number of electrons and holes in the cooperative state is very large (about 10^8 [2]) and the total mass is many times that of a single electron (hole), the ensemble coherence does not vary. In other words, the time of phase relaxation of coherent ensemble must exceed the time T_2 and increase with the number of particles.

In conclusion, note that the *e*–*h*-BCS state, unlike the regular state of Cooper pairs in a superconductor, must be substantially unstable because of recombination of *e*–*h* pairs.

5. CONCLUSIONS

The cooperative state of electrons and holes, whose radiative recombination is observed in the form of high-power femtosecond superradiation pulses, possesses a number of unusual properties. The spectral distribution of this recombination corresponds to the condensation of electrons and holes to the bottom of bands. The lifetime of the cooperative state is several hundred femtoseconds. The experimentally obtained values of *e*–*h* density are close to the critical concentration necessary for the condensation of *e*–*h* pairs at room temperature.

The coherence in the cooperative state persists much longer than the time of phase (intraband) relaxation under normal conditions. Because the high-density mode ($r_s < 1$) has been attained experimentally, the coherent collective state of electrons and holes is analogous to the BCS state of Cooper pairs in a superconductor rather than to the Bose condensate of excitons. In our case, an important and decisive part in pairing electrons and holes and inducing the coherence is played by a resonance electromagnetic field. The value of the parameter Δ_0 (the gap in the energy spectrum of quasiparticles), calculated by the approximation of optical spectra of cooperative recombination, was 2–3 meV. It is worthy of note that a BCS-like gap in a degenerate *e*–*h* system must be less in magnitude than the binding energy of an exciton (in our case, 4 meV) and must decrease as the concentration increases [19]. The latter quantity, in turn, is less than the width of observed spectra of cooperative recombination, which is defined by the lifetime of the cooperative state. This fact, in the case when the energy levels are highly smeared out, makes a direct observation of the energy gap in the spectrum of electrons and holes very problematic.

ACKNOWLEDGMENTS

We are grateful to Yu.M. Popov for numerous valuable discussions.

REFERENCES

1. P. P. Vasil'ev, Kvantovaya Élektron. (Moscow) **21**, 585 (1994).
2. P. P. Vasil'ev, Kvantovaya Élektron. (Moscow) **24**, 885 (1997).
3. P. P. Vasil'ev, Kvantovaya Élektron. (Moscow) **29**, 4 (1999).
4. N. Skribanowitz, I. T. Herman, J. C. MacGillivray, *et al.*, Phys. Rev. Lett. **30**, 309 (1973).
5. M. F. H. Schuurmans, Q. H. F. Vrehen, and D. Polder, Adv. At. Mol. Phys. **17**, 167 (1981).
6. A. V. Andreev, Usp. Fiz. Nauk **160** (12), 1 (1990) [Sov. Phys. Usp. **33**, 997 (1990)].
7. L. V. Keldysh and A. N. Kozlov, Zh. Éksp. Teor. Fiz. **54**, 978 (1968) [Sov. Phys. JETP **27**, 521 (1968)].
8. V. M. Galitskiĭ, S. P. Goreslavskiĭ, and V. F. Elesin, Zh. Éksp. Teor. Fiz. **57**, 207 (1969) [Sov. Phys. JETP **30**, 117 (1970)].

9. C. Comte and G. Mahler, Phys. Rev. B **34**, 7164 (1986).
10. T. Iida, Y. Hasegawa, H. Higashimura, *et al.*, Phys. Rev. B **47**, 9328 (1993).
11. *Bose–Einstein Condensation*, Ed. by A. Griffin, D. W. Snoke, and S. Stringari (Cambridge Univ. Press, Cambridge, 1995).
12. P. P. Vasil'ev, Usp. Fiz. Nauk **171**, 679 (2001).
13. P. Vasil'ev, *Ultrafast Diode Lasers: Fundamentals and Applications* (Artech House, Norwood, 1995).
14. S. G. Tikhodeev, Usp. Fiz. Nauk **145**, 3 (1985) [Sov. Phys. Usp. **28**, 1 (1985)].
15. J. C. Hensel, T. G. Phillips, and G. A. Thomas, in *Solid State Physics*, Ed. by H. Ehrenreich, F. Seitz, and D. Turnbull (Academic, New York, 1977), p. 207.
16. H. Kalt, K. Reimann, W. W. Ruhle, *et al.*, Phys. Rev. B **42**, 7058 (1990).
17. V. D. Kulakovskiĭ, V. B. Timofeev, and V. M. Édel'shteĭn, Zh. Éksp. Teor. Fiz. **74**, 372 (1978) [Sov. Phys. JETP **47**, 193 (1978)].
18. V. S. Dneprovskiĭ, V. I. Klimov, and M. G. Novikov, Zh. Éksp. Teor. Fiz. **99**, 843 (1991) [Sov. Phys. JETP **72**, 468 (1991)].
19. T. J. Inagaki, T. Iida, and M. Aihara, Phys. Rev. B **62**, 10 852 (2000).
20. H. Chu and Y. C. Chang, Phys. Rev. B **54**, 5020 (1996).
21. M. Tinkham, *Introduction to Superconductivity* (McGraw-Hill, New York, 1975; Atomizdat, Moscow, 1980).

Translated by H. Bronstein

Journal of Experimental and Theoretical Physics, Vol. 93, No. 6, 2001, pp. 1296–1301.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 6, 2001, pp. 1495–1502.
Original Russian Text Copyright © 2001 by Kozlov, Aleshkin, Gavrilenko.

SOLIDS
Electronic Properties

Resonant States of Shallow Acceptors in Uniaxially Deformed Germanium

D. V. Kozlov*, V. Ya. Aleshkin, and V. I. Gavrilenko
Institute for Physics of Microstructures, Russian Academy of Sciences, Nizhnii Novgorod, 603600 Russia
**e-mail: dvkoz@ipm.sci-nnov.ru*

Received March 28, 2001

Abstract—The states of shallow acceptors in uniaxially deformed germanium are studied theoretically. A nonvariational numerical computational method is developed for determining the energy and wave functions of localized states of holes in the acceptor field as well as the states of the continuous spectrum (including resonant impurity states). The dependence of the energy of the lower resonant state on strain is studied. It is found that this state is formed from the excited $4\Gamma_8^+$ state with a binding energy of 1.3 meV (in the absence of deformation) and not from the ground state. The results presented in this work may be useful in the study of the conditions for the generation of far IR radiation in deformed *p*-Ge, which involves optical transitions between resonant and localized acceptor states. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In semiconductors with a diamond-like structure (e.g., Ge and Si), the subbands of light and heavy holes are degenerate at the center of the Brillouin zone. Consequently, both these subbands make a contribution to the expansion of the wave functions of states of a shallow acceptor in Bloch functions. In the case of a uniaxial deformation of the crystal, the splitting of light and heavy hole subbands is proportional to the strain. The energy of each acceptor state in this case is displaced towards the subband whose functions mainly form the wave function of this state. In the case of a large strain, impurity states corresponding to the upper subband of holes fall into the continuous energy spectrum of the lower subband of holes (here and below, the hole energy is counted in the "upward direction" from the top of the valence band in the absence of deformation). Thus, a resonant state is formed, whose wave function has two ("localized" and "propagating") components, the latter being responsible for the interaction of the state with the lower subband. The "localized" component is formed from wave functions of the upper subband, while the "propagating" component is formed by states of the lower subband.

At the present time, such resonant states of shallow impurity centers in semiconductors have become objects of intense investigations. This is primarily due to the discovery of stimulated long-wave IR radiation in uniaxially compressed *p*-Ge in a strong electric field at helium temperatures [1–3]. The electric field ionizes shallow acceptors and heats the holes in the valence bands. At a certain threshold pressure ($P \approx 4$ kbar for $\mathbf{P} \parallel [111]$ and $P \approx 3$ kbar for $\mathbf{P} \parallel [001]$), an abrupt increase is observed in the intensity of long-wave IR radiation emitted by the sample. The authors of [1–4] attributed this effect to the emergence of stimulated emission during transition of holes from a resonant state with inverse population to localized acceptor states in the forbidden gap. The approximate estimates of the position of resonant states were made during the interpretation of stimulated emission spectra in [3, 4]. In the large strain limit, each split subband and the impurity levels associated with it were analyzed independently [3]. The spectrum of resonant states was sought in the form of the spectrum of localized states of an impurity in the vicinity of the simple anisotropic band; the calculations were based on the variational method. For small and intermediate strains, the model of zero-radius potential was employed. In this model, only one fourfold degenerate acceptor level exists in an undeformed crystal [4]. However, as observed in [4], such a model cannot claim a high accuracy in a quantitative description of acceptor states in a deformed semiconductor since the zero-radius potential is a rough approximation for the Coulomb potential.

In the present work, localized and resonant states of shallow acceptors in uniaxially deformed Ge are calculated more accurately on the basis of the wave function expansion in plane waves. This method makes it possible to determine the energies of acceptor states with an error lower than 10%. A similar approach was applied earlier for studying the resonant states of acceptors in Ge/GeSi heterostructures, in which such states are formed as a result of size quantization [5, 6].

2. COMPUTATIONAL TECHNIQUE

The energies and wave functions of holes in the field of an acceptor were determined by solving the Schrödinger equation in the effective mass approximation. The Hamiltonian was chosen in the form of the sum of the kinetic energy (Luttinger Hamiltonian), the term describing the deformation effects, and the energy of the Coulomb interaction with the charged acceptor [7]. Following [6], we used the uniaxial approximation; i.e., the dispersion relation for holes was assumed to be isotropic in a plane perpendicular to the deformation axis. For this purpose, we omitted the terms proportional to $\gamma_2 - \gamma_3$ (γ_1, γ_2, and γ_3 are the Luttinger parameters [7]) in nondiagonal elements of the Luttinger Hamiltonian. It should be noted that the correction to energy associated with the discarded terms is equal to zero in the first order of perturbation theory. In the present work, we present the results mainly for the uniaxial compression along the [111] direction. In should be noted that as a result of uniaxial compression, the subband whose states have the total angular momentum component ±3/2 along the deformation axis is displaced upwards relative to the subband whose states have the component ±1/2 along this axis (we are speaking of states with the momentum directed along the deformation axis).

In the uniaxial approximation, the total angular momentum component J_z along the deformation axis is conserved and the acceptor spectrum is doubly degenerate in the sign of this component (i.e., in $\pm J_z$). It should be noted that, in accordance with the group theory, the spectrum of holes in the acceptor field in Ge deformed along [001] or [111] must be doubly degenerate even if we take into account the anisotropy of the dispersion relation for holes in a plane perpendicular to the deformation axis. A "good" quantum number is also the parity for the reflection relative to the plane perpendicular to the symmetry axis [111] and passing through the impurity center (when speaking of the parity of a state in this section, we will mean precisely such a parity, while in subsequent sections, conventional parity relative to inversion is meant). The existence of this integral of motion considerably simplified our calculations.

The wave function for holes in the field of acceptors is sought in the form of an expansion in the eigenfunctions of holes in the absence of an acceptor:

$$\mathbf{F}(\mathbf{r}) = \sum_{n=1}^{2} \iint dk_z d\mathbf{k}_\perp C_J(\mathbf{k}_\perp, k_z, n) \times \mathbf{g}_n(\mathbf{k}_\perp, k_z, z) \exp(i\mathbf{k}_\perp \cdot \boldsymbol{\rho}), \tag{1}$$

where

$$\mathbf{g}_n(\mathbf{k}_\perp, k_z, z) = U \begin{pmatrix} \left(\left(\gamma_1 - \frac{1}{2}\gamma_3\right)k_\perp^2 + (\gamma_1 + \gamma_3)k_z^2 + \frac{1}{2}\Delta - E_n(k_\perp, k_z)\right)\cos(k_z z) \\ -\frac{2}{\sqrt{3}}(2\gamma_2 + \gamma_3)k_z k_\perp \sin(k_z z)\exp(i\alpha) \\ \frac{1}{\sqrt{3}}(\gamma_2 + 2\gamma_3)k_\perp^2 \cos(k_z z)\exp(2i\alpha) \\ 0 \end{pmatrix} \tag{2}$$

is the wave eigenfunction for even hole states,

$$\mathbf{g}_n(\mathbf{k}_\perp, k_z, z) = U \begin{pmatrix} 0 \\ -\frac{1}{\sqrt{3}}(\gamma_2 + 2\gamma_3)k_\perp^2 \cos(k_z z)\exp(-2i\alpha) \\ -\frac{2}{\sqrt{3}}((2\gamma_2 + \gamma_3)k_z k_\perp \sin(k_z z)\exp(-i\alpha)) \\ \left(\left(\gamma_1 - \frac{1}{2}\gamma_3\right)k_\perp^2 + (\gamma_1 + \gamma_3)k_z^2 + \frac{1}{2}\Delta - E_n(k_\perp, k_z)\right)\cos(k_z z) \end{pmatrix} \tag{3}$$

is the wave eigenfunction for odd hole states, k_z is the component of the wave vector of a hole along the deformation axis, $\mathbf{k}_\perp$ is the wave vector of a hole in a plane perpendicular to the deformation axis, $k_\perp$ is the modulus of this vector, and α is the angle between vector $\mathbf{k}_\perp$ and the x axis. Here,

$$U = \left\{\left[\left(\gamma_1 - \frac{1}{2}\gamma_3\right)k_\perp^2 + (\gamma_1 + \gamma_2)k_z^2 + \frac{1}{2}\Delta - E_n(k_\perp, k_z)\right]^2 + \frac{4}{3}(2\gamma_2 + \gamma_3)^2 k_z^2 k_\perp^2 + \frac{1}{3}(\gamma_2 + 2\gamma_3)^2 k_\perp^4\right\}^{-1/2}$$

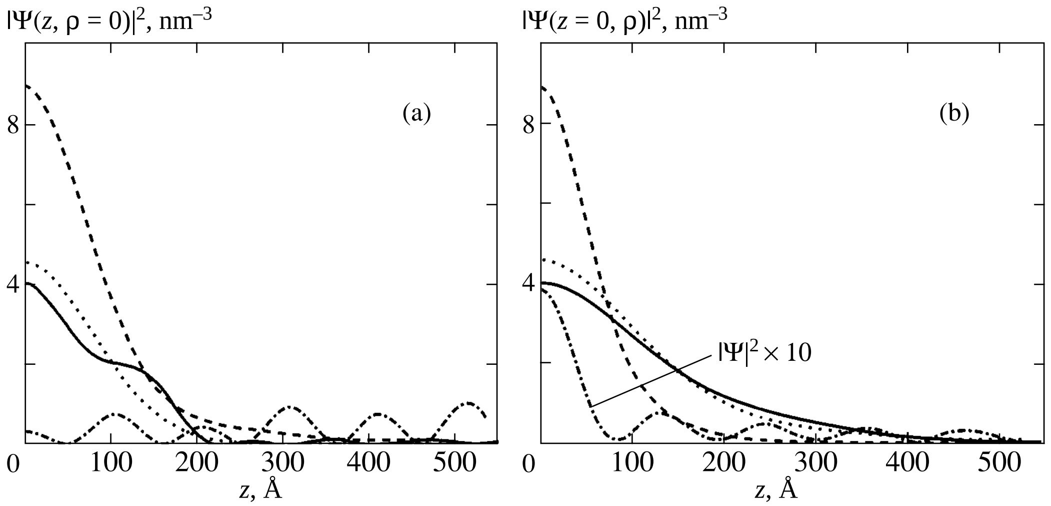


Fig. 1. Dependence of the probability density for the state of a hole in the field of an acceptor in germanium compressed along the [111] direction on *z* (a) and ρ (b). The *z* axis is chosen along the direction of deformation. The pressure is 10 kbar. The solid curve corresponds to the lower resonant state; the dashed curve, to the ground state of the acceptor; the dotted curve represents the "localized" component of the resonant state, formed by the upper subband; and the dot-and-dash curve corresponds to a nonresonant even state of the continuous spectrum with an energy of 100.8 meV (counted from the bottom of the lower subband); $J_z = +1/2$.

is the normalization coefficient,

$$E_n(k_\perp, k_z) = \gamma_1(k_z^2 + k_\perp^2) + (-1)^n\left\{\frac{1}{4}[\gamma_3(k_\perp^2 - 2k_z^2) - \Delta]^2 + \frac{4}{3}(2\gamma_2 + \gamma_3)^2 k_z^2 k_\perp^2 + \frac{1}{3}(\gamma_2 + 2\gamma_3)^2 k_\perp^4\right\}^{1/2}$$

is the dispersion relation for free holes for the two subbands, Δ is the subband splitting for $k_z, k_\perp = 0$, associated with deformation ($\Delta < 0$ for uniaxial compression). It should be noted that the dispersion relation for the lower subband (which corresponds to the minus sign) is transformed into the dispersion relation for heavy holes in the absence of deformation ($\Delta = 0$). Thus, if the strain tends to zero, the lower subband becomes the subband of heavy holes and the upper subband is the subband of light holes.

The dependence of coefficients $C_J(\mathbf{k}_\perp, k_z, n)$ on angle α in the uniaxial approximation has the form [5, 6]

$$C_J(\mathbf{k}_\perp, k_z, n) = C_J(k_\perp, k_z, n)\exp\left[i\alpha\left(J_z - \frac{3}{2}\right)\right], \quad (4)$$

where J_z is the total angular momentum component along the deformation axis. Taking into account relation (4), we can derive the following expression for determining $C_J(\mathbf{k}_\perp, k_z, n)$:

$$[E_n(k_\perp, k_z) - E]C_J(k_\perp, k_z, n) + \sum_{n'=1}^{2}\int_0^\infty dk'_z \times \int_0^\infty dk'_\perp \sqrt{k'_z k'_\perp}\, C_J(k'_\perp, k'_z, n')V(k_\perp, k_z, n, k'_\perp, k'_z, n') = 0, \quad (5)$$

where *E* is the energy of the state of holes in the acceptor field

$$V(k_\perp, k_z, n, k'_\perp, k'_z, n') = -\frac{e^2}{\chi}\int_0^{2\pi} d\beta \exp\left[i\beta\left(J - \frac{3}{2}\right)\right] \times \int d^3r \left\langle \mathbf{g}_{n'}(k_\perp, k_z, r)\left|\frac{1}{r}\right|\mathbf{g}_n(k'_\perp, k'_z, r)\right\rangle, \quad (6)$$

β is the angle between vectors $\mathbf{k}_\perp$ and $\mathbf{k}'_\perp$, χ is the permittivity of the semiconductor, and *e* is the electron charge. It should be noted that the kernel of the integral operator in the equation for $C_J(k_\perp, k_z, s)$ is symmetric relative to the variables and *V* is a real quantity. This allows us to solve Eq. (5) by diagonolizing the symmetric real-valued matrix. Indeed, if the step in $k'_{\perp, z}$ is smaller than the reciprocal Bohr radius, the integrand varies over a step only slightly. In this case, the integral can be presented as a sum in $k'_{\perp, z}$. The quantities $C_J(k_\perp, k_z, s)$ are small for values of $k'_{\perp, z}$ much larger than the reciprocal Bohr radius. Consequently, the series in $k'_{\perp, z}$ can be truncated without introducing a serious error. Since the Bohr radius of the state of a hole in the field of an acceptor in undeformed germanium is of the order

Table 1. Binding energies of some acceptor states in undeformed Ge, meV

States	$J_z = \pm 1/2$	$J_z = \pm 3/2$	$J_z = \pm 5/2$	Theory [8, 9]	Experiment (boron) [10]
Ground state $1\Gamma_8^+$	9.65	9.23	–	9.8 [9]	10.8
$1\Gamma_8^-$ (G-line)	4.5	4.7	–	4.58 [8]	4.6
$2\Gamma_8^+$ (E-line)	3.05	3.3	–	2.9 [9]	3.25
$2\Gamma_8^-$ (D-line)	–	2.8	2.95	2.88 [8]	2.88
Doubly degenerate ($1\Gamma_7^+$ or $1\Gamma_6^+$)	–	2.4	–	–	–
$1\Gamma_7^-$ (C-line)	2.2	–	–	2.13 [8]	2.13
$3\Gamma_8^+$	1.55	1.49	–	–	–
$4\Gamma_8^+$	1.33	1.29	–	–	–

of 40 Å, we calculated the spectrum for pressures leading to splitting of subbands with energies lower than 10 meV, choosing the step in $k_\perp$ as well as in k_z equal to 2×10^{-3} Å^{-1}. For large strains, it is expedient to choose different values of steps in $k_\perp$ and k_z to improve the accuracy of calculations used for studied resonant and localized states. This is associated with a considerable anisotropy of wave functions: the wave functions of resonant states must be compressed along z, while those for localized states must be extended along z. We choose the steps in $k_\perp$ in k_z equal to 1.8×10^{-3} Å^{-1} and 1×10^{-3} Å^{-1}, respectively, for localized states, and 1×10^{-3} Å^{-1} in $k_\perp$ and 2.3×10^{-3} Å^{-1} in k_z for resonant states.

Thus, the problem of determining localized or delocalized states of an acceptor is reduced to diagonalization of a finite-dimensional symmetric matrix (in our calculations, the dimensionality of this matrix was 1000×1000).

The wave functions of acceptors have a complex structure. Expansion (1) contains terms corresponding to two different subbands. If the expansion of the wave function of a certain acceptor state mainly includes the envelopes of the wave functions of a certain hole subband, we assume that such a state corresponds to this subband. In a deformed material, the acceptor ground state corresponds to the lower subband and lies below its bottom. The energies of states corresponding to the upper subband can be lower (localized states) or higher (resonant states) than the energy corresponding to the bottom of the lower subband. The localized states associated with the upper subband become resonant states upon an increase in the uniaxial strain of the crystal.

Figure 1 shows the spatial distribution of the probability density for the lower resonant (solid curve) and localized ground (dashed curve) states of holes in the field of an acceptor in germanium compressed along the [111] axis. It was mentioned above that the wave function of the resonant state has two components: the "localized" component formed by the upper subband and the "propagating" component formed by the lower subband. The dotted curve in Fig. 1 describes the probability density corresponding to the localized component of the wave function of the resonant state. It can be seen that the wave function of the resonant state is similar to the wave function of the ground state at small distances from the impurity center and is a standing wave along the z coordinate for large distances. Comparing the wave function of a resonant state with the wave function of a nonresonant state of the continuous spectrum (dot-and-dash curve in Fig. 1), we note that the delocalized component of the given resonant state is mainly formed by the waves propagating along the z direction since the oscillations of the wave function of this state in the ρ coordinate are manifested only slightly.

3. RESULTS AND DISCUSSION

In order to verify the accuracy of the method, we calculated the energies of the ground state and several excited states of a shallow acceptor in undeformed germanium, corresponding to different momentum components along the symmetry axis. The results are presented in Table 1 (in our calculations, the uniaxial symmetry axis was chosen along the [111] crystallographic direction). The first column of the table presents the classification of states according to the types of representations to which the functions of the determined states in the germanium crystal must correspond. The corresponding transitions in the absorption spectrum of germanium are indicated for several states. Table 1 contains the results of calculations made in [8, 9], which can apparently be regarded at present as the most accurate (fifth column). The sixth column contains the results of experimental determination of the energy of levels in Ge from [10].

Table 2. Binding energies of some acceptor states in uniaxially compressed Ge (**P** || [100], P = 6 kbar)

Classification of states	Binding energy, meV (our calculations)	Binding energy, meV (experiment [12])
1*s*	5.56	6.00
2*s*	1.32	1.385
$2_{p\pm1}$	1.44	1.335
$2p_0$	0.9	0.889
$3_{p\pm1}$	0.621	0.613

It is well known that the ground state $1\Gamma_8^+$ of an acceptor is fourfold degenerate. In calculations, the uniaxial symmetry was used instead of the cubic symmetry. For this reason, the ground state calculated by us turned out to be split. The binding energy of the state corresponding to $J_z = \pm1/2$ amounts to 9.65 meV, while the binding energy of the state corresponding to $J_z = \pm3/2$ is 9.23 meV. Thus, the splitting of the ground state associated with the "wrong" symmetry used in calculations amounts approximately to 5% of the ionization energy. This is also holds for excited fourfold degenerate states (see Table 1).

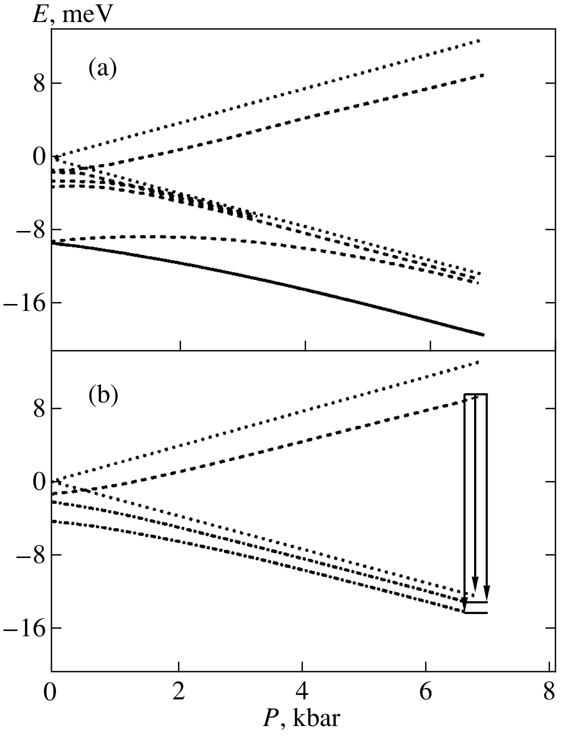


Fig. 2. Pressure dependence of the energy of acceptor states in germanium compressed along the [111] direction. The solid curve corresponds to the acceptor ground state ($J_z = \pm1/2$, even). (a) Ground state and several even excited states. The dashed curves correspond to several states with the same symmetry ($J_z = \pm3/2$, even): the resonant "ground" state and all excited states "emerging" from lower-lying levels at zero pressure. (b) Resonant "ground " state and several localized odd states. The dashed curve corresponds to the resonant "ground " state. Dot-and-dash curves correspond to odd localized states to which optical transitions from the lower resonant state are allowed. The arrows indicate allowed transitions from the resonant ground state, which may be present in the emission spectrum considered in [3]. Dotted curves correspond to the edges of hole subbands.

It should be noted that the largest discrepancy between the calculated and experimental results is observed for the ground (1*s* type) state and for the $2\Gamma_8^+$ (2*s* type) state. The mismatching between the theoretical and experimental result is 12% for the ground state and 7.5% for the $2\Gamma_8^+$ state, while for other excited states the discrepancy between the calculated and experimental values is within 4–5%. This can be explained by the presence of a chemical shift [11] which is significant for *s*-like states and is virtually absent for other states since the wave function for the latter states has a node at the point of location of an acceptor.

In order to verify the accuracy of the approach developed in this work, we compare the results of our computations with the experimental data on the impurity photoconductivity of Ge strongly compressed along the [100] direction [12] (Table 2). The pressure amounts to 6 kbar, which corresponds to a subband splitting of 36 meV.

It can be seen that the discrepancy between the theoretical and experimental results is smaller than in the case of undeformed germanium: the mismatching does not exceed 5% for all energy levels except the 1*s* state (with a mismatching of 7.3%), in which the chemical shift is significant.

In the case of uniaxial deformation, the fourfold degeneracy of the levels is removed. The solid curve in Fig. 2a shows the dependence of the ground state in uniaxially compressed Ge on pressure in the [111] direction. The ground state in deformed Ge is an even state with the total angular momentum component $J_z = \pm1/2$. Dashed curves in Figs. 2a and 2b correspond to the energy of the resonant "ground" state and several lower-lying states with the same symmetry (even states with $J_z = \pm3/2$). A comparison of the data presented in Table 1 and Fig. 1 shows that the resonant "ground" state originates from the excited state $4\Gamma_8^+$ with a binding energy of 1.3 meV (in the absence of deformation) rather than from the ground state as in the model of zero-radius potential (in which it appears in the continuous spectrum for $P \approx 4$ kbar) [3]. This is not surprising since, like in an undeformed crystal, the $4\Gamma_8^+$ energy level is the first state whose wave function is constructed mainly from the wave functions of light holes, while the wave functions of all lower-lying levels receive the main contributions from the states of the subband of heavy holes. It can be seen, in particular, that the upper state split from the ground energy level does not experience anti-crossing and does not appear in the continuous spectrum in contrast to what was assumed earlier [3]. For large strains, this state passes to the $3d_{\pm1}$ level in the model of a simple anisotropic band with an ionization energy of 0.7 meV. Thus, the conclusion drawn in [3]—according to which the minimum pressure under which stimulated emission can still be observed ($P \approx 4$ kbar for $\mathbf{P} \parallel [111]$) is due to the affiliation of the level split from the ground state of the acceptor to the continuous spectrum—appears erroneous. Dot-and-dash curves in Fig. 2b describe the pressure dependence of the positions of the excited localized levels on which radiative transitions from the resonant "ground" state are allowed (odd states). The arrows indicated allowed transition from the resonant "ground" state, which might be present in the emission spectrum considered in [3].

It should be noted that as the pressure increases, the resonant state becomes more pronounced. We estimated the width of this state for several values of pressure. The width of the resonant "ground" state is 0.5 meV for $P = 3$ kbar, 0.4 meV for $P = 4$ kbar, and 0.3 meV for $P = 6.85$ kbar. The lifetime of this state should increase accordingly. The emergence of emission under pressures exceeding 4 kbar is probably due to an increase in the lifetime of resonant states with pressure.

In conclusion, let us consider the effect of an electric field on stationary and resonant states of acceptors in deformed Ge. Since spectral studied of stimulated emission were carried out in an electric field of 3 kV/cm under a pressure of 6.85 kbar ($\mathbf{E} \parallel \mathbf{P}$) [3], we estimated the Stark effect exactly for these conditions. The application of the electric field has two consequences:

(1) It shifts all impurity states downwards, thus reducing the separation between energy levels since excited states have large orbits and experience a stronger effect of the electric field.

(2) The electric fields renders all states resonant states since the probability of tunnelling to the continuous spectrum becomes appreciable. The displacement of deep levels can be estimated in perturbation theory; in this case, the first-order correction is equal to zero. The second-order correction for the ground state corresponding to the lower subband amounts approximately to 1.6 meV (the ionization energy of this state in zero electric field is 6.2 meV). The corrections for excited states are much higher than their ionization energies (the ionization energy of the $2p_{\pm 1}$ state is 2.15 meV); i.e., perturbation theory is violated. Thus, the spectrum of excited states associated with the lower subband changes significantly upon application of an electric field of 3 kV/cm.

The effect of the field on the resonant states associated with the upper subband is much weaker. The second-order correction to the ground state is only 0.2 meV (the binding energy is 3.95 meV), while the correction to the energy of $2p_{\pm 1}$ state amounts to 0.9 meV (the binding energy is 1.2 meV). Such a difference from the states located under the lower subband appears because the localization scale for the wave function of the resonant state is smaller in the direction of the electric field than in the transverse direction, while the situation for localized states is the opposite (see Fig. 1). It should be noted that since the electric field affects the states belonging to the lower subband more strongly, the lifetime of such states may turn out to be shorter than that of the resonant states in zero electric field. This circumstance might be important for creating an inverse population of impurity states in compressed germanium.

ACKNOWLEDGMENTS

This work was financed by the Russian Foundation for Basic Research (project nos. 00-02-16568, 00-02-81022, 01-02-16106), NATO GLG 975592, Interdepartmental Science and Engineering Program "Physics of Solid Nanostructures" (project no. 97-1069), "Physics of Microwaves" (project no. 4.5), National Science and Engineering Program "Fundamental Spectroscopy" (project no. 8/02.08), State Program Supporting Leading Scientific Schools (project no. 00-15-96618), Federal Purpose Program "Integration" (projects nos. 540 and 541), and the Program "Universities of Russia" (project no. 015.01.0194).

REFERENCES

1. I. V. Altukhov, M. S. Kagan, K. A. Korolev, *et al.*, Zh. Éksp. Teor. Fiz. **101**, 756 (1992) [Sov. Phys. JETP **74**, 404 (1992)].
2. V. M. Bondar, L. E. Vorob'ev, A. T. Dalakyan, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **70**, 257 (1999) [JETP Lett. **70**, 265 (1999)].
3. I. V. Altukhov, M. S. Kogan, K. A. Korolev, *et al.*, Zh. Éksp. Teor. Fiz. **115**, 89 (1999) [JETP **88**, 51 (1999)].
4. M. A. Odnoblyudov, A. A. Pakhomov, V. M. Chistyakov, and I. N. Yassievich, Fiz. Tekh. Poluprovodn. (St. Petersburg) **31**, 1180 (1997) [Semiconductors **31**, 1014 (1997)].
5. V. Ya. Aleshkin, B. A. Andreev, V. I. Gavrilenko, *et al.*, Fiz. Tekh. Poluprovodn. (St. Petersburg) **34**, 582 (2000) [Semiconductors **34**, 563 (2000)].
6. V. Ya. Aleshkin, B. A. Andreev, V. I. Gavrilenko, *et al.*, Nanotechnology **11**, 348 (2000).
7. G. L. Bir and G. E. Pikus, *Symmetry and Strain-Induced Effects in Semiconductors* (Nauka, Moscow, 1972, Wiley, New York, 1975), pp. 294, 394.
8. A. Baldereschi, O. Lipari, and M. Altarelli, Izv. Akad. Nauk SSSR, Ser. Fiz. **42**, 1179 (1978).
9. A. Baldereschi and O. Lipari, Phys. Rev. B **8**, 2697 (1973).
10. E. E. Haller and W. Hansen, Solid State Commun. **15**, 687 (1974).
11. S. T. Pantelides, Rev. Mod. Phys. **50**, 797 (1978).
12. E. I. Voevodin, E. M. Gershenzon, G. N. Gol'tsman, and N. G. Ptitsina, Fiz. Tekh. Poluprovodn. (Leningrad) **23**, 1356 (1989) [Sov. Phys. Semicond. **23**, 843 (1989)].

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 93, No. 6, 2001, pp. 1302–1306.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 6, 2001, pp. 1503–1508.
Original Russian Text Copyright © 2001 by Buchel'nikov, Romanov, Vasil'ev, Takagi, Shavrov.

SOLIDS
Electronic Properties

Model of Colossal Magnetostriction in the Martensite Phase of Ni–Mn–Ga Alloys

V. D. Buchel'nikov[a], *, V. S. Romanov[a], A. N. Vasil'ev[b], **, T. Takagi[c], and V. G. Shavrov[d], ***
[a]*Chelyabinsk State University, Chelyabinsk, 454021 Russia*
[b]*Moscow State University, Vorob'evy gory, Moscow, 119899 Russia*
[c]*Institute of Fluid Science, Tohoku University, 980-8577, Sendai, Japan*
[d]*Institute of Radio Engineering and Electronics, Russian Academy of Sciences, ul. Mokhovaya 18, Moscow, 101999 Russia*
**e-mail: buche@cgu.chel.su*
***e-mail: vasil@lt.phys.msu.su*
****e-mail: shavrov@mail.cplire.ru*

Received May 3, 2001

Abstract—A model is proposed for describing the experimentally observed martensite and magnetic domain structures in Heusler ferromagnetic alloys $Ni_{2+x}Mn_{1-x}Ga$. On the basis of this model, the field dependences of magnetization and deformation of the alloys are calculated numerically and an expression for the maximum attainable strains induced by external magnetic fields in these alloys is derived. It is shown that for small values of the effective elastic modulus and demagnetizing factor of alloys, the strains induced by the magnetic field may attain maximum possible values of approximately 5%, which are determined by lattice distortions as a result of the martensite transition in fields of about 1 kOe. *© 2001 MAIK “Nauka/Interperiodica”.*

The experimental observation of structural and magnetic domains in the martensite phase of $Ni_{2+x}Mn_{1-x}Ga$ Heusler alloys has been reported recently. Structural and magnetic domains in $Ni_{2+x}Mn_{1-x}Ga$ single crystals and their behavior in external magnetic fields of various orientations relative to crystallographic axes were studied using microphotography [1] and magnetic force microscopy [2]. The strains induced by a magnetic field in $Ni_{2+x}Mn_{1-x}Ga$ polycrystals and single crystals were also measured experimentally [3–5]. It was found that in fields ≤10 kOe, these strains may attain record-high values of several percent (this effect can be referred to as colossal magnetostriction). The experimental results also proved that not only structural, but also magnetic domains play a significant role in the colossal magnetostriction effect in $Ni_{2+x}Mn_{1-x}Ga$ alloys.

The effect of a magnetic field on structural and magnetic domains in $Ni_{2+x}Mn_{1-x}Ga$ was investigated in [6]; however, the specific experimentally observed domain structure was not considered and the energy of the ferromagnet was not calculated consistently either.

In the present work, we consider a model on a self-consistent structure of martensite and magnetic domains on the basis of the results of experimental investigations [1, 2], a theoretical analysis of the phase diagram of the $Ni_{2+x}Mn_{1-x}Ga$ alloys [7], and the thermodynamic theory of domains [6]. The behavior of the model in a magnetic field is studied and the strains induced by the field in these alloys are calculated.

Upon a transition from the high-temperature cubic phase to the low-temperature tetragonal modification, a $Ni_{2+x}Mn_{1-x}Ga$ single crystal splits into martensite domains of three types, each of which corresponds to deformation (compression or extension) of the crystal lattice along directions of the [100] type [8]. In a ferromagnetic crystal, structural domains in turn split into magnetic domains. The direction of magnetization in each magnetic domain coincides with the principal crystallographic axis of the structural domain. An analysis of the experimental results [1, 2] shows that the martensite and magnetic domain structures in $Ni_{2+x}Mn_{1-x}Ga$ can be described by the model presented in Fig. 1. The tree structure of magnetic domains observed in [2], which is formed at the boundaries of structural and magnetic domains, is disregarded in view of the smallness of the volume occupied by these domains.

Let us assume, for the sake of definiteness, that the tetragonal axis in a structural domain of the first type (see Fig. 1) is directed along the x axis. In domains of the second and third types, tetragonal axes are directed along the y and z axes, respectively. We also assume that each structural domain splits into 180-degree magnetic domains.

Let us study the effect of an external magnetic field directed along the x axis on the domain structure presented in Fig. 1. In zero magnetic field, all structural domains have the same energy and, accordingly, the volumes occupied by different types of domains are identical. The application of the magnetic field increases the volume occupied by the first type of domains at the expense of domains of the second and third types, which exhibit identical behavior relative to the field in view of the symmetry of the model. The volume fractions of different types of domains in the presence of the field can be presented in the form

$$f_1 = \frac{1}{3} + \alpha, \quad f_{2,3} = \frac{2}{3} - \alpha, \qquad (1)$$

where f_1 is the volume fraction of the first type of domains, $f_{2,3}$ is the volume fraction of domains of the second and third types, and α is a parameter taking into account the change in the domain volume in a magnetic field.

It follows from the phase diagram of a cubic ferromagnet in zero magnetic field that, upon a change in temperature, it may pass to tetragonal phases with magnetizations along axes of the {001} type [7]. For example, in a phase with $m_z = 1$, i.e., $\mathbf{M} \parallel [001]$ ($\mathbf{m}$ is the unit vector along the magnetization vector $\mathbf{M}$), the spontaneous deformation tensor has the form

$$e_{ik} = \left\| \begin{matrix} e_{xx} & 0 & 0 \\ 0 & e_{xx} & 0 \\ 0 & 0 & e_{zz} \end{matrix} \right\|, \qquad (2)$$

where

$$e_{xx} = -\frac{B_1}{3(c_{11} + 2c_{12})} - \frac{e_0}{\sqrt{6}},$$

$$e_{zz} = -\frac{B_1}{3(c_{11} + 2c_{12})} + \frac{2e_0}{\sqrt{6}},$$

and the variable e_0 can be determined from the equation

$$ae_0 + be_0^2 + ce_0^3 + \frac{2B_2}{\sqrt{6}} = 0,$$

in which a, b, and c are linear combinations of the elastic moduli components of the second (c_{ik}), third, and fourth orders, respectively, and B_i are the magnetostriction constants. In the remaining two phases with magnetizations directed along the [100] and [010] axes, the spontaneous deformation tensor can be obtained from expression (2) through cyclic permutation of indices. Thus, three tetragonal phases with the same energy may exist in a ferromagnet; as a result, the crystal may contain three types of structural domains. The first terms in the expression for e_{ik}, which describe the change in strain due to bulk magnetostriction, can be neglected as compared to the effects associated with structural distortions of the lattice.

The substitution of the total energy for a cubic ferromagnet [6] into formula (2) leads to the following expression for the energy of each type of domains in the intrinsic crystallographic system of coordinates:

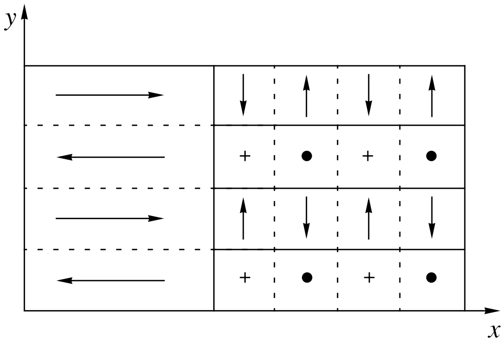


Fig. 1. Model of martensite and magnetic domain structures of the $Ni_{2+x}Mn_{1-x}Ga$ alloy. Solid lines indicate the domain walls between structural domains and dashed lines show domain walls between magnetic domains. Arrows indicate the directions of magnetization vectors in magnetic domains; symbols + and ● correspond to the direction of magnetization in magnetic domains perpendicular to the plane of the sample; the coordinate axes x and y coincide with the crystallographic axes [100] and [010]; the z axis ([001]) is directed along the normal to the plane of the figure.

$$F = \frac{1}{2}ae_0^2 + \frac{1}{3}be_0^3 + \frac{1}{4}ce_0^4 + Km_z^2$$
$$+ K_1(m_x^2m_y^2 + m_y^2m_z^2 + m_z^2m_x^2) - \mathbf{H} \cdot \mathbf{M}, \qquad (3)$$

where $K = \sqrt{6}B_2e_0/2$ is the axial anisotropy induced by magnetostriction, K_1 is the first cubic anisotropy constant, and $\mathbf{H}$ is the intrinsic magnetic field. The measurements of magnetic anisotropy of $Ni_{2+x}Mn_{1-x}Ga$ alloys [4] proved that the axial anisotropy is much stronger than the cubic anisotropy; for this reason, the latter will be disregarded in the subsequent analysis.

Let us write the energy of the domain structure under investigation (see Fig. 1) in the form

$$F = F_e + F_m + F_Z, \qquad (4)$$

where F_e is the energy of the elastic subsystem, F_m is the energy of the magnetic subsystem, and F_Z is the Zeeman energy.

We can write the energy of the elastic subsystem in the form

$$F_e = \frac{1}{2}\tilde{C}_{11}(\langle e_{xx}\rangle^2 + \langle e_{yy}\rangle^2 + \langle e_{zz}\rangle^2)$$
$$+ \tilde{C}_{12}(\langle e_{xx}\rangle\langle e_{yy}\rangle + \langle e_{xx}\rangle\langle e_{zz}\rangle + \langle e_{zz}\rangle\langle e_{yy}\rangle), \qquad (5)$$

where the angle brackets indicate averaging over volume and $\tilde{C}_{ijkl}$ is the tensor of effective elastic moduli of the $Ni_{2+x}Mn_{1-x}Ga$ alloy with the given domain structure [8]. The model of the domain structure presented in Fig. 1 and taking into account Eq. (1) suggests that the average values of the deformation tensor components of the alloy can be written in the form

$$\langle e_{ij}\rangle = \left(\frac{1}{3}+\alpha\right)e_{ij}^{(1)} + \left(\frac{2}{3}-\alpha\right)\left(\frac{1}{2}e_{ij}^{(2)} + \frac{1}{2}e_{ij}^{(3)}\right), \qquad (6)$$

where $e_{ij}^{(k)}$ is the deformation tensor of the *k*th type of domains, defined by formula (2). The substitution of these strains into Eq. (6) leads to the following expression for the deformation tensor averaged over the sample volume (Fig. 1):

$$\langle e_{ij}\rangle = \frac{\sqrt{6}}{4}\begin{pmatrix} 2e_0\alpha & 0 & 0 \\ 0 & -e_0\alpha & 0 \\ 0 & 0 & -e_0\alpha \end{pmatrix}. \qquad (7)$$

It should be noted that while writing the expression for the averaged deformation tensor, we disregarded bulk magnetostriction. The substitution of relation (7) into Eq. (5) leads to the following expression for the elastic energy of the domain structure under investigation:

$$F_e = \frac{9}{8}\tilde{C}e_0^2\alpha^2, \qquad (8)$$

where $\tilde{C} = \tilde{C}_{11} - \tilde{C}_{12}$.

For the chosen direction of the external magnetic field, the magnetization in the first type of domains is parallel or antiparallel to the field, while in the second and third types of domains, the magnetization is perpendicular to the field. For this reason, in the course of magnetization, domains of the first type experience only displacements of domain walls, while domains of the second and third type experience only the rotation of the magnetization vector. In this case, the anisotropy energy and the Zeeman energy can be written in the form

$$F_m = \left(\frac{2}{3}-\alpha\right)K\cos^2\psi, \quad F_z = -H\langle M_x\rangle, \qquad (9)$$

where the magnetization vector component along the magnetic field has the form

$$\langle M_x\rangle = \left(\frac{1}{3}+\alpha\right)M^{(1)} + \left(\frac{2}{3}-\alpha\right)M_0\cos\psi, \qquad (10)$$

and the intrinsic magnetic field has the form $H = H_0 - 4\pi N\langle M_x\rangle$, where H_0 is the external magnetic field and N is the demagnetizing factor; ψ is the angle between the directions of the magnetization vector and the magnetic field in domains of the second and third types; and $M^{(1)}$ is the average magnetization in the first type of domains. It can be written in the form

$$M^{(1)} = \begin{cases} H_0M_0/H_c, & H_0 < H_c, \\ M_0, & H_0 > H_c, \end{cases} \qquad (11)$$

where the phenomenological parameter H_c corresponds to the field in which the displacement of magnetic domain walls in martensite domains of the first type is completed.

Substituting Eqs. (8) and (9) into relation (4), we obtain the following expression for the energy of the domain structure depicted in Fig. 1:

$$\begin{aligned} F = {} & \frac{9}{8}\tilde{C}e_0^2\alpha^2 + \left(\frac{2}{3}-\alpha\right)K\cos^2\psi \\ & - H\left[\left(\frac{1}{3}+\alpha\right)M^{(1)} + \left(\frac{2}{3}-\alpha\right)M_0\cos\psi\right]. \end{aligned} \qquad (12)$$

The equilibrium values of parameters α and ψ for the domain structure under investigation can be determined by minimizing expression (12) in these variables. This leads to the following equations for determining α and ψ:

$$\begin{aligned} & 32\pi NKM_0^2\cos^3\psi - 96\pi NKM_0M^{(1)}\cos^2\psi \\ & + [6\tilde{C}e_0^2(8\pi NM_0^2 + 3K) + 64\pi NKM^{(1)2}]\cos\psi \\ & - 3\tilde{C}e_0^2M_0(3H_0 - 8\pi NM^{(1)}) = 0, \end{aligned}$$

$$\begin{aligned} \alpha = {} & -\frac{4}{3}[(M^{(1)} - M_0\cos\psi) \\ & \times(16\pi NM_0\cos\psi - 3H_0 + 8\pi NM^{(1)}) - 3K\cos^2\psi] \\ & \times[9\tilde{C}e_0^2 + 32\pi N(M^{(1)} - M_0\cos\psi)^2]^{-1}. \end{aligned} \qquad (13)$$

Substituting the solutions of these equations into expressions (7) for the deformation tensor and (10) for the magnetization, we can determine the dependence of these characteristics of the $Ni_{2+x}Mn_{1-x}Ga$ alloy on the magnitude of the applied magnetic field. The results of numerical calculations of the field dependences of magnetization, parameter α, and strain $\langle e_{xx}\rangle$ for various values of the effective elastic modulus $\tilde{C}$ and the demagnetizing factor N are presented in Figs. 2 and 3. In our calculations, we used the following values of parameters for the $Ni_{2+x}Mn_{1-x}Ga$ alloy [3, 4]:

$$x \approx 0, \quad M_0 = 500\ \text{Oe}, \quad e_0 = 0.06,$$

$$K = 2\times 10^6\ \text{erg/cm}^3.$$

The results presented in Fig. 2 were obtained for $N = 0.02$ and $H_c = 2000$ kOe, while those in Fig. 3 correspond to $N = 0.5$ and $H_c = 2500$ kOe. In our calculations, we also took into account the fact that the alloy acquires the monodomain structure upon the attainment of the value $\alpha = 2/3$. In this case, the strain $\langle e_{xx}\rangle \approx 0.06$, which is determined by the ratio of the crystal lattice parameters in the tetragonal phase ($c/a = 0.94$). A further increase in the magnetic field does not lead to any further deformation.

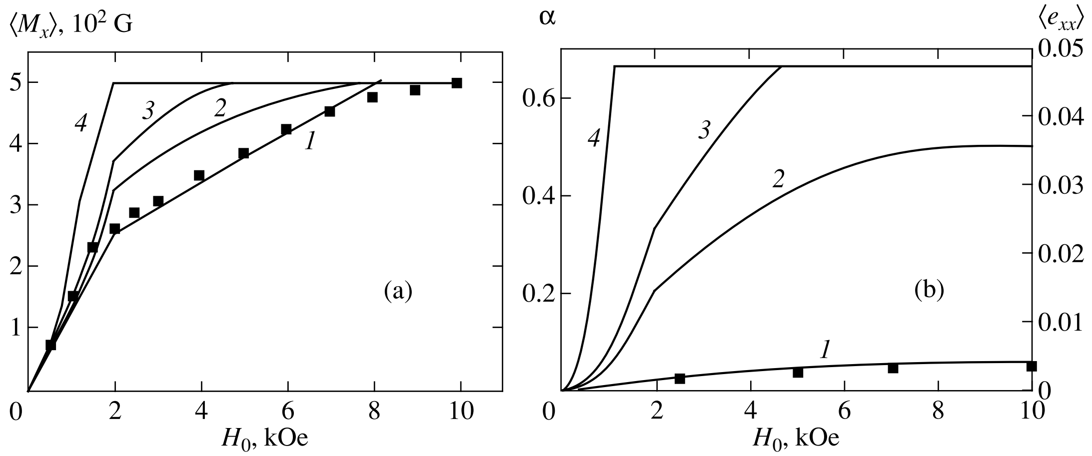


Fig. 2. Field dependences of (a) magnetization, (b) volume fraction and induced strain for $N = 0.02$ and $H_c = 2000$ kOe for various values of the effective elastic modulus $\tilde{C}$ (in erg/cm^3): 5×10^9 (*1*), 5×10^8 (*2*), 3×10^8 (*3*), and 5×10^7 (*4*); the symbols correspond to experimental results obtained in [3].

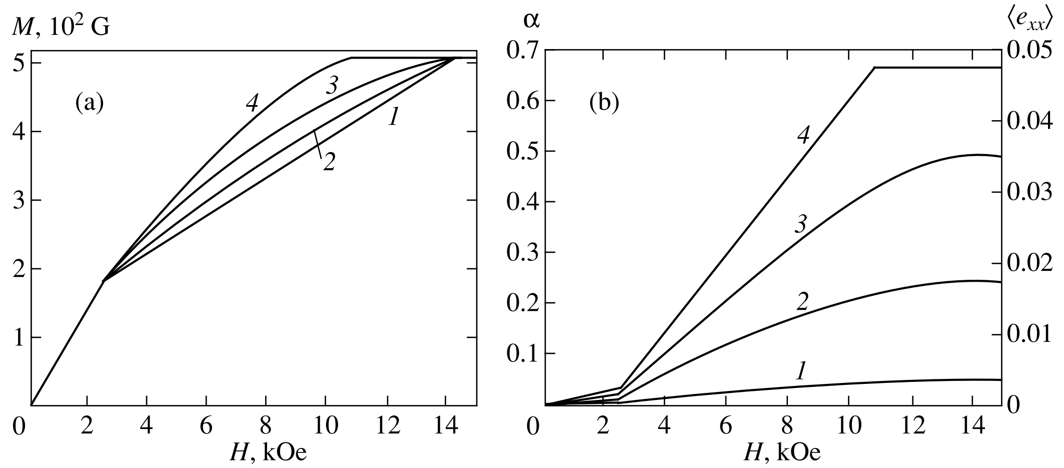


Fig. 3. Field dependences of (a) magnetization, (b) volume fraction and induced strain for $N = 0.5$ and $H_c = 2500$ kOe for various values of the effective elastic modulus $\tilde{C}$ (in erg/cm^3): 5×10^9 (*1*), 1×10^9 (*2*), 5×10^8 (*3*), and 3×10^8 (*4*).

In fields $H_0 \sim H_c$ and $H_0 \sim H_1 = H_a + 8\pi N\langle M_x\rangle$, where $H_a = 2K/M_0$, the field dependences of $\langle M_x\rangle$, α, and $\langle e_{xx}\rangle$ acquire kinks. This can be explained as follows. For $H_0 < H_c$, the magnetization increases due to the displacement of domain walls in structural domains of the first type as well as due to the rotation of magnetization vectors in structural domains of the second and third types. For $H_0 = H_c$, magnetic domains in structural domains of the first type disappear altogether and a further increase in the magnetization in fields $H_c < H_0 < H_1$ is associated only with the rotation of magnetization in structural domains of the second and third types. In the field $H_0 = H_1$, vectors **M** in structural domains of the second and third type are completely reoriented to the direction of the external magnetic field and the magnetization of the sample attains saturation.

It follows from Figs. 2 and 3 that structural domains also experience a rearrangement simultaneously with the rearrangement of magnetic domains under the action of the external field. As the magnetic field increases, structural domains of the first type grow. For $H_0 > H_1$, when the magnetization in structural domains is the same, the Zeeman energies in them also become identical. Since the motion of structural domain walls under the action of the magnetic field occurs precisely due to an increase in the Zeeman energy, a further growth of structural domains of the first type at the expense of domains of the second and third type becomes impossible for $H_0 > H_1$.

The limiting value of the volume fraction α which can be attained upon the saturation of magnetization amounts to

$$\alpha_{\max} = \frac{2M_0H_a}{9\tilde{C}e_0^2}, \tag{14}$$

while the corresponding value of strain $\langle e_{xx} \rangle$ induced by the magnetic field is expressed by the formula

$$\langle e_{xx} \rangle_{\max} = \frac{\sqrt{6}M_0H_a}{9\tilde{C}e_0^2}. \tag{15}$$

Figures 2 and 3 show that for large values of the effective elastic modulus $\tilde{C}$, the values of $\alpha_{\max}$ and $\langle e_{xx} \rangle_{\max}$ are smaller than the maximum values determined by lattice distortions. It follows from Eqs. (14) and (15) that the necessary condition for the attainment of maximum strains induced by a magnetic field in the $Ni_{2+x}Mn_{1-x}Ga$ alloys is that the elastic energy must be of the order of the anisotropy energy. This can be attained for small values of the elastic modulus in the vicinity of the martensite transformation point.

It should be noted that Fig. 2 corresponds to the sample magnetization in the plane of the plate, while Fig. 3 corresponds to the sample magnetization in the transverse direction. A comparison of Figs. 2 and 3 shows that the effect of colossal magnetostriction of the $Ni_{2+x}Mn_{1-x}Ga$ alloys is strongly anisotropic: the maximum strain of the sample for the sample magnetized in the plane of the plate can be attained in much weaker magnetic fields as compared to the case of magnetization perpendicular to the plane of the plate. This result, as well as the calculated field dependences of magnetization and strain for $N = 0.02$, $H_c = 2000$ kOe, and $\tilde{C} = 5 \times 10^9$ erg/cm^3 (Fig. 2), is in good agreement with experimental data [3–5].

Thus, we have considered a thermodynamic model describing the structural and magnetic domains in Heusler ferromagnetic $Ni_{2+x}Mn_{1-x}Ga$ alloys. On the basis of the model, the field dependences of magnetization and strain of the alloys have been calculated and the expression for the maximum attainable strains induced in these alloys by an external magnetic field has been derived.

ACKNOWLEDGMENTS

This work was financially supported by the Russian Foundation for Basic Research (project no. 99-02-18247) and by Ministry of Education of the Russian Federation (project no. E00-3.4-536).

REFERENCES

1. H. D. Chopra, Ch. Ji, and V. V. Kokorin, Phys. Rev. B **61**, 14 913 (2000).
2. Q. Pan and R. D. James, J. Appl. Phys. **87**, 4702 (2000).
3. G. H. Wu, C. H. Yu, L. Q. Meng, *et al.*, Appl. Phys. Lett. **75**, 2990 (1999).
4. R. Tickle and R. D. James, J. Magn. Magn. Mater. **195**, 627 (1999).
5. S. J. Murray, M. Marioni, S. M. Allen, *et al.*, Appl. Phys. Lett. **77**, 886 (2000).
6. R. C. O’Handley, J. Appl. Phys. **83**, 3263 (1998).
7. V. D. Buchelnikov, V. S. Romanov, and A. T. Zayak, J. Magn. Magn. Mater. **191**, 203 (1999).
8. A. L. Roĭtburd, Usp. Fiz. Nauk **113**, 69 (1974) [Sov. Phys. Usp. **17**, 326 (1974)].

Translated by N. Wadwha

Journal of Experimental and Theoretical Physics, Vol. 93, No. 6, 2001, pp. 1307–1313.
Translated from Zhurnal Éksperimental’noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 6, 2001, pp. 1509–1516.
Original Russian Text Copyright © 2001 by Ovchinnikov, Vereshchagin.

NONLINEAR PHYSICS

The Properties of Weakly Collapsing Solutions to the Nonlinear Schrödinger Equation at Large Values of the Free Parameters

Yu. N. Ovchinnikov* and V. L. Vereshchagin**
** Max Planck Institute for Physics of Complex Systems, Dresden, D-01187 Germany*
** Landau Institute for Theoretical Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia*
*** Institute of Mathematics and Computing Center, Ural Division Russian Academy of Sciences, Yekaterinburg, 620219 Russia*

Received July 17, 2001

Abstract—It is shown that there exists an infinite set of weakly collapsing solutions with zero energy. Zero energy solutions are distributed along two lines in the space of parameters (A, C_1). At large values of C_1 ($C_1 \longrightarrow \infty$), the distance between the nearest points on every line tends to a finite limit. Along each of the lines, the amplitude of the oscillating terms is exponentially small with respect to the parameter C_1. © *2001 MAIK “Nauka/Interperiodica”.*

The nonlinear Schrödinger equation in which the nonlinearity corresponds to effective attraction arises in some physical problems [1, 2]. In the d-dimensional space, this equation can be written in the form

$$i\frac{\partial\psi}{\partial t} + \Delta\psi + |\psi|^{2\sigma}\psi = 0, \tag{1}$$

where ψ is a scalar function and Δ is the Laplace operator. Equation (1) has weakly collapsing solutions of the form

$$\psi(r, t) = \lambda^{\nu}\varphi(\rho\lambda)\exp(i\chi(\rho, t)), \tag{2}$$

where $\lambda = l(t)$ is a parameter, $\rho = |r|$, and φ is a real function. Weakly collapsing solutions were studied in [1, 3, 4].

For every solution to Eq. (1), the total number of particles and the total energy of the system are conserved. These two conservation laws lead to the following equations for the parameter ν and the functions λ and χ [5, 6]:

$$\begin{gathered}\nu\sigma = 1, \quad \chi(\rho, t) = \chi_0(t) + \tilde{\chi}(\lambda\rho),\\ \lambda = \frac{C}{\sqrt{t_0 - t}}, \quad \chi_0(t) = \frac{C_1}{2}\ln(t_0 - t),\end{gathered} \tag{3}$$

where C and C_1 are constants and t_0 is the collapse moment. It was shown in [5, 6] that the system of equations for φ and $\tilde{\chi}$ is reduced to the single third-order differential equation:

$$\begin{gathered}Z''' - \frac{(Z'')^2}{2Z'} - \frac{(d-1)(d-3)}{2y^2}Z'\\ - \frac{1}{C^2}\left[C_1 Z' - \frac{y(yZ' + (2/\sigma - d)Z)}{4C^2}\right]\\ - \frac{1}{8C^4 Z'}(yZ' + (2/\sigma - d)Z)^2 + \frac{2(Z')^{\sigma+1}}{y^{(d-1)\sigma}} = 0.\end{gathered} \tag{4}$$

The functions φ and $\tilde{\chi}$ are related to the function Z by the simple equations

$$\varphi = \frac{\sqrt{Z'}}{y^{(d-1)/2}}, \quad \tilde{\chi}' = -\frac{yZ' + (2/\sigma - d)Z}{4C^4 Z'}. \tag{5}$$

Our purpose is to investigate possible types of the asymptotic behavior of solutions to Eq. (4). In this paper, we restrict ourselves to the case $d = 3$, $\sigma = 1$, which is the most interesting from the physical point of view. Then, Eq. (4) is reduced to the form

$$\begin{gathered}Z''' - \frac{(Z'')^2}{2Z'} - \frac{C_1}{C^2}Z' + \frac{2(Z')^2}{y^2}\\ - \frac{(yZ' - Z)^2}{8C^4 Z'} + \frac{y(yZ' - Z)^2}{4C^2} = 0.\end{gathered} \tag{6}$$

The parameter C corresponds to the scaling transformation of Eq. (6):

$$y = C\tilde{y}, \quad Z = C\tilde{Z}(\tilde{y}). \tag{7}$$

Hence, we set $C = 1$ below.

There exists a single one-parameter family of solutions to Eq. (6) satisfying the physical boundary conditions at zero [5, 6]:

$$Z(y) = Ay^3 + Ay^5\left(\frac{C_1}{10} - \frac{3}{5}A\right) + \dots. \tag{8}$$

In order to analyze the asymptotic behavior of the solution as $y \longrightarrow \infty$, we multiply both parts of Eq. (6) by Z'

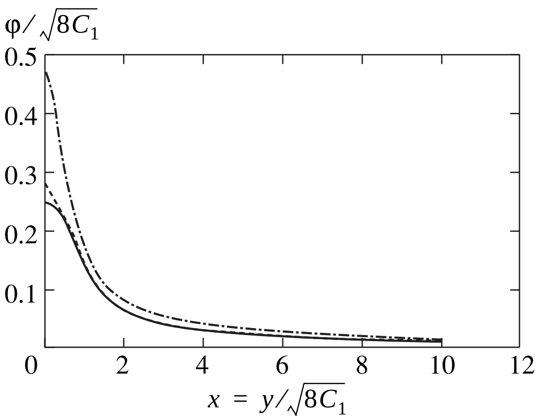


Fig. 1. Functions $\varphi/\sqrt{8C_1}$ at the points with a zero energy: the dot-and-dash curve corresponds to $(A, C_1) = (0.644, 1.09)$ and the dashed curve corresponds to $(A, C_1) = (0.913, 4.41)$. The solid curve shows the limiting function $\varphi_0 = \sqrt{Z_0(x)}/x$ (Eqs. (34), 35), $x = y/\sqrt{8C_1}$

and differentiate it with respect to y. As a result, we obtain the equation

$$4Z'''' + y(Z' + yZ'') - 8C_1Z'' - Z = \frac{16(Z')^2}{y^3} - \frac{24Z'Z''}{y^2}. \tag{9}$$

When $y \gg \max(1, \sqrt{C_1})$, the right-hand side of Eq. (9) can be neglected in the principal approximation. Then, Eq. (9) becomes linear. In the asymptotic domain $y \longrightarrow \infty$, the general solution to this equation is

$$Z = By + \frac{B_1}{y}\left(1 + \frac{16C_1}{9y^2} + \dots\right) - \frac{4}{y^2}(1 + \dots) \times\left[d_1\cos\left(\frac{y^2}{4} - 2C_1\ln y\right) + d_2\sin\left(\frac{y^2}{4} - 2C_1\ln y\right)\right]. \tag{10}$$

Formula (10) provides an expansion in powers of $1/y^2$. The four coefficients $\{B, B_1, d_1, d_2\}$ are related by an equation that can be easily found by substituting (10) into Eq. (6) (see also [7]):

$$B_1 = -2BC_1 - \frac{d_1^2 + d_2^2}{B}. \tag{11}$$

In order for the solution (8), (10), (11) to have zero energy, it is necessary and sufficient that the condition

$$d_1 = d_2 = 0 \tag{12}$$

be satisfied. The coefficients d_1 and d_2 are analytic functions of the parameters $\{A, C_1\}$. Thus, only a discrete set of the points $\{A, C_1\}$ with zero energy can exist. One such point was numerically found in [3]:

$$\{A, C_1\} = \{0.644; 1.09\}.$$

In addition to this point, there exist others corresponding to zero energy. Here are some of them:

$$\{A, C_1\} = \{0.365; 3.113\},$$
$$\{A, C_1\} = \{0.913; 4.41\}, \quad \{A, C_1\} = \{0.924; 5.89\},$$
$$\{A, C_1\} = \{1.2619; 7.26\}, \quad \{A, C_1\} = \{1.435; 8.66\},$$
$$\{A, C_1\} = \{1.689; 10.04\}, \quad \{A, C_1\} = \{1.907; 11.42\},$$
$$\{A, C_1\} = \{2.137; 12.79\}, \quad \{A, C_1\} = \{2.362; 14.16\}.$$

Below we show that the set of points $\{A, C_1\}$ with zero energy is infinite and the points are located almost equidistantly on two curves (see Fig. 3). We also obtain an analytical equation of these curves in the space $\{A, C_1\}$.

First, we investigate the domain $C_1 \gg 1$. For this purpose, consider the linear equation

$$Z'''' - 2C_1Z'' + \frac{1}{4}y(Z' + yZ'') - \frac{1}{4}Z = 0. \tag{13}$$

With the help of the substitution

$$Z = y\phi, \quad \phi' = \mathcal{F}, \tag{14}$$

Eq. (13) is reduced to the third-order equation

$$y\mathcal{F}''' + 4\mathcal{F}'' - 2C_2[2\mathcal{F} + y\mathcal{F}'] + \frac{3}{4}y^2\mathcal{F} + \frac{1}{4}y^3\mathcal{F}' = 0. \tag{15}$$

Equation (15) has the turning point

$$y = 2\sqrt{2C_1}. \tag{16}$$

In the neighborhood of this point, we set

$$y = 2\sqrt{2C_1} + t. \tag{17}$$

We consider four linearly independent solutions to Eq. (13):

$Z_1 = y$ (an exact solution to Eq. (13));

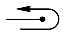
$$Z_2 = \begin{cases} \dfrac{\sqrt{8C_1 - y^2}}{\sqrt{8C_1}} & \text{before the turning point,} \\ -\dfrac{1}{2\sqrt{\pi}(2C_1)^{1/4}}\theta_3(t) & \text{in the neighborhood of the turning point,} \end{cases} \tag{18}$$

$$Z_3 = \begin{cases} \dfrac{1}{(8C_1 - y^2)}\exp\left(-\dfrac{1}{2}\displaystyle\int_0^y dy_1\sqrt{8C_1 - y^2}\right) & \text{before the turning point,} \\ -\dfrac{\exp(-\pi C_1)}{4\sqrt{\pi}(2C_1)^{1/4}}\theta_1(t) & \text{in the neighborhood of the turning point,} \end{cases}$$

$$Z_4 = \begin{cases} \dfrac{1}{(8C_1 - y^2)}\exp\left(\dfrac{1}{2}\displaystyle\int_0^y dy_1\sqrt{8C_1 - y_1^2}\right) & \text{before the turning point,} \\ -\dfrac{\exp(\pi C_1)}{4\sqrt{\pi}(2C_1)^{1/4}}\theta_2(t) & \text{in the neighborhood of the turning point.} \end{cases}$$

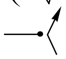
In (18), the functions $\theta_{1,2,3}$ are defined as follows:

$$\theta_1(t) = -\frac{i}{2}\int\frac{dt_1}{t_1^{3/2}}e^{tt_1}\exp\left(\frac{t_1^3}{3\sqrt{2C_1}}\right)$$



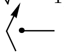
$$= \begin{cases} 2\sqrt{\pi t}, & t \gg (2C_1)^{1/6}, \\ -\dfrac{\sqrt{\pi}\exp\left(\dfrac{2}{3}|t|^{3/2}(2C_1)^{1/4}\right)}{|t|(2C_1)^{1/4}}, & t \ll -(2C_1)^{1/6}, \end{cases}$$

$$\theta_2(t) = i\int\frac{dt_1}{t_1^{3/2}}e^{tt_1}\exp\left(\frac{t_1^3}{3\sqrt{2C_1}}\right)$$



$$= \begin{cases} \dfrac{2\sqrt{\pi}}{t(2C_1)^{1/4}}\cos\left(\dfrac{2}{3}t^{3/2}(2C_1)^{1/4}\right), & t \gg (2C_1)^{1/6}, \\ -\dfrac{\pi}{|t|(2C_1)^{1/4}}\exp\left(-\dfrac{2}{3}|t|^{3/2}(2C_1)^{1/4}\right), & t \ll -(2C_1)^{1/6}, \end{cases} \quad (19)$$

$$\theta_3(t) = \frac{1}{2}\int\frac{dt_1}{t_1^{3/2}}e^{tt_1}\exp\left(\frac{t_1^3}{3\sqrt{2C_1}}\right)$$



$$= \begin{cases} -2\sqrt{\pi}|t|^{1/2}, & t \ll -(2C_1)^{1/6}, \\ \dfrac{\sqrt{\pi}}{t(2C_1)^{1/4}}\sin\left(\dfrac{2}{3}t^{3/2}(2C_1)^{1/4}\right), & t \gg (2C_1)^{1/6}. \end{cases}$$

Solutions (18), (19) make it possible to pass turning point (17) and obtain a solution to Eq. (13) in the entire domain $y \geq 0$.

The general solution to Eq. (19) satisfying the boundary condition at zero can be written in the form

$$Z = A_1\left(Z_4 - Z_3 - \frac{y}{2\sqrt{2C_1}}\right) + D_1 y + D_2 Z_2 + D_3 Z_3 + D_4 Z_4, \quad (20)$$

where $Z_{1,4}$ are solutions (18), (19) to the linear equation (13) and the functions D_{1-4} are expressed in terms of integrals of the derivatives of Z:

$$D_k = -\frac{8}{\sqrt{2C_1}}\int_0^y dy_1 P_{k4}\left(\frac{4(Z')^2}{y_1^3} - \frac{6Z'Z''}{y_1^2}\right). \quad (21)$$

Here,

$$P_{14} = -\det\begin{pmatrix} Z_2 & Z_3 & Z_4 \\ Z_2' & Z_3' & Z_4' \\ Z_2'' & Z_3'' & Z_4'' \end{pmatrix},$$

$$P_{24} = \det\begin{pmatrix} Z_1 & Z_3 & Z_4 \\ Z_1' & Z_3' & Z_4' \\ Z_1'' & Z_3'' & Z_4'' \end{pmatrix},$$

$$P_{34} = -\det\begin{pmatrix} Z_1 & Z_2 & Z_4 \\ Z_1' & Z_2' & Z_4' \\ Z_1'' & Z_2'' & Z_4'' \end{pmatrix}, \quad (22)$$

$$P_{44} = \det\begin{pmatrix} Z_1 & Z_2 & Z_3 \\ Z_1' & Z_2' & Z_3' \\ Z_1'' & Z_2'' & Z_3'' \end{pmatrix}.$$

In the domain $y \ll 1$, we find

$$P_{34} + P_{44} = \frac{\sqrt{2C_1}}{2}y, \quad D_3 + D_4 = 96A^2y^3,$$
$$D_1 = \frac{18A^2y^2}{C_1}, \quad D_2 = -\frac{12A^2y^3}{C_1}. \quad (23)$$

With the help of formulas (8), (20), (23), we find the coefficient A_1:

$$A_1 = \frac{12A}{\sqrt{2C_1}}. \tag{24}$$

Equations that determine the points $\{A, C_1\}$ with a zero energy follow from formulas (20) and (23):

$$\frac{12A}{\sqrt{2C_1}} + D_4(\infty) = 0, \quad D_2(\infty) = 0. \tag{25}$$

In the principle approximation with respect to the parameter C_1^{-2}, we find $D_4(\infty)$ from Eqs. (21), (22):

$$D_4(\infty) = \frac{2}{\sqrt{2C_1}}\int_0^\infty dy \exp(-\sqrt{2C_1}y) \times\left[\frac{4(Z')^2}{y^3} - \frac{6Z'Z''}{y^2}\right]. \tag{25a}$$

In the same approximation, the first equation in (25) can be simplified and reduced to the form [8]

$$6A = \int_0^\infty dy \exp(-\sqrt{2C_1}y)\frac{(Z')^2}{y^2}\left(3\sqrt{2C_1} + \frac{2}{y}\right). \tag{26}$$

Assume that the inequality

$$\left|A - \frac{C_1}{6}\right| \ll C_1 \tag{27}$$

holds. Then the function

$$Z' = \frac{C_1}{2}y^2 + \frac{3y}{\sqrt{C_1}}\left(A - \frac{C_1}{6}\right)\sin(\sqrt{C_1}y) \tag{28}$$

is a solution to Eq. (9) in the domain $y \ll \sqrt{C_1}$. It satisfies the boundary condition at zero for the function Z. Assume that there exists a solution to Eq. (26) satisfying condition (27). Then, formulas (26) and (28) imply

$$\left(A - \frac{C_1}{6}\right)\left\{1 - \frac{1}{2\sqrt{2}}\int_0^\infty dt e^{-t}(3t+2)\sin\left(\frac{t}{\sqrt{2}}\right) - o\left(\frac{A}{C_1} - \frac{1}{6}\right)\right\} = 0. \tag{29}$$

Since

$$\frac{1}{2\sqrt{2}}\int_0^\infty dt e^{-t}(3t+2)\sin\left(\frac{t}{\sqrt{2}}\right) = 1, \tag{30}$$

Eq. (26) has two close solutions,

$$6A = C_1, \tag{31}$$

which split only when the exponentially small terms (with respect to C_1) are taken into account. This fact is proved below. Simultaneously, we find the next terms (with respect to $1/C_1^2$) in Eq. (31).

Let us show that the amplitudes of the oscillating terms are proportional to the amplitudes of the oscillating terms in the domain $1 \ll y \ll \sqrt{C_1}$ as $y \longrightarrow \infty$ with an exponential accuracy with respect to C_1. For this purpose, we set

$$y = 2\sqrt{2C_1}x, \quad Z = (2\sqrt{2C_1})^5\tilde{Z}(x). \tag{32}$$

Using (32), Eq. (9) is reduced to the form

$$\frac{x}{4}(\tilde{Z}' + x\tilde{Z}'') - \frac{1}{4}\tilde{Z}'' - \frac{1}{4}\tilde{Z} + \frac{1}{64C_1^2}\tilde{Z}'''' = \frac{4}{x^3}(\tilde{Z}')^2 - \frac{6}{x^2}\tilde{Z}'\tilde{Z}''. \tag{33}$$

The equation

$$\frac{x}{4}(\tilde{Z}' + x\tilde{Z}'') - \frac{1}{4}\tilde{Z}'' - \frac{1}{4}\tilde{Z} = \frac{4}{x^3}(\tilde{Z}')^2 - \frac{6}{x^2}\tilde{Z}'\tilde{Z}'' \tag{34}$$

has a solution $\tilde{Z}_0$, which is regular at zero:

$$\tilde{Z}_0 = \frac{x^3}{48} - \frac{x^5}{90} + \frac{2x^7}{2835} + \dots. \tag{35}$$

This corresponds to the point $C_1 = 6A$. In the domain $x \gg 1$, we find that

$$\tilde{Z}_0 = Bx + \frac{B_1}{x} \tag{36}$$

is the general solution to Eq. (34). Numerical solution of Eq. (34) yields values for the coefficients B and B_1; as a result, we have

$$\tilde{Z}_0 = 0.0156x - \frac{0.0039}{x}. \tag{37}$$

We seek a solution to Eq. (33) in the form

$$Z = \tilde{Z}_0 + \tilde{Z}_1, \quad |\tilde{Z}_1| \ll |\tilde{Z}_0|. \tag{38}$$

The function $\tilde{Z}_1$ satisfies the equation

$$\tilde{Z}_1''\left(\frac{x^2}{4} - \frac{1}{4} + \frac{6}{x^2}\tilde{Z}_0'\right) + \tilde{Z}_1'\left(\frac{x}{4} - \frac{8}{x^3}\tilde{Z}_0' + \frac{6}{x^2}\tilde{Z}_0''\right) - \frac{1}{4}\tilde{Z}_1 + \frac{1}{64C_1^2}\tilde{Z}_1'''' = -\frac{1}{64C_1^2}\tilde{Z}_0''''. \tag{39}$$

The homogeneous equation (39) has a single smooth solution

$$\tilde{Z}_1^{(1)} = 1 + \frac{x^2}{3} + \dots. \tag{40}$$

To obtain three other solutions in the domain $x \ll 1$, we make the change of variables

$$\tilde{Z}_1 = \tilde{Z}_1^{(1)} f. \tag{41}$$

The function f satisfies the equation

$$\frac{x}{16C^2} f'''' + \frac{x}{2} f'' + f' = 0. \tag{42}$$

Solving Eq. (42), we find

$$\begin{gathered} f'_1 = x\sin(2\sqrt{2C_1}x), \quad f'_2 = \cos(2\sqrt{2C_1}x), \\ f'_3 = 16C_1^2 \int_0^\infty \frac{dx_1 x_1 \exp(-xx_1)}{(x_1^2 + 8C_1^2)^2}. \end{gathered} \tag{43}$$

Thus, the homogeneous equation (39) has two smooth solutions and two rapidly oscillating ones. For the rapidly oscillating solutions, the turning point lies in the complex plane x at a finite distance from the real axis. Therefore, the probability of the transition between the two branches (rapid and slow) is exponentially small. Thus, the problem is reduced to the construction of a solution to Eq. (6) in the domain $y \ll \sqrt{C_1}$ that has no rapidly oscillating terms. We present a procedure that yields such a solution accurate to an arbitrary order with respect to $1/C_1^2$.

The first approximation is chosen as

$$Z = \left(\frac{C_1}{6} + A_1\right) y^3 + Z_1, \tag{44}$$

where $A_1 = o(1/C_1)$. Substituting (44) into (6), we obtain the following equation for Z_1:

$$Z_1''' + \frac{2}{y^2} Z_1' - \frac{2}{y} Z_1'' + C_1 Z_1' = -\frac{C_1 y^4}{18} - 3y^2 C_1 A_1. \tag{45}$$

Integration of this equation yields

$$\begin{gathered} Z_1' = -y\left\{\frac{y^3}{18} + y\left(3A_1 - \frac{1}{3C_1}\right)\right. \\ \left. + \frac{1}{3C_1^{3/2}} \sin(\sqrt{C_1}y)(1 - 9A_1C_1)\right\}. \end{gathered} \tag{46}$$

Here, if we set

$$A_1 = \frac{1}{9C_1}, \tag{47}$$

then the function Z will contain no oscillating terms; i.e.,

$$Z = \frac{y^3}{6}\left(C_1 + \frac{2}{3C_1}\right) - \frac{y^5}{90}. \tag{48}$$

To obtain the next approximation, we use representation (48) for Z as the zero approximation:

$$\begin{gathered} Z = \frac{y^3}{6}\left(C_1 + \frac{2}{3C_1} + A_1\right) - \frac{y^5}{90} + Z_1, \\ A_1 = o(1/C_1^3). \end{gathered} \tag{49}$$

Now, we obtain the following equation for Z_1 from (6) and (49):

$$\begin{gathered} Z_1''' + \frac{2}{y^2} Z_1' - \frac{2}{y} Z_1'' + C_1 Z_1' \\ = \frac{4y^4}{81C_1} + \frac{y^6}{1620} - 2y^2\left(\frac{C_1A_1}{4} + \frac{1}{9C_1^2}\right). \end{gathered} \tag{50}$$

Solving Eq. (50), we obtain the expression

$$\begin{gathered} Z_1' = \frac{y}{C_1}\left\{\frac{y^5}{1620} + \frac{y^3}{27C_1} - 2y\left(\frac{C_1A_1}{4} + \frac{2}{9C_1^2}\right)\right. \\ \left. + \frac{2}{\sqrt{C_1}}\left(\frac{A_1C_1}{4} + \frac{2}{9C_1^2}\right)\sin(\sqrt{C_1}y)\right\} \end{gathered} \tag{51}$$

for Z_1', which is similar to (46). The oscillating correction vanishes if the coefficient A_1 is set to

$$A_1 = -\frac{8}{9C_1^3}. \tag{52}$$

In this approximation, the function Z is written as

$$\begin{gathered} Z = \frac{y^3}{6}\left(C_1 + \frac{2}{3C_1} - \frac{8}{9C_1^3}\right) \\ - \frac{y^5}{90}\left(1 - \frac{2}{3C_1^2}\right) + \frac{y^7}{11340C_1}. \end{gathered} \tag{53}$$

The process described can be continued infinitely. As a result, we construct a function Z that has no rapidly oscillating terms and is written in the form

$$\begin{gathered} Z(y) = (2\sqrt{2C_1})^5 \\ \times\left\{\tilde{Z}_0(x) + \frac{1}{C_1^2}\tilde{Z}_1(x) + \frac{1}{C_1^4}\tilde{Z}_2(x) + \ldots\right\}, \end{gathered} \tag{54}$$

where $y = 2\sqrt{2C_1}x$, and $\tilde{Z}_0$ is determined by formulas (33) and (35). The first terms of the expansions of $\tilde{Z}_1(x)$ and $\tilde{Z}_2(x)$ in powers of x are determined from Eq. (53):

$$\begin{gathered} \tilde{Z}_1(x) = \frac{x^3}{72} + \frac{x^5}{135} + \ldots, \\ \tilde{Z}_2(x) = -\frac{x^3}{54} + \ldots. \end{gathered} \tag{55}$$

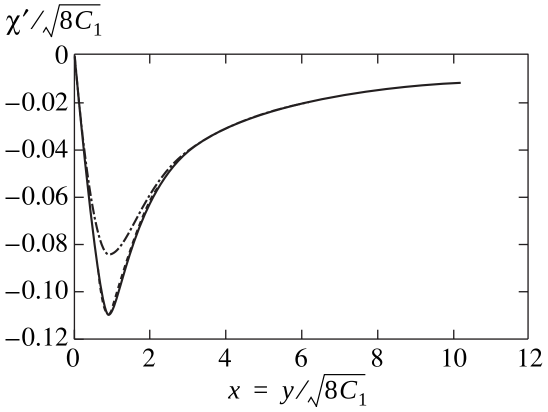


Fig. 2. Functions $\chi'/\sqrt{8C_1}$ at the points with a zero energy: the dot-and-dash curve corresponds to $(A, C_1) = (0.644, 1.09)$ and the dashed curve corresponds to $(A, C_1) = (0.913, 4.41)$. The solid curve shows the limiting function x_0' constructed on the basis of $\tilde{Z}_0(x)$; $x = y/\sqrt{8C_1}$.

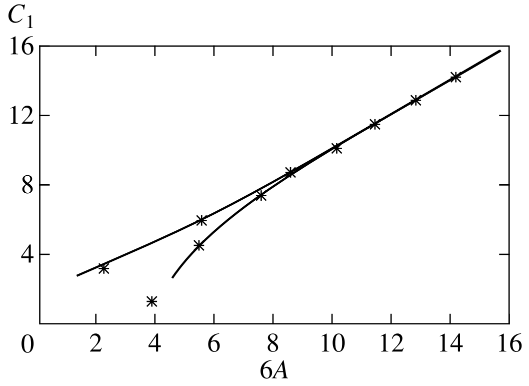


Fig. 3. Asterisks denote zero-energy points in the plane $(6A, C_1)$; the solid curves correspond to Eqs. (57) and (63).

Simultaneously, we obtain an expression for the coefficient A, which improves formula (31):

$$6A = C_1 + \frac{2}{3C_1} - \frac{8}{9C_1^3} + \dots. \tag{56}$$

Expansion (54) is an asymptotic series in C_1. The right-hand sides of (54), (56) include exponentially small terms with respect to C_1; they are associated with tunneling between branches in the linearized Eq. (9). By virtue of Eq. (29), Eq. (56) must split into two equations. It follows from the above reasoning that the splitting magnitude is exponentially small with respect to the parameter C_1:

$$6A - \left(C_1 + \frac{2}{3C_1} - \frac{8}{9C_1^3} + \dots\right) = \mp D C_1^{1.25}\exp(-\beta C_1). \tag{57}$$

The magnitude β is determined by the location of the turning point and will be determined below. On the curves determined by Eq. (57), the amplitudes $d_{1,2}$ are exponentially small with respect to the parameter C_1. The oscillation correction to (54) in the domain $y \gg \sqrt{C_1}$ consists of two terms (see (10)). The linear combination of these terms, which goes to the function $\sin(\sqrt{C_1})$ in the domain $1 \ll y \ll \sqrt{C_1}$, can be eliminated by a small shift (of order $C_1^{-1/2}\exp(-2\beta C_1)$ in Eq. (57). The second linear combination shifted by the phase $\pi/2$ relative to the first one, which goes to $\cos(\sqrt{C_1}y))$ in the domain $1 \ll y \ll \sqrt{C_1}$, appears only due to the tunneling between the rapid and slow branches in the linearized Eq. (9). With this combination, the amplitude is an oscillating function of the parameter C_1. Its zeros yield the states with zero energy. Now, we find the period of this function and β.

The linearized Eq. (9) yields the "momentum" $P(x)$

$$P^2(x) = 2C_1\left(-1 + x^2 + \frac{24}{x^2}\frac{\partial\tilde{Z}_0}{\partial x}\right), \tag{58}$$

where the function $\tilde{Z}_0$ is determined by formulas (33) and (35). Let x^* be the turning point at which the "momentum" $P(x)$ vanishes, i.e.,

$$P(x^*) = 0. \tag{59}$$

Then, the action on the trajectory,

$$\mathcal{P} = 4C_1\int_0^{x^*} dx\sqrt{x^2 + \frac{24}{x^2}\frac{\partial\tilde{Z}_0}{\partial x} - 1} = C_1(\beta_0 + i\beta), \tag{60}$$

determines the amplitude of oscillating terms in formula (10) on the curves (57):

$$d_{1,2} \sim C_1^{5/2}\exp(-2\beta C_1). \tag{61}$$

Each of the curves contains points $\{A, C_1\}$ corresponding to zero energy. These points are located almost at the same distance from each other with respect to C_1 with the period

$$\delta C_1 \approx \frac{2\pi}{\beta_0} \approx 2.6. \tag{62}$$

Numerical integration of Eqs. (34), (60) yields the following values for the constants β, β_0, and x^*:

$$\begin{gathered}\beta = 0.712, \quad \beta_0 = 2.416,\\ x^* = 0.973 + i \times 0.393.\end{gathered} \tag{63}$$

Formula (54) entails an important inequality, which holds along curves (57):

$$\frac{1}{\tilde{Z}_0(x)}\left|\frac{Z(y)}{(2\sqrt{2C_1})^5} - \tilde{Z}_0(x)\right| < \frac{\text{const}}{C_1^2}. \tag{64}$$

Estimate (64) is uniform with respect to the coordinate y. Hence, all zero-energy solutions are almost similar.

Figures 1 and 2 show the dependence of φ and χ' at zero-energy points $\{A, C_1\}$ (0.644, 1.09) and (0.913, 4.41). Solid curves in Figs. 1 and 2 show the functions φ_0 and χ'_0 constructed on the basis of $\tilde{Z}_0$.

Figure 3 shows zero-energy points in the plane (6A, C_1); solid curves correspond to Eqs. (57) and (63).

As the parameter C_1 increases, the accuracy of determining zero-energy points rapidly decreases.

Thus, there exist two curves in the plane $\{A, C_1\}$ on which the amplitude of the oscillating term is exponentially small. On these curves (see Eq. (57)), points corresponding to zero-energy solutions are almost equidistant with respect to C_1. All zero-energy solutions are smooth and almost similar. The function φ corresponding to these solutions slowly decreases with increasing y in the domain $y \gg \sqrt{C_1}$:

$$\varphi(y) \sim \frac{8C_1\sqrt{B}}{y},$$

where B is determined by (36) and (37). It is unjustified to assume that the solution corresponding to the point $\{A, C_1\} = (0.644, 1.09)$ is distinguished among others, and the instability development results in the formation of this very solution. At the present time, the issue of the probability of forming weakly collapsing solutions remains unstudied. It is possible that the existence of specific curves in the plane $\{A, C_1\}$ can be used both for solving this problem and for stability analysis of weakly collapsing solutions.

The integral of the total number of particles is divergent for any weakly collapsing solution. A weak decrease of the solution forces us to truncate the integral for sufficiently large y. Hence, the probability of forming a state leading to collapse is drastically decreased. For points in general position $\{A, C, C_1\}$, there exists a set of parameters such that the deep minimum of φ is at $y \sim 1$ [5, 6]. One can hope that these solutions will be realized in numerical simulation and in real physical objects.

ACKNOWLEDGMENTS

We are grateful to A.B. Shabat and S.V. Iordanskii for helpful remarks.

The work by Yu.N. Ovchinnikov was supported by the US Civilian Research & Development Foundation (project no. RP1-2251) and by the Russian Foundation for Basic Research (project no. 00-02-17729a). The work by V.L. Vereshchagin was supported by INTAS (project no. 99-01782) and by the Russian Foundation for Basic Research (project nos. 99-01-00656 and 01-01-00931).

REFERENCES

1. V. E. Zakharov, Zh. Éksp. Teor. Fiz. **62**, 1746 (1972) [Sov. Phys. JETP **35**, 908 (1972)].
2. V. E. Zakharov and V. S. Sinakh, Zh. Éksp. Teor. Fiz. **68**, 940 (1975) [Sov. Phys. JETP **41**, 465 (1975)].
3. V. E. Zakharov and L. N. Shchur, Zh. Éksp. Teor. Fiz. **81**, 2019 (1981) [Sov. Phys. JETP **54**, 1064 (1981)].
4. V. E. Zakharov and E. A. Kuznetsov, Zh. Éksp. Teor. Fiz. **91**, 1310 (1986) [Sov. Phys. JETP **64**, 773 (1986)].
5. Yu. N. Ovchinnikov, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 387 (1999) [JETP Lett. **69**, 418 (1999)].
6. Yu. N. Ovchinnikov and I. M. Sigal, Zh. Éksp. Teor. Fiz. **116**, 67 (1999) [JETP **89**, 35 (1999)].
7. C. Sulem and P.-L. Sulem, *The Nonlinear Schrödinger Equation: Self-Focusing and Wave Collapse* (Springer-Verlag, New York, 1999).
8. Yu. N. Ovchinnikov and V. L. Vereshchagin, Pis'ma Zh. Éksp. Teor. Fiz. **74**, 76 (2001) [JETP Lett. **74**, 72 (2001)].

Translated by A. Klimontovich

Journal of Experimental and Theoretical Physics, Vol. 93, No. 6, 2001, pp. 1314–1322.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 6, 2001, pp. 1517–1526.
Original Russian Text Copyright © 2001 by Dmitrieva, Ryskin.

MISCELLANEOUS

Complex Dynamics of a Distributed Parametric Oscillator

T. V. Dmitrieva and N. M. Ryskin*
Saratov State University, Saratov, 410026 Russia
**e-mail: ryskinnm@info.sgu.ru*

Received April 12, 2001

Abstract—The dynamics of a system with three parametrically coupled waves with delayed feedback is considered. Results of the detailed numerical simulation of the onset of self-modulation, as well as complex dynamic and chaotic regimes, are presented. The relation of self-modulation regimes with the formation and propagation of solitons is investigated. It is discovered that as the pump parameter increases, the synchronization of phases of the interacting waves, which is characteristic of stationary generation and periodic self-modulation regimes, is violated, and the system goes to a chaotic regime via intermittency. *© 2001 MAIK “Nauka/Interperiodica”.*

1. INTRODUCTION

The study of spatio-temporal chaos in distributed self-oscillatory systems is one of the most important problems in modern nonlinear dynamics. In this paper we consider the chaotic dynamics of a system consisting of three interacting waves that propagate in a quadratically nonlinear medium. Parametric instability plays an important role in nonlinear optics, fluid dynamics, plasma physics, and so on [1–3]. Amplifiers and oscillators based on this phenomenon are extensively used in radio, microwave, and optical bands [4–6]. From the point of view of nonlinear dynamics, distributed parametric oscillators belong to the class of distributed self-oscillatory systems with delayed feedback. It is well known that such systems can demonstrate complex (including chaotic) oscillation regimes (see, e.g., [2, 7, 8]). In [9], it was shown that a system of parametrically coupled waves can demonstrate a chaotic behavior. In that study, the problem of acousto-optical interaction with regard for the reflection of the acoustic wave from the boundaries was considered. In our paper [10], the parametric interaction involving counterpropagating waves, which ensures an internal distributed feedback, was considered. It was discovered that as the pumping amplitude increases, the stationary oscillation regime becomes unstable and is replaced by chaotic oscillations; moreover, the chaotic attractor itself and the sequence of bifurcations preceding its appearance are similar to some well-known finite-dimensional models of nonlinear dynamics including the Lorenz system [1, 2, 8].

However, by and large, the complex dynamics under parametric interaction is poorly understood. For systems with concurrent waves, which are of particular practical interest, no thorough investigations of self-modulation and chaotic regimes in a wide range of control parameters have been conducted. Even the scenario of the onset of the chaotic regime remains unknown. Of particular interest is the comparison of such a system with other distributed self-oscillatory systems with well-understood dynamics; these are oscillators based on beam and plasma-beam traveling wave amplifiers with external feedback [11–13], free electron lasers [14–16], and microwave backward wave oscillators [17–19]. In recent years, such systems have been studied and have been shown to demonstrate a rich set of various dynamic regimes including the transition to chaos by all known scenarios characteristic of finite-dimensional dynamical systems [17, 18]. Thus, a thorough study of the complex dynamics of wave parametric interaction will be important for revealing general laws of chaotic behavior in distributed self-oscillatory systems. Investigations of nonstationary oscillation regimes are also of considerable importance since they make it possible to study mechanisms of various secondary instabilities in parametric oscillators. In addition, generation of chaotic signals is of independent interest in relation to their use in communication systems, information processing, etc. (see, e.g., [20–23]).

The purpose of this paper is a detailed theoretical treatment of the complex dynamics of a distributed parametric oscillator with delayed feedback. In Section 2, a model of the oscillator is discussed and basic equations of its dynamics are presented. The system under study is a resonator filled with a quadratically nonlinear substance in which three parametrically coupled waves propagate. Self-excitation conditions and stationary generation regimes on the principal mode are briefly discussed in Sections 3 and 4, respectively. In Section 5, results of numerical simulation are presented. Mechanisms of the onset of self-modulation and chaos are thoroughly investigated.

2. STATEMENT OF THE PROBLEM AND BASIC EQUATIONS

Consider a system of three parametrically coupled waves whose frequencies and wave numbers satisfy the resonance conditions

$$\omega_1 + \omega_2 = \omega_3,$$

$$k_1 + k_2 = k_3.$$

For slowly varying complex amplitudes of the waves $A_{1,2,3}$, one can write the well-known system of equations [1–3]

$$\frac{\partial A_1}{\partial t} + v_1 \frac{\partial A_1}{\partial x} = \sigma_1 A_2^* A_3, \quad (1)$$

$$\frac{\partial A_2}{\partial t} + v_2 \frac{\partial A_2}{\partial x} = \sigma_2 A_1^* A_3, \quad (2)$$

$$\frac{\partial A_3}{\partial t} + v_3 \frac{\partial A_3}{\partial x} = -\sigma_3 A_1 A_2, \quad (3)$$

where $v_{1,2,3}$ are group velocities and $\sigma_{1,2,3}$ are real nonlinear interaction coefficients. Equations (1)–(3) describe the interaction of the high-frequency pump wave (A_3) with the low-frequency signal (A_1) and idler (A_2) waves. We assume that $v_{1,2,3} > 0$; i.e., all waves propagate in the same direction, and $\sigma_{1,2,3} > 0$, which means that waves with positive energy are considered. In this case, the instability is convective (see [1–3]), and to turn the system into a self-oscillatory one, feedback is required, which can be realized by using reflections from the boundaries of the system or at the expense of an external circuit. We consider the simplest case of the signal wave feedback. This case is described with the help of the boundary condition

$$A_1(0, t) = R A_1(l, t - \Delta t), \quad (4)$$

where $R = \rho e^{i\psi}$ is the complex feedback parameter, Δt is the delay time in the feedback circuit, and l is the extent of the nonlinear medium. For example, in the presence of reflections, we have

$$\rho = |R_0 R_l|, \quad \psi = \arg(R_0 R_l) - 2k_1 l,$$

$$\Delta t = \frac{l}{v_1},$$

where $R_{0,l}$ are the complex reflection coefficients from the left and the right boundaries, respectively. We assume that the pump wave has constant amplitude on the left boundary and the idler wave is absent. Then, the boundary conditions for these waves have the form

$$A_2(0) = 0, \quad A_3(0) = A_0 = \text{const}. \quad (5)$$

Let us introduce the dimensionless coordinate and time

$$\xi = \frac{x}{l}, \quad \tau = \frac{v_1 v_2 (t - x/v_2)}{l(v_2 - v_1)},$$

and the new dependent variables

$$A_1' = \sqrt{\frac{\sigma_3 v_1}{\sigma_1 v_3}} \frac{A_1}{A_0}, \quad A_2' = \sqrt{\frac{\sigma_3 v_2}{\sigma_2 v_3}} \frac{A_2}{A_0}, \quad A_3' = \frac{A_3}{A_0}.$$

Then, the original equations (1)–(3) are reduced to the form (primes at the dimensionless variables are omitted)

$$\frac{\partial A_1}{\partial \tau} + \frac{\partial A_1}{\partial \xi} = \alpha A_2^* A_3, \quad (6)$$

$$\frac{\partial A_2}{\partial \xi} = \alpha A_1^* A_3, \quad (7)$$

$$(1 + u)\frac{\partial A_3}{\partial \tau} + \frac{\partial A_3}{\partial \xi} = -\alpha A_1 A_2, \quad (8)$$

where

$$\alpha = A_0 l \sqrt{\frac{\sigma_1 \sigma_2}{v_1 v_2}}$$

is the parameter interpreted as the normalized amplitude of the pump wave at the entry of the system and the parameter

$$u = \frac{v_2(v_1 - v_3)}{v_3(v_2 - v_1)}$$

characterizes the detuning of group velocities of the interacting waves.

In terms of the new variables, the boundary conditions (4), (5) have the form

$$A_1(0, \tau) = R A_1(1, \tau - \delta), \quad (9)$$

$$A_2(0, \tau) = 0, \quad A_3(0, \tau) = 1, \quad (10)$$

where

$$\delta = \frac{v_1 v_2 (\Delta t + l/v_2)}{l(v_2 - v_1)}$$

is the dimensionless delay time. We set the initial conditions

$$A_3(x, 0) = 1, \quad |A_{1,2}(x, 0)| \ll 1,$$

which corresponds to the oscillator excitation due to small fluctuations of the idler and signal waves.

Thus, the model of the parametric oscillator under study is a resonator (one-dimensional or annular) filled with a nonlinear substance in which parametrically interacting waves propagate. The eigenfrequencies of this resonator are easy to determine. Indeed, in the absence of pumping, Eq. (6) for the signal wave takes the form

$$\frac{\partial A_1}{\partial \tau} + \frac{\partial A_1}{\partial \xi} = 0.$$

Consider the excitation of the resonator by the harmonic input signal $A_0 \exp(i\Omega\tau)$. Here Ω is the detuning of frequency from the exact resonance ω_1. Then, we have

$$A_1(0, \tau) = A_0 e^{i\Omega\tau} + RA_1(1, \tau - \delta)$$

instead of boundary condition (9). We seek the solution in the form

$$A_1 = \hat{A}\exp[i\Omega(\tau + \xi)].$$

As a result, we obtain

$$\hat{A} = \frac{A_0}{1 - \rho\exp(-i\theta)}$$
$$= \frac{A_0 \exp[i\arctan(\rho\sin\theta/(1 - \rho\cos\theta))]}{\sqrt{1 + \rho^2 - 2\rho\cos\theta}},$$

where $\theta = \Omega(\delta + 1) - \psi$. The frequencies that maximize the amplitude of the forced oscillations can be interpreted as the frequencies of the resonator's eigenmodes. They can be found from the condition $\cos\theta = 1$ and have the form

$$\Omega_n = \frac{2\pi n + \psi}{1 + \delta}. \qquad (11)$$

Thus, in our approximation, the spectrum of eigenfrequencies is equidistant and the distance between modes is $2\pi/(1 + \delta)$. It is clear that the resonance properties of the system are the more pronounced, the larger ρ.

3. SELF-EXCITATION CONDITIONS

The linear theory of parametric oscillator has been thoroughly studied (see, e.g., [5, 6, 24, 25]). Here we present only the basic concepts that are necessary for further presentation. First of all, we must determine the gain $G(\Omega)$ of the parametric amplifier without feedback. Then, the condition of the oscillator self-excitation is written in the form

$$G(\Omega)R = 1.$$

At the first stage, when depletion of the pump has no effect, we can set $A_3 = 1$ and obtain the following linearized system of equations from Eqs. (6), (7):

$$\frac{\partial A_1}{\partial \tau} + \frac{\partial A_1}{\partial \xi} = \alpha A_2^*, \qquad (12)$$

$$\frac{\partial A_2}{\partial \xi} = \alpha A_1^*. \qquad (13)$$

Setting $A_{1,2} \propto \exp[i(\Omega\tau - \kappa\xi)]$, we find the dispersion equation

$$(\Omega - \kappa)\kappa = \alpha^2,$$

with the roots

$$\kappa_\pm = \frac{\Omega}{2} \pm \sqrt{\frac{\Omega^2}{4} - \alpha^2}.$$

Hence, it is seen that the amplification band of the parametric amplifier is determined by the condition $\Omega^2 < 4\alpha^2$. In this frequency band, we have the pair of conjugate wave numbers

$$\kappa_\pm = \frac{\Omega}{2} \pm i\lambda,$$

where

$$\lambda = \sqrt{\alpha^2 - \frac{\Omega^2}{4}} \qquad (14)$$

is the increment of parametric instability. Thus, solutions to system (12), (13) can be represented as a superposition of growing and decaying waves

$$A_1 = [a_+\exp(\lambda\xi) + a_-\exp(-\lambda\xi)]\exp\left[i\Omega\left(\tau - \frac{\xi}{2}\right)\right],$$

$$A_2^* = [b_+\exp(\lambda\xi) + b_-\exp(-\lambda\xi)]\exp\left[i\Omega\left(\tau - \frac{\xi}{2}\right)\right].$$

The constants $a_\pm$ and $b_\pm$ can be determined from the boundary conditions, which have the form

$$A_1(\xi = 0) = A_0 e^{i\Omega\tau}, \quad A_2^*(\xi = 0) = 0$$

for the amplifier. Here A_0 is the amplitude of the input signal. Upon simple computations, we find that

$$A_1 = A_0\left[\cosh(\lambda\xi) - \frac{i\Omega}{2\lambda}\sinh(\lambda\xi)\right]\exp\left[i\Omega\left(\tau - \frac{\xi}{2}\right)\right],$$

$$A_2^* = \frac{\alpha A_0}{\lambda}\sinh(\lambda\xi)\exp\left[i\Omega\left(\tau - \frac{\xi}{2}\right)\right].$$

Then, the gain factor is written as

$$G(\Omega) = \frac{A_1(\xi = 1)}{A_1(\xi = 0)} = \left[\cosh\lambda - \frac{i\Omega}{2\lambda}\sinh\lambda\right]\exp\left(-\frac{i\Omega}{2}\right).$$

With regard for the boundary condition for the signal wave (9), we obtain the self-excitation condition in the form

$$1 = \rho\left[\cosh\lambda - \frac{i\Omega}{2\lambda}\sinh\lambda\right]\exp\left[i\left(\psi - \delta\Omega - \frac{\Omega}{2}\right)\right]. \qquad (15)$$

This yields the value of the pump parameter $\alpha = \alpha_{st}$ and the oscillation frequency.

It is obvious that the self-excitation is facilitated when one of the eigenfrequencies of resonator (11) is close to zero; indeed, for $\Omega = 0$, the increment of the parametric instability (14) attains its maximum $\lambda = \alpha$. Thus, the values $\psi = 2\pi n$ are optimal. It is seen from Eq. (15) that in this case the self-excitation condition takes the simple form

$$\rho\cosh\alpha_{st} = 1. \qquad (16)$$

On the contrary, for $\psi \approx 2\pi n + \pi$, when the frequencies of two adjacent modes are located at an equal distance from the amplification band, self-excitation is hampered. A numerical analysis of Eq. (15) shows that the dependence of α_{st} on ψ is rather weak. The oscillation frequency is close to the frequency of the resonator's zero eigenmode:

$$\Omega \approx \frac{\psi}{1+\delta}.$$

In the region of adverse phases, self-excitation can occur at the frequencies of two adjacent modes simultaneously.

4. STATIONARY GENERATION REGIMES

Solutions that describe stationary generation regimes on the principal mode can also be analyzed analytically. For simplicity, consider the case $\psi = 0$ (i.e., the parameter R is real and positive), which is the optimal condition for self-excitation. In this case, the oscillation frequency is zero. Assuming that the amplitudes $A_{1,2,3}$ in Eqs. (6)–(8) depend only on the coordinate ξ, we obtain the system of ordinary differential equations

$$\frac{dA_1}{d\xi} = \alpha A_2^* A_3, \quad \frac{dA_2}{d\xi} = \alpha A_1^* A_3, \quad \frac{dA_3}{d\xi} = -\alpha A_1 A_2. \qquad (17)$$

Note that the parameter u does not appear in these equations; i.e., stationary solutions are independent of the detuning of group velocities (under the adopted normalization of variables). Solutions to Eqs. (17) in an infinite medium are well known [1, 3]. In our case, we must add the boundary conditions

$$A_1(0) = \rho A_1(1), \quad A_2(0) = 0, \quad A_3(0) = 1.$$

Changing to real amplitudes and phases in Eqs. (17), we have

$$A_j = a_j e^{i\varphi_j}.$$

It is easy to show (the details of the computations are omitted, since they are similar to those used in [10]) that $\sin\Phi = 0$, where $\Phi = \varphi_3 - \varphi_2 - \varphi_1$; i.e., the phases of the waves in the stationary regime are synchronized. Furthermore, using the conservation law (the Manley–Rowe relation [1–4])

$$a_2^2(\xi) + a_3^2(\xi) = a_2^2(0) + a_3^2(0) = 1$$

and changing the variables

$$a_1 = \frac{1}{2\alpha}\frac{d\theta}{d\xi}, \quad a_2 = \sin\frac{\theta}{2}, \quad a_3 = \cos\frac{\theta}{2},$$

we obtain the pendulum equation

$$\frac{d^2\theta}{dz^2} = \sin\theta, \qquad (18)$$

where $z = \alpha\xi$, with the boundary conditions (the lower subscript denotes differentiation with respect to z)

$$\theta(0) = 0, \quad \theta_z(0) = \rho\theta_z(\alpha). \qquad (19)$$

Note that Eq. (18) describes a pendulum that has an unstable equilibrium at the point $\theta = 0$; hence, the stationary solutions we are interested in correspond to rotational oscillations (idling motions).

Let us write the energy integral

$$\frac{\theta_z^2}{2} + \cos\theta = E,$$

for Eq. (18), where

$$E = 1 + \frac{\theta_z^2(0)}{2}$$

is the total energy. Then, using the boundary conditions (19), we find that

$$\frac{(1-\rho^2)\theta_z^2(\alpha)}{2} = 1 - \cos\theta(\alpha). \qquad (20)$$

Thus, on the phase plane, stationary solutions correspond to portions of the phase trajectories that begin at the vertical axis and end on the curve determined by Eq. (20). Since $a_1 = \theta_z/2$, this fact entails that the amplitude of the output signal a_1 ($z = \alpha$) attains the maximal value

$$a_{\max} = \frac{1}{\sqrt{1-\rho^2}} \qquad (21)$$

at $\theta(\alpha) = \pi$, i.e., at $z = T/4$, where T is the period of the pendulum oscillations. Using the well-known formula for the period of oscillations (see, e.g., [26]), we find the optimal value of the parameter α,

$$\alpha_{\max} = \sqrt{1-\rho^2}K(\sqrt{1-\rho^2}), \qquad (22)$$

where K is the complete elliptic integral of the first kind. Thus, as the pump parameter increases, the amplitude of the output signal also increases until α exceeds $\alpha_{\max}$. Then, the inverse process of energy transfer of the signal and idler waves to the pump wave starts, and the amplitude of the signal decreases.

One also can find the spatial distribution of wave amplitudes in the stationary regime. Writing the exact solution for the rotational oscillations [26], with regard for the fact that the pendulum has an unstable state at $\theta = 0$, we have

$$\frac{\psi_z}{2} = \frac{1}{m}\mathrm{dn}\left(\frac{z}{m} + K(m);\; m\right) = \sqrt{\frac{1-m^2}{m^2}}\frac{1}{\mathrm{dn}(z/m;\; m)},$$

where $\mathrm{dn}(z;\, m)$ is the elliptic Jacobi function and

$$m^2 = \frac{1}{1+\psi_z^2(0)/4} = \frac{1}{1+a_1^2(0)}.$$

Thus, we find the dependence of the signal wave amplitude on the coordinate

$$a_1(\xi) = \frac{a_1(0)}{\mathrm{dn}(\alpha\xi\sqrt{1+a_1^2(0)};\; (1+a_1^2(0))^{-1/2})}.$$

This relation must be considered in combination with the boundary condition $a_1(0) = \rho a_1(1)$. For this reason, we cannot find exact stationary solutions explicitly as was done for the backward-wave oscillator in [10]. However, the equations can be solved graphically. This approach is similar to the analysis based on the return map proposed in [24]. In the case under consideration, this map has the form

$$a_{n+1} = \frac{\rho a_n}{\mathrm{dn}(\alpha\sqrt{1+a_n^2};\; (1+a_n^2)^{-1/2})}, \qquad (23)$$

where $a_n = a_1$ ($\xi = 0$ and $\tau = n(1+\delta)$). The fixed points of mapping (23) correspond to the amplitudes of the stationary generation regime. This mapping is also useful for the stability analysis of stationary regimes and transient oscillation processes [24]. However, it does not allow one to determine stability conditions. For this purpose, it is appropriate to use direct numerical simulation of nonstationary equations; results of such a simulation are presented in the following section.

5. RESULTS OF NUMERICAL SIMULATION AND THEIR DISCUSSION

5.1. Onset of Self-Modulation

A numerical integration of the nonstationary equations (6)–(8) with the boundary conditions (9), (10) was performed using a finite difference scheme of second-order accuracy similar to that used in [10]. First of all, we investigated the self-excitation of oscillations. The results of computations (the threshold value of the pump parameter and the oscillation frequency) are in complete agreement with the theoretical results (see Section 3). It must be noted that in the region of adverse phases of the feedback parameter $\psi \approx \pi$, self-excitation can occur at the frequencies of two adjacent modes. The frequency of the stationary oscillation depends on the boundary conditions; more precisely, the mode with the greater amplitude “survives.” Such a picture is typical of competition of strongly coupled modes [1].

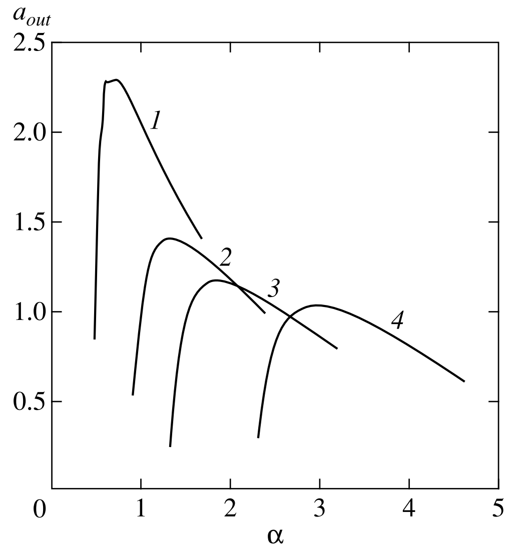


Fig. 1. Stationary generation amplitude as a function of α for various ρ: ρ = 0. 9 (*1*), 0.7 (*2*), 0.5 (*3*), and 0.2 (*4*).

As the system under study is distributed and is characterized by a large number of control parameters, we consider only the case $\psi = 0$ below. Recall that this corresponds to the situation when the signal wave carrier frequency, for which the conditions of the three-wave parametric resonance are satisfied exactly, coincides with one of the resonator’s eigenfrequencies (11). In this case, a_{st} is determined by Eq. (16) and the fundamental frequency is zero. We also assume that the group velocities of the pump and idler waves coincide; i.e., the detuning parameter of group velocities is $u = -1$. Such a choice of parameters corresponds to that used in [9].

If the excess of the generation threshold is not very large, single-frequency stationary oscillation regimes are realized. At the first stage of the transient process, the state $\sin\Phi = 0$ quickly settles; in this state, the wave phases are synchronized. Then, the oscillation amplitude stabilizes. Note that after the phases have been synchronized, the wave amplitudes can be considered real. The computed dependences of the output signal amplitudes $a_{\mathrm{out}} = a_1$ ($\xi = 1$) on the parameter α are presented in Fig. 1. They are in complete agreement with the theoretical results (see Section 4). In particular, the maximal value of the amplitude and the corresponding value of α satisfy formulas (21) and (22).

As α increases, the single-frequency regime becomes unstable and self-modulation is brought about; i.e., periodic oscillation of the output signal amplitude occurs. A pair of satellites, which are symmetric about the fundamental frequency, appear in the

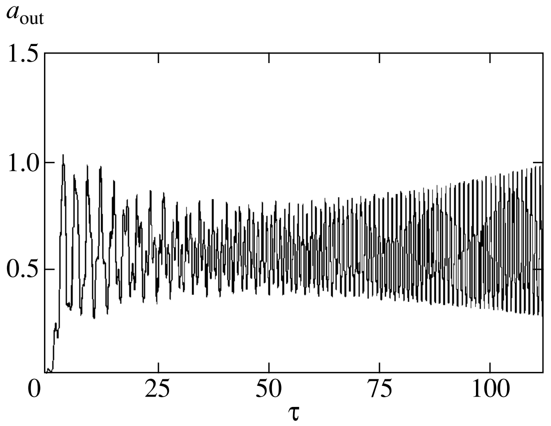


Fig. 2. Output signal a_{out} amplitude as a function of time in the course of the transient process: $\rho = 0.2$, $\alpha = 4.75$, $\delta = 1$.

spectrum. An analysis shows that self-modulation is explained by the excitation of oscillations at higher eigenmodes of the resonator. The values of self-modulation frequencies Ω_{sm} are rather close to the eigenmode frequencies (11). For small δ, the modes $n = \pm 1$ are excited; the frequencies of those modes are the closest to the frequency at which the self-modulation instability increment is maximal. Computations show that this frequency is about 2π. As δ increases, the eigenfrequencies (11) decrease. Hence, the instability increment decreases, and, as a result, the self-modulation threshold α_{sm} slightly increases. Conversely, the frequencies $\Omega_{\pm 2}$ approach the optimal value and the increment of these modes begins to increase. In the region $\delta \approx 0.8$–0.9, two pairs of secondary modes are excited, so that the self-modulation at its beginning is quasi-periodic. However, as α increases, one of the pairs of secondary modes is suppressed in the competition process and the self-modulations become periodic again.

For $\delta > 0.9$, the second mode is under more favorable conditions than the first one; hence, the second mode suppresses the first one (in this region, a_{sm} slightly decreases). Figure 2 illustrates the mode competition effect. It shows a typical picture of the transient process (for $\delta = 1$). At the first stage, the principal (zero) mode quickly increases; then, the modes with $n = \pm 1$ are excited, which results in large oscillation of the amplitude. Gradually, this oscillation is replaced by a higher frequency one, which corresponds to the excitation of the modes with $n = \pm 2$.

After the self-modulation threshold has been exceeded, the transient process illustrated in Fig. 2 is completed by forming a soliton-like pulse that periodically propagates along the system. Figure 3 illustrates the corresponding spatio-temporal dynamics. A similar behavior was described in [9]. Note that the onset of self-modulation is hard, since the deep periodic modulation regime is settled immediately. This regime is strongly nonlinear and is characterized by the appearance of a large number of harmonics with the self-modulation frequency Ω_{sm}. For the chosen values of the parameters, $\Omega_{\text{sm}} \approx 2\pi$. Since the spectrum of the resonator's eigenmodes is equidistant, the higher harmonics are close to the frequencies of the modes Ω_n ($n = 4, 6, \ldots$). Thus, the self-synchronization of modes is realized, which is also typical of other resonance oscillators [2, 15, 16]. Simultaneous excitation of a large number of resonator modes with strongly coupled frequencies

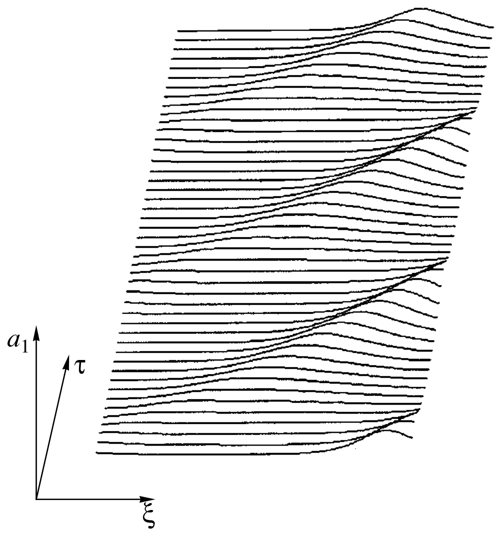


Fig. 3. The spatio-temporal dynamics of the signal wave amplitude in the periodic self-modulation regime. The values of the parameters are the same as in Fig. 2.

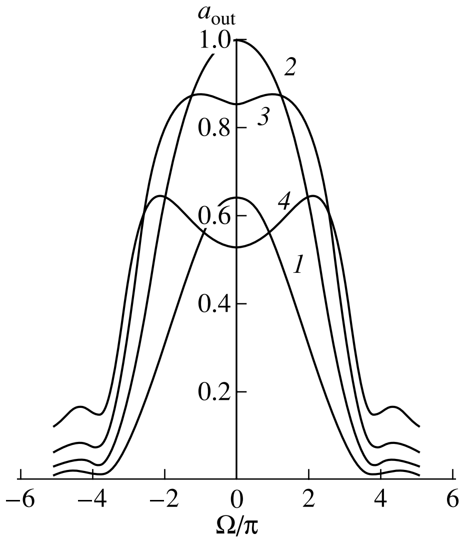


Fig. 4. The parametric amplifier's output signal amplitude as a function of frequency at $\alpha = 5.0$ and various amplitudes of the input signal: $A_0 = 0.01$ (*1*), 0.025 (*2*), 0.05 (*3*), and 0.1 (*4*).

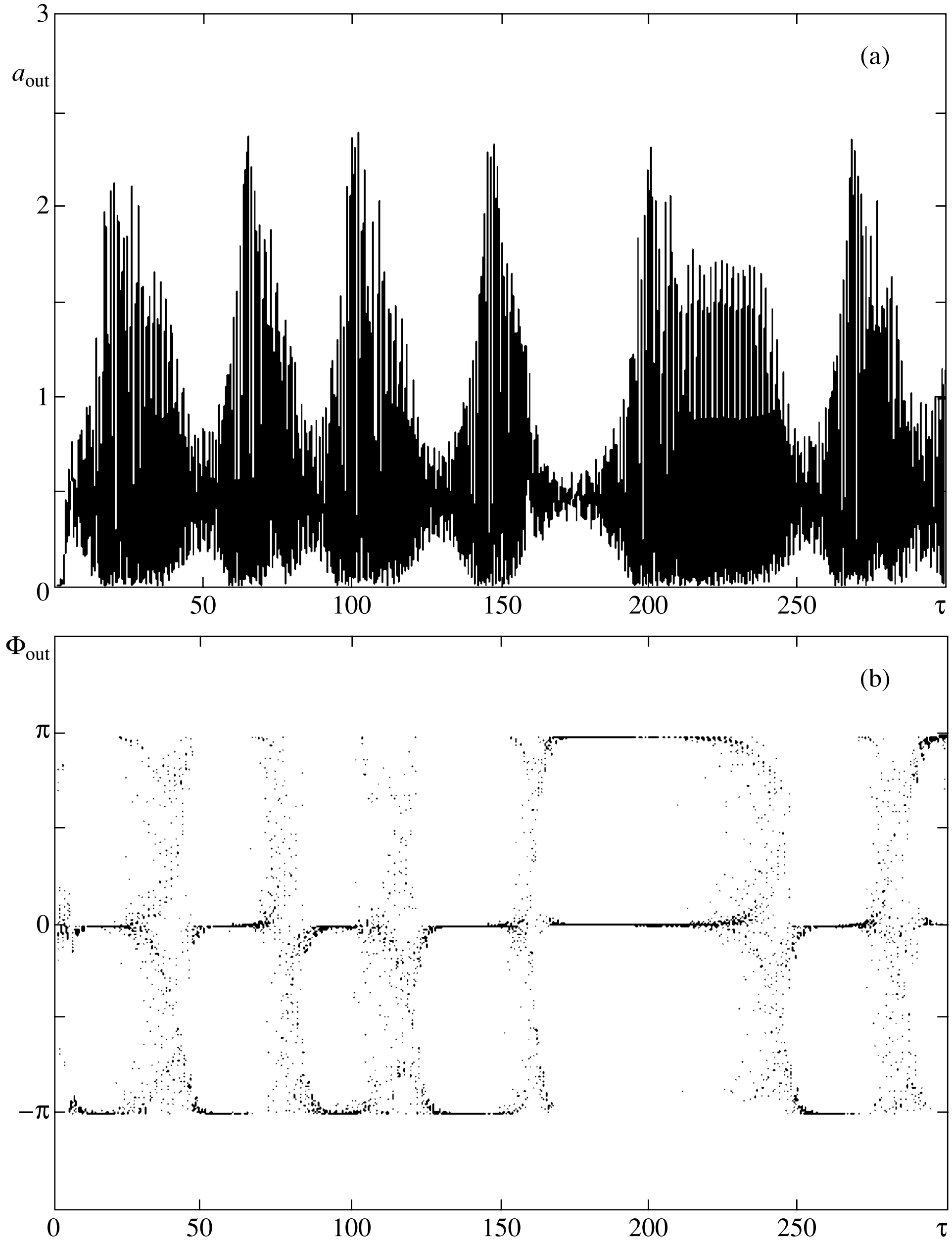


Fig. 5. Temporal realization of the amplitude (a) and phase (b) of the output signal in the intermittency regime: $\rho = 0.2$, $\alpha = 5.2$, $\delta = 1$.

leads to the formation of a soliton. Such processes are of practical interest in connection with the generation of ultrashort pulses [27].

Computations show that a further increase in the delay time δ gradually leads to self-modulation based on modes with higher indexes; moreover, a greater number of modes are excited. Clearly, this is explained by condensation of the spectrum of the resonator's eigenfrequencies.

Using terminology common in the theory of distributed self-oscillatory systems [7, 11, 12, 14], the self-modulation mechanism described above is called the frequency or phase mechanism since the phase oscillation is dominant at the first stage of the instability development (certainly, this does not mean that there is no amplitude oscillation). The frequency mechanism is characteristic of many distributed self-oscillatory systems such as beam and beam-plasma oscillators with delay [11–13] and free electron lasers [15, 16]. Another mechanism called amplitude is realized in the presence of a steep decrease section on the amplifier gain characteristic [7, 11, 12, 14]. The amplitude self-modulation develops softly, and its period is close to the double characteristic feedback time (for the system under con-

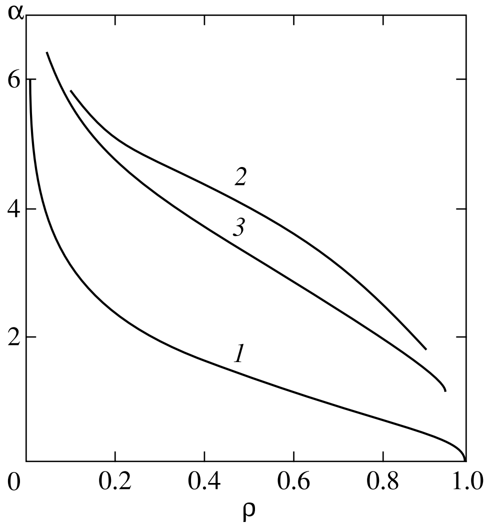


Fig. 6. The map of dynamic regimes on the plane of control parameters: *1* is the boundary of oscillation self-excitation; *2* is the boundary of self-modulation occurrence; *3* is the boundary of the chaotic behavior onset via intermittency.

sideration, this time is $1 + \delta$). In the case of the frequency mechanism, several eigenmodes of the oscillator are excited, whereas the amplitude mechanism is associated with the modulation of a single mode.[1] In this case, secondary modes are excited before the amplitude instability mechanism manifests itself. This fact is strongly confirmed both by the frequencies of self-modulation satellites and by the hard character of their excitation. However, it must be noted that the synchronization of phases is still established at the first stage of the transient process, so that only amplitudes oscillate. For this reason, the terminology adopted seems not quite appropriate.

It is well known that the frequency mechanism manifests itself when the gain–frequency characteristic of the amplifier in the vicinity of the fundamental frequency has a concave saddle-shaped form [7, 11, 12]. Figure 4 shows the results of computations for open loop amplifier operating in the harmonic signal amplification mode $A_1(\xi = 0) = A_0\exp(i\Omega\tau)$.

For small A_0 (curve *1*), the gain–frequency characteristic is convex in a rather large region near the maximum, which falls on the exact resonance frequency $\Omega = 0$ and corresponds to the linear theory (Section 3). As A_0 increases, the amplitude of the output signal also increases (curve *2*) until it reaches the maximal possible value of $\sqrt{1 + |A_0|^2}$, which corresponds to the total energy transfer to the signal wave. Then, the inverse process of pump regeneration on the fundamental frequency begins, and the output signal decreases; on adjacent frequencies, where the instability increment is smaller, this effect does not manifest itself yet. Thus, the gain–frequency characteristic becomes concave near $\Omega = 0$ (curves *3* and *4*), which confirms the frequency mechanism of stability loss.

5.2. Violation of the Phase Synchronization and the Onset of Chaos

The regimes of periodic self-modulation described above are stable in a certain range of variation of the pump parameter α above the instability threshold. However, as α increases, two solitons can occur that fit within the system extent. Furthermore, the signal wave amplitude is zero at certain points ($a_1(\xi) = 0$), which leads to the appearance of states with rapidly varying phases. As a result, the phase synchronization is violated; i.e., $\sin\Phi \neq 0$. Rapid phase oscillation leads to a sharp decrease of the interaction efficiency, the amplitude oscillation decreases, and the phase synchronization recovers. Thus, at certain moments in time the phase trajectory goes off the manifold $\sin\Phi = 0$ and then returns to it. A similar behavior is characteristic of a system with counterpropagating waves [10].

The behavior described above is illustrated in Fig. 5. It shows temporal realizations of the output signal amplitude and phase $\Phi_{\text{out}} = \Phi(\xi = 1)$. Regions of increasing amplitude oscillations, in which the phases are synchronized, are clearly seen. In these regions, as a soliton passes through the right-hand boundary of the system, Φ_{out} changes jumpwise by π. Regions of irregular oscillation corresponding to violated phase synchronization are also clearly seen. Thus, the oscillation becomes chaotic, and the system goes to the chaotic regime via intermittency. The spectrum of the output signal becomes continuous even though it contains well-defined components at the frequencies of the resonator's eigenmodes. We note that the transition via intermittency is also characteristic of systems considered in [11–13].

Since $\sin\Phi \neq 0$ in chaotic regimes, one cannot assume (as it was done in [9]) that the amplitudes of interacting waves are real. For this reason, the chaotic regimes described in [9] are generally unstable. Nevertheless, since the transient processes are rather long, these regimes can be considered metastable. Possibly, they can be realized under pulse pumping.

The overall picture is illustrated by a map of dynamic regimes on the plane of the parameters ρ and α shown in Fig. 6. In this figure, the boundaries of self-excitation (see formula (16)), self-modulation, and transition to chaos are depicted. However, it must be stressed that in the regions of chaos there exist numerous "windows" of regular behavior. A thorough study of the structure of chaotic regimes is of independent interest.

---

[1] This instability could be approximately described by mapping (23) for $da_{n+1}/da_n < -1$.

6. CONCLUSION

The distributed parametric oscillator considered in this paper provides a typical example of a distributed delayed self-oscillatory system. It demonstrates a transition from single-frequency to multifrequency to chaotic oscillations as the nonequilibrium parameter increases. In the present paper, the mechanism of the onset of self-modulation is thoroughly studied. It is shown that self-modulation is caused by the excitation of one more or several more pairs of the resonator's eigenmodes with the indexes depending on the delay time δ. This is the so-called frequency mechanism; it is related to the appearance of a dip in the gain–frequency characteristic of the amplifier in the vicinity of the fundamental frequency. The development of the instability leads to the formation of a soliton-like pulse that periodically propagates along the system. The onset of the self-modulation is hard; i.e., strong oscillations of the output signal appear immediately.

As the pump parameter increases, the system goes to a chaotic regime via intermittency due to violation of the phase synchronization; the phase synchronization is violated due to the appearance of states with quickly oscillating phases. A comparison with the results of the studies [11–13] suggests that such a behavior is typical for distributed self-oscillatory systems with the frequency self-modulation mechanism; under the amplitude mechanism, the transition to chaos usually occurs according to Feigenbaum's scenario [7, 11, 12, 17, 18].

ACKNOWLEDGMENTS

This work was supported by the program"Universities of Russia: Basic Research," project no. 992604, and the U.S. Civilian Research and Development Foundation, project no. REC-006.

REFERENCES

1. M. I. Rabinovich and D. I. Trubetskov, *Oscillations and Waves in Linear and Nonlinear Systems* (Nauka, Moscow, 1984; Kluwer, Dordrecht, 1989).
2. P. S. Landa, *Nonlinear Oscillations and Waves in Dynamical Systems* (Kluwer, Dordrecht, 1996; Nauka, Moscow, 1997).
3. N. M. Ryskin and D. I. Trubetskov, *Nonlinear Waves* (Nauka, Moscow, 2000).
4. W. H. Louisell, *Coupled Mode and Parametric Electronics* (Wiley, New York, 1960; Inostrannaya Literatura, Moscow, 1963).
5. S. A. Akhmanov and R. V. Khokhlov, Usp. Fiz. Nauk **88**, 439 (1966) [Sov. Phys. Usp. **9**, 210 (1966)].
6. V. G. Dmitriev and L. V. Tarasov, *Applied Nonlinear Optics* (Sov. Radio, Moscow, 1985).
7. S. P. Kuznetsov, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **25**, 1410 (1982).
8. Yu. I. Neimark and P. S. Landa, *Stochastic and Chaotic Oscillations* (Nauka, Moscow, 1987; Kluwer, Dordrecht, 1992).
9. G. N. Burlak and K. Ishkabulov, Zh. Éksp. Teor. Fiz. **109**, 774 (1996) [JETP **82**, 416 (1996)].
10. T. V. Dmitrieva and N. M. Ryskin, Zh. Éksp. Teor. Fiz. **116**, 1871 (1999) [JETP **89**, 1015 (1999)].
11. Yu. P. Bliokh, A. V. Borodkin, M. G. Lyubarskiĭ, *et al.*, Izv. Vyssh. Uchebn. Zaved., Prikl. Nelineĭnaya Din. **1** (1–2), 34 (1993).
12. Yu. P. Bliokh, M. G. Lyubarskiĭ, V. O. Podobinskiĭ, and Ya. B. Faĭnberg, Fiz. Plazmy **20**, 718 (1994) [Plasma Phys. Rep. **20**, 648 (1994)].
13. Yu. P. Bliokh, M. G. Liubarskii, V. O. Podobinskii, *et al.*, Phys. Plasmas **5**, 4061 (1998).
14. T. M. Antonsen and B. Levush, Phys. Fluids B **1**, 1097 (1989).
15. N. S. Ginzburg and M. I. Petelin, Int. J. Electron. **59**, 291 (1985).
16. N. S. Ginzburg and M. I. Petelin, Izv. Vyssh. Uchebn. Zaved., Prikl. Nelineĭnaya Din. **2** (6), 3 (1994).
17. N. M. Ryskin, V. N. Titov, and D. I. Trubetskov, Dokl. Ross. Akad. Nauk **358**, 620 (1998) [Phys. Dokl. **43**, 90 (1998)].
18. N. M. Ryskin and V. N. Titov, Izv. Vyssh. Uchebn. Zaved., Prikl. Nelineĭnaya Din. **6** (1), 75 (1998).
19. N. M. Ryskin and V. N. Titov, J. Commun. Technol. Electron. **45**, S46 (2000).
20. L. Kocarev, K. S. Halle, K. Eckert, *et al***.,** Int. J. Bifurcation Chaos Appl. Sci. Eng. **2**, 709 (1992).
21. S. Hayes, C. Grebogy, and E. Ott, Phys. Rev. Lett. **70**, 3031 (1993).
22. A. S. Dmitriev, Radiotekh. Élektron. (Moscow) **38**, 1 (1993).
23. M. Hasler, Int. J. Bifurcation Chaos Appl. Sci. Eng. **8**, 647 (1998).
24. M. M. Sushchik, V. M. Fortus, and G. I. Freĭdman, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **13**, 631 (1970).
25. A. Yariv, *Quantum Electronics* (Wiley, New York, 1975; Sov. Radio, Moscow, 1980).
26. R. Z. Sagdeev, D. A. Usikov, and G. M. Zaslavsky, *Nonlinear Physics: From the Pendulum to Turbulence and Chaos* (Nauka, Moscow, 1988; Harwood, Chur, 1988).
27. E. Ibragimov, A. A. Struthers, D. J. Kaup, *et al.*, Phys. Rev. E **59**, 6122 (1999).

Translated by A. Klimontovich

Journal of Experimental and Theoretical Physics, Vol. 93, No. 6, 2001, pp. 1323–1329.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 6, 2001, pp. 1527–1534.
Original Russian Text Copyright © 2001 by Kravtsov, Bilchinskaya, Butkovskii, Rychka, Surovyatkina.

MISCELLANEOUS

Prebifurcational Noise Rise in Nonlinear Systems

Yu. A. Kravtsov[a], *, S. G. Bilchinskaya[b], O. Ya. Butkovskii[c], **,
I. A. Rychka[b], and E. D. Surovyatkina[b], ***
[a]Institute for Space Research, Russian Academy of Sciences, ul. Profsoyuznaya 84/32, GSP-7, Moscow, 117810 Russia
Space Research Center, Polish Academy of Sciences, Warsaw, 00716 Poland
[b]Kamchatka State Technical University, Petropavlosk-Kamchatskii, 683003 Russia
[c]Vladimir State University, Vladimir, 600029 Russia
**e-mail: kravtov@asp.iki.rssi.ru*
***e-mail: oleg@vpti.vladimir.ru*
**** e-mail: elena@marine.kamchatka.su*

Received April 13, 2001

Abstract—The phenomenon of prebifurcational noise increase in nonlinear systems in the process of period-doubling bifurcation is investigated. The study is conducted for a discrete system (quadratic mapping); however, many of the laws discovered apply to more general systems. Estimates of the fluctuation variance are obtained both for the linear (away from the bifurcation threshold) and for the nonlinear mode (in the vicinity of the bifurcation threshold). It is shown that the variance of forced fluctuations in the strongly nonlinear mode is proportional to the root-mean-square of the noise intensity rather than to the variance. The possibility of measuring the noise in nonlinear systems on the basis of the prebifurcational noise amplification factor is demonstrated. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

It is well known that the noise level in a system increases when approaching the generation threshold (the phenomenon of pregeneration noise increase [1]). Similar phenomena might be expected near bifurcation points. It was shown in [2–4] that weak signals can be amplified when approaching the critical values of the bifurcation parameter; this phenomenon is explained by a decrease of the damping decrement (possibly, down to zero at the critical point). It is clear that the prebifurcation amplification of weak signals must be accompanied by an increase of weak noise. This phenomenon, which can be naturally called the prebifurcation noise increase, was first analyzed in paper [5], devoted to the analysis of noisy precursors of nonlinear instabilities. Taking into account a similarity between phase transitions and bifurcation transitions in nonlinear systems [1], the prebifurcation noise increase phenomenon can be considered as an analog of the amplification of fluctuations near the phase transition point.

In a linear approximation, the prebifurcation noise was studied in [5] for three bifurcation types—saddle-point bifurcation, bifurcation of the spontaneous symmetry violation, and period-doubling bifurcation. However, the linear theory developed in [5] does not take into account nonlinear effects and predicts an unbounded increase of fluctuations near the bifurcation point.

The main purpose of the present study is to perform a nonlinear analysis of the prebifurcation noise increase and estimate the variance of forced fluctuations exactly at the bifurcation point. The analysis is performed for a specific case of the period-doubling bifurcation in a discrete system (a quadratic mapping); however, the features of this phenomenon are characteristic of other bifurcation types as well.

The paper is organized as follows. Section 2 contains the statement of the problem. Sections 3 and 4 expand the linear and the nonlinear theory, respectively. In Section 5, the variance of the postbifurcation noise, i.e., the noise occurring after the transition through the bifurcation point, is estimated. In Section 6, a new feature of the prebifurcation and postbifurcation noise is discovered; this feature manifests itself in a quick transition through the bifurcation point. Then, discrete time of the fluctuation stabilization n_{trans} is estimated, which is inversely proportional to the root-mean-square value of the fluctuation intensity. Estimates obtained in Sections 3–6 are illustrated in Section 7 by results of the numerical modeling. Finally, Section 8 describes the application of the prebifurcation noise increase phenomenon to measuring weak noise in nonlinear systems.

2. STATEMENT OF THE PROBLEM

Consider the quadratic mapping

$$x_{n+1} = F(x_n), \quad F(x_n) = \mu - x_n^2. \tag{1}$$

It is known that this mapping admits a period-doubling bifurcation for the critical value of the control parameter $\mu_{c1} = 3/4$ [6]. As the critical value $\mu = \mu_{c1}$ is reached, the fixed point

$$\bar{x}(\mu) = -\frac{1}{2} + \sqrt{\frac{1}{4} + \mu} \tag{2}$$

of mapping (1) becomes unstable. Ultimately, the loss of stability at the critical point causes the prebifurcation noise increase.

Let us introduce a fluctuation f_n into mapping (1) and analyze the properties of the perturbed quadratic mapping

$$x_{n+1} = \mu - x_n^2 + f_n. \tag{3}$$

Set

$$\mu = \mu_{c1} - \lambda = \frac{3}{4} - \lambda, \tag{4}$$

$$x_n = \bar{x}(\mu) + \xi_n, \tag{5}$$

where $\lambda = \mu_{c1} - \mu$ is the “stability margin” in the vicinity of the bifurcation threshold. Then, near the critical point $\mu = \mu_{c1}$, the fixed point (2) can be represented in the form

$$\bar{x}(\mu) = -\frac{1}{2} + \sqrt{1-\lambda} \approx \frac{1}{2}(1-\lambda) + \frac{1}{8}\lambda^2 + \dots . \tag{6}$$

Substituting (4) and (5) into Eq. (2), we obtain the perturbed mapping for the perturbation $\xi_n = x_n - \bar{x}(\mu)$:

$$\xi_{n+1} = -2\bar{x}(\mu)\xi_n - \xi_n^2 + f_n. \tag{7}$$

The purpose of this study is to estimate the variance σ_ξ^2 and the mean square of the fluctuation $\langle \xi_n^2 \rangle$ both away from the bifurcation threshold (the linear mode) and in its immediate vicinity (nonlinear mode).

3. LINEAR THEORY

Let us neglect the quadratic term ξ_n^2 in the perturbation $\xi_n = x_n - \bar{x}(\mu)$ in (7) to obtain the linear equation

$$\xi_{n+1} = -\gamma\xi_n + f_n, \tag{8}$$

where

$$\gamma = 2\bar{x}(\mu) \approx 1 - \lambda + \lambda^2/4 + \dots \tag{9}$$

is the multiplicator of mapping (1) in the vicinity of the fixed point $\bar{x}(\mu)$. The solution to the difference linear equation (8) with the initial value $\xi_0 = 0$ can be obtained by the iteration method:

$$\begin{aligned} \xi_1 &= f_0, \\ \xi_2 &= -\gamma\xi_1 + f_1 = -\gamma f_0 + f_1, \\ \xi_3 &= -\gamma\xi_2 + f_2 = f_2 - \gamma f_1 + (-\gamma)^2 f_0, \\ &\vdots \\ \xi_{n+1} &= f_n + (-\gamma) f_{n-1} + (-\gamma)^2 f_{n-2} + \dots + (-\gamma)^n f_0. \end{aligned} \tag{10}$$

The last formula can be written in the form

$$\xi_{n+1} = \sum_{k=0}^{n} (-\gamma)^{n-k} f_k. \tag{11}$$

Assume that the functions f_j and f_k are independent; i.e.,

$$\langle f_j f_k \rangle = \sigma_f^2 \delta_{jk}, \tag{12}$$

where δ_{jk} is the Kronecker delta and $\sigma_f^2 = \langle f_k^2 \rangle$ is the variance of the fluctuations. Then, the mean square of the fluctuations $\langle \xi_{n+1}^2 \rangle$ is written as

$$\begin{aligned} \langle \xi_{n+1}^2 \rangle &= \left\langle \sum_{k=0}^{n} (-\gamma)^{n-k} f_k \right\rangle^2 \\ &= \sum_{k=0}^{n} \gamma^{2(n-k)} \sigma_f^2 = \sigma_f^2 \frac{1-\gamma^{2(n+1)}}{1-\gamma^2}. \end{aligned} \tag{13}$$

In the stationary state (as $n \longrightarrow \infty$), the term $\gamma^{2(n+1)}$ in (13) can be neglected since $|\gamma| < 1$; as a result, the mean square of the fluctuations $\langle \xi_{n+1}^2 \rangle$ is written as

$$\langle \xi^2 \rangle = \langle \xi_{n+1}^2 \rangle_{n \to \infty} = \frac{\sigma_f^2}{1-\gamma^2}. \tag{14}$$

Using expansion (9), we can represent the mean square of the fluctuations $\langle \xi^2 \rangle$ near the bifurcation threshold $\gamma = \gamma_c = 1$ in the form

$$\langle \xi^2 \rangle = \frac{\sigma_f^2}{2\lambda}, \quad 2\lambda \approx 1 - \gamma^2. \tag{15}$$

It is seen from (14) and (15) that away from the bifurcation threshold, say for $\gamma \le 1/2$, the mean square of the fluctuations $\langle \xi^2 \rangle$ is comparable with the variance of the external noise σ_f^2; however, when approaching the threshold $\gamma = \gamma_c$, fluctuations in the system increase without limit.

4. ESTIMATES OF THE FLUCTUATION LEVEL WITH REGARD FOR NONLINEAR EFFECTS

Formulas (14) and (15) derived from the linearized Eqs. (8) are inapplicable in the vicinity of the bifurcation threshold (i.e., as $\gamma \longrightarrow 1$ or, which is the same, as $\lambda \longrightarrow 0$). However, they allow us to obtain an estimate of fluctuations at the bifurcation threshold ($\gamma = 1$, $\lambda = 0$) using the following simple reasoning. In the linear mode, the fluctuations ξ_n consist of a sum of a large number of independent terms (11). Hence, by virtue of the central limit theorem, the sequence ξ_n is nearly Gaussian. This enables us to evaluate the fluctuation part $\eta_n = \xi_n^2 - \langle\xi_n^2\rangle$ of the nonlinear term ξ_n^2 in Eq. (7) using the Gaussian statistics. As a result, the standard deviation $\sigma_\eta = \langle\eta^2\rangle^{1/2}$ is estimated as

$$\sigma_\eta \approx \langle\xi^2\rangle \sim \sigma_f^2/2\lambda. \tag{16}$$

The smallness condition for the standard deviation σ_η compared to the external fluctuations characterized by the standard deviation σ_f has the form

$$\sigma_\eta \sim \frac{\sigma_f^2}{2\lambda} \le \sigma_f, \text{ or } \lambda \ge \lambda_{\min} \sim \frac{\sigma_f}{2}. \tag{17}$$

Substituting the nonlinear estimate $\lambda = \lambda_{\min} = \sigma_f/2$ into linear formula (15), we obtain the desired estimate of the variance of fluctuations σ_ξ^2 in the vicinity of the bifurcation threshold:

$$\langle\xi^2\rangle_{\max} \sim \frac{\sigma_f^2}{2\lambda_{\min}} \sim \sigma_f. \tag{18}$$

The ratio

$$K = \frac{\langle\xi^2\rangle}{\sigma_f^2} \tag{19}$$

is a natural measure of the prebifurcation fluctuation amplification. According to (18), the maximum fluctuation "amplification factor" $K_{\max} = \langle\xi^2\rangle_{\max}/\sigma_f^2$ is

$$K_{\max} \sim \frac{\langle\xi^2\rangle_{\max}}{\sigma_f^2} \sim \frac{1}{\sigma_f}. \tag{20}$$

For example, for $\sigma_f^2 = 10^{-8}$, the amplification factor $K_{\max}$ can become as large as $K \approx 10^4$. Such a large value can make it possible to measure weak fluctuation actions σ_f in various physical systems (see Section 8).

Another phenomenon associated with the prebifurcation fluctuation increase is a shift of the mean value $\langle\xi\rangle$ with respect to zero. The magnitude of this effect can be determined by averaging Eq. (7):

$$\langle\xi\rangle = -\gamma\langle\xi\rangle - \langle\xi^2\rangle.$$

This equation implies that the mean shift of the fixed point from the unperturbed value $\bar{x}(\mu)$ in the close neighborhood of the bifurcation point $\gamma = 1$ can be estimated as

$$\langle\xi\rangle \sim \frac{1}{2}\langle\xi^2\rangle \sim \frac{\sigma_f^2}{4\lambda}.$$

This is the shift of the mean value $\langle\xi\rangle$ from zero.

The maximum shift $\langle\xi\rangle_{\max}$ is attained at $\lambda \sim \lambda_{\min} \sim \sigma_f/2$:

$$\langle\xi\rangle_{\max} \sim \sigma_f/2. \tag{21}$$

Although the shift of the mean value is usually small (about 0.5×10^{-3} for $\sigma_f = 10^{-3}$), it can also indicate the prebifurcation rise of fluctuations. Since $\langle\xi\rangle^2$ is small compared to $\langle\xi^2\rangle$, the variance σ_ξ^2 is almost equal to $\langle\xi^2\rangle$; i.e.,

$$\sigma_\xi^2 = \langle(\xi - \langle\xi\rangle)^2\rangle \approx \langle\xi^2\rangle.$$

This fact will be used in subsequent estimates.

5. FLUCTUATIONS AFTER THE TRANSITION THROUGH THE BIFURCATION POINT

For $\mu > \mu_{c1} = 0.75$, system (1) undergoes a period-doubling bifurcation. Stationary oscillations with the period 2 are described (in the absence of noise) by the system of two equations

$$\begin{aligned} x_{n+1} &= \mu - x_n^2, \\ x_{n+2} &= \mu - x_{n+1}^2. \end{aligned} \tag{22}$$

Eliminating x_{n+1} from (22) and equating x_{n+2} to x_n (which corresponds to the period $T = 2$), we obtain the following equation for the fixed point $\hat{x}(\mu)$ of the 2-period mode:

$$\hat{x} = \mu - (\mu - \hat{x}^2)^2. \tag{23}$$

Denote by

$$\delta = \hat{x} - \bar{x}(\mu) \tag{24}$$

the deviation of $\hat{x}$ from the fixed point $\bar{x}(\mu)$ (2) (for $\mu > \mu_{c1}$, the latter becomes unstable); by

$$\nu = \mu - \mu_{c1} > 0 \tag{25}$$

we denote the supercriticality factor, i.e., the excess of μ over the bifurcation threshold μ_c. In terms of the variables δ and ν, Eq. (23) takes the form

$$\delta^2 = (\nu - \delta - \delta^2)^2 \tag{26}$$

and has a two-valued solution

$$\delta = \pm\sqrt{\nu}, \tag{27}$$

which has the form of a typical bifurcation fork. In terms of the original variables, this relation reads as

$$\hat{x}^{\pm} = \bar{x} \pm \sqrt{\mu - \mu_{c1}}. \tag{28}$$

Using relation (27), we can estimate the extent of the region of strong fluctuations to the right of the bifurcation point μ_{c1}. For this purpose, we use the following reasoning. It was shown in Section 4 that the mean-square value of fluctuations near the bifurcation point is $\sigma_\xi \sim \sqrt{\sigma_f}$. Clearly, for $\sigma_\xi \geq \delta$, external fluctuations f_n can move points from on branch of the bifurcation diagram (27) to the other, for example, from $\hat{x}^+$ to $\hat{x}^-$ and back. Thus, the vertical size of the fluctuation region can be estimated as

$$\delta \sim \sigma_\xi \sim \sqrt{\sigma_f}. \tag{29}$$

By virtue of (27), the horizontal size of this region, $\nu_{\min}$, is estimated as

$$\nu_{\min} \sim \delta^2 \sim \sigma_f. \tag{30}$$

Estimates (29) and (30) conform with the previous estimate (18) for $(\sigma_\xi^2)_{\max}$ and estimate (17) for $\lambda_{\min}$.

As the supercriticality factor $\nu = \mu - \mu_{c1}$ increases, the variance of fluctuations decreases according to the law

$$\sigma_\xi^2 = \frac{\sigma_f^2}{2\nu}, \tag{31}$$

which is similar to (15). This variance decrease law for the postbifurcation noise can be derived using the linear theory much in the same way as it was done in Section 3 for the prebifurcation noise.

Note that the variances of fluctuations near the states $\hat{x}^+$ and $\hat{x}^-$, σ_+^2 and σ_-^2, respectively, are different. Denote by $\Pi^\pm$ the values of the derivative of the mapping function $f(x) = \mu - x^2$ at the points $\hat{x}^\pm$:

$$\Pi^\pm = \frac{df}{dx} = -2\hat{x}^\pm.$$

Taking into account the fact that $\Pi^\pm$ is interpreted as the slope of the tangents to $f(x)$ at the points $\hat{x}^\pm$ and $\hat{x}^-$, we obtain, for sufficiently small fluctuations,

$$\sigma_- = \Pi^+\sigma_+, \quad \sigma_+ = \Pi^-\sigma_-.$$

As a result, we arrive at the estimate

$$\frac{\sigma_+^2}{\sigma_-^2} = \left|\frac{\Pi^-}{\Pi^+}\right| = \frac{\hat{x}^-}{\hat{x}^+} \approx 1 - 2\frac{\sqrt{\mu - \mu_{c1}}}{\bar{x}(\mu)}, \tag{32}$$

which shows that fluctuations near the upper branch $\hat{x}^+$ are always smaller than those near the lower branch.

6. WEAKENING OF THE AMPLIFICATION EFFECT ON QUICK TRANSITION THROUGH THE BIFURCATION POINT, AND THE FLUCTUATION SETTLING TIME

On a quick transition through the bifurcation point, the fluctuation amplification effect must diminish; indeed, fluctuations are not accumulated in this case. In order to verify this fact, we interpret n in Eq. (13) as the time over which the control parameter changes. For $n = 1$, (13) implies that

$$\sigma_\xi^2 = \sigma_f^2, \tag{33}$$

whereas for $n = 2$ we have

$$\sigma_\xi^2 = \sigma_f^2\frac{1-\gamma^4}{1-\gamma^2} = \sigma_f^2(1+\gamma^2) \approx 2\sigma_f^2. \tag{34}$$

According to (33), when the bifurcation point is passed very quickly (in a single step), no rise in the fluctuations is observed. If the number of steps increases to $n = 2$, then the fluctuations double; i.e., the variation σ_ξ^2 is twice as large as the variation of the noise σ_f^2.

In order to estimate the fluctuation transient time n_{trans}, we resort to formula (13), which gives the mean square of the fluctuations after n iterations. By virtue of (13), the fluctuations settle when $\gamma^{2(n+1)} \ll 1$. If we approximately replace γ by $1 - \lambda \approx \exp(-\lambda)$ for small λ, then the inequality $\gamma^{2(n+1)} \ll 1$ can be written as $\exp[-2\lambda(n+1)] \ll 1$, which yields the estimate

$$n_{\text{trans}} \geq \frac{1}{2\lambda}. \tag{35}$$

Away from the bifurcation threshold ($\mu = \mu_{c1}$), when λ is comparable with unity, the time n_{trans} is about several iteration steps. In the immediate vicinity of μ_{c1}, 2λ is comparable with $2\lambda_{\min} \sim \sigma_f$; as a result, the lower value of the fluctuation transient time is estimated as

$$n_{\text{trans}} \geq \frac{1}{2\lambda} \sim \frac{1}{\sigma_f}. \tag{36}$$

For example, for $\sigma_f = 10^{-3}$, in order to estimate the steady-state fluctuation variance σ_ξ^2, one must take a sample consisting of not less than 1000 iterations; for $\sigma_f = 10^{-5}$, the sample length must be not less than 10^5 iterations.

7. RESULTS OF NUMERICAL MODELING

The system described in the preceding sections was numerically investigated for the initial value of the control parameter $\mu_0 = 0.2$ and the final value $\mu_f = 0.9$ with the step $\Delta\mu = 0.001$. Such a choice of the range of variation for μ is explained by the location of the bifurcation points of the quadratic mapping: the first period-doubling bifurcation occurs at $\mu = \mu_{c1} = 0.75$ and the second at $\mu = \mu_{c2} = 1.25$. Thus, the initial value μ_0 was taken below μ_{c1} and the final μ_f above μ_{c1} but below μ_{c2}.

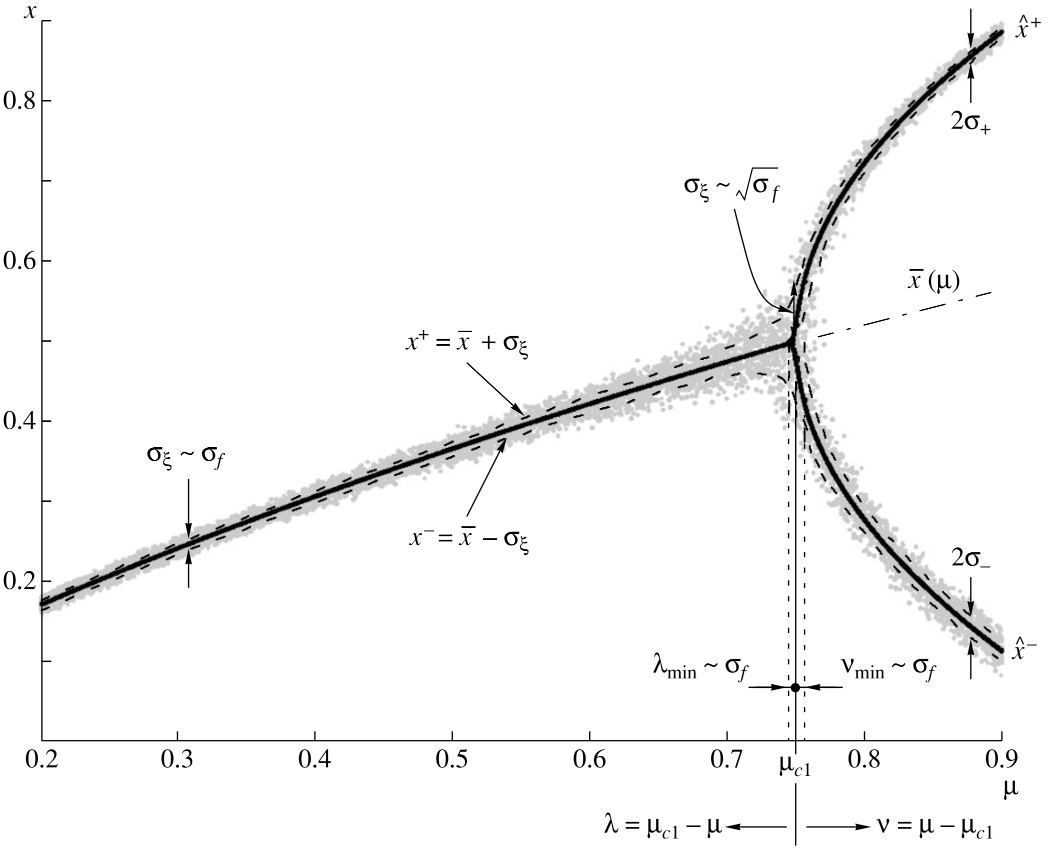


Fig. 1. The area of the fluctuation concentration (shaded) on the bifurcation diagram in the presence of noise with the standard deviation $\sigma_f = 5.7 \times 10^{-3}$. Dashed lines correspond to the standard deviation σ_ξ. The prebifurcation amplification of fluctuations is observed near the bifurcation point μ_{c1}.

The random number generator produced uniformly distributed independent (see (12)) fluctuations f_n within the interval $[-\beta, +\beta]$

$$w(f_n) = \begin{cases} \dfrac{1}{2\beta}, & |f_n| \le \beta, \\ 0, & |f_n| > \beta. \end{cases} \tag{37}$$

For the uniform distribution (35), the variance of the noise σ_f^2 is related to the parameter β as $\sigma_f^2 = \beta^2/3$. The use of other random number generators, for example, normal ones, yielded qualitatively similar results.

The initial value x_0 was always set to the fixed point of the mapping $\bar{x}(\mu)$. For every value of μ, at least $1/\sigma_f$ iteration steps of mapping (3) were performed (see Section 6). To obtain more reliable results, the number of iterations was set to $100/\sigma_f$ in the numerical experiments; as a result, the fluctuation variance σ_ξ^2 was determined with a relative error $(100/\sigma_f)^{-1/2}$. For example, for $\sigma_f = 10^{-2}$, the relative error was 1%, while for $\sigma_f = 10^{-4}$ it was 0.1%.

Figure 1 presents the bifurcation diagram of the quadratic mapping (3) under the influence of noise with the standard deviation $\sigma_f = 5.7 \times 10^{-3}$. The shaded area in this diagram shows the location of fluctuations of the perturbed values of $\bar{x}(\mu)$ and $\hat{x}^{\pm}(\mu)$. Dashed lines correspond to the standard deviation σ_ξ from the unperturbed values. Figure 1 illustrates all the specific features of the problem described above. Away from the bifurcation point $\mu = \mu_{c1}$, the variance of fluctuations σ_ξ^2 is comparable with the variance of the external action σ_f^2. While approaching the critical point, the variance of fluctuations σ_ξ^2 first increases according to Eqs. (14), (15), and takes the bounded value (20) just before reaching the value $\mu = \mu_{c1}$. In this case, the maximal fluctuation amplification factor

$$K_{\max} \sim \frac{(\sigma_\xi^2)_{\max}}{\sigma_f^2}$$

is about 84, which is of the same order of magnitude as estimate (20). Nonlinear effects play a determining role in the interval $|\mu - \mu_{c1}| \le \sigma_f \approx 0.005$.

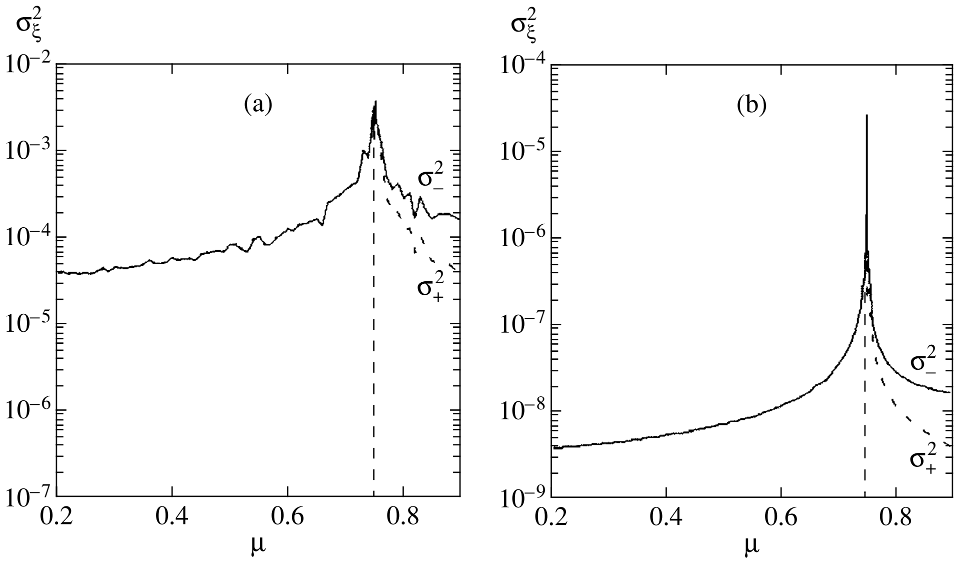


Fig. 2. Dependence of the variance of fluctuations σ_ξ^2 on the control parameter μ in the vicinity of the critical point μ_{c1} for the noise intensity $\sigma_f^2 = 3.3 \times 10^{-5}$ (a) and $\sigma_f^2 = 3.3 \times 10^{-9}$ (b). The solid (dashed) curves to the right of the critical point correspond to fluctuations near the lower (respectively, upper) branch of the bifurcation diagram.

Figure 2 shows the dependences of the variance of fluctuations σ_ξ^2 on the control parameter μ for two values of the noise variance $\sigma_f^2 = 3.3 \times 10^{-5}$ (Fig. 2a) and $\sigma_f^2 = 3.3 \times 10^{-9}$ (Fig. 2b). The curves on the plots are not quite smooth, which is explained by the fact that the sample used to determine σ_ξ^2 was finite; more precisely, σ_ξ^2 shown in Fig. 2a was calculated using 10^4 iteration steps, while 10^6 iteration steps were used for Fig. 2b. In the vicinity of the critical point μ_{c1}, σ_ξ^2 drastically increases compared to σ_f^2. For $\sigma_f^2 = 3.3 \times 10^{-5}$, the amplification factor is $K_{\max} \sim 10^2$; for $\sigma_f^2 = 3.3 \times 10^{-9}$, we have $K_{\max}$ approximately 10^4. In the first case, σ_ξ^2 attains the value $(\sigma_\xi^2)_{\max} \approx 2.8 \times 10^{-3}$; i.e., it differs from the theoretical estimate $(\sigma_\xi^2)_{\max} \sim \sigma_f$ by a factor of 0.49. In the second case, the maximal variance $(\sigma_\xi^2)_{\max} \approx 2.7 \times 10^{-5}$, which differs from the estimate $\sigma_f = 10^{-4}$ by a factor of 0.47.

After the transition through the critical value μ_{c1}, the variance of fluctuations σ_ξ^2 becomes a two-valued function: for $\mu > \mu_{c1}$, the solid curve in Fig. 2 corresponds to the lower branch of the bifurcation diagram (the variance of fluctuations is σ_-^2), and the dashed curve corresponds to the upper branch (the variance is σ_+^2). As would be expected on the basis of estimate (32), fluctuations near the upper branch $\hat{x}^+$ are always less than those near the lower branch $\hat{x}^-$.

Figure 3 illustrates the dependence of the maximal variance of fluctuations $(\sigma_\xi^2)_{\max}$ near the bifurcation threshold on the mean-square fluctuation action σ_f. Points on this plot can be approximated by the dependence

$$(\sigma_\xi^2)_{\max} \approx 0.5\sigma_f, \tag{38}$$

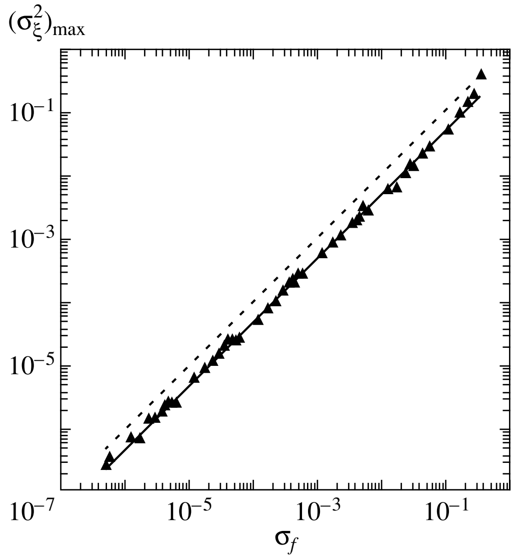


Fig. 3. Dependence of the maximal fluctuation variance $(\sigma_\xi^2)_{\max}$ on the standard deviation of the noise action σ_f. The dashed line corresponds to estimate (18).

which is in good agreement with the theoretical estimate (18) (the latter is shown in the dashed line in Fig. 3). Numerical computations show that formula (38) can be used not only for the uniform distribution of the fluctuation intensity f_n, but also for the normal distribution.

For large noise, when the standard deviation σ_f is close to unity, the fluctuation amplification factor K is also close to unity; in this case, the nonlinear mechanisms of bounding the noise variance are replaced by the dependence $\sigma_\xi^2 \sim \sigma_f^2$, which is characteristic of linear systems.

8. ON MEASURING WEAK NOISE

A noticeable increase in the variation of fluctuations σ_ξ^2 near the bifurcation threshold can be used as the basis of a method for measuring weak noise in the system under study. In contrast to the highly unstable pregeneration noise, which appears due to sporadic transitions of the system into the generation mode, the prebifurcation noise is stabler and ensures more reliable measurements. This is true not only for period-doubling bifurcations, but also for many other bifurcation types that do not involve a large increase in the oscillations when transiting through the critical point as is the case for Landau–Hopf bifurcations. Our approach is based on a comparison of the maximal variation $(\sigma_\xi^2)_{\max} \sim \sigma_f$ at the bifurcation point with the variation $\sigma_\xi^2 \approx \sigma_f^2$ away from that point. An estimate of the noise variation σ_f^2 on the basis of the mean shift $\langle\xi\rangle$, which appears due to a sort of detection of fluctuations on the nonlinearity of the system, seems less reliable. The possibility of measuring weak noise by measuring $(\sigma_\xi^2)_{\max}$ is limited by condition (36) on the duration of measurements. According to (36), the duration of the sample must be not less than the transient time $n_{\text{trans}} \sim 1/\sigma_f$.

To complete the picture, we note that the transition of a nonlinear system through the period-doubling bifurcation presents one more possibility for estimating the internal noise using the magnitude of the noise-dependent hysteresis curve, which occurs when the critical point is transited in the forward and backward direction [7].

9. CONCLUSIONS

In this paper, for a specific case of the period-doubling bifurcation, we investigate the phenomenon of the prebifurcation amplification of fluctuations, which is similar in nature to the rise of fluctuations near phase transition points. Analytic estimates of the prebifurcation fluctuation amplification both away from the critical point and in its immediate vicinity are obtained. These estimates are in satisfactory agreement with the results of numerical modeling. On the basis of the phenomenon under study, a method for measuring the variance of weak noise in nonlinear systems is proposed. The applicability of this method is limited by the necessity to perform rather long-term observations.

ACKNOWLEDGMENTS

This work was supported in part by the Russian Foundation for Basic Research, project nos. 99-02-16625 and 00-02-17741, and by the federal program “Integration”, project no. A-0030.

REFERENCES

1. S. A. Akhmanov, Yu. E. D’yakov, and A. S. Chirkin, *Introduction to Statistical Radio Physics and Optics* (Nauka, Moscow, 1981).
2. K. Wiesenfeld, Phys. Rev. A **32**, 1744 (1985).
3. K. Wiesenfeld and N. F. Pedersen, Phys. Rev. A **36**, 1440 (1987).
4. K. Wiesenfeld and B. McNamara, Phys. Rev. A **33**, 629 (1986).
5. K. Wiesenfeld, J. Stat. Phys. **38**, 1071 (1985).
6. H. G. Schuster, *Deterministic Chaos* (Physik-Verlag, Weinheim, 1984; Mir, Moscow, 1988).
7. O. Ya. Butkovskiĭ, Yu. A. Kravtsov, and E. D. Surovyatkina, Zh. Tekh. Fiz. **67** (9), 128 (1997) [Tech. Phys. **42**, 1099 (1997)].

Translated by A. Klimontovich

Journal of Experimental and Theoretical Physics, Vol. 93, No. 6, 2001, pp. 1330–1336.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 6, 2001, pp. 1535–1543.
Original Russian Text Copyright © 2001 by Murtazaev, Kamilov, Magomedov.

MISCELLANEOUS

Cluster Algorithms of the Monte Carlo Method, Finite-Size Scaling, and Critical Exponents of Complex Lattice Models

A. K. Murtazaev, I. K. Kamilov, and M. A. Magomedov

Institute of Physics, Dagestan Research Center, Russian Academy of Sciences, Dagestan, 367003 Russia

e-mail: kamilov@datacom.ru

Received June 14, 2001

Abstract—For the first time, cluster algorithms of the Monte Carlo method are used to investigate critical properties of microscopic models of real ferromagnetic gadolinium. On the basis of the finite-size scaling theory, the critical exponents of the heat capacity α, magnetization β, susceptibility γ, and Fisher index η are calculated. Specific features, character, and the degree of influence of two types of weak relativistic interactions on the critical properties of gadolinium models are determined when both these interactions are taken into account. It is shown that cluster algorithms of the Monte Carlo method provide an effective tool for studying critical properties of complex models involving crossovers. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The modern theory of phase transitions and critical phenomena is mainly based on the ideas of scaling, universality, and renormalization group theory [1, 2]. On the basis of these ideas, the majority of important results of the modern phase transition and critical phenomena theory were obtained, basic laws established, relations between critical exponents and critical amplitudes were obtained, equations of state were derived, and the values of critical exponents and critical amplitudes were calculated. The numerical values obtained on the basis of renormalization group theory and ε-expansion are believed to be the most accurate and reliable among those available [3–6]. The current stage in the study of phase transitions and critical phenomena is characterized by the investigation of more complex and realistic models [7–11]. The approach underlying renormalization group theory faces considerable difficulties when studying models that involve crossovers; besides, it is not quite microscopic [6–12].

These facts and some other reasons led to the wide use of Monte Carlo methods for studying phase transitions [6–11, 13–16]. Quantitative analysis of the critical region by Monte Carlo methods has become possible only in recent years. At the present time, in terms of accuracy, the results obtained by the Monte Carlo method compare well with the results obtained by other methods and sometimes outperform them [6, 13–16].

Naturally, this progress could not have been made by increasing the power of computers without the use of additional ideas and methods. Among them, one must distinguish powerful cluster algorithms of the Monte Carlo method [17–20], the use of ideas involved in the theory of finite-size scaling (for the computation of critical parameters) [6, 21, 22] and histogram methods for the analysis of Monte Carlo data [14].

Until the present time, these algorithms and ideas were applied only to the simplest first-order approximation models (the classical Ising and Heisenberg models etc.). Much less attention was paid to more complex and realistic models that admit crossover transitions.

Another important aspect of the investigation of models of real magnetic materials by Monte Carlo techniques consists in the possibility of comparing the results of numerical simulation not only with theoretical but also with experimental data. This is especially important when the results of laboratory investigations of critical phenomena are contradictory and do not permit an unambiguous answer to certain important questions.

We propose microscopic models of the real ferromagnetic gadolinium. These models are investigated using cluster Monte Carlo algorithms; on the basis of the finite-size scaling theory, the main static critical exponents are calculated.

The interest in gadolinium models is explained by the following reasons.

First, the critical behavior of gadolinium can be influenced by weak relativistic interactions, such as anisotropy and dipole–dipole interactions [5, 23–26]. Earlier, the efficiency of Monte Carlo cluster algorithms in the investigation of models in which weak relativistic interactions are taken into account along with strong exchange interactions (in such models, crossover effects can occur) has never been studied.

Second, in the models under study, both types of weak relativistic interactions are simultaneously taken into account on the background of each other and strong exchange interactions. To our knowledge, the applicability and effectiveness of the finite-size scaling theory in the study of such models has never been verified [6].

Third, the sensitivity and resolution of the Monte Carlo method, as well as its capability of revealing the influence of such weak factors on the critical behavior, remain practically unstudied.

Fourth, the phase transition temperature $T_c \approx 293$ K of the rare-earth metal gadolinium is convenient for laboratory experiments, and this metal has been thoroughly studied. However, laboratory investigations of the critical properties of gadolinium do not provide a complete and strict description of the gadolinium critical behavior [26, 27].

2. CRITICAL PROPERTIES OF GADOLINIUM

Gadolinium is a rare-earth metal with a close-packed hexagonal structure. In the range of temperatures $232 \text{ K} < T < T_c \approx 293$ K, gadolinium demonstrates simple ferromagnetic ordering. Magnetic and neutron radiography investigations show that anisotropy in gadolinium is explained both by the single-ion and two-ion mechanisms, while in the paramagnetic phase, it is caused by the uniaxial anisotropy of the short-range magnetic order [22–25]. On the one hand, gadolinium is a uniaxial weakly anisotropic ferromagnetic; hence, its critical behavior at temperatures close to T_c can be of the Ising character. On the other hand, spherically symmetric distribution of the electron density and the absence of the orbital moment lead to the isotropic exchange interaction, which implies the Heisenberg character of the critical behavior.

The static critical behavior of gadolinium was studied in a number of papers [25–35]. Measurements of the heat expansion [25, 28], heat capacity [27, 29, 30], magnetic properties [26, 31–34], and Mössbauer studies [35] conducted for various mono and polycrystal samples helped determine a set of static critical exponents α, β, γ, and δ. In [5, 26, 35], tables containing the values of these exponents are presented. The comparison of these data with theoretical predictions obtained in the framework of the three-dimensional Ising and Heisenberg models shows their inconsistency. The values of the critical exponents of heat capacity α, heat expansion a, and spontaneous magnetization β testify that gadolinium is either a Heisenberg or isotropic dipole magnetic. At the same time, the values of the critical index γ are close to those characteristic of the Ising model. The index δ complies neither with the predictions of microscopic theories nor with those of molecular field theory.

An analysis of the available experimental data shows that these inconsistencies can be caused by the following reasons.

1. The methodology used to determine the exponents [5, 26]. In the greater part of studies, the critical exponents β, γ, and δ were determined by fitting the experimental m–H–T data to the scaling equation of state for magnetization, which assumes that the law of similarity $\gamma = \beta(\delta - 1)$ is satisfied. In this case, the critical exponents must obey the laws of similarity involving the same exponents β, γ, and δ; however, individually their values can be inconsistent with the actual asymptotic critical behavior.

2. In real crystals, there are always certain additional interactions that perturb the original critical behavior. For example, the presence of isotropic dipole interaction in Heisenberg magnets results in the dependence of the index γ on reduced temperature [26].

3. Experiments conducted on various gadolinium samples showed that defects can substantially change the critical behavior [27].

4. It turned out that the critical properties of gadolinium, in particular, the width and height of the heat capacity peak, depend on the sample preparation technique [27].

5. Theoretical estimates of the exponents are obtained for static models with fixed magnitudes of the geometric parameters of the lattice (angles, location of atoms, etc.). However, when investigating real samples in laboratory experiments, these magnitudes can vary; as a result, the exchange parameters can also vary, which sometimes leads to inconsistencies in the theoretical and experimental data.

Note that a thorough experimental investigation of static critical properties of gadolinium were performed in [26], where the exponents β, γ, and δ were determined. Specific features of the heat capacity behavior were studied in [27]. The results obtained in this paper show that the critical behavior of the heat capacity is very sensitive to the sample preparation technique, its purity, and chemical composition.

3. MODELS

When constructing models of gadolinium, one must take into account the following features of this element:

(a) The distribution of the electron density is spherically symmetric, and the orbital moment is nil.

(b) The energy of the magnetic crystallographic anisotropy is much less than for other rare-earth elements.

(c) The isotropic dipole–dipole interaction can play a substantial role in the critical region.

With regard for these features, the Hamiltonian of the system can be written in the form

$$H = -\frac{1}{2}\sum_{i,j} J(\mathbf{S}_i \cdot \mathbf{S}_j) - \frac{1}{2}D_a\sum_i (\mathbf{S}_i^z)^2 - D_d\sum_i (\langle \mathbf{m} \rangle \cdot \mathbf{S}_i), \quad (1)$$

where $\mathbf{m}$ is the magnetization, $\mathbf{S}$ is the classical three-dimensional spin, $|\mathbf{S}_i| = 1$, the first term accounts for the exchange interaction of each of the ions Gd^{3+} with its nearest neighbors ($J > 0$), the second term accounts for the one-ion anisotropy (D_a), and the third term accounts for the isotropic dipole–dipole interaction (D_d). According to the data obtained on the basis of molecular field theory [23, 26, 36, 37], the parameters of anisotropy D_a and isotropic dipole forces D_d have the values $D_a/J = 1.41 \times 10^{-4}$ and $D_d/J = 1.35 \times 10^{-3}$. Note that all the physical quantities are expressed in dimensionless units.

Calculations were performed for cubic samples of size $L \times L \times L$ (L = 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32) with periodic boundary conditions. In the construction of the model systems, all magnetic and crystallographic features of real gadolinium were taken into account. To reveal the degree of influence of dipole forces on the critical behavior, two gadolinium models were considered. The Γ1 model took into account the exchange interaction with the closest neighbors and the uniaxial anisotropy, while the Γ2 model additionally took into account the isotropic dipole–dipole interaction. Earlier, cluster Monte Carlo algorithms were not used for the analysis of the effect of weak relativistic interactions of various types against the background of strong exchange interactions and simultaneously against the background of each other on the critical behavior; the effectiveness and specific features of cluster algorithms as applied to complex models involving crossovers were not studied.

4. INVESTIGATION TECHNIQUES

Cluster Monte Carlo algorithms [17–28] proved to be a very effective tool for the investigation of critical phenomena in various systems and models [6, 15, 16, 38, 39]. The critical parameters calculated on the basis of data obtained with the use of cluster algorithms are very accurate and reliable [6]. Among all variants of cluster Monte Carlo algorithms, the Wolf algorithm [17, 18] is the most effective. We used the following form of this algorithm to investigate the models Γ1 and Γ2.

1. A direction of the unit vector $\mathbf{r}$ is chosen at random.

2. One of the spins of the lattice $\mathbf{S}_i \in A$, where A the set of all lattice sites, is chosen at random; below, following [38], we will call this spin central.

3. A new direction of the spin $R\mathbf{S}_i \longrightarrow \mathbf{S}'_i$, $R\mathbf{S}_i = \mathbf{S}_i - 2(\mathbf{S}_i \cdot \mathbf{r})\mathbf{r}$ is specified. Actually, the operation R consists in a specular reflection of the spin with respect to the plane perpendicular to the vector $\mathbf{r}$.

4. All nearest neighbors j of the chosen ith spin are visited. The bond $\langle ij \rangle$ is activated with the probability

$$P = 1 - \exp\{\min[0, 2J\beta(\mathbf{r} \cdot \mathbf{S}_i)(\mathbf{r} \cdot \mathbf{S}_i)]\}, \quad \beta = 1/k_B T. \tag{2}$$

5. If the bond $\langle ij \rangle$ is activated, then the spin at the site j changes its direction $R\mathbf{S}_j \longrightarrow \mathbf{S}'_j$; in this case, the spin j is included in the cluster.

6. After checking all nearest neighbors of the spin i, the first reversed spin j becomes central and the process of establishing its bonds with the closest neighbors is started. This process proceeds until the boundaries of the system are reached.

Thus, the set of all reversed spins forms a “cluster.” One and the same spin can be reversed only once; however, it can be multiply checked.

Note that the activation of a bond can be represented as

$$P = 1 - \exp\{\min[0, dU]\},$$

where $dU = dU_j - dU_i$ and U is the internal energy.

The initial configuration was chosen so that all spins were arranged along the axis z. To guide the system to the equilibrium state, 10^4 Monte Carlo steps per spin were made (one Monte Carlo step per spin corresponds to one reversal of the cluster), which is at least by a factor of 20 more than the nonequilibrium section even in the vicinity of the critical point. Thermodynamic quantities were averaged using the Markov chain of length up to 3×10^6 Monte Carlo steps per spin. We note that, for $T \approx T_c$, the average size of the reversed cluster increases with the size of the system L, which is quite natural. At the same time, the relative size of the reversed cluster gradually decreases with increasing L. For example, for our models, if $L = 8$ ($N = 512$ spins), a cluster contains 51 spins on average, which is about 10%. For a system with $L = 32$ ($N = 32768$ spins), a cluster contains 796 spins on average, which is about 2.5%. Note that at $T \ll T_c$, the relative average size of a cluster is the same for all systems at a fixed temperature.

5. FINITE-SIZE SCALING

The theory of finite-size scaling proposed by Ferdinand and Fisher in [21, 40] accounts for the influence of the finite size of the system on its critical properties. The ideas underlying this theory make it possible to extrapolate the results obtained for finite-size systems by the Monte Carlo method to the thermodynamic limit $N = L^3 \longrightarrow \infty$ and have found wide application [6, 8, 9, 13–16, 39]. According to this theory, the free energy of a sufficiently large system with periodic boundary conditions at a temperature T close to the critical temperature T_c of the infinite system can be represented in the form

$$F(T, L) \propto L^{-d} F_0(tL^{1/\nu}), \tag{3}$$

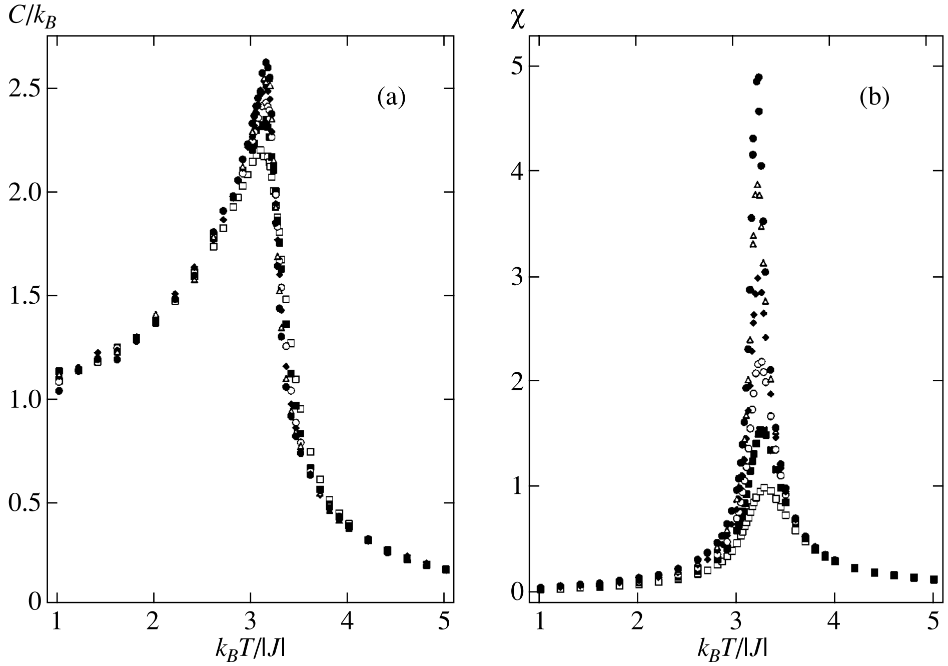


Fig. 1. Heat capacity (a) and susceptibility (b) as a function of temperature for the model Γ2 for $N = 512$ (□), 1000 (■), 1728 (○), 2744 (◆), 4096 (△), and 5832 (●).

where $t = |T - T_c|/T_c$, $T_c = T_c(L = \infty)$, and ν is the static critical index of the correlation radius of the infinite system ($L = \infty$). The shift of the "effective transition temperature" occurs in accordance with the relation

$$\frac{k_BT_c(L)}{J} = \frac{k_BT_c}{J} + aL^{-1/\nu}, \tag{4}$$

where a is a constant. Relation (3) yields similar relations of the heat capacity, susceptibility, and spontaneous magnetization per spin:

$$C(T, L) \propto L^{\alpha/\nu}C_0(tL^{1/\nu}), \tag{5}$$

$$\chi(T, L) \propto L^{\gamma/\nu}\chi_0(tL^{1/\nu}), \tag{6}$$

$$m(T, L) \propto L^{-\beta/\nu}m_0(tL^{1/\nu}), \tag{7}$$

where α, β, and γ are the static critical exponents for the system with $L = \infty$ satisfying the hyperscaling equation $2 - \alpha = d\nu = 2\beta + \gamma$ [1].

Equations (5)–(7) give an adequate description of the critical behavior of infinite systems for $t \ll 1$ as $L \longrightarrow \infty$. The theory of finite-size scaling was successfully applied to a number of simple well-known models (see [6]). However, the applicability of this theory to models involving crossovers was tested only on a very limited number of models [6, 8–10]. As far as we know, models that admit several crossover transitions have never been investigated by cluster Monte Carlo algorithms with the use of finite-size scaling theory.

We call the reader's attention to the following fact. When calculating the critical exponents of susceptibility (γ) and magnetization (β), the relations

$$\chi \propto L^{\gamma/\nu}, \tag{8}$$

$$m \propto L^{-\beta/\nu} \tag{9}$$

are used, which are derived from (6) and (7) at $T = T_c$. However, the heat capacity cannot be described following this scheme. For this reason, the following relation is practically used to scale the heat capacity in the calculation of the index α (see [6, 8–10]):

$$C_{\max}(L) = C_{\max}(L = \infty) - aL^{\alpha/\nu}, \tag{10}$$

where a is a coefficient.

Equation (4) is also of little use in practice because it yields an inaccurate value of T_c. The cumulant fourth-order method proposed by Binder [41] is much more accurate:

$$U_L = 1 - \frac{\langle m^4\rangle_L}{3\langle m^2\rangle_L^2}, \tag{11}$$

where m is the magnetization of the system of the linear size L. This method makes it possible to determine T_c up to a high accuracy. If the system undergoes a first-order phase transition, then the energy of the system E appears in (11) instead of m (see [6]). In the following section, we demonstrate a high efficiency of this method for models under study.

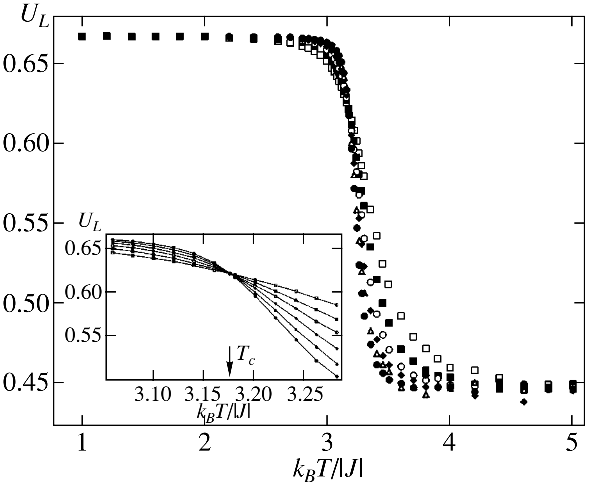


Fig. 2. The cumulant U_L as a function of temperature for the model Γ2. The notation is the same as in Fig. 1.

6. SIMULATION RESULTS

To observe the temperature dependence of the heat capacity and susceptibility, we use the fluctuation relations

$$C = (NK^2)(\langle U^2\rangle - \langle U\rangle^2), \tag{12}$$

$$\chi = (NK)(\langle m^2\rangle - \langle m\rangle^2), \tag{13}$$

where $K = |J|k_BT$.

Figure 1 presents the dependence of the heat capacity C and susceptibility χ on temperature for the more complex model Γ2. Here and in what follows, the error of the data does not exceed the size of the symbols used in the figures. We note the well-defined maxima for systems of all sizes and the fact that these maxima correspond to the same temperature up to the accuracy of computations. It is seen that no dependence of T_c on L, which would be expected from Eq. (4), is observed. This proves the effectiveness of the method used to add periodic boundary conditions and shows that many parameters under study are saturated with respect to N. Since Fig. 1 illustrates the data for the lower part of the investigated range of L, our results for greater values of L are not less accurate and reliable.

In order to determine the critical parameters on the basis of the finite-size scaling theory, one must first of all determine the critical temperature T_c; for this purpose, we used Binder's cumulant method. According to the finite-size scaling theory, all cumulants U_L determined by formula (11) for systems of various sizes L meet at the point T_c. Figure 2 depicts the dependence of U_L on temperature for the model Γ2. The inset in this figure demonstrates the accuracy with which the critical temperature can be determined. Similar computations were performed for the model Γ1. Qualitatively, all specific features characteristic of Γ2 and shown in Figs. 1 and 2 are characteristic for Γ1 as well.

The analysis of data for susceptibility and magnetization based on relations (8) and (9) allowed us to determine the exponents β and γ. For this purpose, we plotted the dependences m and χ on linear size L of the lattice on the log-log scale (Fig. 3). For the model Γ1, these plots are characterized by the ratios $\beta/\nu = 0.5081$ and $\gamma/\nu = 1.9716$. Assuming that (1) is the Heisenberg Hamiltonian and setting $\nu = 0.706$ [3, 4], we obtain $\beta = 0.359(1)$ and $\gamma = 1.392(1)$. These values of the exponents are in good agreement with the theoretical data obtained for the Heisenberg model: $\beta = 0.368$ and $\gamma = 1.390$ [3, 4].

Similar dependences for the model Γ2 yield the values $\beta/\nu = 0.5066$ and $\gamma/\nu = 1.9233$. Since the model Γ2 takes into account dipole–dipole interactions along with the exchange one, it is useful to determine the exponents both for $\nu = 0.706$ (the Heisenberg model) and for $\nu = 0.692$ (the three-dimensional dipole model [5]). Thus, $\beta = 0.358(1)$, $\gamma = 1.358(1)$ for $\nu = 0.706$ and $\beta = 0.351(1)$, $\gamma = 1.331(1)$ for $\nu = 0.692$. It is interesting to note that the value of γ for this model shifted in the direction of the value corresponding to the three-dimensional dipole model ($\gamma = 1.37$ [5]); at the same time, β remained practically unchanged. This feature is characteristic of the transition from the Heisenberg crit-

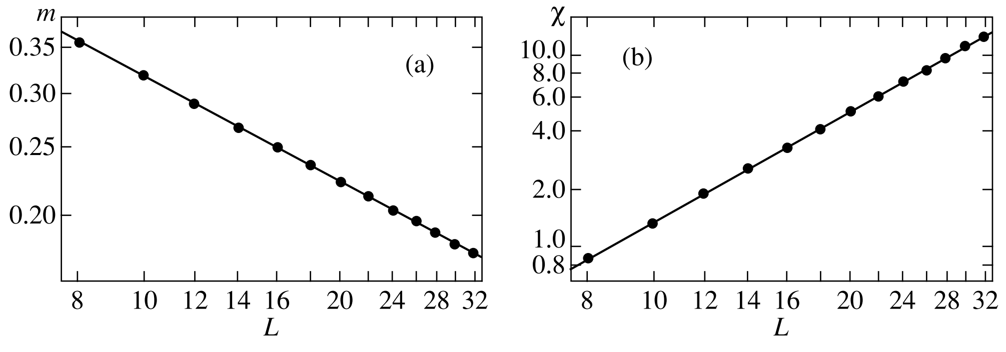


Fig. 3. Magnetization (a) and susceptibility (b) as a function of the linear size of the system for the model Γ1 at $T = T_c$: (a) $\beta/\nu = 0.50808$, $\beta = 0.35870$; (b) $\gamma/\nu = 1.97164$, $\gamma = 1.39198$.

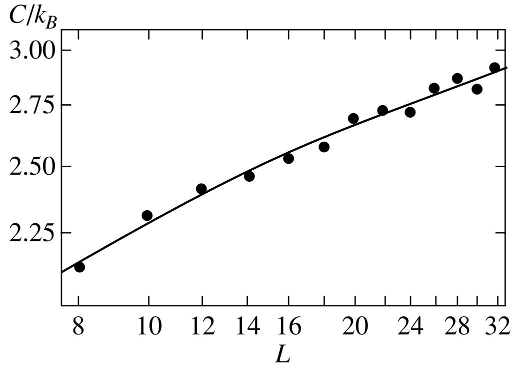


Fig. 4. Heat capacity as a function of the system's linear size for the model Γ1 at $T = T_c$: $\alpha/\nu = -0.16347$ and $\alpha = -0.11540$.

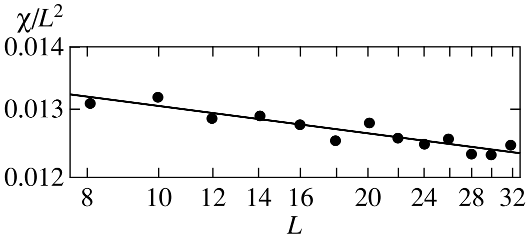


Fig. 5. χ/L^2 as a function of the system's linear size for the model Γ1 at $T = T_c$; $\eta = 0.04514$.

ical behavior with isotropic short-range forces to the three-dimensional dipole model.

As has already been noted, when processing heat capacity data, formula (10) rather than (5) is used in practice. Our data for heat capacity show that the dependence $C(L)$ on the log-log scale is nonlinear (Fig. 4). The approximation performed by the nonlinear least-square method on the basis of Eq. (10) yields the value of the critical index $\alpha = -0.115(2)$ for the model Γ1 and $\alpha = -0.120(2)$ for the model Γ2. The recalculation of α for the model Γ2 at $\nu = 0.692$ (the dipole model) yields $\alpha = -0.118(2)$, which coincides (within the accuracy of computations) with the value obtained for $\nu = 0.706$.

Sometimes, the critical index γ is determined using the value $\chi_{\max}$ instead of $\chi(T_c)$. The values of γ determined in such a way for the models Γ1 and Γ2 are equal to 1.379(1) and 1.366(1), respectively.

Our data can also be used to determine the Fisher index η. Using the relation between the susceptibility χ and the correlation radius ξ (see [16]),

$$\chi \propto \xi^{\gamma/\nu}, \tag{14}$$

and the relation between the critical exponents $\eta = 2 - \gamma/\nu$, we obtain

$$\ln(\chi/\xi^2) = C - \eta \ln \xi, \tag{15}$$

where C is a constant.

For systems of a finite size $\xi = L$, we obtain at $T = T_c$

$$\ln(\chi/L^2) = C - \eta \ln L. \tag{16}$$

The dependence of χ/L^2 on L for the model Γ1 is shown in Fig. 5. According to these data, the value of the Fisher index is $\eta = 0.045(5)$. It is seen from Fig. 5 that the values obtained are well fitted by a line for small L, which is also confirmed in other studies [16]. It is possible that the asymptotic critical regime for the correlation radius ξ and susceptibility χ is attained already at $L \geq 8$. The value of η determined above is in a rather good agreement with the data obtained on the basis of field theory ($\eta \approx 0.033$–0.038 [3, 4]) and by Monte Carlo methods ($\eta \approx 0.027$ [6, 16]). For the model Γ2, we found that $\eta = 0.048(5)$ coincides with the value for Γ1 within the accuracy computations.

7. CONCLUSIONS

The investigation results of complex models of real gadolinium obtained on the basis of the one-cluster Wolf algorithm of the Monte Carlo method show that this algorithm is very efficient. The use of this algorithm in the vicinity of the critical point reduces the computation time at least by an order of magnitude compared to the classical one-spin Metropolis algorithm while providing the same accuracy. This is especially important for the investigation of complex models in which crossover effects manifest themselves only when a sufficiently rich statistics is accumulated and highly accurate results are obtained.

Note that the isotropic dipole–dipole interactions taken into account in the model Γ2 are only a weak perturbation factor on the background of strong exchange interactions. No investigations of the influence of such forces on the critical behavior with a simultaneous account for another weak perturbation factor—the uniaxial anisotropy—were earlier conducted by cluster Monte Carlo algorithms. The effectiveness, efficiency, and resolution of cluster algorithms in revealing these factors have not been studied earlier. For this reason and in order to compare the results with those obtained by classical algorithms, all experiments were conducted strictly following a unified methodology. Specific features revealed in the analysis of the model Γ2 demonstrate a high resolution of cluster Monte Carlo algorithms.

The basic static critical exponents

$$\alpha = -0.115(1), \quad \beta = 0.359(1),$$
$$\gamma = 1.392(1), \quad \eta = 0.045(5),$$

calculated on the basis of the data obtained by the Monte Carlo method with the use of the finite-size scaling theory for the model Γ1 coincide (within the accuracy of computations) with the values predicted for the three-dimensional isotropic Heisenberg model with short-range forces; they also coincide with the values that were obtained by Monte Carlo methods for this model.

For the model Γ2, the following set of index values was obtained:

$$\alpha = -0.120(1), \quad \beta = 0.368(1),$$
$$\gamma = 1.358(1), \quad \eta = 0.048(5).$$

The comparison of the exponents for the models Γ1 and Γ2 shows that changes greater than the accuracy of computations occurred only in the index γ; for Γ2, γ shifts closer to the value theoretically predicted for the dipole model (in fact, γ coincides with the theoretical value within the accuracy of computations). A similar decrease of γ is characteristic of the change of the critical behavior from Heisenberg to dipole.

Our data also confirm that the finite-size scaling theory can be used for the investigation of complex models involving crossovers.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project no. 01-02-16103, and by the Commission of the Academy of Sciences for Supporting Young Scientists.

REFERENCES

1. A. Z. Patashinskiĭ and V. A. Pokrovskiĭ, *Fluctuation Theory of Phase Transitions* (Nauka, Moscow, 1982, 2nd ed.; Pergamon, Oxford, 1979).
2. S. Ma, *Modern Theory of Critical Phenomena* (Benjamin, Reading, 1976; Mir, Moscow, 1980).
3. J. J. C. Le Guillou and J. J. Zinn-Justin, Phys. Lett. **46**, L157 (1985).
4. S. A. Antonenko and A. I. Sokolov, Phys. Rev. E **51**, 1894 (1995).
5. I. K. Kamilov and Kh. K. Aliev, *Static Critical Phenomena in Magnetic-Ordered Crystals* (Dagest. Nauchn. Tsentr Ross. Akad. Nauk, Makhachkala, 1993).
6. I. K. Kamilov, A. K. Murtazaev, and Kh. K. Aliev, Usp. Fiz. Nauk **169**, 773 (1999).
7. A. K. Murtazaev, I. K. Kamilov, Kh. K. Aliev, and V. A. Mutaĭlamov, Zh. Éksp. Teor. Fiz. **117**, 559 (2000) [JETP **90**, 488 (2000)].
8. A. K. Murtazaev, Fiz. Nizk. Temp. **25**, 469 (1999) [Low Temp. Phys. **25**, 344 (1999)].
9. A. K. Murtazaev, I. K. Kamilov, and Kh. K. Aliev, J. Magn. Magn. Mater. **204**, 151 (1999).
10. A. K. Murtazaev, I. K. Kamilov, and K. Sh. Khizriev, Fiz. Tverd. Tela (St. Petersburg) **43**, 659 (2001) [Phys. Solid State **43**, 685 (1998)].
11. V. V. Prudnikov, S. V. Belim, E. V. Osintsev, and A. A. Fedorenko, Zh. Éksp. Teor. Fiz. **114**, 972 (1998) [JETP **87**, 527 (1998)].
12. G. A. Martynov, Usp. Fiz. Nauk **169**, 600 (1999).
13. D. P. Landau, Physica A (Amsterdam) **205**, 41 (1994).
14. A. M. Ferrenberg and D. P. Landau, Phys. Rev. B **44**, 5081 (1991).
15. K. Chen, A. M. Ferrenberg, and D. P. Landau, Phys. Rev. B **48**, 3249 (1993).
16. Ch. Holm and W. Janke, Phys. Rev. B **48**, 936 (1993).
17. U. Wolff, Phys. Rev. Lett. **62**, 361 (1989).
18. U. Wolff, Nucl. Phys. B **322**, 759 (1989).
19. A. M. Ferrenberg and R. N. Swendsen, Phys. Rev. Lett. **61**, 2635 (1988).
20. A. M. Ferrenberg and R. N. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989).
21. A. E. Ferdinand and M. E. Fisher, Phys. Rev. E **185**, 832 (1969).
22. A. L. Tseskis, Zh. Éksp. Teor. Fiz. **106**, 1089 (1994) [JETP **79**, 591 (1994)].
23. K. P. Belov, M. A. Belyanchikova, R. Z. Levitin, and S. A. Nikitin, *Rare-Earth Ferromagnetics and Antiferromagnetics* (Nauka, Moscow, 1965).
24. V. M. Kuchin, V. A. Somenko, S. Sh. Shil'shteĭn, and Yu. B. Patrikiev, Zh. Éksp. Teor. Fiz. **55**, 1241 (1968) [Sov. Phys. JETP **28**, 649 (1969)].
25. R. H. Child, Phys. Rev. B **18**, 1247 (1978).
26. Kh. K. Aliev, I. K. Kamilov, and O. M. Omarov, Zh. Éksp. Teor. Fiz. **94** (11), 153 (1988) [Sov. Phys. JETP **67**, 2262 (1988)].
27. C. Bednarz, D. J. W. Geldart, and Mary Anne White, Phys. Rev. B **47**, 14247 (1993).
28. D. A. Doleisi and S. A. Swenson, Phys. Rev. B **24**, 6326 (1981).
29. E. A. Lewis, Phys. Rev. B **1**, 4368 (1970).
30. D. S. Simons and M. B. Salamon, Phys. Rev. B **10**, 4680 (1974).
31. G. H. J. Wantenaar, S. L. Compbell, and D. N. Chaplin, Phys. Rev. B **29**, 1419 (1984).
32. P. Molho and J. L. Portosseill, J. Magn. Magn. Mater. **31-34**, 1023 (1983).
33. A. J. Saleh and N. H. Saunders, J. Magn. Magn. Mater. **29**, 197 (1982).
34. P. Heller, Rep. Prog. Phys. **30**, 731 (1967).
35. A. R. Chowdhury, C. S. Collins, and Ch. Hohenemser, Phys. Rev. B **33**, 6231 (1986).
36. S. V. Vonsovskiĭ, *Magnetism* (Nauka, Moscow, 1971; Wiley, New York, 1974).
37. K. P. Belov, A. K. Zvezdin, A. M. Kadomtseva, and R. Z. Levitin, *Orientational Transitions in Rare-Earth Magnets* (Nauka, Moscow, 1979).
38. O. A. Vasil'ev and L. N. Shchur, Zh. Éksp. Teor. Fiz. **117**, 1110 (2000) [JETP **90**, 964 (2000)].
39. W. Janke and K. Nather, Phys. Rev. B **48**, 7419 (1993).
40. M. E. Fisher and M. N. Barber, Phys. Rev. Lett. **28**, 1516 (1972).
41. K. Binder, Phys. Rev. Lett. **47**, 693 (1981).

Translated by A. Klimontovich