THEORETICAL AND MATHEMATICAL PHYSICS

The Theory of Planar Bragg Resonators

P. V. Petrov

All-Russia Research Institute of Technical Physics, Russian Federal Nuclear Center,
Snezhinsk, Chelyabinsk oblast, 456770 Russia
e-mail: p.v.petrov@vniitf.ru
Received February 12, 2001; in final form, June 8, 2001

Abstract—The reflecting properties of one-dimensional planar Bragg gratings are studied. A coupled resonator model for studying the diffraction of electromagnetic waves in an arbitrarily corrugated waveguide is suggested. It is based on exact relationships that follow from the two-dimensional boundary-value problem stated in terms of the Helmholtz equation. The specific relationships for the rectangular corrugation of the grating-forming plates are presented. The reflection coefficients of the Bragg gratings vs. corrugation length and incident radiation frequency are calculated. An analytical solution for the "narrow" corrugation is obtained. © 2002 MAIK "Nauka/Interperiodica".

INTRODUCTION

The use of Bragg resonators, which provide distributed feedback and the spatial coherency of radiation, seems to be a promising approach to implementing electrodynamic systems for free-electron masers [1–4]. Bragg resonators are waveguides with a single- or a double-periodic corrugation (Bragg gratings) [1, 5].

Mathematical models [1–5] used for describing the spectra of the modes and reflection coefficients in Bragg gratings are based on the coupled mode theory, within which the mode coupling coefficients are found by the perturbation method [6, 7]. Namely, the waveguide surface deformation (corrugation) defined by some function $l(\mathbf{r})$ is replaced by the "equivalent" boundary condition on the undisturbed surface of a regular waveguide [7]:

$$\mathbf{E}_\tau = \nabla(l(\mathbf{r})\mathbf{E} \cdot \mathbf{n}) - i\frac{\omega}{c}l(\mathbf{r})[\mathbf{n}, \mathbf{H}], \qquad (1)$$

where $\mathbf{n}$ is the unit vector of the outer normal to the waveguide surface; $\mathbf{E}$, $\mathbf{H} \sim \exp(-i\omega t)$ are the electric and magnetic fields; ω is the radiation circular frequency; and ∇ is the Hamiltonian operator. The parenthesized and bracketed vectors mean the scalar product and the vector product, respectively.

Approximation (1) applies [7] if (i) the deformation is small (the corrugation height is much smaller than the wavelength of the incident radiation and the corrugation length) and (ii) the angle between the deformed and virgin surfaces is small. As a rule, waveguides used in experiments with Bragg gratings have a rectangular corrugation [1–5]. In this case, the deformation is usually small, while the angle is not. Even for a sinusoidal corrugation on the surface, which is widely used in numerical simulation, the angle between the surfaces is $\pi/4$ and can hardly be considered small. Because of this, it would be appropriate to develop a physicomath-

ematical model that adequately describes the diffraction of electromagnetic waves by one- and two-dimensional Bragg gratings without invoking the perturbation theory, has an accuracy close to that of direct numerical calculations, and is free of awkward mathematics.

In this work, we suggest a coupled resonator model for the analysis of the electromagnetic fields in one-dimensional planar Bragg gratings and report their reflecting properties.

EQUATION FOR MAGNETIC FIELD

Consider a hollow planar waveguide with perfectly conducting walls that is infinitely long in the $0Y$ direction and symmetric about the plane $0X$. Its undisturbed (regular) boundary is defined by the function $x = \pm L_x$. Let a finite number M of corrugations be patterned on the side walls of the waveguide in the interval $0 < z < L$. The corrugations are also symmetric about the plane $0X$. Let the contour of the qth corrugation be described by the function $x = l_q(z)$ (Fig. 1)

$$x = \begin{cases} l_q(z), & z \in [a_{q,1}, a_{q,2}] \\ 0, & z < a_{q,1}, \quad z > a_{q,2}, \end{cases} \quad q = 1, M. \qquad (2)$$

We assume that the Bragg grating is excited by a natural mode of the regular waveguide from the side of negative z (i.e., from left to right). The mode has a longitudinal wave number h_p and a transverse wave number $g_p = \sqrt{k^2 - h_p^2}$. Let the *TM* modes of the waveguide be excited (for them, only the components E_x, E_z, and H_y are other than zero). Then, the initial problem of finding the electromagnetic fields is reduced to the boundary-value problem that, in terms of the Helm-
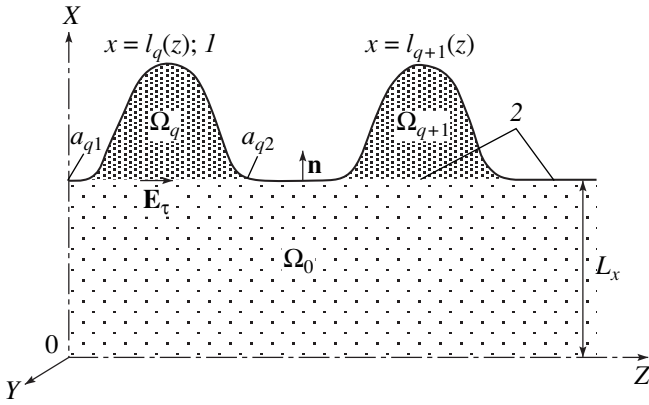
Fig. 1. Planar Bragg grating: *1*, contour of the *q*th corrugation; *2*, contour of the undisturbed (regular) waveguide.

holtz scalar equation for the magnetic field $H \equiv H_y(x, z)$, is stated as

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} + k^2\right)H = 0 \qquad (3)$$

in the domain

$$(x, z) \in \Omega = \bigcup_{q=1}^{M} \Omega_q \bigcup \Omega_0,$$

where $\Omega_0 = \{0 < x < L_x, -\infty < z < \infty\}$ is the region of the regular waveguide and $\Omega_q = \{L_x < x < l_q(z), 0 < x < L\}$ is the region of the *q*th corrugation ($q = 1, \ldots, M$). The boundary conditions are as follows: the Neumann homogeneous condition

$$\left.\frac{\partial H}{\partial \mathbf{n}}\right|_{S = \bigcup_q \{x = l_q(z)\} \bigcup \{x = 0\}} = 0 \qquad (4)$$

on the metallic surface of the grating and on the plane of symmetry 0Z

$$\frac{\partial H}{\partial z} + ih_p H = 2ih_p \cos(g_p x) \qquad (5)$$

at $z = 0$ if the grating is excited by the *TM* mode, and

$$\frac{\partial H}{\partial z} - ih_v H = 0, \quad h_v = \sqrt{k^2 - g_v^2},$$

$$g_v = \frac{\pi}{L_x}v; \quad v = 0, 1, \ldots \qquad (6)$$

at $z = L$ if the modes can freely leave the system.

The electric field components are given by the equations

$$E_z = -\frac{1}{ik}\frac{\partial H}{\partial x}, \quad E_x = \frac{1}{ik}\frac{\partial H}{\partial z}. \qquad (7)$$

The solutions to Eqs. (7) under boundary conditions (4)–(6) will be sought separately for the regular waveguide region Ω_0 and for each of the corrugations

Ω_q ($q = 1, M$) (in Ω_0, the magnetic field can be represented as the sum of the incident and scattered fields); that is,

$$H = \begin{cases} H_0 + H_s, & \mathbf{r} \in \Omega_0 \\ H_q, & \mathbf{r} \in \Omega_q, \end{cases} \qquad (8)$$

where H_0 meets boundary conditions (4)–(6).

At the boundary $x = L_x$ with the regular waveguide, the tangential field components are continuous,

$$H_q(z) = H_0(z) + H_s(z) \qquad (9)$$

and the scattered field H_s meets homogeneous boundary conditions (4) and the condition that the modes leave the system at $z = 0, L$.

At the interface between the corrugation and the regular waveguide, the electric field component along the 0Z axis is assumed to be some function $E_z(x = L_x, y, z) = E_\tau(z)$. Then, in the region Ω_0, H_s is obtained by solving the boundary-value problem

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} + k^2\right)H_s = 0, \qquad (10)$$

$$\left[\pm\frac{\partial H_s}{\partial z} + ih_v H_s\right]\bigg|_{z = 0, L} = 0 \quad (v = 0, 1, 2, \ldots),$$

$$\left.\frac{\partial H_s}{\partial x}\right|_{x = 0} = 0, \quad \left.\frac{\partial H_s}{\partial x}\right|_{x = L_x} = -ikE_\tau(z). \qquad (11)$$

Its solution can be represented through the Green function [8]:

$$H_s(x, z) = -\frac{ik}{4\pi}\int_0^L dz' G_k(x, z, x' = L_x, z')E_\tau(z'), \qquad (12)$$

$$G_k(x, z, x', z') = \frac{2\pi i}{L_x}\sum_{v=0}^{\infty}\frac{\varepsilon_v}{h_v}\cos(g_v x)\cos(g_v x')$$

$$\times \exp(ih_v|z - z'|), \qquad (13)$$

$$g_v = \frac{\pi v}{L_x}, \quad h_v = \sqrt{k^2 - g_v^2}, \quad \varepsilon_v = \begin{cases} 2, & v = 0 \\ 1, & v \neq 0. \end{cases}$$

On the surface of the regular waveguide, expression (12) for $H_s^s(z) \equiv H_s(x = L_x, z)$ can be written in the compact operator form

$$H_s^s = -ik\hat{G}_R E_\tau, \qquad (14)$$

$$\hat{G}_R = \frac{i}{2L_x}\sum_{v=0}^{\infty}\frac{\varepsilon_v}{h_v}\int_0^L dz' \exp(ih_v|z - z'|). \qquad (15)$$

Let us find a relation between the electric field $E_\tau(z)$ and the magnetic field H_q at the interface between the

qth corrugation region Ω_q and the regular waveguide: $x = L_x$, $a_{q,1} \leq z \leq a_{q,2}$. Irrespective of the corrugation shape, the magnetic field H_q inside the region Ω_q is the solution of Helmholtz equation (3) with Neumann homogeneous conditions (4) on the metallic surface $x = l_q(z)$, $a_{q,1} \leq z \leq a_{q,2}$ of the corrugation and must obey relationship (9) at the boundary with the regular waveguide. Using the Green function [8] for determining $H_q(x, z)$ through the field value at the boundary $S_q = \{x = l_q(z)\} \cup \{x = L_x\}$, $a_{q,1} \leq a_{q,2}$ between the qth corrugation region Ω_q and the regular waveguide, we find

$$
\begin{aligned}
H_q(x, z) &= \oint_{S_q} d\mathbf{S}_q H(\mathbf{r}_s') G_q(x, z, \mathbf{r}_s') \\
&= \int_{a_{q,1}}^{a_{q,2}} dz' (H_0^s(z') + H_s^s(z')) G_q(x, z, z'),
\end{aligned}
\tag{16}
$$

where G_q is the surface Green function for the Helmholtz equation in the region Ω_q. G_q satisfies homogeneous Neumann conditions (4) on the metallic surface $\{x = l_q(z), a_{q,1} \leq z \leq a_{q,2}\}$ and also the inhomogeneous Dirichlet conditions at the boundary $x = L_x$ with the regular waveguide.

In view of Eqs. (7), we obtain from (16) the expression for $E_z^q(x, z)$ in the qth corrugation region:

$$
\begin{aligned}
E_z^q(x, z) &= -\frac{1}{ik} \frac{\partial H_q(x, z)}{\partial x} \\
&= -\frac{1}{ik} \int_{a_{q,1}}^{a_{q,2}} dz' (H_0^s(z') + H_s^s(z')) \frac{\partial}{\partial x} G_q(x, z, z').
\end{aligned}
\tag{17}
$$

Accordingly, at the interface between the qth corrugation and the regular waveguide, we have $E_\tau^q(z) = E_\tau(z) = E_z(x = L_x, z)$ ($z \in [a_{q,1}, a_{q,2}]$) or, in the operator form,

$$
E_\tau^q = -\frac{1}{ik} \hat{G}_{in}^q (H_0^s + H_s^s).
\tag{18}
$$

Here, $\hat{G}_{in}^q$ is the conjugation operator for the electric and magnetic fields in the qth corrugation:

$$
\begin{aligned}
\hat{G}_{in}^q &= \int_{a_{q,1}}^{a_{q,2}} dz' \frac{\partial}{\partial x} G_q(x, z, z')\big|_{x = L_x}, \\
&\quad z, z' \in [a_{q,1}, a_{q,2}].
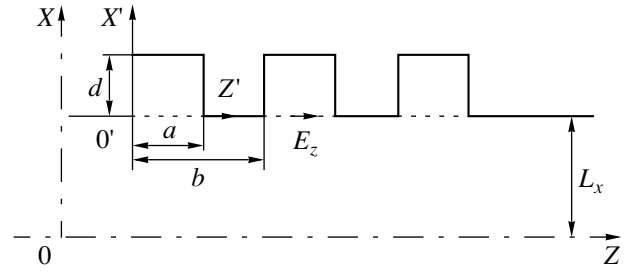\end{aligned}
\tag{19}
$$

Fig. 2. Periodic rectangular corrugation in a planar waveguide.

Taking into account the corrugation geometry, one can express $E_\tau(z)$ on the regular waveguide surface $z \in [0, L]$ through the Heaviside function $\eta(x) = \{0, x < 0; 1, x > 0\}$ in the operator form

$$
\begin{aligned}
E_\tau(z) &= -\frac{1}{ik} \hat{G}_{in} (H_0^s + H_s^s), \\
\hat{G}_{in} &= \sum_{q=1}^{M} (\eta(z - a_{q,1}) - \eta(z - a_{q,2})) \hat{G}_{in}^q.
\end{aligned}
\tag{20}
$$

Using relationships (14) and (20), we come to the Fredholm integral equation of the first kind for $H_s^s(z)$ on the regular waveguide surface:

$$
H_s^s - \hat{G}_R \hat{G}_{in} H_s^s = \hat{G}_R \hat{G}_{in} H_0^s.
\tag{21}
$$

Its solution has the form

$$
H_s^s = (\hat{I} - \hat{G}_R \hat{G}_{in})^{-1} \hat{G}_R \hat{G}_{in} H_0^s.
\tag{22}
$$

Then, knowing $H_s^s(z)$, one can find $E_\tau(z)$ by formula (20) and determine the scattered magnetic field throughout the waveguide with (12).

CONJUGATION OPERATOR FOR RECTANGULAR CORRUGATION

The operator $\hat{G}_{in}$ specifies the dependence of the scattered radiation parameters on the corrugation shape and on the corrugation distribution over the Bragg lattice length. The form of $\hat{G}_{in}$ is the simplest for a rectangular corrugation (Fig. 2). Let it have a period b.

We introduce a local coordinate system $X'0'Z'$ for the initial rectangular corrugation (Fig. 2). In the rectangular domain of width a and height d subject to the homogeneous Neumann conditions on the metallic walls of the corrugation ($z' = [0, a]$; $x' = d$) and the inhomogeneous Dirichlet condition on the interface with the regular waveguide ($x' = 0$), the Green function for

Helmholtz equation (3) has the form [8]

$$G_q(x, z, z') = \frac{2}{a}\sum_n \cos(\chi_n z)\cos(\chi_n z')\frac{\sinh\beta_n(x-d)}{\varepsilon_n\cosh\beta_n d}, \quad (23)$$

$$\chi_n = \frac{\pi n}{a}, \quad \beta_n = \sqrt{\chi_n^2 - k^2}.$$

According to (19), we obtain

$$\hat{G}_{in}^q(z, z')$$
$$= -\frac{2}{a}\sum_n \frac{\beta_n}{\varepsilon_n}\tanh(\beta_n d)\cos(\chi_n z)\int_0^a dz'\cos(\chi_n z'), \quad (24)$$

$$\hat{G}_{in}(z, z') = -\frac{2}{a}\sum_{q=0}^{M-1}[\eta(z_q) - \eta(z_q - a)]$$
$$\times\left(\sum_n \gamma_n\cos(\chi_n z_q)\int_{z_q}^{z_q+a} dz'\cos(\chi_n(z'-bd))\right), \quad (25)$$

$$\gamma_n = \frac{\beta_n}{\varepsilon_n}\tanh(\beta_n d), \quad z_q = z - qb.$$

The explicit form of the operators $\hat{G}_R$ and $\hat{G}_{in}$ [see (15) and (25)] makes it possible to numerically obtain $H_s^s(z)$ from (21) and to determine the reflection coefficients of the grating:

$$R = \frac{1}{S_0}\int_0^{L_x} dx S_z(x, z_0), \quad (26)$$

$$S_z(x, z_0) = E_x(x, z_0)H_y^*(x, z_0) = \frac{1}{ik}H_y^*\frac{\partial H_y}{\partial z}(x, z_0).$$

Here, $H_y(x, z_0)$ is determined from (12) for points with the coordinate z_0 outside the corrugation, $E_\tau(z)$ is calculated by (20), and S_0 is the electromagnetic wave energy flux incident on the Bragg grating.

ANALYTICAL SOLUTION FOR ONE-MODE SCATTERING BY NARROW CORRUGATION

Generally, Eq. (21) is solved numerically; however, if a corrugation is narrow (its width a is much smaller than the incident wavelength λ), this equation is analytically solvable. In this case, one can leave only the first term in expression (25) for $\hat{G}_{in}$. Then, relationship (18) between the electric and magnetic fields on the surfaces of the corrugation and the regular waveguide is reduced to

$$E_z^q(z) = \frac{i}{a}\tan(kd)\int_{bq}^{bq+a} dz'(H_0^s(z') + H_s^s(z')). \quad (27)$$

In the infinitely narrow corrugation limit ($a \longrightarrow 0$), (27) transforms into the boundary condition for the impedances

$$E_\tau^q = i\tan(kq)(H_0^s(bq) + H_s^s(bq)), \quad (28)$$

which was used in the analysis of surface waves over an interdigital structure [9], and the operator $\hat{G}_{in}$ degenerates into the sum of delta functions:

$$\hat{G}_{in} = -ka\tan(kd)\sum_{q=0}^{M-1}\delta(z-qb). \quad (29)$$

We will restrict our analysis by the mutual scattering of the *TEM* modes (one-mode scattering): $\nu = 0$, $H_0^s = \exp(ikz)$. Then, Eq. (21) takes the form

$$H_s^s(z) = i\beta\sum_{q=0}^{M-1}\int_0^L dz'\exp(ik|z-z'|)$$
$$\times(H_s^s(z') + \exp(ikz'))\delta(z'-qb), \quad (30)$$

where the dimensionless parameter

$$\beta = \frac{a}{2L_x}\tan(kd) \quad (31)$$

defines the dependence of the solution on the parameters of the corrugation and the waveguide. For the specific case $h_p = k = \pi/b$, which corresponds to the Bragg resonance condition [1]

$$h_p + h_0 = h, \quad (32)$$

the solution of (30) can be written in the analytical form

$$H_s^s(z) = \frac{i\beta}{1-i\beta M}\sum_{q=0}^{M-1}(-1)^q\exp(ik|z-qb|). \quad (33)$$

Accordingly, the reflection coefficients becomes

$$R = \frac{\beta^2 M^2}{1 + \beta^2 M^2}. \quad (34)$$

Within such an approximation, we can find the resonance region. Here, it is appropriate to pass from integral equation (30) to the second-order differential equation [8]. To determine the resonance region, it is sufficient to consider the homogeneous equation for a grating with the infinite number of corrugations:

$$\frac{d^2 H_s^s(z)}{dz^2} + (k^2 + F(z))H_s^s(z) = 0,$$
$$\quad (35)$$
$$F(z) = 2k\beta\sum_{q=-\infty}^{\infty}\delta(z-qb).$$

The function $F(z)$ has a period b and can be expanded into a Fourier series. Passing to the variable $x = k_0 z$, we obtain

$$\frac{d^2 H_s^s}{dx^2} + \left(\frac{k^2}{k_0^2} + 2\frac{\beta k}{\pi k_0} \right.$$

$$\left. \times (1 + 2\cos(2x) + 2\cos(4x) + \ldots) \right) H_s^s = 0. \tag{36}$$

It is known that Eq. (36) describes parametric resonance [10]. If one considers only the first (the most intense) resonance, which takes place at $k/k_0 \simeq 1$ [10], and ignores higher order resonances, Eq. (36) is reduced to the Mathieu equation [11]

$$\frac{d^2 H_s^s}{dx^2} + (s(k) - 2p(k)\cos(2x))H_s^s = 0,$$

$$s(k) = \frac{k^2}{k_0^2} + 2\frac{\beta k}{\pi k_0}; \quad p(k) = -2\frac{\beta k}{\pi k_0}. \tag{37}$$

The conditions for parametric resonance correspond to the instability regions of the Mathieu functions and, for $p(k) \leq 0$, lie between the curves [11]

$$s_{2r+1}(p) \leq s \leq c_{2r+1}(p). \tag{38}$$

Here, s_r and c_r are the eigenvalues of the Mathieu equation that correspond to the even and odd solutions at a fixed p. For the first resonance $r = 0$ at $k \approx k_0$, the functions $s_1(p)$ and $c_1(p)$ can be represented as the power series in p [11]:

$$s_1(p) = 1 + p - \frac{p^2}{8} + \ldots, \quad c_1(p) = 1 - p - \frac{p^2}{8} + \ldots. \tag{39}$$

Substituting expressions (39) with the functions $s(k)$, $p(k)$ from (37) into (38), we find the range of k for which the parametric resonance arises. Up to the terms of the order of β, we have

$$k_0\left(1 - \frac{2}{\pi}\beta\right) \leq k \leq k_0. \tag{40}$$

Thus, Bragg resonance condition (32), used in [1–5], sets the resonance position only approximately. This is corroborated by calculating the coefficient of electromagnetic wave reflection [see (22) for the one-mode regime] as a function of the wave vector for narrow-corrugation Bragg gratings ($L_x = 0.5$ cm, $a = 0.01$ cm, $d = 0.09$ cm, and $\beta = 6.3 \times 10^{-2}$) of length $L = 5$ and 10 cm shown in Fig. 3. Here, the reflection coefficient obtained from solution (30) with $\beta \cong 6.3 \times 10^{-2}$ for the delta corrugation is also shown.
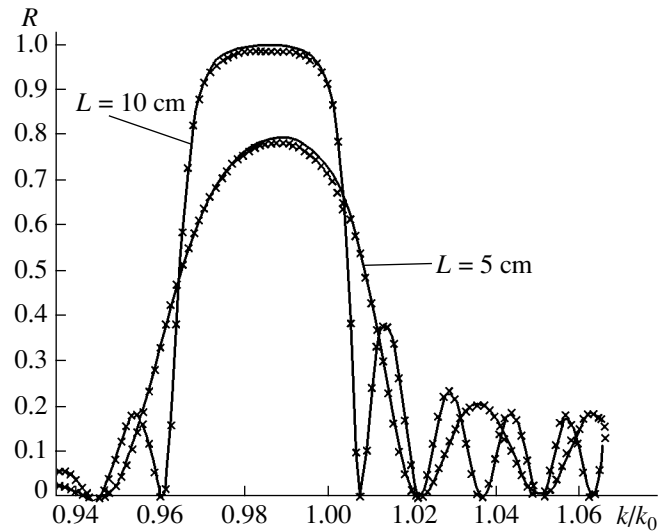
Fig. 3. Reflection coefficient vs. wave vector for the single-mode scattering of electromagnetic wave from narrow-corrugation ($a = 0.01$ cm, $d = 0.09$ cm) Bragg gratings with various lengths. Continuous curve refers to Eq. (20) with $v = 0$, $a = 0.02$ cm, and $d = 0.09$ cm; $\times$, to Eq. (29) with $\beta = 6.3 \times 10^{-2}$.
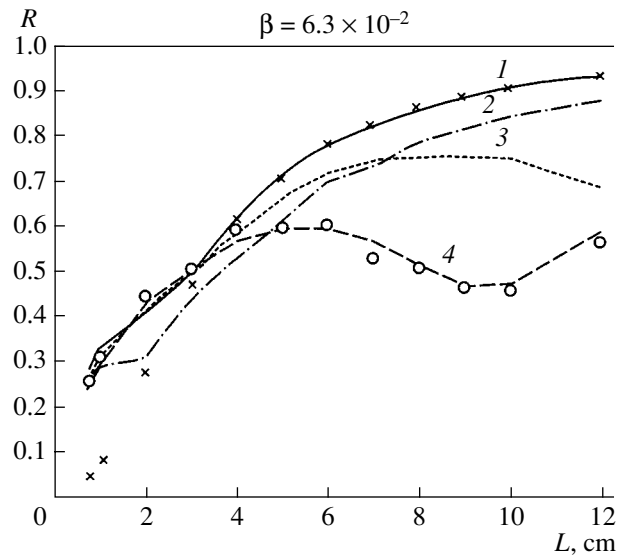

Fig. 4. *TEM* mode ($v = 75$ GHz) reflection coefficient vs. length of the one-dimensional Bragg grating with a period 0.2 cm for various corrugation depths and widths ($\beta = 6.3 \times 10^{-2}$). (*1*) $a = 0.004$ cm, $d = 0.096$ cm; (*2*) $a = 0.01$ cm, $d = 0.09$ cm; (*3*) $a = 0.04$ cm, $d = 0.064$ cm; and (*4*) $a = 0.1$ cm, $d = 0.036$ cm. $\times$, analytical representation (32); $\circ$, finite-element method ($a = 0.1$ cm, $d = 0.036$ cm).

REFLECTION COEFFICIENTS OF ONE-DIMENSIONAL BRAGG GRATINGS

The practical use of Bragg gratings is usually associated with their reflecting properties. Of most interest is the dependence of the reflection coefficient on the corrugation length and incident radiation frequency for
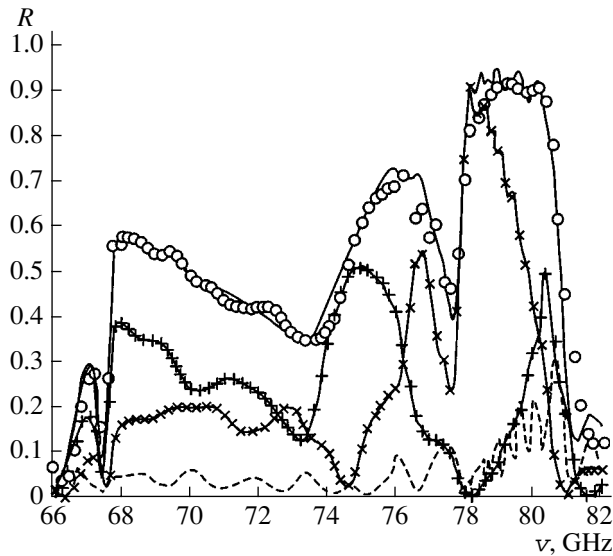
Fig. 5. Frequency dependence of the *TEM* mode reflection coefficient for the one-dimensional Bragg grating of length $L = 5$ cm with a period 0.2 cm and a rectangular corrugation ($a = 0.01$ cm, $d = 0.09$ cm). Continuous curve, R_{tot}; dashed curve, $R_{TEM_{02}}$; $\circ$, R_{tot} obtained by finite-element method; $+$, R_{TEM}; and $\times$, $R_{TEM_{04}}$.
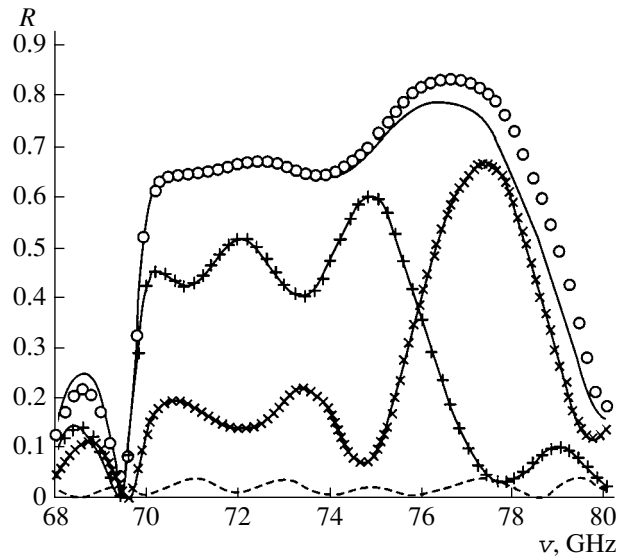
Fig. 6. Frequency dependence of the *TEM* mode reflection coefficient for the one-dimensional Bragg grating of length $L = 5$ cm with a period 0.2 cm and a rectangular corrugation ($a = 0.1$ cm, $d = 0.03$ cm). Designations are the same as in Fig. 5.

given corrugation period and depth. We determined the reflecting properties of one-dimensional Bragg gratings for the case of a planar waveguide with the plate spacing $2L_x = 1$ cm and a rectangular corrugation with a period of 0.2 cm.

Figure 4 depicts the reflection coefficient vs. grating length for the wave frequency $v = 75$ GHz and various corrugation width and depth. The parameter β obtained by numerically solving Eq. (21) with the operator $\hat{G}_{in}$

in form (25) was kept the same, $\beta \cong 6.3 \times 10^{-2}$. For comparison, the reflection coefficients for an infinitely narrow corrugation [see (34)] and those calculated by the finite-element method [12] are also shown.

It is seen that formula (34) describes the dependence of the reflection coefficients R on the corrugation length with a fairly good accuracy if the gratings are long. For the short gratings, the discrepancy between the values calculated and those predicted by the analytical for-

Mode composition of the reflected radiation when the *TEM* mode of the planar waveguide with $L_x = 0.5$ cm strikes the Bragg grating of length L

L, cm	Mode composition, %, for narrow corrugation: $a = 0.01$ cm, $d = 0.09$ cm			Mode composition, %, for wide corrugation: $a = 0.1$ cm, $d = 0.36$ cm		
	TEM	TEM_{02}	TEM_{04}	*TEM*	TEM_{02}	TEM_{04}
0.8	30.1	60.3	9.7	28.6	61.0	10.4
1	33.5	64.4	2.2	31.8	65.7	2.4
2	46.4	46.3	7.2	36.3	58.0	5.6
3	62.5	32.5	5.0	49.9	43.0	7.1
4	75.5	23.8	0.7	66.5	28.3	5.2
5	84.4	15.4	0.2	80.2	16.9	3.0
6	89.9	9.0	1.1	87.9	10.0	2.1
7	92.7	6.2	1.1	87.9	10.7	1.4
8	94.2	5.5	0.2	76.6	22.7	0.7
9	95.6	4.3	0.0	55.2	43.7	1.1
10	96.9	2.8	0.3	45.0	53.1	1.9
12	97.9	1.9	0.1	71.4	24.9	3.7

mula is associated with the different mode compositions of the radiation. Formula (34) was derived under the assumption that both incident and reflected modes are the *TEM* modes. It turned out, however, that, when the corrugation is short or "wide," a considerable part of the energy is reflected into the TM_{02} and TM_{04} modes (see table). This part of the energy decreases with increasing grating length, and eventually the one-mode scattering regime $TEM \Rightarrow TEM$ sets up.

As follows from the calculations, for a narrow corrugation ($a \ll \lambda$), where boundary condition (28) for the impedances is quite appropriate, the reflecting properties of the grating are governed by the dimensionless parameter β [see (31)], which is consistent with the theoretical prediction. As the corrugation widens, conditions (28) fail and higher order terms must be left to correctly include diffraction in the operator $\hat{G}_{in}$. This is supported by the direct computation of electromagnetic wave scattering with the finite-element method [12].

Figures 5 and 6 demonstrate the dependences of the total reflection coefficient and of the partial reflection coefficients (for the *TM* modes) on the frequency of the *TEM* mode incident on Bragg gratings with narrow ($a = 0.01$ cm, $d = 0.09$ cm) and wide ($a = 0.1$ cm, $d = 0.03$ cm) corrugations for $L = 5$ cm.

CONCLUSION

We presented the theory of electromagnetic field excitation by one-dimensional Bragg resonators in terms of the coupled resonator method. It applies the Green function to the two-dimensional boundary-value problem for the Helmholtz equation, is rigorous (i.e., does not use any approximation), and can thus be employed for analyzing corrugated structures of arbitrary shape and size. For a narrow corrugation, we obtained an analytical solution, which is governed by a single dimensionless parameter depending on the corrugation shape and sizes. The coefficients of the reflection of the *TEM* mode from one-dimensional Bragg gratings with different lengths were numerically calculated for different radiation frequencies. The data obtained are compared with the analytical results for a narrow corrugation and with the direct numerical calculations by the finite-element method.

REFERENCES

1. G. G. Denisov and M. G. Reznikov, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **25**, 562 (1982).

2. V. L. Bratman, N. S. Ginzburg, and G. G. Denisov, Pis'ma Zh. Tekh. Fiz. **7**, 1320 (1981) [Sov. Tech. Phys. Lett. **7**, 565 (1981)].

3. N. S. Ginzburg, N. Yu. Peskov, and A. S. Sergeev, Pis'ma Zh. Tekh. Fiz. **18** (9), 23 (1992) [Sov. Tech. Phys. Lett. **18**, 285 (1992)].

4. N. S. Ginzburg, N. Yu. Peskov, A. S. Sergeev, *et al.*, Nucl. Instrum. Methods Phys. Res. A **358**, 189 (1995).

5. N. Yu. Peskov, N. S. Ginzburg, G. G. Denisov, *et al.*, Pis'ma Zh. Tekh. Fiz. **26** (8), 72 (2000) [Tech. Phys. Lett. **26**, 348 (2000)].

6. N. F. Kovalev, I. M. Orlova, and M. I. Petelin, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **11**, 783 (1968).

7. B. Z. Katsenelenbaum, *Theory of Irregular Waveguides with Slowly Varying Parameters* (Akad. Nauk SSSR, Moscow, 1961).

8. P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953; Inostrannaya Literatura, Moscow, 1958).

9. L. A. Vaĭnshteĭn, *Electromagnetic Waves* (Radio i Svyaz', Moscow, 1988).

10. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics,* Vol. 1: *Mechanics* (Nauka, Moscow, 1988; Pergamon, New York, 1988).

11. *Handbook of Mathematical Functions*, Ed. by M. Abramowitz and I. A. Stegun (National Bureau of Standards, Washington, 1964; Nauka, Moscow, 1979).

12. *Partial Differential Equation Toolbox: User's Guide* (The Mathworks Inc., 1997).

Translated by V. Isaakyan

THEORETICAL AND MATHEMATICAL PHYSICS

The Cantor Distribution and Fractal Transition Scattering

V. N. Bolotov

Institute of Electromagnetic Research, Kharkov, 61022 Ukraine
e-mail: renic@iemr. vl.net.ua
Received January 9, 2001

Abstract—A new distribution is introduced for the elaboration of a theory of fractal electrodynamics. The construction of the new distribution is based on the Cantor set structure and Cantor measures. The basic properties of this Cantor distribution are discussed. The semigroup of operators defined in terms of Cantor measures is examined. As an example, the theory of transition scattering of a dissipative permittivity burst by a stationary charge is considered. © 2002 MAIK "Nauka/Interperiodica".

INTRODUCTION

The electrodynamics of fractal media examines electromagnetic processes in a space filled with a material forming a fractal structure. Like any macroscopic theory, the electrodynamics of fractal media operates on physical quantities averaged over "physically infinitesimal" volume elements and does not touch upon the atomic (molecular) structure of the material. Fractal structures are fractal clusters, fractal surfaces, percolation clusters, and other material formations observed in experiments. Of special importance in the development of fractal electrodynamics is averaging, which can be made with fractal and multifractal measures.

Thus, the electrodynamics of fractal media relies on the fundamental concept of fractal and multifractal or, more specifically, on the Cantor set structure and Hausdorff–Bezikovich fractional-dimension spaces. The mathematical apparatus of fractal electrodynamics is taking several paths. It can be based on fractional integration/differentiation, fractal wavelets, or iterative functional systems. In this work, we will consider the apparatus of distributions. It is the author's opinion that such an approach is the most general and includes fractional integration/differentiation and wavelets as specific cases.

As an example, we will examine the formation of the fractal spectrum (the localization of the radiation frequencies at points of the Cantor set) upon transition scattering. The permittivity wave is specified by a function of dissipative burst $\gamma_\xi(x)$. The complex analytic form of the dielectric burst can be formed by shock waves in the dielectric, as well as by turbulent flows in plasmalike media and Kerr liquids.

CANTOR SET

The structure of the conventional triadic Cantor set with a ratio ξ is well known. The unit segment $E_0 = [0, 1]$ is divided into three parts, and the middle open

interval $1 - 2\xi$ is rejected. In this way, the set E_1 of two closed segments of length ξ is obtained. Then, the same is done with the segments $[0, \xi]$ and $[1 - \xi, 1]$, which enter into the set E_1, etc. At the nth step, we come to the set E_n consisting of 2^n segments of length ξ^n. The compact set E_n is called a pre-Cantor set (Fig. 1), and the intersection of pre-Cantor sets produces a Cantor set $E = \bigcap_{n=0}^{\infty} E_n$.

The Hausdorff–Bezikovich dimension of a Cantor set E is $d_H = \ln 2/|\ln \xi|$, $0 < \xi < 1/2$ [1]. A Cantor set is characterized by points of the first and second kind. Points of the set E that are the (left or right) ends of intervals adjacent to the Cantor discontinuum are called points of the first kind of the Cantor set (left, E^L, or right, E^R, respectively). Similar sets E_n^R and E_n^L can be constructed for any of the pre-Cantor sets E_n. In Fig. 1, the points belonging to the set E_n^R are marked by circles and those belonging to the set E_n^L, by squares. These points of the first order are ordered in a natural way and produce a countable set. All other points of the Cantor set are points of the second kind. The set of the points of the second kind is similar to that of irrational numbers and has the power of continuum [1].

Any point of the Cantor set E with a ratio ξ can be written as

$$t = \sum_{n=1}^{\infty} \omega_n \xi^n, \qquad (1)$$

where $\omega_n = \{0, \xi^{-1} - 1\}$ takes only two different values.

Numbers $t \in E$ for which all ω_n equal each other, starting with some number, belong to points of the first kind of Cantor set. Thus, the set of right points of the first kind E_n^R is defined by expansion (1) where all coefficients, starting with the $(n + 1)$th term, are zero:

$\omega_{n+1} = \omega_{n+2} = \ldots = 0$. The set of left points of the first kind E_n^L is defined by expansion (1) where all coefficients, starting with the $(n+1)$th term, equal $\xi^{-1} - 1$: $\omega_{n+1} = \omega_{n+2} = \ldots = \xi^{-1} - 1$.

To enumerate all points of a set, for example, E_n^R, it is necessary that each ω_k ($k = 1, 2, \ldots, n$) be assigned two values: 0 and $\xi^{-1} - 1$. In doing so, we obtain 2^n points of the set E_n^R; that is, $N(E_n^R) = 2^n$. Clearly, $N(E_n^R) = N(E_n^L)$.

FREE SEMIGROUP

A triadic Cantor set with a ratio ξ is generated by an alphabet A consisting of two letters: $A = \{a, \bar{a}\}$ ($\bar{a}$ can be viewed as the negation of a). This is due to the fact that the one-to-one correspondence maps $a \longrightarrow 0$, $\bar{a} \longrightarrow \xi^{-1} - 1$ and only one point t_α from the Cantor set corresponds to any word $\alpha = a_1 a_2 \ldots a_k \ldots$ from a set Ω of (finite and infinite) words. The approach using a word set is common in the theory of fractals and multifractals [2]. In this case, any a_k entering into the word α equals either a or $\bar{a}$. Thus, $a_k \longrightarrow \omega_k$ and

$$t_\alpha = t(\alpha) = \sum_{k=1}^{\infty} \omega_k \xi^k. \tag{2}$$

For example, to the word $\alpha = a\bar{a}a\bar{a}\ldots(a\bar{a})$, where the combination $(a\bar{a})$ is repeated, there corresponds the Cantor number $t_\alpha = \sum_{k=1}^{\infty}(\xi^{-1} - 1)\xi^{2n} = \xi/(1 + \xi)$. For $\xi = 1/3$, $t_\alpha = 1/4$. In the other case, to the word $\beta = \bar{a}a\bar{a}a\bar{a}\ldots(\bar{a}a)$, there corresponds the Cantor number $t_\beta = 1/(1 + \xi)$. For $\xi = 1/3$, $t_\beta = 3/4$.

The numbers 1/3 and 3/4 are the points of the second kind of the Cantor set [3]. In fact, in the above examples and in both the word α and word β the combination of two letters, rather than one letter, from the alphabet A is repeated. In the other case, we would deal with points of the first kind.

The word set Ω can be divided into subsets Ω_n ($n = 0, 1, 2, \ldots$) of words of finite length. The number of elements $N(\Omega_k)$ in a set of words of length k is 2^k. The element Ω_0 is an empty word Λ. Two words $x_1 x_2 \ldots x_m$ and $y_1 y_2 \ldots y_n$ are considered to be equal if they coincide as sequences: $m = n$, $x_1 = y_1$, $x_2 = y_2$, $\ldots$, $x_m = y_m$.

Note also that the bijective mapping $t: \Omega_n \longrightarrow E_n^R$ establishes a correspondence between a word of length n and a point of the pre-Cantor set E_n^R: $\alpha \longrightarrow t_\alpha$, $\alpha \in \Omega_n$, $t_\alpha \in R_n^R$.
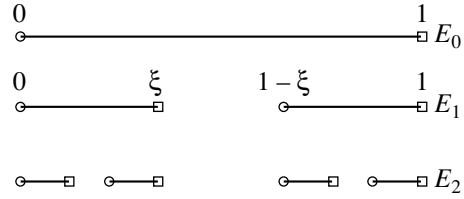
Fig. 1. Pre-Cantor sets.

In the set Ω of finite-length words in the alphabet A, one can separate the operation of occurrence or concatenation (*):

$$x_1 x_2 \ldots x_m * y_1 y_2 \ldots y_n = x_1 x_2 \ldots x_m y_1 y_2 \ldots y_n. \tag{3}$$

Thus, the concatenation of two words $\alpha_1 \in \Omega_n$ and $\alpha_2 \in \Omega_k$ (of length n and k, respectively) gives the word of length $n + k$: $\alpha_1 * \alpha_2 = \alpha_1 \alpha_2 \in \Omega_{n+k}$.

The operation of concatenation is associative by definition. Therefore, a set Ω of finite words with concatenation forms the free semigroup $\Omega^* = (\Omega, *)$ [4].

Note that each word from Ω admits the single expansion in the product of elements from A. Recall that by a semigroup Ω^* we mean a nonempty set Ω with the binary operation * that meets the associative law $x * (y * z) = (x * y) * z$ for any $x, y, z \in \Omega$.

Let the set E^R contain a binary operation ($\circ$) such that

$$t_\alpha \circ t_\beta = t_\alpha + \xi^k t_\beta \tag{4}$$

for $t_\alpha \in E_k^R$ and $t_\beta \in E_p^R$.

It is easy to check by direct calculation [see (2)] that $t_\alpha \circ t_\beta = t_{\alpha\beta} \in E_{k+p}^R$. On the other hand, $t_\beta \circ t_\alpha = t_\beta + \xi^p t_\alpha = t_{\beta\alpha} \in E_{k+p}^R$. Thus, $t_{\alpha\beta} \neq t_{\beta\alpha}$ by construction, so that this binary operation is noncommutative but associative. The associativity property is checked by direct calculation.

Thus, the set E^R with binary operation (4) forms the semigroup $(E^R, \circ)$.

Theorem. If Ω^* is a free semigroup and $(E^R, \circ)$ is the semigroup of right points of the first kind in a Cantor set, then there exists the isomorphism $t : \Omega^* \longrightarrow (E^R, \circ)$.

The proof is straightforward: $t(\alpha * \beta) = t(\gamma)$, since $\alpha * \beta = \gamma$, where $\alpha \in \Omega_k$, $\beta \in \Omega_n$, and $\gamma \in \Omega_{n+k}$. On the other hand, $t_\alpha \circ t_\beta = t_\alpha + \xi^k t_\beta = t_\gamma$, where $t_\alpha \in E_k^R$, $t_\beta \in E_n^R$, and $t_\gamma \in E_{n+k}^R$. However, since there exists a one-to-one correspondence between the sets Ω_{n+k} and E_{n+k}^R, then $t(\gamma) = t_\gamma$.

CANTOR DISTRIBUTION

The introduction of distributions is associated with the need for extending the representation of functions, for example, in theoretical physics. A function can be considered as a mathematical object influencing a trial function rather than as a set of values at various points. Of most interest in this respect is the approach that deals with distributions generated by measures [5]. As has been noted in the Introduction, a theory of electrodynamics of fractal media will be based on fractal or, more accurately, Cantor measures. In this work, such a measure is based on the Cantor function α, which is called "devil's ladder." A Cantor function is a monotonic nondecreasing function bounded from above and from below that is the limit of the sequence of pre-Cantor functions α_n; that is, $\alpha(x) = \lim\limits_{n \to \infty} \alpha_n(x)$.

The population of the pre-Cantor functions $\{\alpha_n\}$ is constructed in the segment [0, 1] using the totality of sets $\{E_n\}$. Figure 2 shows the function $\alpha_n(x)$ for $n = 2$. It is seen that the function $\alpha_2(x)$ is constant in the open intervals rejected and linearly grows in the closed intervals, which belong to E_2 in our case. As n increases, so does the slope of the straight lines $(2\xi)^{-n}$ (or the derivative of $\alpha_n(x)$ on E_n). In the limit $n \longrightarrow \infty$, the slope tends to infinity. Thus, the Cantor function α is a function that changes stepwise at points of the Cantor set. In this work, all constructions use a triadic Cantor set with a ratio ξ. If this is quite clear, the parameter ξ is omitted for simplicity. For example, $\alpha_n(x; \xi) \equiv \alpha_n(x)$.

On each of the pre-Cantor sets E_n, one can specify measures that are totally additive functions and, which is of special importance, are generated by the pre-Cantor functions $\alpha_n(x)$:

$$\mu(E_n) = \int\limits_{E_n} d\alpha_n. \tag{5}$$

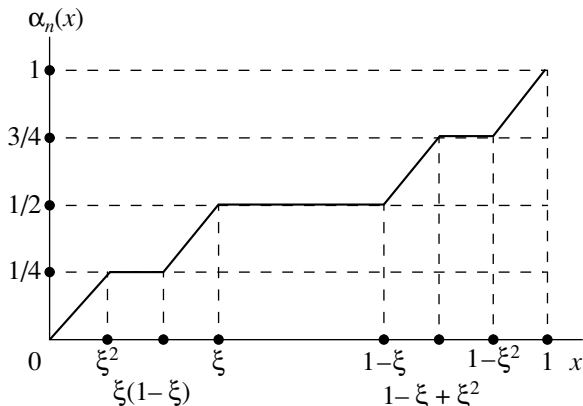The set of the measures $\{\mu(E_n)\}$ form a sequence converging to a measure specified on the Cantor set;

Fig. 2. Cantor function for $n = 2$.

that is, $\lim\limits_{n \to \infty} \mu(E_n) = \mu(E)$. The measure $\mu(E_n)$ on the pre-Cantor set E_n defines the distribution Δ_n ($n = 0, 1, 2, \ldots$) by the formula

$$(\Delta_n, \varphi) = \int \varphi d\alpha_n. \tag{6}$$

As before, we will use the shorthand writing $\Delta_n(x; \xi) \equiv \Delta_n(x)$. Integral (6) should be read as the Lebesgue–Stieltjes integral [3]. Note also that if $n = 0$, integral (6) becomes the Riemann integral. Thus, the pre-Cantor functions $\alpha_n(x)$ are generating functions for the measures $\mu(E_n)$.

The distribution Δ_n should be read as a functional acting over the space of the basic functions $D(Q)$, $\varphi \in D(Q)$. This space also covers all finite functions that are indefinitely differentiable in Q: $D(Q) = C_0^\infty$. In our case, $Q = R$ is the real axis. The convergence $\varphi_k \longrightarrow \varphi$ in the space of the basic functions $D(Q)$ depends on the convergence of the functional sequence $\{\varphi_k\}$ and all the derivatives $\varphi_k^{(n)} \longrightarrow \varphi^{(n)}$. For the distributions Δ_n, the functional (Δ_n, φ) has the form

$$\int \varphi d\alpha_n = \int \Delta_n \varphi dx \equiv (\Delta_n, \varphi). \tag{7}$$

Thus, the pre-Cantor distributions Δ_n can be explicitly expressed through the derivatives of $\alpha_n(x)$:

$$\frac{d}{dx}\alpha_n(x) = \begin{cases} 0, & x \notin E_n \\ \dfrac{1}{(2\xi)^n}, & x \in E_n. \end{cases} \tag{8}$$

From (7) and (8), it follows that

$$(\Delta_n, \varphi) = \int\limits_{E_n \subset R} \frac{1}{(2\xi)^n}\chi_{E_n}(x)\varphi(x)dx, \tag{9}$$

where χ_{E_n} is the characteristic function of the set E_n:

$$\chi_{E_n} = \begin{cases} 1, & x \in E_n \\ 0, & x \notin E_n. \end{cases} \tag{10}$$

Expression (9) can easily be transformed to the form

$$(\Delta_n, \varphi) = \frac{1}{2^n}\int\limits_0^1 \sum\limits_{\alpha \in \Omega_n} \varphi(\xi^n x + x_\alpha)dx, \tag{11}$$

where $x_\alpha \in E_n^R$. Note that $(\Delta_n, 1) = 1$ for any n.

Functional (11) shows the result of action of the distributions Δ_n on the trial function $\varphi(x)$ from the space of basic functions. This result is virtually the averaging of $\varphi(x)$ on E_n. The supports of the distributions $\{\Delta_n\}$ are the pre-Cantor sets $\{E_n\}$; that is, $\mathrm{supp}\Delta_n = E_n$. Thus, the distributions $\{\Delta_n\}$ are functionals on $D(Q)$ and $\Delta_n \in D'(R)$

for any n. Here, $D'(R)$ is the linear space of all distributions that is conjugate to $D(R)$. Convergence in $D'(R)$ is introduced as the weak convergence of the functionals. Thus, the sequence Δ_n defines the Cantor distribution

$$\lim_{n \to \infty} (\Delta_n, \varphi) = (\Delta_\xi, \varphi). \tag{12}$$

Using (11) and (12), we represent the Cantor distribution in the form

$$(\Delta_\xi, \varphi) = \lim_{n \to \infty} \frac{1}{2^n} \sum_{\alpha \in \Omega_n} \varphi(x_\alpha), \quad (\Delta_\xi, 1) = 1. \tag{13}$$

Clearly, $\mathrm{supp}\Delta_\xi = E$.

Turning back to the introduction of the distributions through measures (5), we can represent Δ_ξ in the form that is the most convenient for calculations and shows the relation between the Cantor function α, generating the measure $\mu(E)$, and the Cantor distribution Δ_ξ, which is a functional:

$$(\Delta_\xi, \varphi) = \int_R \varphi d\alpha. \tag{14}$$

From (13), it readily follows that the distribution Δ_ξ is a singular distribution. Note that both the Dirac generalized function δ and Δ_ξ are singular distributions and their supports are sets of measure zero: $\mathrm{supp}\delta = \{0\}$ and $\mathrm{supp}\Delta_\xi = E$.

The Cantor distribution Δ_ξ defines the Cantor structure of a point. The author believes that it can be applied for elaborating physical theories not on smooth manifolds but on manifolds where each point is a Cantor set (for example, on fractal space-time).

SEMIGROUP OF OPERATORS ON SET OF CANTOR MEASURES $\mu(E_n)$

Let us introduce a set of operators $\{A_n\}$ that act on the measure set $\{\mu(E_n)\}$ ($n = 0, 1, 2, \ldots$) following the law

$$A_k \mu(E_n) = \mu(E_{k+n}). \tag{15}$$

The operators $\{A_k\}$ defined by action (15) form a semigroup G, since it is easy to check that $A_k A_n = A_{k+n}$ and that the associative law $A_k(A_n A_j) = (A_k A_n)A_j$ is fulfilled. The operator A_0, which leaves the measure unchanged, is the unity of the semigroup. The action of the operator $\{A_n\}$ can be extended for functional spaces, for example, for the space of the basic functions $D(Q)$. To obtain the operators $\{A_k\}$ in the explicit form, we assume that any $A_k \in G$ has the property

$$(A_k \Delta_n, \varphi) = (\Delta_n, A_k \varphi). \tag{16}$$

This property of the operators $\{A_k\}$ is identical to the action of the differentiation operator on the distributions from $D'(Q)$: $(D^n f, \varphi) = (-1)^n (f, D^n \varphi)$, where n is the

order of derivative. The operator of Fourier transformation (29) acts in a similar way. To express the operators $\{A_k\}$ in the explicit form, we put $n = 0$ in expression (16). Then, $(A_k \Delta_0, \varphi) = (\Delta_0, A_k \varphi)$. In another form,

$$\int_R \varphi d\alpha_k = \int_R A_k \varphi dx. \tag{17}$$

Using expressions (7), (11), and (17), we come to the explicit form of the operators $\{A_k\}$ ($k = 0, 1, 2, \ldots$):

$$A_k \varphi(t) = \frac{1}{2^k} \sum_{\alpha \in \Omega_k} \varphi(\xi^k t + t_\alpha). \tag{18}$$

At $k = 0$, $A_0 \varphi = \varphi$. The operators A_k map the space of the basic functions into itself. They perform the transformation $t \longrightarrow t' = \xi^k t + t_\alpha$ ($t \in R, t' \in R, t_\alpha \in E_k^R$) and averaging over the right points of the first kind in the Cantor set.

The operators $\{A_k\} \in G$ retain the structure of a semigroup, since it is easy to check by direct calculations using (18) that $A_k A_n \varphi = A_{k+n}\varphi$. Indeed, taking into account that $t_\beta + \xi^k t_\alpha = t_\gamma$, where $\alpha \in \Omega_n$, $\beta \in \Omega_k$, $\gamma \in \Omega_{n+k}$, we find

$$A_n A_k \varphi(t) = \frac{1}{2^n 2^k} \sum_{\alpha \in \Omega_n} \sum_{\beta \in \Omega_k} \varphi(\xi^{n+k} t + \xi^k t_\alpha + t_\beta)$$
$$= \frac{1}{2^{n+k}} \sum_{\gamma \in \Omega_{n+k}} \varphi(\xi^{n+k} t + t_\gamma). \tag{19}$$

From (15) and (16) at $n \longrightarrow \infty$, one important property of invariance that is necessary for calculating the functionals Δ_ξ follows; namely,

$$A_k \mu(E) = A_n \mu(E), \quad (\Delta_\xi, A_k \varphi) = (\Delta_\xi, A_n \varphi). \tag{20}$$

Equality (20) is valid for any n ($k = 0, 1, 2, \ldots$). As an example, let us calculate the specific integral

$$I(v) = (\Delta_\xi, x^v) = \int_0^1 x^v d\alpha. \tag{21}$$

Using relationship (21) for A_1 and expression (18), we easily obtain the functional $I(v)$:

$$I(v) = \int_0^1 A_1 x^v d\alpha = \int_0^1 \frac{(\xi x)^v + (1 - \xi + \xi x)^v}{2} d\alpha. \tag{22}$$

Thus, we can find the functional relationship for the functionals $I(v)$:

$$I(v) = \frac{(1-\xi)^v}{2-\xi^v} \sum_{k=0}^{\infty} \binom{v}{k} (\xi^{-1}-1)^{-k} I(k),$$

$$\binom{v}{k} = \frac{v(v-1)\ldots(v-k+1)}{k!}. \tag{23}$$

For $v = m$, $m \in Z$, and in view of the fact that $I(0) = 1$, expression (23) takes the form

$$I(m) = \frac{(1-\xi)^m}{2(1-\xi^m)} \sum_{k=0}^{m-1} \binom{m}{k} (\xi^{-1}-1)^{-k} I(k). \tag{24}$$

If $m = 1$, $I(1) = 1/2$. In this case, Lebesgue–Stieltjes integral (21) does not differ from the Riemann integral of the function x^m. However, for $m = 2$, $I(2) = 1/[2(1 + \xi)]$. This result is markedly different from the value of the Riemann integral, $1/3$. The results coincide for $\xi = 1/2$. Note also that $1/[2(1 + \xi)] > 1/3$ if $0 < \xi < 1/2$.

Below are the expressions for several functionals calculated with the above procedures:

$$(\Delta_\xi, e^{\eta t}) = e^{\eta/2} \prod_{n=0}^{\infty} \cosh\left[\frac{(1-\xi)\xi^n}{2}\eta\right], \tag{25}$$

$$(\Delta_\xi, e^{i\pi a t}) = e^{i\pi a/2} \gamma_\xi(\pi a),$$

$$\gamma_\xi(\pi a) = \prod_{n=0}^{\infty} \cos\left[\frac{(1-\xi)\xi^k}{2}\pi a\right]. \tag{26}$$

From expressions (26) and (13) for the functionals, we arrive at the expression that will be used later:

$$\lim_{n \to \infty} \frac{1}{2^n} \sum_{\alpha \in \Omega_n} e^{it_\alpha \omega} = e^{i\omega/2} \gamma_\xi(\omega). \tag{27}$$

Thus, invariant (20) provides a simple way of calculating the functionals Δ_ξ.

FOURIER TRANSFORM FOR CANTOR DISTRIBUTION

Let φ be a function locally integrable on R^n. The Fourier transform of this function is given by

$$F[\varphi](\xi) = \frac{1}{(2\pi)^n} \int_{R^n} e^{-i(\xi, x)} dx. \tag{28}$$

Let $x = (x_1, \ldots, x_n)$ and $\xi = (\xi_1, \ldots, \xi_n)$ be pseudo-Euclidean coordinates in the general case. The scalar product (ξ, x) of these vectors is introduced, as usual, by means of the pseudo-Euclidean metrics. Integral (28) is assumed to be convergent in principal value.

To construct the Fourier transform of the distributions, we introduce the space of rapidly decreasing basic functions $\mathscr{L}$ and the set of slowly increasing distributions $\mathscr{L}'$ [5]. The former space, $\mathscr{L} = \mathscr{L}(R^n)$, consists of indefinitely differentiable (on R^n) functions that decrease, along with all derivatives, faster than any power function of $|x|^{-1}$ at $|x| \longrightarrow \infty$. In this case, $\mathscr{L} \supset D$. Any linear continuous functional on the set of basic functions $\mathscr{L}$ is called a slowly increasing distribution. The space of the distributions $\mathscr{L}' = \mathscr{L}'(R^n)$ consists of the set of all slowly increasing distributions. It is obvious that $\mathscr{L}' \subset D'$.

Let $f(x) \in \mathscr{L}$ and $\varphi(x) \in \mathscr{L}$; then, from the definition of the Fourier transform for distributions, we have

$$(F[f], \varphi) = (f, F[\varphi]). \tag{29}$$

Since the operation $\varphi \longrightarrow F[\varphi]$ is linear and continuous in going from $\mathscr{L}$ to $\mathscr{L}$, the functional on the right-hand side of (29) represents a distribution from $\mathscr{L}'$. The operation $f \longrightarrow F[f]$ is linear and continuous in going from $\mathscr{L}'$ to $\mathscr{L}'$ [5].

If f is a distribution with a compact support, it is a slowly increasing distribution and its Fourier transform exists [5]. It is this case that is of interest, because the Cantor distribution has the compact support $\text{supp}\Delta_\xi = E$. It should be noted that the distributions $\{\Delta_n\}$ also have the compact carriers $\text{supp}\Delta_n = E_n$; hence, they also have their associated Fourier transforms. Based on the theorem proved in [5], the Fourier transform $F[f]$ for $f \in \mathscr{L}'$ and, hence, for $\Delta_\xi \in \mathscr{L}'$ can be represented as

$$F[\Delta_n](k) = \left(\Delta_n, \eta(x)\frac{e^{ikx}}{2\pi}\right), \tag{30}$$

where $\eta \in D$ and equals unity in the vicinity of $\text{supp}\Delta_n$.

Using the algorithm for calculating functional (11), we easily find the Fourier transform for Δ_n:

$$F[\Delta_n] = \frac{1}{2^n}\frac{1}{2\pi} \sum_{\alpha \in \Omega_n} e^{ikx_\alpha} e^{ik\xi^n/2} \frac{\sin(k\xi^n/2)}{k\xi^n/2}. \tag{31}$$

Thus, the Fourier transform of the Cantor distribution has the form

$$F[\Delta_\xi] = \lim_{n \to \infty} F[\Delta_n] = \lim_{n \to \infty} \frac{1}{2^n}\frac{1}{2\pi} \sum_{\alpha \in \Omega_n} e^{ikx_\alpha}, \tag{32}$$
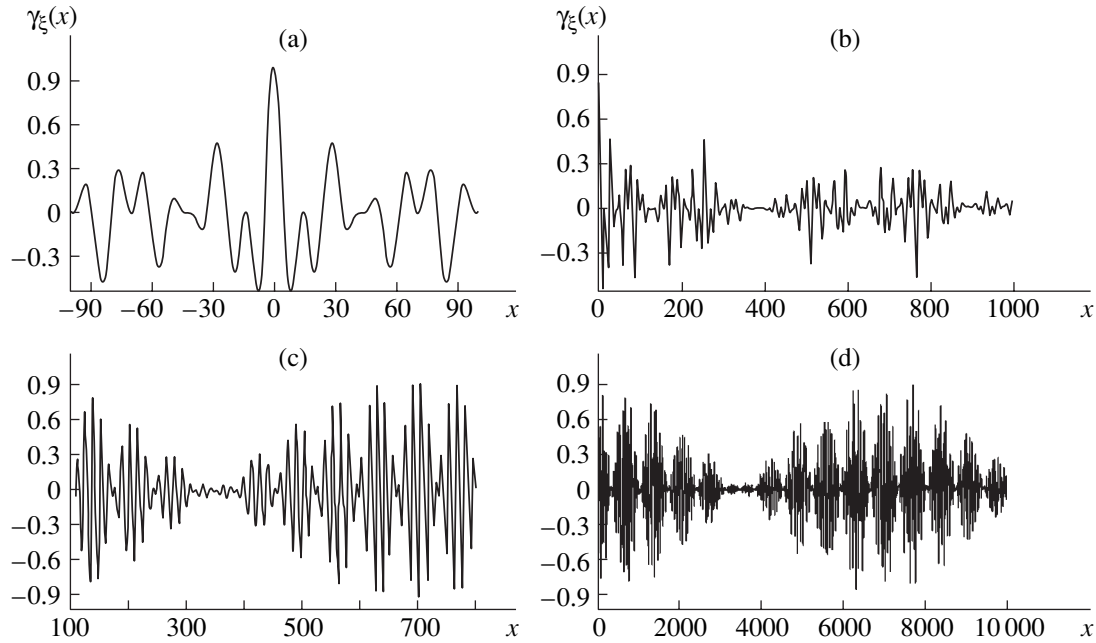
where $x_\alpha \in E_n^R$.

With regard for previous result (27), $F[\Delta_\xi]$ can be rearranged to

$$F[\Delta_\xi](k) = \frac{e^{ik/2}\gamma_\xi(k)}{2\pi}. \tag{33}$$

Let us define the inverse Fourier transform F^{-1} in $\mathscr{L}'$ as

$$F^{-1}[f] = F[f(-x)](2\pi)^n, \tag{34}$$

where $f(-x)$ is the image of $f(x)$. The operation F^{-1} is linear and continuous from $\mathscr{L}'$ to $\mathscr{L}'$ and also is inverse

Fig. 3. Plots of the function $\gamma_\xi(x)$ for $\xi =$ (a, b) 1/2 and (c, d) 1/10. x varies from (a) –90 to +90, (b) 0 to 1000, (c) 100 to 800, and (d) 0 to 10 000.

relative to F; that is,

$$F^{-1}[F[f]] = f, \quad F[F^{-1}[f]] = f, \quad f \in \mathscr{L}'. \quad (35)$$

Thus, the Cantor distribution can be expressed through the inverse Fourier transform:

$$F^{-1}[e^{ik/2}\gamma_\xi(k)] = 2\pi\Delta_\xi. \quad (36)$$

On the other hand, it is easy to check that

$$F\left[\Delta_\xi\left(x + \frac{1}{2}\right)\right](k) = \frac{1}{2\pi}\gamma_\xi(k). \quad (37)$$

Figure 3 plots the function $\gamma_\xi(x)$ for various ξ.

FRACTAL TRANSITION SCATTERING

It is known that permittivity waves scattered by a charge generate electromagnetic (transition) radiation. This process is called a transition process for two reasons. First, the effect can also be observed if the charge is immobile; second, radiation of this type takes place over the entire space, not only in a region localized at the boundary. This causes the electromagnetic waves to interfere [6, 7]. The transition radiation that occurs when the permittivity ε varies by a periodic law has been well studied. The permittivity can depend only on the density N of the medium. If a longitudinal acoustic wave propagates in the medium, $N = N_0 + \Delta N \cos(\mathbf{k}_0\mathbf{r} - \omega_0 t)$. Thus,

$$\varepsilon = \varepsilon_0 + \varepsilon_1 \cos(\mathbf{k}_0\mathbf{r} - \omega_0 t), \quad (38)$$

where ε_1 is a change in ε due to a change in N ($\varepsilon_1 \sim \Delta N$).

The electrostatic field of a charge embedded in a medium with a permittivity ε_0 is characterized by the induction $\mathbf{D}$:

$$\mathbf{D}(\mathbf{r}) = \varepsilon_0\mathbf{E}_0(\mathbf{r}), \quad \mathbf{E}_0(\mathbf{r}) = \frac{q\mathbf{r}}{\varepsilon_0 r^3}. \quad (39)$$

When a permittivity wave appears, $\mathbf{D}(\mathbf{r}) \longrightarrow \mathbf{D}(\mathbf{r}, t) = \varepsilon(\mathbf{r}, t)\mathbf{E}(\mathbf{r}, t)$. In the first approximation $|\varepsilon_1| \ll |\varepsilon_0|$, the variable polarization appears around the charge [7]:

$$\delta\mathbf{P} = \frac{\delta\mathbf{D}}{4\pi} = \frac{\varepsilon_1\mathbf{E}_0}{4\pi}\cos(\mathbf{k}_0\mathbf{r} - \omega_0 t). \quad (40)$$

This polarization generates a wave with a frequency ω_0 radiating from the charge.

We will study the transition scattering due to a nonperiodic dissipative perturbation of the permittivity:

$$\varepsilon(\mathbf{r}, t) = \varepsilon_0 + \varepsilon_1\gamma_\xi(\mathbf{k}_0\mathbf{r} - \omega_0 t), \quad (41)$$

where $\mathbf{k}_0\mathbf{r}$ is the scalar product.

Dielectric burst (41) (the perturbation of the density of the medium) can be caused, for example, by shock waves in this medium or by concentration nonuniformities in a plasma. We simulate the transition scattering using the function $\gamma_\xi(x)$, whose Fourier image is related to the Cantor set. The behavior of this function for various ξ is shown in Fig. 3.

The Fourier transform $\mathbf{D}(\mathbf{k}, \omega)$ for the induction $\mathbf{D}(\mathbf{r}, t)$ has the form of (28):

$$\mathbf{D}(\mathbf{k}, \omega) = \frac{1}{(2\pi)^4}\int d^3\mathbf{r}\,dt\,e^{-i(\mathbf{k}\mathbf{r} - \omega t)}\varepsilon(\mathbf{r}, t)\mathbf{E}(\mathbf{r}, t). \quad (42)$$

Substituting $\varepsilon(\mathbf{r}, t)$ from (41) into (42) and using (36) yields

$$\mathbf{D}(\mathbf{k}, \omega) = \varepsilon_0 \mathbf{E}(\mathbf{k}, \omega)$$

$$+ 2\pi\varepsilon_1 \int_{-1/2}^{1/2} dv \mathbf{E}(\mathbf{k} - v\mathbf{k}_0, \omega - v\omega_0)\Delta_\xi\left(v + \frac{1}{2}\right). \quad (43)$$

Consider the transition scattering if the charge is stationary. The Fourier transforms of the induction $\mathbf{D}(\mathbf{k}, \omega)$ and the polarization vector $\delta\mathbf{P}(\mathbf{k}, \omega)$ are related by the well-known formula

$$\mathbf{D}(\mathbf{k}, \omega) = \varepsilon_0 \mathbf{E}(\mathbf{k}, \omega) + 4\pi\delta\mathbf{P}(\mathbf{k}, \omega). \quad (44)$$

Hence,

$$\delta\mathbf{P}(\mathbf{k}, \omega) = \frac{\varepsilon_1}{2} \int_{-1/2}^{1/2} dx \Delta_\xi\left(x + \frac{1}{2}\right)\mathbf{E}(\mathbf{k} - x\mathbf{k}_0, \omega - x\omega_0). (45)$$

If the charge is stationary, the equation for the field is written as follows:

$$\text{curl curl}\mathbf{E} + \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\mathbf{D} = 0. \quad (46)$$

From these equations, it is easy to derive the expression for the Fourier images in the form of (43):

$$\left(k^2\delta_{ij} - k_ik_j - \frac{\omega^2}{c^2}\varepsilon_0\delta_{ij}\right)E_j(\mathbf{k}, \omega) = \frac{2\pi\omega^2}{c^2}\varepsilon_1$$

$$\times \int_{-1/2}^{1/2} dv E_i(\mathbf{k} - v\mathbf{k}_0, \omega - v\omega_0)\Delta_\xi\left(v + \frac{1}{2}\right). \quad (47)$$

As usual, the summation is over the repeating indices. It is assumed that $|\varepsilon_1| \ll |\varepsilon_0|$. Then, the right-hand side of (47) can be considered as a perturbation due to a change in the polarization near the charge. In the first order of perturbation theory, the Fourier components of the stationary charge field are given by

$$\mathbf{E}_{\mathbf{k}, \omega} = -\frac{4\pi iq\mathbf{k}}{(2\pi)^3\varepsilon_0 k^2}. \quad (48)$$

Thus, the perturbation of the polarization near the charge can be obtained by substituting (48) into (45):

$$\delta\mathbf{P}(\mathbf{k}, \omega) = \frac{iq\varepsilon_1}{4\pi^2\varepsilon_0\omega_0}\frac{(\mathbf{k} - \omega\mathbf{k}_0/\omega_0)}{(\mathbf{k} - \omega\mathbf{k}_0/\omega_0)^2}\Delta_\xi\left(\frac{\omega}{\omega_0} + \frac{1}{2}\right). \quad (49)$$

If the waves scattered are assumed to be transverse (relative to the vector $\mathbf{k}$), expression (47) takes the form

$$\left(k^2 - \frac{\omega^2}{c^2}\varepsilon_0\right)\mathbf{E}^\tau(\mathbf{k}, \omega) = \frac{4\pi\omega^2}{c^2}\frac{[\mathbf{k}, [\mathbf{k}, \delta\mathbf{P}, (\mathbf{k}, \omega)]]}{k^2}. (50)$$

From (49) and (50), one can calculate the field of the transverse waves scattered:

$$\mathbf{E}^\tau(\mathbf{k}, \omega) = -\frac{iq\omega^3[\mathbf{k}, [\mathbf{k}_0, \mathbf{k}]]}{\pi(\omega_0 kc)^2(k^2 - \omega^2\varepsilon_0/c^2)(\mathbf{k} - \omega\mathbf{k}_0/\omega_0)^2}$$

$$\times \frac{\varepsilon_1}{\varepsilon_0}\Delta_\xi\left(\frac{\omega}{\omega_0} + \frac{1}{2}\right). \quad (51)$$

Expression (51) is identical to those for the Fourier components of the field that were obtained in [6] for permittivity waves like (38). The basic difference is that formula (51) derived in our work involves the Cantor distribution Δ_ξ rather than the δ function. This means that, in our case, the electromagnetic radiation has some frequency range specified by the Cantor distribution.

It has been shown [8] that the transverse wave intensity is given by

$$W\left(\omega, \frac{\mathbf{k}}{k}\right)d\omega do = (2\pi)^6\frac{\omega^4}{c^3}\sqrt{\varepsilon_0}\left|\delta\mathbf{P}^\tau(\mathbf{k}, \omega)\right|^2 d\omega do. \quad (52)$$

Here, $\delta\mathbf{P}^\tau(\mathbf{k}, \omega)$ is the transverse polarization, which is readily found from (49):

$$\delta\mathbf{P}^\tau(\mathbf{k}, \omega) = \frac{[\mathbf{k}, [\mathbf{k}, \delta\mathbf{P}, (\mathbf{k}, \omega)]]}{k^2}, \quad (53)$$

where $\mathbf{k} = \mathbf{k}/k\sqrt{\varepsilon_0}\,\omega/c$.

Thus, expressions (49), (51), and (52) imply that the electromagnetic spectrum of the fractal transition scattering is concentrated in the frequency range $\omega = \omega_0(\omega_\alpha - 1/2)$. The numbers ω_α belong to points of the Cantor set and are defined by expression (2).

CONCLUSION

This work elaborates on the theory of fractal transition scattering [9]. The results obtained in this article are more general and can be applied to developing the theory of fractal electrodynamics. Our considerations rely upon the Cantor distribution based on Cantor measures. The algorithm for calculating the functionals with the Cantor distributions was described, and the Fourier transform of the Cantor distribution was derived.

By way of example, we considered the transition scattering of a dissipative burst of permittivity (41) incident on a charge. Being scattered by the charge, the space-time permittivity function of this type forms the wide fractal radiation spectrum (the radiation frequencies are localized at points of the Cantor set). A dielectric burst of a complex shape may be generated by shock waves in an insulator, as well as by turbulent flows in plasmalike media and Kerr liquids.

Fractal transition scattering can serve as a source of wide-band electromagnetic radiation, on the one hand,

or be used for the diagnostics of chaotic dynamics in various media, on the other.

REFERENCES

1. P. S. Alexandroff, *Introduction to the Theory of Sets and General Topology* (Nauka, Moscow, 1997).
2. E. Olivier, Nonlinearity **12**, 1571 (1999).
3. A. N. Kolmogorov and S. V. Fomin, *Elements of the Theory of Functions and Functional Analysis* (Nauka, Moscow, 1968; Graylock Press, Rochester, 1957).
4. G. Lallement, *Semigroups and Combinatorial Applications* (Wiley, New York, 1983; Mir, Moscow, 1985).
5. V. S. Vladimirov, *Generalized Functions in Mathematical Physics* (Nauka, Moscow, 1979).
6. V. L. Ginzburg and V. N. Tsytovich, Zh. Éksp. Teor. Fiz. **65**, 1818 (1973) [Sov. Phys. JETP **38**, 909 (1973)].
7. V. L. Ginzburg and V. N. Tsytovich, *Transient Radiation and Transient Scattering* (Nauka, Moscow, 1984).
8. V. V. Kolesov, Zh. Éksp. Teor. Fiz. **106**, 77 (1994) [JETP **79**, 39 (1994)].
9. V. N. Bolotov, Zh. Tekh. Fiz. **70** (12), 98 (2000) [Tech. Phys. **45**, 1604 (2000)].

Translated by V. Isaakyan

THEORETICAL AND MATHEMATICAL PHYSICS

A Neuron as a Quantum-Optical Device: A Model

S. L. Grigor'ev

Received April 4, 2001

Abstract—Inversion in the structure of polyaminoacids and proteins is discussed, and the electrodynamic quantum-optical model of a neuron is constructed. The feasibility of the quantum-logical summation of exciting pulses by a nanomembrane is considered. The model is consistent with the physiology foundations and assumes implementation by means of techniques used in microelectronics, integrated optics, and nanooptics. © *2002 MAIK "Nauka/Interperiodica".*

The physical properties of neurons depend on those of proteins and amino acids that enter into the composition of the membrane. These building blocks, in turn, may contain polar molecular groups (peptide, etc.) [1] and fit into the class of polar dielectrics, whose resonance frequencies lie typically in the microwave and far-IR ranges [2, 3]. At microwaves, the classical approach [4, 5] to analyzing the atomic oscillation dynamics in these materials applies up to the resonance frequencies of polar molecular groups (50–200 cm^{-1}).

The peptide groups incorporated into protein chains are shown in Fig. 1. Their dipole moment is about $0.3D$ and arises because of the shift of the electron density from a nitrogen atom to an oxygen one. The peptide groups can readily rotate about the C_1–C_2 and C_2–N bonds.

Proteins and amino acids have the property of self-assembling into secondary structures like crystals or even higher order structures [6]. Among them, the "folded leaf" structure is one of the simplest (Fig. 2). The secondary structure consists of the primary polymer molecules with hydrogen bonds. Such structures readily polarize and have a high permittivity in a wide frequency range [1].

The polarization mechanism is well described by the Pauling model [7] (Fig. 3). By analogy with this model, the folded leaf structure can have two, base and inverse, states in which the bonds lie in the plane of the leaf (Fig. 2a) and between the leaves (Fig. 2b). This is a specific case of conformation typical of many high-molecular materials [8].

Let us consider this structure from the viewpoint of electrodynamics and microwave spectroscopy. In the secondary structure of the base type (Fig. 2a), the elementary cell has four dipoles forming an electroneutral system (Fig. 2, *a*1). The dipoles are arranged at different angles to each other, so that they variously respond to an external field. In addition, in the substances considered, the rocking (pendular) vibration of the peptide group prevails (Fig. 4). The frequency of the vibration is the frequency of a torsion pendulum, which is found

from the relationship $T = 2\pi\sqrt{I/D}$, where T is the pendulum period, I is the moment of inertia of the pendulum, and D is the torsional stiffness of the bond. The moment of inertia is $I = mr^2$, where $m = 3 \times 10^{-26}$ kg is the mass of an oxygen atom and $r = 10^{-10}$ m is the length of a C_1–O bond. The torsional stiffness depends on the C_1–C_2 valence bond stiffness, O–H hydrogen bond stiffness, and dipole–dipole Coulomb interaction [9]. Experimental data for molecular motion in crystals of the simplest amino acid, glycine, is available from [4].

Consider the behavior of a rotatable pendular dipole subjected to an electromagnetic field (Fig. 5). Obviously, its response to the field will differ from that of a Hertzian classical dipole with longitudinally bound charges [10]. The maximum of the radiation scattered (reflected) by a pendular dipole is observed when the incident wave is directed along the dipole vector and the electric field vector of the incident wave is orthogonal to the dipole vector. If the dipole moment of the peptide molecule is reversed with the external field direction remaining the same, the total reflection of the wave is changed to its total transmission (Fig. 6). Thus, the conformation of the structure causes the effect of "optical switching" (but in the microwave range). Since the microwave component in the earth's electromag-
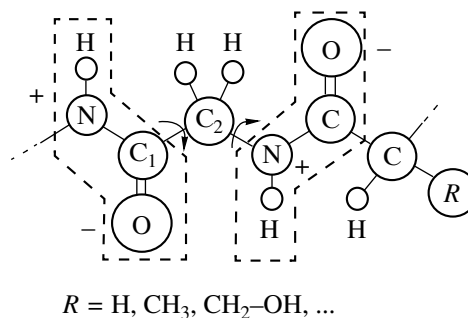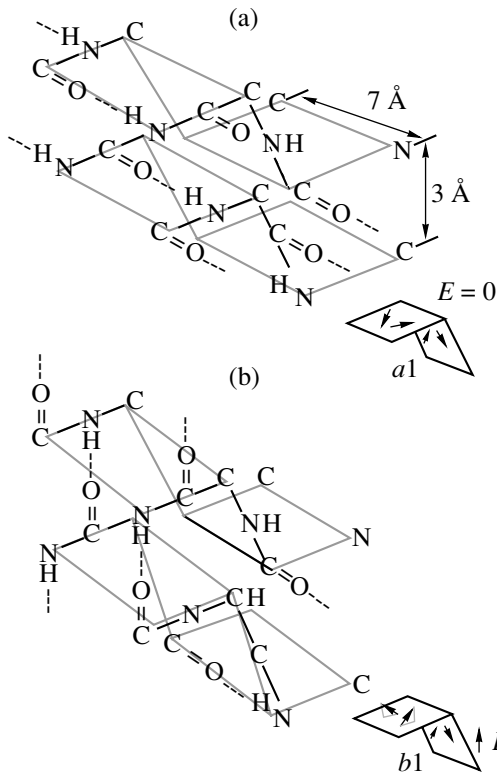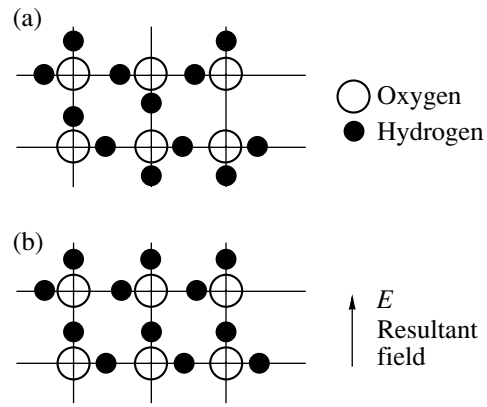

R = H, CH$_3$, CH$_2$–OH, ...

Fig. 1. H–N–C–O peptide groups in the protein chain.
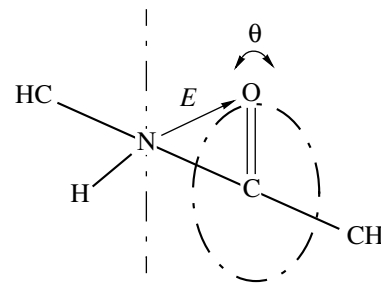
Fig. 2. (a) Base and (b) inverse secondary structures of polyalanine. $a1$ and $b1$, dipole configurations.

**Fig. 3.** Pauling model for polar molecules of water and ice: (a) disordered and (b) ordered structures.

**Fig. 4.** Rocking vibration (θ) of the peptide group. E is the electric vector of the dipole.

netic field is absent, this switching effect seems to be used in the biological memory of animals [3].

Now let us turn to an elementary cell of four dipoles and take advantage of electromechanical analogy to construct its model. We represent a peptide group as a torsional pendulum having two charged poles that are attracted to or repel from the poles of other pendulums according to the relation between the signs of the charges (Fig. 7a). For such a four-pendulum system, the stable equilibrium state is lacking if the initial dipole directions are shifted by a quarter-period ($\pi/2$). As is known from quantum electrodynamics, this system may exist only in motion [11]. The Coulomb forces provide the potential energy, while the kinetic energy is provided by the rotating masses. If the axes of the pendulums are parallel to each other, the system loses the anisotropy of interaction with the external electromagnetic field: its electric component accelerates two pendulums and slows down the two other. The anisotropy arises when the axes of one pair of pendulums are rotated relative to the other, as in the case of the folded leaf structure (Fig. 7b). Here, two of four Coulomb bonds are replaced by torsional bonds for the analogy with the polyalanine structure to be complete.

Let us trace the interaction of the dipoles with a weak external electric field (wave) for a short time interval (much shorter than the vibration period). In this
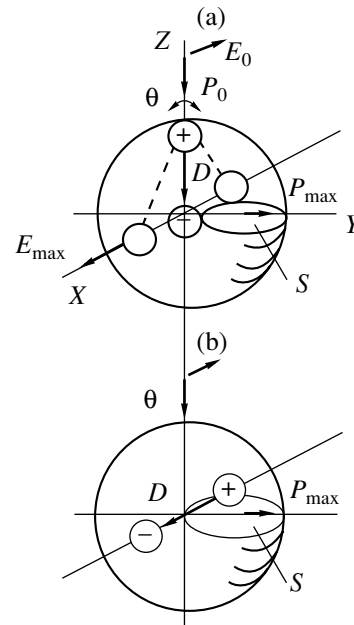
Fig. 5. Scattering of the incident wave P_0, E_0 by (a) pendular and (b) Hertzian dipole (D). P_{max}, E_{max} are the direction and the electric field vector of the wave scattered and S is the cross section of the toroid where the radiation scattered is concentrated.
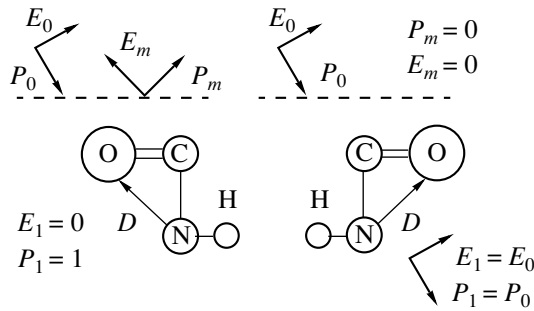
Fig. 6. Reflection and transmission of the incident wave in the different states of the structure. 0, incident wave; m, reflected wave; and 1, transmitted wave. Other designations are the same as in Fig. 5.

case, the following simple rule holds: if the direction of the wave (P) and the instantaneous direction of rotation of a pendular dipole (ω) coincide, the field transfers the energy to the dipole (absorption takes place); if not, the dipole gives the energy to the field (enhances it). Applying this rule and summing over all dipoles, we find that the maximal absorption is observed when the directions of the external field and one of the dipoles coincide; these maxima arise in 90° degree intervals. Such a phenomenon has been discovered in thin protein films in the IR range [1].

Now consider the inverse structure, where some dipoles are rotated through 90° and fixed by hydrogen bonds between the leaves (Fig. 2b). The electromechanical model for this case is depicted in Fig. 8. Here, a weak external wave directed along the vector of one

dipole is not absorbed; instead, it gains the energy from the material. The material, a polar dielectric, has the directional internal field and surface charge. Thus, the incident wave can be enhanced by induced radiation due to the conformation (inversion) of the structure. Consequently, a nanofilm consisting of several folded polypeptide leaves can act as a maser (the microwave and IR analog of a laser) [12]. Similar properties may be found in proteins and polyaminoacids with the helical structure. This simple qualitative model can be extended up to the mathematical one.

Next, we will treat an invertible ferroelectric membrane in terms of the quantum-mechanical theory of coherent radiation. The time of observation is assumed to be shorter than the time of excitation relaxation into thermal vibrations of the molecules. Then, the state of a charged particle on the membrane can be represented as the superposition of the wave functions of its two possible states [13]:

$$\psi = \psi_1(r)\exp((-i/h)E_1 t) + \psi_2(r)\exp((-i/h)E_2 t), \quad (1)$$

where r is the position of the particle, t is time, and E_1 and E_2 are the energies of the states.

The polarization of the membrane by a pulse causes an electric field of potential energy U to appear. This field affects the energies of the states:

$$E_1 = E_1^0 - U, \quad E_2 = E_2^0 + U. \quad (2)$$

The polarized membrane may spontaneously pass into the initial (base) state by radiation. The probability
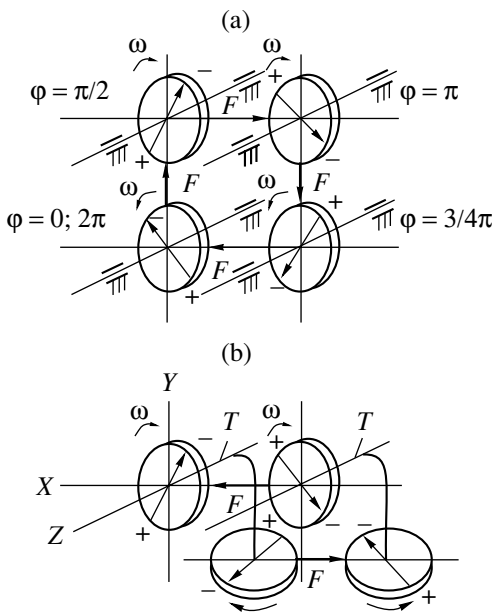
Fig. 7. Electromechanical configuration of the vibratory system: (a) isotropic and (b) anisotropic. F, Coulomb forces; T, torsional bonds; and ω, instantaneous direction of rotation.
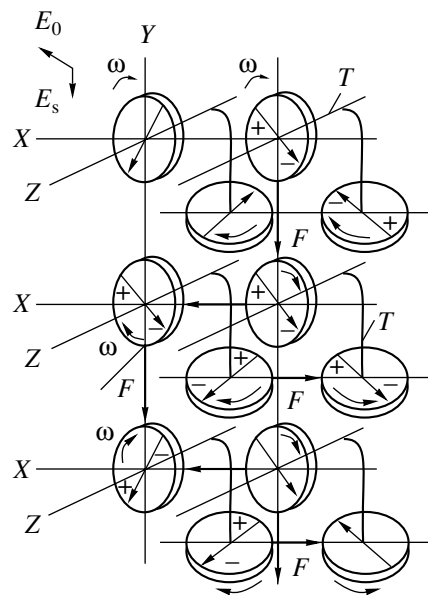
Fig. 8. Electromechanical model of the inverse structure. E_0, external field; E_s, self-field of the system charge. Other designations are the same as in Fig. 7.
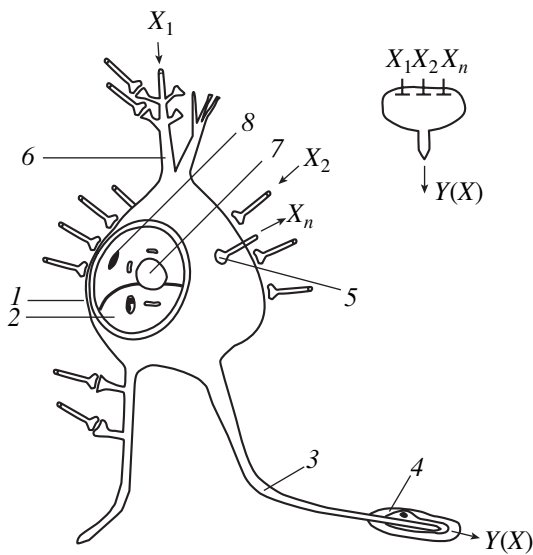
Fig. 9. Pyramidal neuron and its configuration (on the right): *1*, neuron enclosure; *2*, axon membrane; *3*, axon (peripheral nerve fiber); *4*, myelin enclosure (Schwann's cell); *5*, synapses (ends of other nerve fibers); *6*, dendrite of neuron (fiber bundle); *7*, neuron nucleus; and *8*, cytoplasm and organelles. The nucleus is shown in the cross-sectional view.

density of this radiation is

$$S = 1/2m(\psi^* p\psi - \psi p\psi^*), \qquad (3)$$

where p is the operator of the particle momentum, ψ^* is the complex conjugate wave function of the particle, and m is the mass of the particle.

Substituting (1) and (2) into (3), leaving only the time-dependent terms, and separating the real part, we arrive at the expression for the probability density:

$$S \sim [\cos(U + (E_2 - E_1)) - \cos(U - (E_2 - E_1))]. \quad (4)$$

It follows that there exists some polarization potential of a ferroelectric membrane at which the probability of coherent radiation is maximal.

Importantly, our model implies that a ferroelectric membrane can serve as a logic adder of quantized polarizing pulses. Each pulse increases or decreases the membrane potential according to the polarization vector. Once the radiation potential has been attained, the membrane emits a coherent pulse.

It is obvious that a pyramidal neuron [14] is the real device implementing our model (Fig. 9). Here, the axon membrane, containing high-molecular polar substances, serves as a ferroelectric membrane. These substances are all amino acids acting as neuromediators [14]. Here, the membrane is pumped by pulses coming from many synapses that are the ends of the nerve fibers of dendrites of the given and other neurons. Nerve fibers are waveguides, as was shown in the previous article of the author [15]. In them, the receptors of sense organs and other neurons of various types serve as sources of quantized pulses (solitons) [14, 15].

CONCLUSIONS

(1) The nervous system of higher animals and humans is a holographic neural network (see [16]) in which signals are processed not only by the conventional electrochemical technique but also by the quantum-optical way.

(2) A neuron as a quantum-optical device is akin to a parallel-pumped laser in which the ferroelectric membrane is a lasing medium.

(3) The quantum-chemical properties of ferroelectric nanotubes and nanomembranes can lay the groundwork for creating artificial quantum-optical devices for energy and data conversion.

REFERENCES

1. M. V. Vol'kenshteĭn, *Biophysics* (Nauka, Moscow, 1981).

2. J. S. Blakemore, *Solid State Physics* (Cambridge Univ. Press, Cambridge, 1985; Mir, Moscow, 1988).

3. O. V. Betskiĭ and N. D. Devyatkov, Radiotekhnika, No. 9, 4 (1996).

4. K. N. Trueblood, in *Accurate Molecular Structures*, Ed. by A. Domenicano and I. Hargittai (Oxford Univ. Press, Oxford, 1992; Mir, Moscow, 1997).

5. B. P. van Eijck, in *Accurate Molecular Structures*, Ed. by A. Domenicano and I. Hargittai (Oxford Univ. Press, Oxford, 1992; Mir, Moscow, 1997).

6. V. M. Stepanov, *Molecular Biology* (Vysshaya Shkola, Moscow, 1996).

7. J. M. Ziman, *Models of Disorder: the Theoretical Physics of Homogeneously Disordered Systems* (Cambridge Univ. Press, Cambridge, 1979; Mir, Moscow, 1982).

8. S. M. Shevchenko, *Molecule in Space: Problems of Modern Chemistry* (Khimiya, Leningrad, 1986).

9. V. F. Antonov *et al.*, *Biophysics: Textbook* (Arktos-Vika-Press, Moscow, 1996).

10. Ya. P. Terletskiĭ and Yu. P. Rybakov, *Electrodynamics: Textbook* (Vysshaya Shkola, Moscow, 1990, 2nd ed.).

11. E. Wichmann, *Quantum Physics* (McGraw-Hill, New York, 1967; Nauka, Moscow, 1986).

12. S. G. Ryabov, G. N. Toropkin, and I. F. Usol'tsev, *Devices of Quantum Electronics*, Ed. by M. F. Stel'makh (Radio i Svyaz', Moscow, 1985, 2nd ed.).

13. A. I. Akhiezer, *Atomic Physics: Reference Book* (Naukova Dumka, Kiev, 1988).

14. *Human Physiology*, Ed. by R. F. Schmidt and G. Thews (Springer-Verlag, Berlin, 1989; Mir, Moscow, 1996), Vol. 1.

15. S. L. Grigor'ev, Zh. Tekh. Fiz. **71** (11), 110 (2001) [Tech. Phys. **46**, 1457 (2001)].

16. *Problems of Optical Holography*, Ed. by Yu. N. Denisyuk (Nauka, Leningrad, 1981).

Translated by V. Isaakyan

<div style="text-align: center">

===== **THEORETICAL AND MATHEMATICAL PHYSICS** =====

The Chaotic Layer of a Nonlinear Resonance under Low-Frequency Perturbation

V. V. Vecheslavov

</div>

Budker Institute of Nuclear Physics, Siberian Division, Russian Academy of Sciences,
pr. Akademika Lavrent'eva 11, Novosibirsk, 630090 Russia

Received May 10, 2001

Abstract—Conditions whereby the chaotic layer of a nonlinear resonance is described in terms of low-frequency separatrix mapping are discussed. In this case, the accurate estimation of the size of the layer requires the arrangement of resonances at its edge to be known. The resonance picture is constructed using the separatrix mapping invariants of the first three orders. The variation of the layer size with the mapping amplitude is traced with the criterion for resonance overlapping. Results obtained by direct calculation and by invariants analysis are compared. Issues that remain to be solved are noted. © 2002 MAIK "Nauka/Interperiodica".

INTRODUCTION

The occurrence and the characterization of a chaotic layer appearing near the nonlinear resonance separatrix in the presence of other resonances is a central and still poorly understood problem in modern nonlinear dynamics [1–3]. In many cases, the Hamiltonian of the problem can be represented as a pendulum (fundamental resonance) and two perturbing resonances:

$$H(y, p, t) = \frac{p^2}{2} + \omega^2 \cos(y) + V(y, t),$$

$$V(y, t) = \varepsilon_1 \cos(y - \Omega_1 t) + \varepsilon_2 \cos(y - \Omega_2 t). \tag{1}$$

A perturbation is assumed to be weak (ε_1, $\varepsilon_2 \ll 1$), and the frequencies $|\Omega_1|$, $|\Omega_2| > \omega$ are not necessarily high. They may have both the same and opposite signs. It is also assumed that the resonances of system (1) do not overlap.

Let the chaotic layer of the fundamental resonance of system (1) be called the fundamental chaotic layer. It can be subdivided into three parts: upper (the phase y rotates at the top, $p > 0$), middle (the phase oscillates), and lower (the phase rotates at the bottom, $p < 0$). It is important that, in the general case of an asymmetric perturbation, the sizes w_u, w_m, and w_l of these parts may be much different (hereafter, $w = p^2/2\omega^2 + \cos y - 1$ is the dimensionless energy deviation from the unperturbed separatrix) [4–6].

The width of any of the parts can be determined by two methods. The former is finding the minimal period of motion $T_{\min}$ (T is the time interval between successive instants the orbit crosses the stable phase $y = \pi$) and the calculation of the maximal deviation from the unperturbed separatrix by the formula $w_{\max} \approx 32 \exp(-\omega T_{\min})$ [1]. The second one is the so-called separatrix mapping

[7], which makes it possible to approximately describe the system dynamics near the separatrix at instants of stable equilibrium. Note that the second way is much more feasible and is used much more frequently.

To date, the correct procedures for separatrix mapping and determining the sizes of the fundamental chaotic layer have been worked out only for a symmetric high-frequency perturbation $\varepsilon_1 = \varepsilon_2 = \varepsilon$, $\Omega_1/\omega = -\Omega_2/\omega = \lambda \gg 1$. Chirikov [1], using his resonance overlap criterion and the properties of standard mapping, has shown that separatrix mapping in the symmetric case has the form

$$\bar{w} = w + W \sin(x), \quad \bar{x} = x + \lambda \ln \frac{32}{|\bar{w}|} \bmod 2\pi, \tag{2}$$

where the amplitude is given by

$$W = 8\pi\varepsilon\lambda^2 e^{-\pi\lambda/2} \tag{3}$$

and the three parts of the layer are of the same size:

$$w_u = w_m = w_l \approx \lambda W\left(1 + \frac{4\delta}{\lambda}\right), \quad 0 \le \delta \le 1. \tag{4}$$

The correction $4\delta/\lambda$ appearing in (4) is a very complicated function of W and λ (see below), and its value depends on the resonance overlap at the edge of the layer. If $\lambda \gg 1$, it is neglected; if, however, a perturbation is low-frequency or accurate estimates are needed, it should be taken into account and the resonance arrangement becomes much more difficult to visualize and analyze. It should be noted that the case when the frequencies Ω_1 and Ω_2 of system (1) are low is seldom encountered and has not attracted much attention until now. However, the situation noticeably changed after it had been discovered that so-called secondary harmonics have an anomalously strong effect on the formation

of the fundamental chaotic layer. Relevant data will be discussed in the next section.

SECONDARY HARMONICS IN SEPARATRIX MAPPING

The early study of the general asymmetric perturbation of system (1) [4, 5] has shown that, along with the primary harmonics of frequencies Ω_1 and Ω_2 (entering into Hamiltonian (1) in the explicit form), the separatrix mapping spectrum contains also secondary harmonics with frequencies $\Omega_+ = \Omega_1 + \Omega_2$ and $\Omega_- = \Omega_2 - \Omega_1$. The still more intriguing thing is that even very weak low-frequency harmonics, which exponentially depend on the frequency (3), may, under certain conditions, almost completely define the width of any of the parts of the chaotic layer (for details, see [4–6]). Because of this, the separatrix mapping of this part can be represented in one-frequency form (2) but the parameter λ can no longer be considered to be large. It appears as if a one low-frequency harmonic, instead of two perturbing high-frequency ones, exists (this assumption has been confirmed numerically in [5]). Accordingly, it is necessary to construct and analyze the low-frequency separatrix mapping in order to determine the chaotic layer size.

The first difficulty here is that the amplitudes of the secondary harmonics are unknown. A rigorous theory is still lacking, but, within the approximate approach suggested in [4], the separatrix mapping amplitudes W at the sum and difference of the primary harmonic frequencies are given by the analytic expressions

$$W_+ = \frac{4\pi}{3}a_+\varepsilon_1\varepsilon_2\left[\frac{1}{\lambda_1^2} + \frac{1}{\lambda_2^2}\right]\frac{e^{\pi\lambda_+/2}}{\sinh(\pi\lambda_+)}\lambda_+^2(\lambda_+^2 - 2), \quad (5)$$

$$W_- = -\frac{\pi a_-\varepsilon_1\varepsilon_2}{\cosh(\pi\lambda_-/2)}\left\|\left[\frac{1}{\lambda_1^2} - \frac{1}{\lambda_2^2}\right]\right\|, \quad (6)$$

where $\lambda_1 = \Omega_1/\omega$, $\lambda_2 = \Omega_2/\omega$, $\lambda_+ = (\Omega_1 + \Omega_2)/\omega$, and $\lambda_- = (\Omega_2 - \Omega_1)/\omega$; a_+ and a_- are adjusting coefficients.

It has been shown [5] that formulas (5) and (6) fit well the numerical data if we put $a_+ = 0.473$ and $a_- = 1.35$. Yet, the nature of these corrections remains unclear. The parameter ranges where the secondary harmonic completely specifies the size w_u of the upper part of the chaotic layer were also reported [5]. Note that the outer parts of the layer are of the greatest practical interest, since they define the resonance overlap and contribute to the formation of global chaos.

As was noted above, the accurate estimation of the layer size requires the knowledge of the resonance arrangement at its edges. The arrangement can be found without determining the trajectories of motion if so-

called resonant invariants are known. The construction of the invariants for system (2) will be discussed below.

RESONANCE INVARIANTS

Separatrix mapping (2), as well as the well-known Chirikov standard mapping, falls into the class of explicit rotation mapping of type [2]

$$\bar{J} = J - \xi\frac{dQ(\Theta)}{d\Theta}, \quad \bar{\Theta} = \Theta + 2\pi\nu(\bar{J}), \quad (7)$$

where J and Θ are action and angular variables, $\nu(J)$ is the frequency of unperturbed motion, and $Q(\Theta + 2\pi) = Q(\Theta)$ is a trigonometric polynomial in the general case.

Difference equations (7) are known to be totally equivalent to Hamilton continuous equations with a kick-like perturbation:

$$\dot{\Theta} = \nu(J), \quad \dot{J} = -\xi\delta_*(t)\frac{dQ(\Theta)}{d\Theta}, \quad (8)$$

where $\delta_* = [1 + 2\sum_{n \geq 1}\cos(nt)]/2\pi$ is the periodic delta function [1, 2].

To separatrix mapping (2), there correspond the relationships

$$Q(\Theta) = \cos(\Theta), \quad \nu(J) = \frac{\lambda}{2\pi}\ln\frac{32}{|J|}, \quad (9)$$

whereas for the standard mapping, we have

$$Q(\Theta) = \cos(\Theta), \quad \nu(J) = J. \quad (10)$$

Recall that the arrangements of integer resonances ($\nu(J_n) = n$, where n is an integer) on the phase plane (J, Θ) for mappings (9) and (10) diverge. For the standard mapping, the resonance spacing $\delta J_n = J_{n+1} - J_n$ is J_n-independent (hence, standard mapping is also referred to as a homogeneous model), while for the separatrix mapping, it depends on J_n.

If

$$\delta J_n \ll J_n, \quad \lambda \gg 1, \quad (11)$$

the separatrix mapping locally (near the resonance J_n) transforms into the standard one (this fact was taken into account in deriving formulas (4) in [1]).

In [8], the function of dynamic variables and time of type

$$S_m(J(t), \Theta(t), t, \xi) = r_0(J) + \sum_{n=1}^{m}\frac{\xi^n}{n!}G_n(J, \Theta, t), \quad (12)$$

was called the invariant of order m. For this function, the condition

$$\frac{D}{Dt}S_m(t) = \mathcal{O}(\xi^{m+1}) \quad (13)$$

is met along the true trajectories of system (8). Here, D/Dt is the generalized derivative that is the sum of the ordinary derivative and that due to component discontinuities [9].

It is general practice to express the periodic orbit and the associated resonance through the ratio $N : M$, where M is the number of iterations of mapping onto N orbit periods [2]. In [8], the invariants of the first three orders for the standard mapping were constructed. They describe the integer, $M = 1$, and fractional, $M = 2$ and 3, resonances. It was also shown in that work that the terms involved in (12) retain their values in the intervals between kicks and undergo jumps only at the instants of kicks. Consequently, the ordinary derivative of $S_m(t)$ in (12) is everywhere zero and condition (13) should be realized as the restriction imposed on the discontinuity of the first kind in the step function.

The potentials of systems (9) and (10) are the same, $Q(\Theta) = \cos(\Theta)$, so that the expressions obtained in [8] can be rewritten in the form suitable for both mappings:

$$G_1(J, \Theta, t) = \frac{r_0'(J)}{2\sin(\pi\nu(J))}\cos\Omega, \qquad (14)$$

$$G_2(J, \Theta, t) = \frac{a_1'(J)}{4\sin(2\pi\nu(J))}\cos 2\Omega, \qquad (15)$$

$$\begin{aligned}G_3(J, \Theta, t) &= \frac{r_0'''(J) + 3a_2'(J)}{16\sin(3\pi\nu(J))}\cos 3\Omega \\ &\quad - 3\frac{r_0'''(J) + a_2'(J)}{16\sin(\pi\nu(J))}\cos\Omega.\end{aligned} \qquad (16)$$

In these formulas, the primed quantities are derivatives with respect to J and the following dependences are used:

$$\Omega = \Theta + (\pi - t_*)\nu(J), \quad a_1(J) = r_0'(J)\cot(\pi\nu(J)),$$
$$a_2(J) = a_1'(J)\cot(2\pi\nu(J)), \qquad (17)$$

where $0 < t_* \le 2\pi$ is the local time between kicks.

Expressions (14)–(16) for the invariants of the first three orders have small denominators. Following the Dunnet–Laing–Taylor (DLT) method [10], they can be suppressed by appropriately selecting the zero-order invariant $r_0(J)$ in formula (12) [more strictly, its first derivative $r_0'(J)$]. It turned out that the selection of $r_0'(J)$ is not universal and depends on the number m. The DLT invariant of order m adequately reflects the topology of only "its own" resonances with $M = m$, which appear for the first time just in this order and are responsible for "their own" part of the phase space (for details, see [8]). We have shown that the specific dependences $r_0'(J)$ suggested in [8] for the standard mapping can also be used for separatrix mapping (2). Accurate

formulas in the initial variables w and x for this case are given in the Appendix.

It is the author's opinion that he was the first to introduce the invariants S_1 and S_2 for the separatrix mapping of the first two orders [6]. In this work, we will also discuss S_3 and provide full information on all the three invariants.

DIRECT CALCULATION AND ANALYSIS OF INVARIANTS

The multiple iteration of mapping (2) allows one to find the sizes of all the parts of the fundamental chaotic layer and to trace their dependence on the amplitude W and frequency λ. However, it would be appropriate to considerably simplify this process and, moreover, make correct predictions. Formulas (1A)–(3A) for the separatrix mapping invariants (see the Appendix) combined with the resonance overlap criterion seem to furnish such an opportunity. Below is an example of their use.

In [5], system (1) was investigated for the following parameter values: $\omega = 1.0$, $0.01 \le \varepsilon_1 = \varepsilon_2 \le 0.5$, $8.0 \le \Omega_1 \le 18.0$, and $\Omega_2 = -10.0$. The direct calculation and the spectral analysis of the separatrix mapping indicated that the secondary harmonic is the most intense at the sum of the frequencies $\Omega_+ = \Omega_1 + \Omega_2 \approx 3.0$. In this case, its contribution to the separatrix mapping of the upper part of the layer is several hundred times larger than the net contribution from the primary frequencies. For this reason, we will subsequently be interested in mapping (2) with $\lambda = 3.0$. Clearly, the correction $4\delta/\lambda$ in formula (4) cannot be neglected. The object of investigation will be the upper part of the layer, which is w_u thick. Based on formula (4), we introduce the quantity

$$D_u(W) = \frac{w_u(W)}{\lambda W}, \qquad (18)$$

which is the normalized size of this part [1]. In the high-frequency limit ($\lambda \longrightarrow \infty$), $D_u(W)$ does not depend on W and tends to unity.

Figure 1 shows the dependence $D_u(W)$ obtained by numerically iterating mapping (2) (the number of iterations for each of the points is 10^8). The dependence has many discontinuities. Some of them are clearly seen and others can be visualized on an enlarged scale. Within the modern dynamic theory, such a pattern is explained by the fact that, as W grows, the invariant curves with irrational numbers of rotation successively break down and are replaced by so-called Cantori [11]. If an invariant curve is the boundary between the fundamental chaotic layer and the $N : M$ resonance nearest to it, these objects merge together when the curve breaks down and the size of the layer increases by a finite quantity, i.e., by the phase volume of the resonance attached. The exact value of the mapping amplitude W at which the layer and the resonance merge is called the threshold value and is designated as $W(N : M)$.
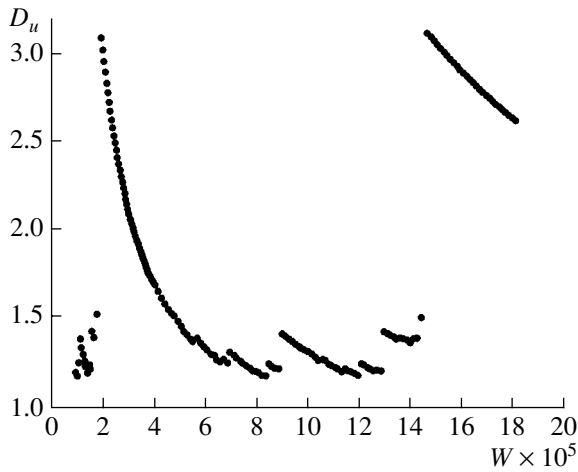
Fig. 1. Normalized width (18) of the upper part of the fundamental chaotic layer of system (1) as a function of the amplitude W of separatrix mapping (2) for $\lambda = 3.0$.

It was noted [1] that the size increases to the greatest extent when the fundamental layer overlaps with an integer resonance of the separatrix mapping. Figure 1 shows two such cases near $W \approx 1.81 \times 10^{-5}$ (resonance

6 : 1) and $W \approx 1.45 \times 10^{-4}$ (resonance 5 : 1). Let us consider the former in greater detail. On the left of Fig. 2, the fundamental chaotic layer is depicted immediately after the integer resonance 6 : 1 has been attached to it. The islands of the 19 : 3, 13 : 2, and 7 : 1 stable resonances are seen in the chaotic sea. On the right of Fig. 2, the levels of the invariants for these four resonances constructed by formulas (A1)–(A3) of the Appendix are demonstrated (the resonance 20 : 3 is omitted, since the lines of its invariant cross those of the invariants 13 : 2 and 7 : 1 and the picture becomes very tangled).

From the picture of the invariants, one can make some estimates. For example, the normalized size of the upper part of the fundamental layer estimated from the maximal ordinate of the boundary invariant 6 : 1 is $D_u \approx 3.06$ with the exact value being 3.10. The line of the invariant 19 : 3 allows one to estimate the size of the layer prior to overlapping: 1.34 versus the exact value 1.53. The latter estimate is much cruder, which is explained as follows. Invariants, like any other analytical description, are basically unable to reflect the chaotic component of motion (in particular, to take into account the widths of the chaotic layers of resonances). We performed our measurements inside, rather than at
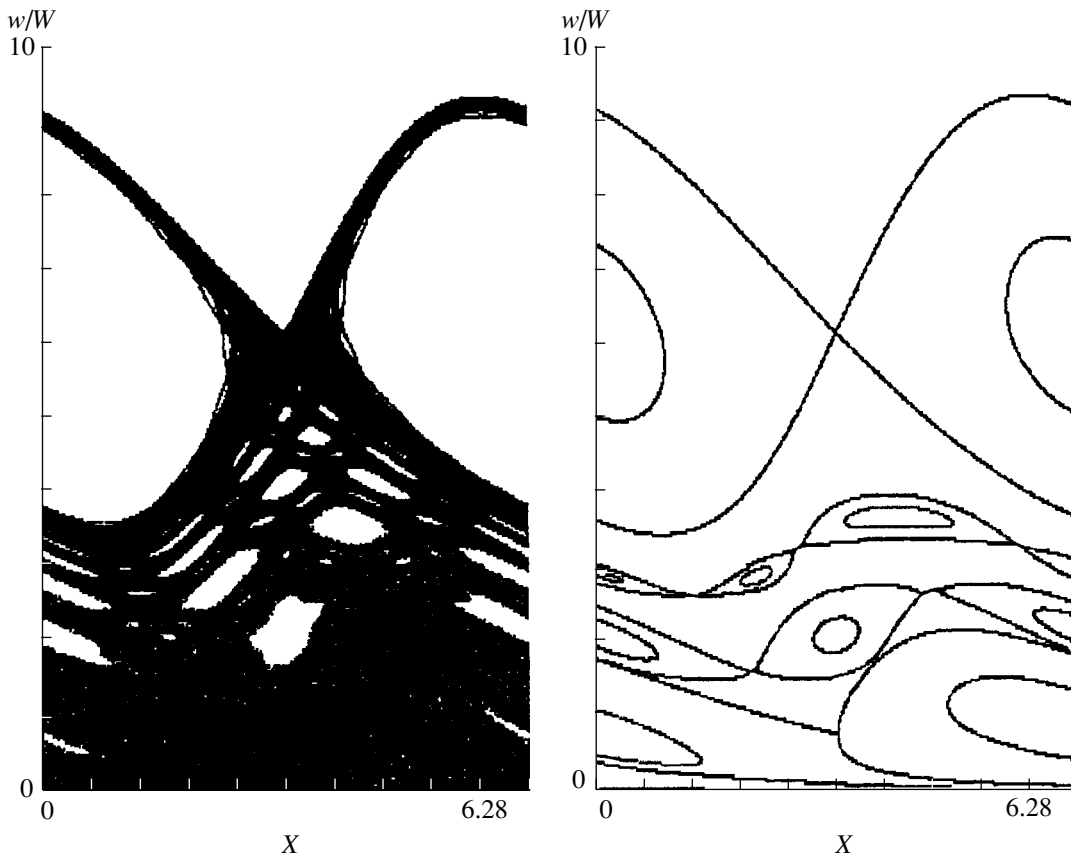
Fig. 2. (Left) the upper part of the fundamental chaotic layer immediately after its overlapping with the integer resonance 6 : 1; (right) resonance invariants (from top to bottom) 6 : 1, 19 : 3, 13 : 2, and 7 : 1. $W = 1.83 \times 10^{-5}$, $\lambda = 3.0$.
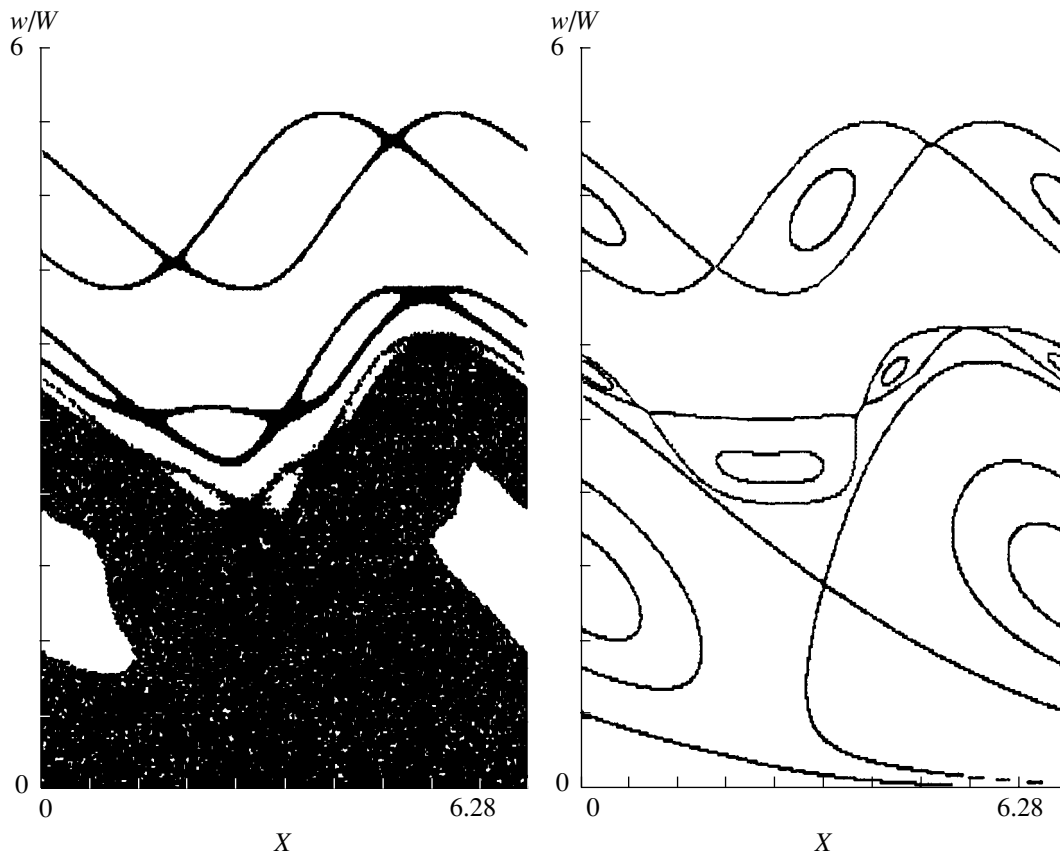
Fig. 3. (Left) three chaotic trajectories: the lower one is the fundamental chaotic layer; the other two are the chaotic layers of the resonances 17 : 3 and 11 : 2 of the separatrix mapping. The situation immediately before the fundamental layer overlaps with the resonance 17 : 3. (Right) the invariants of the resonances 6 : 1, 17 : 3, and 11 : 2 (from bottom to top). $W = 6.7 \times 10^{-5}$, $\lambda = 3.0$.

the edge, of the fundamental layer, where the effect of this component is significant, and obtained the inaccurate value. Figure 2 shows two adjacent resonances (the lower 7 : 1 and the upper 6 : 1), so that we can say that it depicts the instant of their overlapping. These resonances are seen to greatly differ, which is a direct consequence of the separatrix mapping nonuniformity mentioned above (recall that for the standard mapping, all integer resonances are identical). Our nearest goal is to trace the overlapping evolution for the integer resonances 6 : 1 and 5 : 1 by comparing the direct calculation and the analysis of the invariant arrangement.

Far above the upper boundary in Fig. 2, the resonances 17 : 3, 11 : 2, 16 : 3, and 5 : 1 of the separatrix mapping are arranged (from bottom to top). As W grows, their phase regions will "sink" in the fundamental chaotic layer and the attachment of the integer resonance 5 : 1 to this layer will be accompanied by a large discontinuity (the right of Fig. 1).

Figure 3 shows the situation with $W = 6.7 \times 10^{-5}$ immediately before the fundamental layer merges with the resonance 17 : 3. Using the invariants, one can, as before, estimate the size of the layer prior to overlapping and the amount of the discontinuity expected, both

values being in reasonable agreement with the data in Fig. 1. Analyzing the arrangement of the invariants near $W = 6.7 \times 10^{-5}$ and applying the resonance overlap criterion, we can expect overlapping in the interval $6.0 \times 10^{-5} < W < 7.5 \times 10^{-5}$. The accurate value turns out to be $W(17 : 3) = 6.954 \times 10^{-5}$. It is also seen that this value is far from that at which the overlapping with the resonance 11 : 2 will occur ($W(11:2) = 9.009 \times 10^{-5}$, Fig. 1).

Figure 4 illustrates the situation before the fundamental layer merges with the next integer resonance 5 : 1. The invariant picture depicts both integer resonances (5 : 1 at the top and 6 : 1 at the bottom). It follows then that Fig. 4 describes the last stage in their overlapping. Here, we also can indicate the interval ($1.35 \times 10^{-4} < W < 1.5 \times 10^{-4}$) in which the threshold value $W(5 : 1) = 1.453 \times 10^{-4}$ was found.

The above analysis was restricted by the only frequency value $\lambda = 3.0$, and only one stage of overlapping the integer resonances of separatrix mapping (2) was considered. It seems to be of interest to consider the transition from low frequencies to the asymptotic limit $\lambda \longrightarrow \infty$. It would also be useful to gain some experience in applying the effective resonance overlap criterion to the problem discussed and to devise a procedure
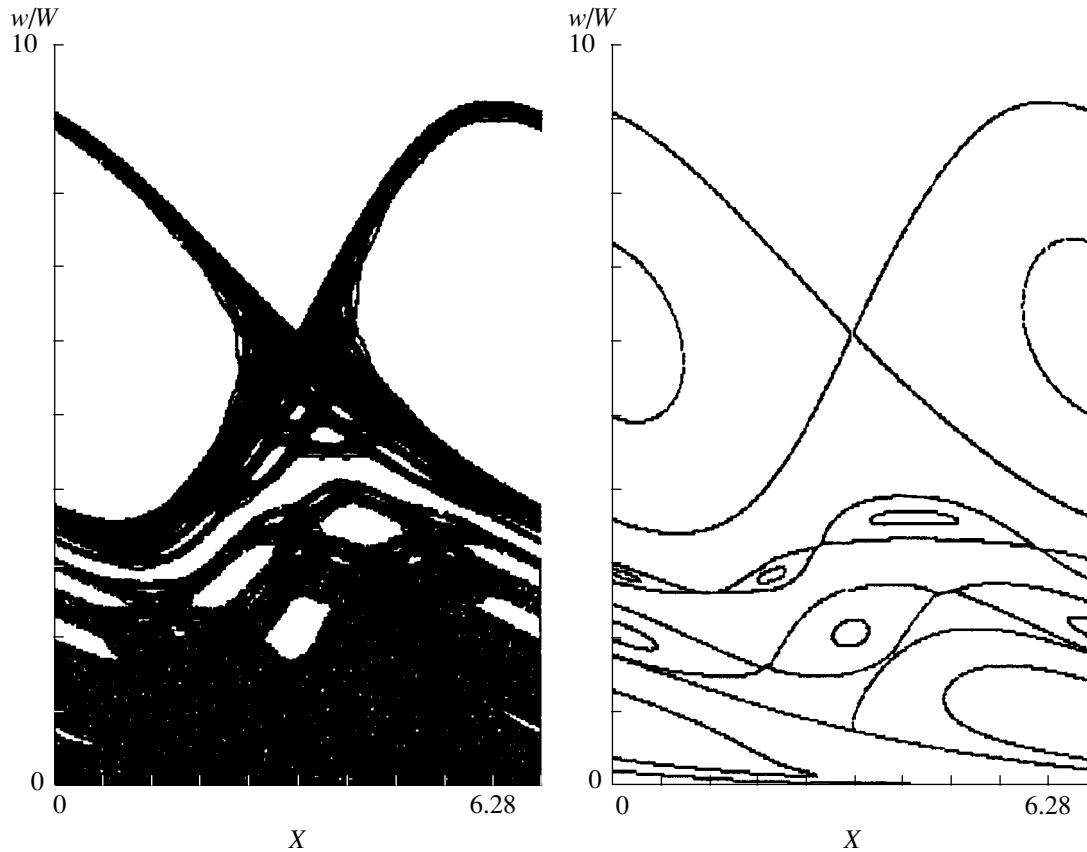
Fig. 4. (Left) two chaotic trajectories: the lower one is the fundamental chaotic layer; the other is the chaotic layer of the integer resonance 5 : 1 of the separatrix mapping. The situation immediately before the layers overlap. (Right) the invariants of the resonances 5 : 1, 16 : 3, 11 : 2, and 6 : 1 (from top to bottom). $W = 1.44 \times 10^{-4}$, $\lambda = 3.0$.

for more accurately determining the threshold values of the mapping parameter $W(N : M)$.

CONCLUSION

From the previous section, we can conclude that DLT invariants reflect the topology of "their" resonances fairly adequately and can be used for reliably estimating the sizes of the fundamental chaotic layer. Using the well-known Chirikov resonance overlap criterion and knowing the arrangement of the invariants, one can find the interval of the parameter W where the chaotic layer may merge with the nearest resonance $N : M$ of the separatrix mapping. To find the exact value of the threshold $W(N : M)$ corresponding to this event, it will suffice to perform numerical analysis only within this interval.

Note that the construction of the complete invariant picture on the phase plane may be unnecessary. For rough estimates, it would suffice to find the length and the arrangement of segments belonging to different resonances on the dominant line $x = 0$, which is the locus of the centers of all resonances.

It should be emphasized that the concept of invariants may lay the foundations of an advanced theory. In the Appendix, $S_m(w, x)$ is written in the starting variables w and x. However, one can construct the generating functions $\mathscr{F}(I, x)$ of the canonic transformation into other variables I and ψ and find the invariant $S_m(I)$ as a function of the new action I. First, this simplifies the study of the invariant itself, since the level line admits the explicit representation in the form $I(w, x) = \text{const}$. Second, this allows one to construct the Hamiltonians for specific resonances and find, for example, the frequencies of motion. The associated procedure for the standard mapping was described in [8].

Unfortunately, the DLT theory of invariants has never come to the attention of mathematicians. Because of this, the construction of invariants is to a great extent a matter of guessing and intuition (the singularity-eliminating dependences $r'_0(\omega)$ in the Appendix have been found just in this way). This circumstance makes the advance to higher m very difficult. Another obstacle is the fact that the analytic expressions become much more awkward as m grows. It has been reported [2, 8] that the DLT method requires the elimination of all the poles of all resonances $M = m$ so that the invariant S_m becomes an analytic function throughout the phase space. It seems plausible that the abandonment of the

global analyticity requirement might facilitate the solution to the problem [8]. Note that the construction of the first invariant for the case when the singularities are eliminated only from some of neighboring resonances (of the entire countable set of integer resonances) is exemplified in [12].

It was the aim of this work to draw attention to the DLT method for constructing resonance invariants, which, in our opinion, deserves further development and refinement.

ACKNOWLEDGMENTS

APPENDIX

First Three Invariants of Mapping (2)

In constructing the invariants of separatrix mapping (2), it is convenient to return from the variables J, Θ to the early variables w, x. Therefore, primed quantities in the Appendix refer to derivatives with respect to w. It was found that the direct application of formulas (14)–(17) to this mapping is difficult because of numerical instability involved in the construction of the contour curves for the invariants. The situation is improved when the new variable

$$z(w) = \frac{\lambda}{2}\ln\frac{32}{|w|}$$

is introduced instead of w.

The resonance picture on the phase plane (w, x) is correct if the instant of observation t_* appearing in (17) is set equal to 2π.

The specific dependences $r_0'(w)$ suggested in [8] with the aim of suppressing small denominators in the case of the standard mapping proved to be applicable also for separatrix mapping (2). In terms of our designations, we have

$$r_0' = \sin z,$$

for the first-order invariant,

$$r_0' = \sin^2 z \sin 2z = \frac{2\sin 2z - \sin 4z}{4}$$

for the second-order invariant, and

$$r_0' = \sin^2 z \sin^2 3z \sin 3z$$

$$= \frac{1}{16}(\sin z + 3\sin 3z - \sin 5z - 2\sin 7z + \sin 9z)$$

for the third-order invariant.

Below we give the complete expressions for the first three invariants of separatrix mapping (2):

$$S_1(z(w), x) = e^{-2z/\lambda}\left[\sin z + \frac{\lambda}{2}\cos z\right]$$
$$+ \frac{W}{64}\left[1 + \frac{\lambda^2}{4}\right]\cos(x - z), \tag{A1}$$

$$S_2(z(w), x) = 16e^{-2z/\lambda}$$
$$\times\left[\frac{\sin 2z + \lambda\cos 2z}{1 + \lambda^2} - \frac{\sin 4z + 2\lambda\cos 4z}{2(1 + 4\lambda^2)}\right]$$
$$+ \frac{W}{2}\sin z \sin 2z\cos(x - z) \tag{A2}$$
$$- \lambda e^{2z/\lambda}\left(\frac{W}{16}\right)^2\cos 2z\cos 2(x - z),$$

$$S_3(z(w), x) = 4e^{-2z/\lambda}f_0(z)$$
$$+ \frac{W}{2}\sin z \sin^2 2z \sin 3z\cos(x - z) - 3\lambda e^{2z/\lambda}$$
$$\times\left(\frac{W}{32}\right)^2\sin z \sin 2z(2 + 3\cos 2z + 3\cos 4z) \tag{A3}$$
$$\times\cos 2(x - z) + \frac{\lambda}{6}\left(\frac{e^{2z/\lambda}}{512}\right)^2$$
$$\times W^3[f_1(z)\cos 3(x - z) - 3f_2(z)\cos(x - z)].$$

In the last expression, we used the designations

$$f_0(z) = g(z) + 3g(3z) - g(5z) - 2g(7z) + g(9z),$$

$$g(nz) = \frac{2\sin nz + n\lambda\cos nz}{1 + (n\lambda/2)^2},$$

$$f_1(z) = \lambda(27 + 35\cos 2z + 49\cos 4z + 81\cos 6z)$$
$$+ 14\sin 2z + 14\sin 4z + 18\sin 6z,$$

$$f_2(z) = \lambda(23 + 51\cos 2z + 69\cos 4z + 49\cos 6z)$$
$$+ 10\sin 2z + 22\sin 4z + 14\sin 6z.$$

It is worth noting that if conditions (11) are met, the invariants of separatrix mapping pass to those of standard mapping.

REFERENCES

1. B. V. Chirikov, Phys. Rep. **52**, 263 (1979).
2. A. Lichtenberg and M. Lieberman, *Regular and Chaotic Dynamics* (Springer-Verlag, New York, 1992).
3. R. Z. Sagdeev, D. A. Usikov, and G. M. Zaslavsky, *Nonlinear Physics: from the Pendulum to Turbulence and Chaos* (Harwood, Chur, 1988); G. M. Zaslavsky and R. Z. Sagdeev, *Introduction to Nonlinear Physics: from*

the Pendulum to Turbulence and Chaos (Nauka, Moscow, 1988).

4. V. V. Vecheslavov, Zh. Éksp. Teor. Fiz. **109**, 2208 (1996) [JETP **82**, 1190 (1996)].

5. V. V. Vecheslavov, Pis'ma Zh. Éksp. Teor. Fiz. **63**, 989 (1996) [JETP Lett. **63**, 1047 (1996)].

6. V. V. Vecheslavov, Physica D (Amsterdam) **131**, 55 (1999).

7. G. M. Zaslavskiĭ and N. N. Filonenko, Zh. Éksp. Teor. Fiz. **54**, 1590 (1968) [Sov. Phys. JETP **27**, 851 (1968)].

8. V. V. Vecheslavov, Zh. Tekh. Fiz. **58**, 20 (1988) [Sov. Phys. Tech. Phys. **33**, 11 (1988)].

9. I. M. Gel'fand and G. E. Shilov, *Generalized Functions* (Fizmatgiz, Moscow, 1958; Academic, New York, 1964).

10. D. A. Dunnet, E. W. Laing, and J. B. Taylor, J. Math. Phys. **9**, 1819 (1968).

11. R. S. MacKay, J. D. Meiss, and I. C. Percival, Physica D (Amsterdam) **13**, 55 (1984).

12. V. V. Vecheslavov, Zh. Tekh. Fiz. **57**, 896 (1987) [Sov. Phys. Tech. Phys. **32**, 544 (1987)].

Translated by V. Isaakyan

THEORETICAL AND MATHEMATICAL PHYSICS

Structural Factor for Bose–Einstein Double-Component Condensate

I. V. Kazinets*, B. G. Matisov*, and I. E. Mazets**

* St. Petersburg State Technical University,
ul. Politekhnicheskaya 29, St. Petersburg, 194251 Russia
e-mail: quark@citadel.stu.neva.ru
** Ioffe Physicotechnical Institute, Russian Academy of Sciences,
ul. Politekhnicheskaya 26, St. Petersburg, 194021 Russia
Received June 27, 2001

Abstract—The structural factors for the single- and double-component Bose–Einstein condensates of neutral atoms in a spherically symmetric harmonic trap are derived. The approach adopted in this paper is applicable in the low excitation energy limit, where the well-known Stringari dispersion relation is valid. Our expression is compared with the quasi-classical one, which is shown to overestimate the structural factor. It is argued that the second component first splits the frequency of collective condensate oscillations, which corresponds to the energy transferred by scattering, into two branches and then multiplies the single-component structural factor by some coefficient that can be greater than unity. The analysis is used to estimate the properties of an experimentally implemented double-component condensate of ^{87}Rb atoms. © 2002 MAIK "Nauka/Interperiodica".

INTRODUCTION

The dynamic structural factor is an important parameter that specifies the behavior of many-particle systems. In particular, it provides information on the collective excitation spectrum upon small transferred momentum scattering and characterizes the momentum distribution in the case that a relatively large momentum is transferred; i.e., that the system response is governed by single-particle effects. For rarefied gases, the structural factor can be measured experimentally, for example, by inelastic scattering of light by a boson gas confined by a magnetic trap [1].

As a many-particle system, we consider the Bose–Einstein condensate (BEC) of a rarefied gas in a spherically symmetric harmonic trap. Particle interaction is assumed to be weak. In this case, the well-known Bogoliubov theory for a weakly interacting degenerate boson gas [2] and its modifications taking into account the external trapping potential [1, 3] apply.

There are several works dealing with the calculation of the structural factor for the BEC of a slightly imperfect gas. Both spatially homogeneous [5, 6] and spatially inhomogeneous gases [7–9] have been considered. The calculations for the former case are rather straightforward, while for a BEC in a harmonic trap the problem cannot be solved exactly and requires approximations to be used, such as the quasi-classical [7] or the local-density or momentum approximations [9]. The above approximations apply if the momentum transferred is relatively large. For example, in [9], $q \gg 1/R$, where R is the BEC size and $\hbar q$ is the momentum transferred by scattering. For small momenta, the response of the system is sensitive to the discrete modes of its collective oscillations. For small transferred momenta, the dynamic structural factor of a single-component BEC has been theoretically analyzed in [8], where the formalism of quantizing the hydrodynamic equations that describe collective oscillations in a BEC and the Thomas–Fermi approximation were used.

In this work, we will derive the dynamic structural factors for single- and double-component BECs in the Thomas–Fermi approximation. Our formalism is simple and more illustrative than that used in [8]. Unlike [8], where only the general approach is stated, we derive formulas for the structural factors in a closed analytical form. The expression for the single-component condensate is compared with the quasi-classical one obtained in [7]. Within the domain of its applicability, the quasi-classical expression is an overestimation. However, both our expression and the quasi-classical one [7] equal zero at the same values of the transferred momentum $\hbar \mathbf{q}$. Therefore, within the domain of its applicability, the latter expression can be used for calculating those values of the transferred momentum $\hbar \mathbf{q}$ making the structural factor vanish.

For a double-component condensate, the q dependence of the partial structural factor (photon scattering by atoms of one component) does not change, but the energy transferred by scattering splits into two branches, for both of them the structural factor being multiplied by either the coefficient β^+ or β^-. It is noteworthy that β may exceed unity for a certain combination of the scattering amplitudes of the components,

suggesting that the presence of the second component may enhance the photon scattering by the other.

SINGLE-COMPONENT CONDENSATE

As is known, the differential scattering cross section $d^2\sigma_N/(d\Omega d\omega)$ for a particle being scattered by a sample of N atoms is expressed through the cross section $d\sigma_1/d\Omega$ of one-atom scattering as [7, 10]

$$\frac{d^2\sigma_N}{d\Omega d\omega} = S(\mathbf{q}, \omega)\frac{d\sigma_1}{d\Omega},$$

where $\hbar\mathbf{q}$ is the momentum imparted to the sample by a particle scattered into a solid angle $d\Omega$; $\hbar\omega$ is the associated transferred energy; and $S(\mathbf{q}, \omega)$ is the dynamic structural factor of the sample, which is given by [10]

$$S(\mathbf{q}, \omega) = \sum_f \tilde{S}_f(\mathbf{q})\delta(\omega - \omega_f), \qquad (1)$$

$$\tilde{S}_f(\mathbf{q}) = \left|\langle f|\int d\mathbf{r}\,\hat{\psi}^\dagger(\mathbf{r})\exp(i\mathbf{qr})\hat{\psi}(\mathbf{r})|0\rangle\right|^2.$$

Here, $|0\rangle$ is the ground state of the BEC; $\langle f|$ and $\hbar\omega_f$ are the excited state and the excitation energy of the BEC, respectively; and $\hat{\psi}(\mathbf{r})$ is the field operator introduced in the secondary quantizing formalism for a many-particle system (BEC). (The designation $(\mathbf{qr})$ means the scalar product.) By a particle being scattered, we mean a photon; however, it may also be another object (for example, a neutron) in the general case. For simplicity, we adopt the formalism $T = 0$, since modern experimental equipment makes it possible to produce an ultracold atomic ensemble with a negligibly small above-condensate fraction [11].

According to the Bogoliubov approximation, we represent the field operator of an atom as the sum of the c-numerical part (corresponding to the BEC) and a small operator part corresponding to collective oscillations of the BEC. The latter, in turn, can be represented through the operators of production, $\hat{\alpha}_f^\dagger$, and annihilation, $\hat{\alpha}_f$, of quasi-particles responsible for the disturbance of a condensate with an energy $\hbar\omega_f$ (see, e.g., [4]):

$$\hat{\psi}(\mathbf{r})$$

$$= \exp(i\phi_0)\left[\psi_0(\mathbf{r})\sqrt{N_0} + \sum_f (u_f(\mathbf{r})\hat{\alpha}_f - v_f^*(\mathbf{r})\hat{\alpha}_f^\dagger)\right], \quad (2)$$

where ϕ_0 is the constant phase of the condensate; ψ_0 is the real function satisfying the stationary Gross–Pitaevski equation and normalized to unity,

$$\int d\mathbf{r}\,\psi_0^2(\mathbf{r}) = 1, \qquad (3)$$

and N_0 is the number of atoms in the condensate.

The Bogoliubov coefficients u_f and v_f are normalized as

$$\int d\mathbf{r}[u_f(\mathbf{r})u_f^*(\mathbf{r}) - v_f(\mathbf{r})v_f^*(\mathbf{r})] = \delta_{ff}. \qquad (4)$$

This condition yields the standard boson commutation relations for the quasi-particle operators. Substituting expansion (2) into (1), we have

$$\tilde{S}_f(\mathbf{q}) = N_0\left|\int d\mathbf{r}\exp(i\mathbf{qr})[u_f^*(\mathbf{r}) - v_f^*(\mathbf{r})]\psi_0(\mathbf{r})\right|^2. (5)$$

Thus, to calculate the dynamic structural factor, one must know analytical expressions for u_f and v_f, which are usually found by solving a set of differential equations. This set can be solved numerically [12]. However, the analytical solutions require additional approximations to be used. In particular, the Thomas–Fermi approximation [13, 14] is applicable if the number N_0 of condensate atoms is large; that is, $N_0 \gg a_{HO}/a$, where

$$a_{HO} = \sqrt{\hbar/(M\omega_0)} \qquad (6)$$

is the length of a spherically symmetric harmonic trap with a frequency ω_0, M is the atomic mass, and a is the amplitude of s-scattering of ultracold atoms by each other in the Born approximation. In this approximation, the kinetic energy is assumed to be much smaller than the energy of particle interaction and the energy of interaction with the external potential.

At temperatures much lower than that of Bose–Einstein condensation, the mean value of the boson field operator $\langle\hat{\psi}(\mathbf{r})\rangle = \Psi(\mathbf{r}, l)\exp[i\phi(\mathbf{r}, l)]$ satisfies the Gross–Pitaevski equation [15]:

$$i\hbar\frac{\partial}{\partial t}\Psi\exp(i\phi)$$

$$= \left[-\frac{\hbar^2\nabla^2}{2M} - \mu + V_{\text{ext}}(\mathbf{r}) + g\Psi^2\right]\Psi\exp(i\phi), \qquad (7)$$

where $g = 4\pi\hbar^2a/M$ and μ is the chemical potential of the BEC ground state and ∇ is the Hamiltonian operator.

Now we turn to the formalism [13], within which low-energy excitations are described in terms of the collisionless hydrodynamic approach that operates on the density perturbation $\delta n(\mathbf{r}, t)$ and hydrodynamic velocity $\mathbf{v}(\mathbf{r}, t)$. They are defined as

$$N_0\psi_0^2(\mathbf{r}) + \delta n(\mathbf{r}, t) = \Psi^2(\mathbf{r}, t),$$
$$M\mathbf{v}(\mathbf{r}, t) = \hbar\nabla\phi(\mathbf{r}, t). \qquad (8)$$

Considering δn and $\mathbf{v}$ as small parameters, we obtain, from (7), the linearized equation

$$\frac{\partial^2}{\partial t^2}\delta n(\mathbf{r}, t) = \frac{gN_0}{M}\nabla[\psi_0^2(\mathbf{r})\cdot\nabla\delta n(\mathbf{r}, t)]. \qquad (9)$$

When deriving (9), we assumed that the term

$$\frac{\hbar^2}{2M\Psi(\mathbf{r}, t)}\nabla^2\Psi(\mathbf{r}, t)$$

of the kinetic energy is small compared with the energy of particle interaction $g\Psi^2(\mathbf{r}, t)$ and the external potential $V_{ext}(\mathbf{r})$. This assumption is consistent with the Thomas–Fermi approximation and is valid at large N_0.

The function $\psi_0(\mathbf{r})$ is determined from (7) and, in the Thomas–Fermi approximation for a spherically symmetric trap ($V_{ext}(\mathbf{r}) = M\omega_0^2 r^2/2$), has the form

$$\psi_0^2(\mathbf{r}) = \frac{M}{gN_0}\frac{\omega_0^2}{2}R^2\left[1 - \left(\frac{1}{R}\right)^2\right] \qquad (10)$$

if $r \le R$ and $\psi_0^2(\mathbf{r}) = 0$ if $r > R$. Here, $R = \sqrt{2\mu/(M\omega_0^2)}$ is the BEC radius defined by the condition $\mu = V_{ext}(R)$. From normalization (3), we find

$$R = a_{HO}\left(15N_0\frac{a}{a_{HO}}\right)^{1/5}, \qquad (11)$$

where a_{HO} is given by (6).

Let us expand the real function $\delta n(\mathbf{r}, t)$ in elementary perturbations with a frequency ω_f:

$$\delta n(\mathbf{r}, t) = \sum_f [\delta n_f(\mathbf{r})\exp(-i\omega_f t) + \delta n_f^*(\mathbf{r})\exp(i\omega_f t)].$$

Then, (9) in view of (10) yields

$$-2\left(\frac{\omega_f}{\omega_0}\right)^2\delta n_f(\mathbf{r}) = \nabla_{\left(\frac{r}{R}\right)}\left\{\left[1 - \left(\frac{r}{R}\right)^2\right]\nabla_{\left(\frac{r}{R}\right)}\delta n_f(\mathbf{r})\right\}. (12)$$

For the trap considered, the appropriate quantum numbers for the elementary excitation of the fth mode are radial quantum number n, as well as orbital moment l and its projection m onto the z axis:

$$\delta n_f(\mathbf{r}) = \delta n_{nlm}(\mathbf{r}) = F_{nl}(r/R)Y_{lm}(\Theta, \phi),$$

$$F_{nl}(r) = \sum_{k=0}^n G_k^{(nl)}r^{l+2k}, \qquad (13)$$

where $Y_{lm}(\Theta, \phi)$ is the eigenfunction of the moment operator.

Substituting this expansion into (12), we come to the recurrence relation

$$C_{k+1}^{(nl)} = C_k^{(nl)}\frac{K_{kl} - (\omega_{nlm}/\omega_0)^2}{(k+1)(2l + 2k + 3)}, \qquad (14)$$

$$K_{kl} = 2k^2 + 2kl + 3k + l.$$

The boundedness condition for $\delta n_f|_{r=R}$ requires that the series be truncated at some finite n: $C_{n+1}^{(nl)} = 0$. As a result, we will come to the expression for the frequencies of collective oscillations in a large atomic ensem-

ble confined by a trap and having a temperature much less than that of Bose–Einstein condensation: $\omega_{nl} = \omega_0\sqrt{K_{nl}}$. Eventually, for an elementary perturbation of a condensate with a frequency ω_{nl}, we have

$$\delta n_{nlm}(\mathbf{r}) = \left[\sum_{k=0}^n C_k^{(nl)}\left(\frac{r}{R}\right)^{l+2k}\right]Y_{lm}(\Theta, \phi), \qquad (15)$$

where

$$C_k^{(nl)} = C_0(-1)^k\prod_{k'=0}^{k-1}\frac{(n-k')(2n+3+2l+2k')}{(k'+1)(3+2l+2k')}$$
$$= C_0(-1)^k\frac{n!}{k!(n-k)!}\frac{(1+2n+2l+2k)!!}{(1+2n+2l)!!}\frac{(1+2l)!!}{(1+2l+2k)!!}, \qquad (16)$$

R is defined by (11), and C_0 is a normalizing constant that is set equal to unity. Expression (15) applies to lower order modes with energies much lower than the BEC chemical potential.

To relate the Bogoliubov coefficients $u_f(\mathbf{r})$ and $v_f(\mathbf{r})$ with $\delta n_f(\mathbf{r})$, we assume that the ensemble of ultracold atoms is so excited that only the fth mode, which is the coherent state, is present; that is, only $\langle\hat{\alpha}_f\rangle = \lambda_f\exp(-i\omega_f t)$ and $\langle\hat{\alpha}_f^\dagger\rangle = \lambda_f^*\exp(i\omega_f t)$ are other than zero. Thus, we have

$$\langle\hat{\psi}(\mathbf{r})\rangle = \Psi(\mathbf{r}, l)\exp[i\phi(\mathbf{r}, l)]$$
$$= \exp(i\phi_0)[\psi_0(\mathbf{r})\sqrt{N_0} + \lambda_f u_f(\mathbf{r})\exp(-i\omega_f t) \qquad (17)$$
$$- \lambda_f^* v_f^*(\mathbf{r})\exp(i\omega_f t)].$$

On the other hand, using (8), we find

$$\Psi(\mathbf{r}, t)\exp[i\phi(\mathbf{r}, t)]$$
$$= \sqrt{N_0\psi_0^2(\mathbf{r}) + \delta n(\mathbf{r}, t)}\exp\{i[\phi_0 + \delta\phi(\mathbf{r}, t)]\}$$
$$\simeq \sqrt{N_0}\psi_0\left(1 + \frac{\delta n(\mathbf{r}, t)}{2N_0\psi_0^2(\mathbf{r})} + i\delta\phi(\mathbf{r}, t) + \ldots\right)\exp(i\phi_0) \qquad (18)$$
$$= \exp(i\phi_0)\left[\sqrt{N_0}\psi_0(\mathbf{r})\right.$$

$$+ \left(\frac{\delta n_f(\mathbf{r})}{2\sqrt{N_0}\psi_0(\mathbf{r})} + i\sqrt{N_0}\psi_0(\mathbf{r})\delta\phi(\mathbf{r}) + \ldots\right)\exp(-i\omega_f t)$$

$$\left. + \left(\frac{\delta n_f^*(\mathbf{r})}{2\sqrt{N_0}\psi_0(\mathbf{r})} + i\sqrt{N_0}\psi_0(\mathbf{r})\delta\phi_f^*(\mathbf{r}) + \ldots\right)\exp(i\omega_f t)\right].$$

Here, we used the Taylor expansion where $\delta n(\mathbf{r}, t)$ and $\delta\phi(\mathbf{r}, t)$ are assumed to be small in the first order. Dots cover the terms of the second and higher orders of smallness, which will be rejected. Also, we used the representation $\phi(\mathbf{r}, t) = \phi_0 + \delta\phi(\mathbf{r}, t) = \phi_0 + \delta\phi_f(\mathbf{r})\exp(-i\omega_f t) + \delta\phi_f^*(\mathbf{r})\exp(i\omega_f t)$.

Using the real part of (7) in the form

$$-\hbar\frac{\partial}{\partial t}\phi \ = \ \frac{\hbar^2}{2M}(\nabla\phi)^2 + g\delta n$$

and neglecting the terms with $(\nabla\phi)^2$ as the terms of the second order of smallness, we arrive at

$$\delta\phi_f(\mathbf{r}) \ = \ \frac{g}{i\hbar\omega_f}\delta n_f(\mathbf{r}). \qquad (19)$$

Comparing expressions (17) and (18), we have

$$u_f(\mathbf{r}) \ = \ \frac{\delta n_f(\mathbf{r})}{\lambda_f}\left(\frac{1}{2\sqrt{N_0}\psi_0(\mathbf{r})} + \frac{g\sqrt{N_0}\psi_0(\mathbf{r})}{\hbar\omega_f}\right),$$

$$v_f(\mathbf{r}) \ = \ \frac{\delta n_f(\mathbf{r})}{\lambda_f}\left(-\frac{1}{2\sqrt{N_0}\psi_0(\mathbf{r})} + \frac{g\sqrt{N_0}\psi_0(\mathbf{r})}{\hbar\omega_f}\right), \qquad (20)$$

where $\delta n_f(\mathbf{r}) \equiv \delta n_{nlm}(\mathbf{r})$ is given by (15).

Taking into account that the operator on the right-hand side of Eq. (12) is Hermitean and that its eigenvalues $(-2K_{nl})$ depend both on n and l, we can write

$$\int d\mathbf{r}\,\delta n_{nl}(\mathbf{r})\delta n^*_{n'l'}(\mathbf{r}) = 0 \text{ if } n \neq n' \text{ or } l \neq l'.$$

Thus, from (4), we find for the normalizing constant

$$|\lambda_f|^2 = 2gI/(\hbar\omega_f),$$

where

$$I = \int_0^R [F_{nl}(r/R)]^2 r^2 dr$$

$$= \frac{2g}{\hbar\omega_f}R^3 \sum_{k=0}^{n}\sum_{j=0}^{n}\frac{C_k^{(nl)}C_j^{(nl)}}{2(k+j+l)+3}. \qquad (21)$$

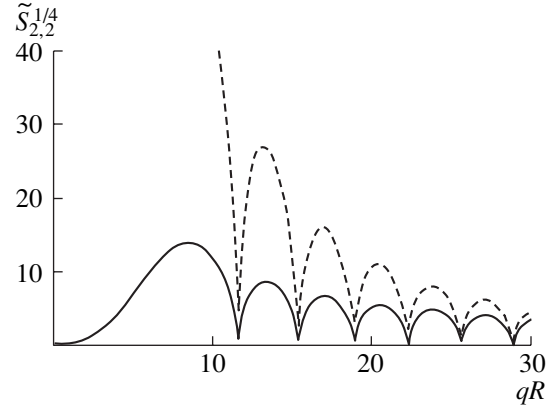Using the well-known expansion of a plane wave [16, 17],

$$\exp(i\mathbf{q}\mathbf{r}) \equiv \exp(iqr\exp\vartheta) =$$

$$\sum_{l=0}^{\infty} i^l\sqrt{4\pi(2l+1)}\,Y_{l0}(\vartheta,\varphi)j_l(qr),$$

where $j_l(r) = \sqrt{\pi/(2r)}J_{l+\frac{1}{2}}(r)$ is the spherical Bessel function, we find the final expression for the structural factor of the single-component BEC in the Thomas–Fermi approximation:

$$\tilde{S}_{n,l}^{TF}(q) \ = \ \frac{2l+1}{2}\frac{a_{HO}}{a}\left(15N_0\frac{a}{a_{HO}}\right)^{3/5}\sqrt{K_{nl}}$$

$$\times\frac{\left|\int_0^1 j_l(qR\bar{r})F_{nl}(\bar{r})\bar{r}^2 d\bar{r}\right|^2}{\displaystyle\sum_{k=0}^{n}\sum_{j=0}^{n}C_k^{(nl)}C_j^{(nl)}/(2k+2j+2l+3)}. \qquad (22)$$

Fig. 1. Structural factor as a function of transferred momentum (continuous curve, formula (22); dashed curve, quasi-classical approximation [7]). The structural factor is plotted in the power 1/4 for better visualization.

The summation is over all nonnegative integers n and l except for the pair $(n, l) = (0, 0)$, which corresponds to the ground (unexcited) state of the condensate. The quantities K_{nl}, $F_{nl}(\bar{r})$, and $C_k^{(nl)}$ are defined by (13), (14), and (16); $\bar{r} = r/R$.

Figure 1 shows the dependence of the structural factor $\tilde{S}_{n,l}(q)$ calculated by (22) and by the quasi-classical formula [7] on the dimensionless parameter qR, where R is the condensate size defined by (11), for $N_0 = 10^7$ atoms, $\frac{a_{HO}}{a} \simeq 432$, and $n = l = 2$. It is seen that the quasi-classical expression estimates $\tilde{S}_{n,l}(q)$ to 2–200 times higher than expression (22) for qR between 10 and 30. However, in both approximations, the function $\tilde{S}_{n,l}(q)$ vanishes at the same values of q (this is also true for other parameters). Thus, in the domain of its applicability, the quasi-classical expression can be used for finding the zeros of the function $\tilde{S}_{n,l}(q)$.

The evidence for the validity of expression (22) is the fact that it obeys the rule of sum for the structural factor [10] with a great accuracy. According to this rule,
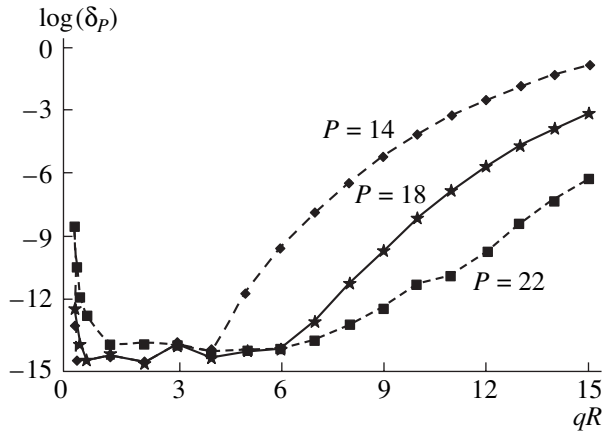
$$m_1(q) \ = \ \int \omega S(\mathbf{q},\omega)d\omega$$

$$= \ \sum_{n,l}\omega_{n,l}\tilde{S}_{n,l}(q) \ = \ N_0\hbar q^2/(2M). \qquad (23)$$

Figure 2 plots the common logarithm of

$$\delta_P(q) \ = \ \sum_{n,l=(0,0)}^{2n+l=P}\frac{\omega_{n,l}\tilde{S}_{n,l}^{TF}(q)}{m_1(q)} - 1 \qquad (24)$$

at $P = 14$, 18, and 22; the function $\tilde{S}_{n,l}^{TF}(q)$ is calculated by (22). The dependence $\delta_P(q)$ is the successive

Fig. 2. Semilogarithmic plot of the $\delta_p(q)$ vs. qR dependence. The higher qR, the greater the number of the series terms necessary for estimating the series limit.

approximation to the relative deviation of the quantity $\sum_{n, l} \omega_{n, l} \tilde{S}_{n, l}^{TF}(q)$ from $m_1(q) = N_0 \hbar q^2/(2M)$: $\delta_P(q) \xrightarrow{P \to \infty}$ $\sum_{(0, 0)}^{\infty} \omega_{n, l} \tilde{S}_{n, l}^{TF}(q)/([m_1(q)] - 1)$. Clearly, the larger q, the larger P (a larger number of the series terms) must be taken to estimate the limit of the series. One may conclude from Fig. 2 that $\delta_P(q) \xrightarrow{P \to \infty} 10^{-14}$. Note that the quasi-classical expression [7] does not obey the rule of sum (there is no quadratic dependence on q). This suggests that the quasi-classical expression considerably overestimates the BEC structural factor, which seems to be natural in view of difficulties associated with the quasi-classical estimation of matrix elements [16].

DOUBLE-COMPONENT CONDENSATE

The simultaneous condensation of two different atomic systems has been first demonstrated in [18]. The two atomic systems were ^{87}Rb atoms occupying the hyperfine sublevel $|F, m_F\rangle = |1, -1\rangle$ and ^{87}Rb atoms on the $|2, 2\rangle$ sublevel. The systems formed slightly overlapping clouds with the separated centers of gravity. Subsequently [19], a double-component BEC with the coinciding centers of gravity of the components has been implemented. In this case, the $|1, -1\rangle$ and $|2, 1\rangle$ hyperfine sublevels of ^{87}Rb, which have the same magnetic dipole moment, was used. We will consider the situation where the centers of the condensates coincide and the trapping potential for either component has the form $V_i = M_i \Omega_i^2 r^2/2$, $i = 1, 2$. The ground state of a double-component BEC has been theoretically considered in [20]; and collective excitations, in [21–23]. For a purely double-phase condensate (when the clouds of both components completely overlap) in a spherically symmetric trap, it has been shown [21] that, in terms of the Thomas–Fermi approximation, the dependence of the collective excitation frequency on the radial and

orbital quantum numbers n and l is similar to that for a single-component condensate:

$$\omega_{nl}^{\pm} = \omega_0^{\pm} \sqrt{2n^2 + 2nl + 3n + l}. \tag{25}$$

However, the factor $\omega_0^{\pm}$, in this case, takes two values corresponding to two branches of the velocity of sound (two branches of the dispersion relation). This factor depends on the parameters a_1, a_2, a_{12}, Ω_1, and Ω_2. Here, a_1 and a_2 are the s-scattering amplitudes for atoms of the first and second components (atoms I and II), respectively; a_{12} is the amplitude of scattering of an atom I by an atom II; and Ω_1 and Ω_2 are the trap frequencies for the components I and II, respectively.

Thus, a double-component condensate has two types of excitation, and, instead of (2), we can write

$$\hat{\psi}_1 = \exp(i\phi_{01}) \left[\psi_{01} \sqrt{N_1} \right.$$
$$\left. + \sum_f (u_{11_f} \hat{\alpha}_{1_f} - v_{11_f}^* \hat{\alpha}_{1_f}^\dagger + u_{12_f} \hat{\alpha}_{2_f} - v_{12_f}^* \hat{\alpha}_{2_f}^\dagger) \right], \tag{26a}$$

$$\hat{\psi}_2 = \exp(i\phi_{02}) \left[\psi_{02} \sqrt{N_2} \right.$$
$$\left. + \sum_f (u_{21_f} \hat{\alpha}_{1_f} - v_{21_f}^* \hat{\alpha}_{1_f}^\dagger + u_{22_f} \hat{\alpha}_{2_f} - v_{22_f}^* \hat{\alpha}_{2_f}^\dagger) \right]. \tag{26b}$$

Here, $\hat{\psi}_1(\mathbf{r})$ and $\hat{\psi}_2(\mathbf{r})$ are the field operators for the components I and II, respectively; $\hat{\alpha}_{1_f}^\dagger$ and $\hat{\alpha}_{1_f}$ are, respectively, the production and annihilation operators for quasi-particles with an energy $\hbar\omega_f^+$; $\hat{\alpha}_{2_f}^\dagger$ and $\hat{\alpha}_{2_f}$ are the production and annihilation operators for quasi-particles with an energy $\hbar\omega_f^-$; ϕ_{01} and ϕ_{02} are the phase constants for the components I and II, respectively; ψ_{01} and ψ_{02} are real functions that satisfy the set of stationary coupled Gross–Pitaevski equations and are normalized to unity:

$$\int d\mathbf{r} \psi_{01}^2(\mathbf{r}) = \int d\mathbf{r} \psi_{02}^2(\mathbf{r}) = 1; \tag{27}$$

and N_1 and N_2 are the number of atoms in the components I and II of the condensate.

The Bogoliubov coefficients u_{11_f}, v_{11_f}, u_{12_f}, v_{12_f} and u_{21_f}, v_{21_f}, u_{22_f}, v_{22_f} are normalized as

$$\int d\mathbf{r} [u_{11_f} u_{11_f}^* - v_{11_f} v_{11_f}^* $$
$$+ u_{12_f} u_{12_f}^* - v_{12_f} v_{12_f}^*] = \delta_{ff'}, \tag{28a}$$

$$\int d\mathbf{r} [u_{21_f} u_{21_f}^* - v_{21_f} v_{21_f}^*$$
$$+ u_{22_f} u_{22_f}^* - v_{22_f} v_{22_f}^*] = \delta_{ff'}. \tag{28b}$$

These conditions yield the standard boson commutation relations for the operators of quasi-particles of the first and second type (the operators for quasi-particles of the first type commute with those for quasi-particles of the second type).

We will also assume that a photon interacting with a BEC resonates with and is scattered by atoms I and does not interact with atoms II. Then, $\hat{\psi}(\mathbf{r}) = \hat{\psi}_1(\mathbf{r})$ should be substituted in expression (1) for the structural factor. In view of expansion (26), we have for the partial structural factor:

$$S_1(\mathbf{q}, \omega) = N_1 \sum_f \left\{ \left| \int d\mathbf{r} \exp(i\mathbf{q}\mathbf{r}) \right. \right.$$

$$\times [u_{11_f}^*(\mathbf{r}) - v_{11_f}^*(\mathbf{r})] \psi_{01}(\mathbf{r}) \Big|^2 \delta(\omega - \omega_f^+)$$
$$+ \left| \int d\mathbf{r} \exp(i\mathbf{q}\mathbf{r}) [u_{12_f}^*(\mathbf{r}) - v_{12_f}^*(\mathbf{r})] \psi_{01}(\mathbf{r}) \right|^2 \tag{29}$$

$$\times \delta(\omega - \omega_f^-) \Big\}.$$

A set of Gross–Pitaevski coupled equations that describe the evolution of the mean field operators $\langle \hat{\psi}_1(\mathbf{r}) \rangle = \Psi_1(\mathbf{r}, t) \exp[i\phi_1(\mathbf{r}, t)]$ and $\langle \hat{\psi}_2(\mathbf{r}) \rangle = \Psi_2(\mathbf{r}, t) \exp[i\phi_2(\mathbf{r}, t)]$ for the components of a double-component BEC has the form [21]

$$i\hbar \frac{\partial}{\partial t} \Psi_1 \exp(i\phi_1) = \left[-\frac{\hbar^2 \nabla^2}{2M_1} - \mu_1 + V_1(\mathbf{r}) \right.$$
$$\left. + g_1 \Psi_1^2 + g_{12} \Psi_2^2 \right] \Psi_1 \exp(i\phi_1), \tag{30a}$$

$$i\hbar \frac{\partial}{\partial t} \Psi_2 \exp(i\phi_2) = \left[-\frac{\hbar^2 \nabla^2}{2M_2} - \mu_2 + V_2(\mathbf{r}) \right.$$
$$\left. + g_2 \Psi_2^2 + g_{12} \Psi_2^2 \right] \Psi_2 \exp(i\phi_2). \tag{30b}$$

Here, $g_i = 4\pi\hbar^2 a_i / M_i$ ($i = 1, 2$) and $g_{12} = 4\pi\hbar^2 a_{12} / \sqrt{M_1 M_2}$ are the interaction parameters, V_i is the trap potential for the ith component (for a spherical trap, $V_i = M_i \Omega_i^2 r^2 / 2$), μ_i is the chemical potential of the ith component in the ground state of the BEC.

Passing to the variables $\delta n_i(\mathbf{r}, t)$ and $\mathbf{v}_i(\mathbf{r}, t)$ ($i = 1, 2$) defined by (8) and using the Thomas–Fermi approxi-

mation (neglecting the kinetic energy), we come to the linearized equations

$$\frac{\partial^2}{\partial t^2} \delta n_1(\mathbf{r}, t)$$
$$= \frac{N_1}{M_1} \nabla \{ \psi_{01}^2(\mathbf{r}) \nabla [g_1 \delta n_1(\mathbf{r}, t) + g_{12} \delta n_2(\mathbf{r}, t)] \}, \tag{31a}$$

$$\frac{\partial^2}{\partial t^2} \delta n_2(\mathbf{r}, t)$$
$$= \frac{N_2}{M_2} \nabla \{ \psi_{02}^2(\mathbf{r}) \nabla [g_2 \delta n_2(\mathbf{r}, t) + g_{12} \delta n_1(\mathbf{r}, t)] \}. \tag{31b}$$

The ground state can be defined by substituting $\Psi_i(\mathbf{r}, t) = \sqrt{N_i} \psi_{0i}(\mathbf{r})$ and $\phi_i(\mathbf{r}, t) = \phi_{0i}(\mathbf{r})$ ($i = 1, 2$) into set (30) and neglecting the kinetic energy term. Eventually, we find

$$\psi_{01}^2(\mathbf{r})$$
$$= \frac{g_2\mu_1 - g_{12}\mu_2 - (g_2 M_1 \Omega_1^2 / 2 - g_{12} M_2 \Omega_2^2 / 2) r^2}{(g_1 g_2 - g_{12}^2) N_1}, \tag{32a}$$

$$\psi_{02}^2(\mathbf{r})$$
$$= \frac{g_1\mu_2 - g_{12}\mu_1 - (g_1 M_2 \Omega_2^2 / 2 - g_{12} M_1 \Omega_1^2 / 2) r^2}{(g_1 g_2 - g_{12}^2) N_2}. \tag{32b}$$

The radii R_i of the condensates are defined by the condition $\psi_{0i}|_{r = R_i} = 0$ ($i = 1, 2$). Normalization (27) yields

$$R_1 = a_{HO_1} \left(\frac{15 N_1 (a_1 a_2 - a_{12}^2)/a_{HO_1}}{a_2 - a_{12} \Omega_{12}^{-2} M_{12}^{-3/2}} \right)^{1/5}, \tag{33a}$$

$$R_2 = a_{HO_2} \left(\frac{15 N_2 (a_1 a_2 - a_{12}^2)/a_{HO_2}}{a_1 - a_{12} \Omega_{12}^2 M_{12}^{3/2}} \right)^{1/5}, \tag{33b}$$

where $\Omega_{12} = \Omega_1/\Omega_2$, $M_{12} = M_1/M_2$, and $a_{HO_i} = \sqrt{h/(M_i \Omega_i)}$ ($i = 1, 2$).

Equating R_1 and R_2 from (33), we obtain the component particle number ratio that provides a purely binary condensate:

$$\frac{N_2}{N_1} = \Omega_{12}^{-2} M_{12}^{-2} \frac{a_1 - a_{12} \Omega_{12}^2 M_{12}^{3/2}}{a_2 - a_{12} \Omega_{12}^{-2} M_{12}^{-3/2}}. \tag{34}$$

Ultimately, for the ground state of a double-component BEC in the Thomas–Fermi approximation, we have

$$\psi_{01}^2(\mathbf{r}) = \psi_{02}^2(\mathbf{r}) = \frac{15}{8\pi} R_1^{-3} [1 - (r/R_1)^2], \tag{35}$$

where R_1 is defined by (33a) and the component ratio satisfies (34).

Let us expand, as before, the real functions $\delta n_i(\mathbf{r}, t)$ ($i = 1, 2$) in elementary perturbations with a frequency ω_f:

$$\delta n_i(\mathbf{r}, t) = \sum_f [\delta n_{i_f}(\mathbf{r})\exp(-i\omega_f t)$$
$$+ \delta n_{i_f}^*(\mathbf{r})\exp(i\omega_f t)].$$

Then, set (31) takes the form

$$-\omega_f^2 \delta n_{1_f}(\mathbf{r}) \tag{36a}$$
$$= X_1 \nabla_{\bar{r}}\{[1 - \bar{r}^2]\nabla_{\bar{r}}[g_1\delta n_{1_f}(\mathbf{r}) + g_{12}\delta n_{2_f}(\mathbf{r})]\},$$

$$-\omega_f^2 \delta n_{2_f}(\mathbf{r}) \tag{36b}$$
$$= X_2 \nabla_{\bar{r}}\{[1 - \bar{r}^2]\nabla_{\bar{r}}[g_2\delta n_{2_f}(\mathbf{r}) + g_{12}\delta n_{1_f}(\mathbf{r})]\},$$

where $X_i = (15/4\pi)N_i/(R_i^5 M_i)$ ($i = 1, 2$) and $\bar{r} = r/R_1$.

Since functions (15) and (16) represent the complete set of the eigenfunctions of operator (12), a solution of set (36) for the same quantum numbers n, l, and m is a linear combination of functions (15). Then, up to the normalizing constant, we can write

$$\delta n_{1f}(\mathbf{r}) = \delta n_{1_{nlm}}(\mathbf{r}) = F_{nl}(r/R_1)Y_{lm}(\Theta, \phi),$$

$$\delta n_{2f}(\mathbf{r}) = \delta n_{2_{nlm}}(\mathbf{r}) = \Theta_l F_{nl}(r/R_1)Y_{lm}(\Theta, \phi),$$

where Θ_{nl} is a constant and $F_{nl}(r/R_1)Y_{lm}(\Theta, \phi) = \delta n_{lmn}(\mathbf{r})$ satisfies Eq. (12), where $(\omega_f/\omega_0)^2 = K_{nl}$ [see (14)]; in other words, $F_{nl}(r)$ is defined by (13) and (16). Then, from (36), we have

$$[\omega_f^2 - X_1g_1K_{nl}] - X_1g_{12}K_{nl}\Theta_{nl} = 0, \tag{37a}$$

$$-X_2g_{12}K_{nl} + [\omega_f^2 - X_2g_2K_{nl}]\Theta_{nl} = 0. \tag{37b}$$

From the solvability condition for (37), we determine the collective excitation frequencies of a double-component BEC in the Thomas–Fermi approximation:

$$\omega_f^2 = (\omega_{nl}^\pm)^2$$

$$= \frac{X_1g_1}{2}\left\{1 + \xi \pm \left[(1 - \xi)^2 + 4\xi\frac{a_{12}^2}{a_1a_2}\right]^{\frac{1}{2}}\right\}K_{nl}$$

$$= \Omega_1^2\frac{a_1(a_2 - a_{12}\Omega_{12}^{-2}M_{12}^{-3/2})}{2(a_1a_2 - a_{12}^2)} \tag{38}$$

$$\times\left\{1 + \xi \pm \left[(1 - \xi)^2 + 4\xi\frac{a_{12}^2}{a_1a_2}\right]^{\frac{1}{2}}\right\}K_{nl},$$

where

$$\xi = \frac{N_2M_1^2a_2}{N_1M_2^2a_1} = \frac{a_1 - a_{12}\Omega_{12}^2M_{12}^{3/2}}{a_2\Omega_{12}^2 - a_{12}M_{12}^{-3/2}}\frac{a_2}{a_1}. \tag{39}$$

Substituting (38) into (37a) yields

$$\Theta_{nl}^\pm = \frac{g_1}{2g_{12}}\left\{\xi - 1 \pm \left[(1 - \xi)^2 + 4\xi\frac{a_{12}^2}{a_1a_2}\right]^{\frac{1}{2}}\right\}. \tag{40}$$

To find a relation of the Bogoliubov coefficients $u_{1i_f}(\mathbf{r})$, $v_{1i_f}(\mathbf{r})$, $u_{2i_f}(\mathbf{r})$, and $v_{2i_f}(\mathbf{r})$ with $\delta n_{1_f}(\mathbf{r})$ and $\delta n_{2_f}^\pm(\mathbf{r})$ ($i = 1$ and 2 refers to "+" and "–", respectively), we assume that the ensemble of ultracold atoms is so excited that there exists only the fth mode with a frequency ω_f^+ (i.e., only $\langle\hat{\alpha}_{1_f}\rangle = \lambda_f^+\exp(-i\omega_f^+ t)$ and $\langle\hat{\alpha}_{1_f}^\dagger\rangle = (\lambda_f^+)^*\exp(i\omega_f^+ t)$ are other than zero) or the fth mode with a frequency ω_f^- (i.e., only $\langle\hat{\alpha}_{2_f}\rangle = \lambda_f^-\exp(-i\omega_f^- t)$ and $\langle\hat{\alpha}_{2_f}^\dagger\rangle = (\lambda_f^-)^*\exp(i\omega_f^- t)$ are other than zero). Then, for $j = 1, 2$, we have

$$\langle\hat{\psi}_j(\mathbf{r})\rangle = \Psi_j\exp(i\phi_j)$$
$$= \exp(i\phi_{0j})[\psi_{0j}\sqrt{N_j} + \lambda_f^\pm u_{j1_f}\exp(-i\omega_f^\pm t) \tag{41}$$
$$- (\lambda_f^\pm)^* v_{j1_f}^*\exp(i\omega_f^\pm t)].$$

On the other hand, for the mean field operator of the jth component, we can use an expansion like (18). A relationship between the phase perturbation $\delta\phi_{j_f}(\mathbf{r})$ and the density perturbation $\delta n_{j_f}(\mathbf{r})$ for the jth component is found from the jth equation of set (30). By analogy with (19), we have

$$\delta\phi_{1_f}^\pm(\mathbf{r}) = \frac{g_1}{i\hbar\omega_f^\pm}\delta n_{1_f}(\mathbf{r}) + \frac{g_{12}}{i\hbar\omega_f^\pm}\delta n_{2_f}^\pm(\mathbf{r}),$$

$$\delta\phi_{2_f}^\pm(\mathbf{r}) = \frac{g_2}{i\hbar\omega_f^\pm}\delta n_{2_f}^\pm(\mathbf{r}) + \frac{g_{12}}{i\hbar\omega_f^\pm}\delta n_{1_f}(\mathbf{r}).$$

Eventually, we come to

$$u_{1i_f}(\mathbf{r}) = \frac{\delta n_{1_f}(\mathbf{r})}{\lambda_f^\pm}$$

$$\times\left(\frac{1}{2\sqrt{N_1}\psi_{01}(\mathbf{r})} + \frac{\sqrt{N_1}\psi_{01}(\mathbf{r})(g_1 + g_{12}\Theta_{nl}^+)}{\hbar\omega_f^\pm}\right), \tag{42a}$$

$$v_{1i_f}(\mathbf{r}) = \frac{\delta n_{1_f}(\mathbf{r})}{\lambda_f^{\pm}}$$

$$\times \left(-\frac{1}{2\sqrt{N_1}\psi_{01}(\mathbf{r})} + \frac{\sqrt{N_1}\psi_{01}(\mathbf{r})(g_1 + g_{12}\Theta_{nl}^+)}{\hbar\omega_f^{\pm}} \right), \quad (42b)$$

$$u_{2i_f}(\mathbf{r}) = \frac{\delta n_{1_f}(\mathbf{r})}{\lambda_f^{\pm}}$$

$$\times \left(\frac{\Theta_{nl}^+}{2\sqrt{N_2}\psi_{02}(\mathbf{r})} + \frac{\sqrt{N_2}\psi_{02}(\mathbf{r})(\Theta_{nl}^+ g_2 + g_{12})}{\hbar\omega_f^{\pm}} \right), \quad (42c)$$

$$v_{2i_f}(\mathbf{r}) = \frac{\delta n_{1_f}(\mathbf{r})}{\lambda_f^{\pm}}$$

$$\times \left(-\frac{\Theta_{nl}^+}{2\sqrt{N_2}\psi_{02}(\mathbf{r})} + \frac{\sqrt{N_2}\psi_{02}(\mathbf{r})(\Theta_{nl}^+ g_2 + g_{12})}{\hbar\omega_f^{\pm}} \right). \quad (42d)$$

Here, $\delta n_{1_f}(\mathbf{r}) \equiv \delta n_{nlm}(\mathbf{r})$ are given by (15) and the constants $\lambda_f^{\pm}$ are determined from normalization condition (28) with regard for the orthogonality of $\delta n_{nlm}(\mathbf{r})$:

$$|\lambda_f^{\pm}|^2 = \frac{2g_1 I}{\hbar\omega_f^{\pm}}$$

$$\times \frac{2\left(1 - \frac{a_{12}^2}{a_1 a_2}\right)\sqrt{(1-\xi)^2 + 4\xi\frac{a_{12}^2}{a_1 a_2}}}{\pm\left(1 - \xi - 2\frac{a_{12}^2}{a_1 a_2}\right) + \sqrt{1 - 2\xi + 4\frac{a_{12}^2}{a_1 a_2}\xi + \xi^2}}, \quad (43)$$

where I is given by (21) with R_1 substituted for R.

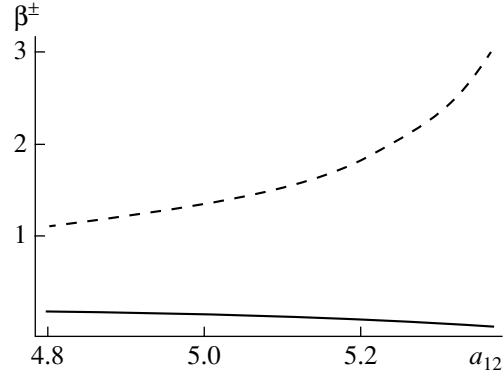The final expression for the partial structural factor of a double-component condensate is

$$S_1^{TF}(\mathbf{q}, \omega)$$

$$= \sum_{n,l} \tilde{S}_{n,l}^{TF}(q)[\beta^+\delta(\omega - \omega_{nl}^+) + \beta^-\delta(\omega - \omega_{nl}^-)], \quad (44)$$

where

$$\beta^{\pm} = \left[\frac{a_2 - a_{12}^2/a_1}{a_2 - a_{12}\Omega_{12}^{-2}M_{12}^{-3/2}} \right]^{\frac{1}{10}}$$

$$\times \left[\frac{\xi + 1}{2} \pm \sqrt{\frac{(1-\xi)^2}{4} + \xi\frac{a_{12}^2}{a_1 a_2}} \right]^{\frac{1}{2}} \quad (45)$$
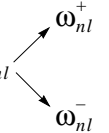
Fig. 3. β^+ (dashed curve) and β^- (continuous curve) vs. amplitude a_{12} of scattering of atoms I by atoms II. ^{87}Rb atoms occupying the hyperfine sublevels $|1, -1\rangle$ and $|2, 1\rangle$.

$$\times \frac{\pm\left(1 - \xi - 2\frac{a_{12}^2}{a_1 a_2}\right) + \sqrt{1 - 2\xi + 4\frac{a_{12}^2}{a_1 a_2}\xi + \xi^2}}{2\left(1 - \frac{a_{12}^2}{a_1 a_2}\right)\sqrt{(1-\xi)^2 + 4\xi\frac{a_{12}^2}{a_1 a_2}}},$$

$\tilde{S}_{n,l}^{TF}(q)$ is defined by (22) if $a = a_1$, $a_{HO} = a_{HO_1}$, $N_0 = N_1$, and ξ is given by (39).

Comparing expression (44) for the partial structural factor of a double-component BEC with expressions (1) and (22) for a single-component condensate allows the following conclusions. The presence of the second component, first, changes the frequencies (i.e., the energy transferred by scattering) so that $\omega_{nl} \nearrow^{\omega_{nl}^+} \searrow_{\omega_{nl}^-}$ and, second, introduces the factors β^+ and β^- for the frequencies ω_{nl}^+ and ω_{nl}^-, respectively.

Consider a double-component condensate at $\Omega_1 = \Omega_2$ and $M_1 = M_2$. Such a situation is experimentally realized for sets of ^{87}Rb atoms occupying the hyperfine sublevels $|1, -1\rangle$ and $|2, 1\rangle$ [19]. According to [23], we put $a_1 = 5.68$ nm and $a_2 = 5.36$ nm. The value of a_{12} will be varied from 1.8 to 5.36 nm. For $a_2 < a_{12} < a_1$, the purely binary condensate cannot be produced, as follows from (34). Figure 3 plots $\beta^{\pm}$ against a_{12}. It follows from this figure that photon scattering may be enhanced in the presence of the other component. This is because the fluctuation amplitude of the BEC density grows as a_{12} approaches the critical value $\sqrt{a_1 a_2}$ above which the components are spatially separated [24]. At anomalously high density fluctuations, an associated ultracold quantum system behaves physically similarly to ordinary liquids near the critical point.

REFERENCES

1. J. Stenger, S. Inouye, A. P. Chikkatur, *et al.*, Phys. Rev. Lett. **82**, 4569 (1999); D. M. Stamper-Kurn, A. P. Chikkatur, A. Görlitz, *et al.*, Phys. Rev. Lett. **83**, 2876 (1999).

2. N. N. Bogoliubov, J. Phys. (Moscow) **11**, 23 (1947); Izv. Akad. Nauk SSSR **11**, 77 (1947).

3. E. P. Gross, Nuovo Cimento **20**, 454 (1961).

4. A. L. Fetter, Ann. Phys. (N. Y.) **70**, 67 (1972).

5. R. Graham and D. Walls, Phys. Rev. Lett. **76**, 1774 (1996).

6. A. Gorlitz, A. P. Chikkatur, and W. Ketterle, cond-mat/0008067.

7. A. Csordas, R. Graham, and P. Szepfalusy, Phys. Rev. A **54**, R2543 (1996); cond-mat/9711140.

8. Wen-Chin Wu and A. Griffin, Phys. Rev. A **54**, 4204 (1996).

9. F. Zambelli, L. Pitaevski, D. M. Stamper-Kurn, *et al.*, Phys. Rev. A **61**, 063608 (2000).

10. D. Pines and P. Nozieres, *The Theory of Quantum Liquids* (Benjamin, New York, 1966; Mir, Moscow, 1967), Vol. 1.

11. M. R. Andrews *et al.*, Phys. Rev. Lett. **79**, 553 (1997); M. R. Matthews *et al.*, Phys. Rev. Lett. **81**, 243 (1998).

12. K. G. Singh and D. S. Rokshar, Phys. Rev. Lett. **77**, 1667 (1996); M. Edwards *et al.*, Phys. Rev. Lett. **77**, 1671 (1996); F. Dalfovo *et al.*, Phys. Rev. A **56**, 3840 (1997).

13. S. Stringari, Phys. Rev. Lett. **77**, 2360 (1996).

14. G. Baym and C. J. Pethick, Phys. Rev. Lett. **76**, 6 (1996).

15. D. Pines and P. Nozieres, *The Theory of Quantum Liquids* (Addison-Wesley, Redwood City, 1990), Vol. 2.

16. 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.).

17. *Handbook of Mathematical Functions*, Ed. by M. Abramowitz and I. A. Stegun (National Bureau of Standards, Washington, 1964; Nauka, Moscow, 1979).

18. C. J. Myatt, E. A. Burt, R. W. Ghrist, *et al.*, Phys. Rev. Lett. **78** (4), 586 (1997).

19. M. R. Matthews, D. S. Holl, D. S. Jin, *et al.*, cond-mat/9803310; D. S. Holl, M. R. Matthews, J. R. Ensher, *et al.*, cond-mat/9804138; D. S. Holl, M. R. Matthews, C. E. Wieman, *et al.*, cond-mat/9805327.

20. T.-L. Ho and V. B. Shenoy, Phys. Rev. Lett. **77**, 3276 (1996).

21. R. Graham and D. Walls, Phys. Rev. A **57**, 484 (1998).

22. B. D. Esry and C. H. Greene, Phys. Rev. A **57**, 1265 (1998).

23. D. Gordon and C. M. Savage, Phys. Rev. A **58**, 1440 (1998).

24. Yu. A. Nepomnyashchiĭ, Zh. Éksp. Teor. Fiz. **70**, 1070 (1976) [Sov. Phys. JETP **43**, 559 (1976)].

Translated by V. Isaakyan

Expansion of the Distance between Two Points in Spheroidal Functions as Applied to Problems of Mathematical Physics

A. S. Baranov

Pulkovo Astronomic Observatory, Russian Academy of Sciences,
nab. Makarova 6, St. Petersburg, 196140 Russia

Received August 21, 2000, in final form, July 9, 2001

Abstract—A complete set of biharmonic functions is constructed in spheroidal coordinates. The distance between two points and its inverse value are expanded into a double series in terms of such functions. Possible applications in the theory of elasticity, in astrophysics, and other fields of mathematical physics are pointed out. © 2002 MAIK "Nauka/Interperiodica".

INTRODUCTION

It is known [1, 2] that in some problems of mathematical physics, functions constructed in the basis of spheroidal coordinates are more convenient to use than spherical functions. Examples are often encountered in engineering mechanics, geophysics, and astrophysics. In a number of cases, not only harmonic but also biharmonic functions are of interest [3–5]. Under these conditions, the construction of elementary biharmonic functions is rather simple, but the expansion of a given biharmonic function in elementary biharmonic functions poses problems. A key point here is the expansion of the distance D between two arbitrary points and its inverse value D^{-1} in these elementary functions.

BASIC FORMULAS

First, we consider harmonic functions of type

$$V_n^k(t, \tau, \varphi) = P_n^k(t)p_n^k(\tau)e^{ik\varphi},$$
$$W_n^k(t, \tau, \varphi) = P_n^k(t)q_n^k(\tau)e^{ik\varphi}, \tag{1}$$

which are of prime concern in this paper. Here, t, τ, and φ are spheroidal coordinates related to Cartesian coordinates x_1, x_2, and x_3 as

$$x_1 = R\cos\varphi, \quad x_2 = R\sin\varphi, \quad R = c\sqrt{(1+\tau^2)(1-t^2)}, \tag{2}$$
$$x_3 = c\tau t, \quad \varphi = \operatorname{arccot}(x_2/x_1)$$

(the x_3 axis is taken as the polar axis). The presence of some reference spheroid with the semiaxes $a_1(=a_2)$ and a_3 with $a_1 > a_3$ is assumed (i.e., the spheroid is oblate). The parameter c introduced in (2) is defined as the focal distance: $c = \sqrt{a_1^2 - a_3^2}$.

In formulas (1), P_n^k is the standard designation of the associated Legendre functions, $p_n^k(\tau) = i^{k-n}P_n^k(i\tau)$, and $q_n^k(\tau)$ is the Legendre function of imaginary argument of the second kind:

$$q_n(\tau) = \frac{i^{-n}}{2}\int_{-1}^{1}\frac{P_n(x)}{\tau - ix}dx(k = 0),$$

$$q_n^k(\tau) = (-1)^k(1+\tau^2)^{k/2}\frac{d^k}{d\tau^k}q_n(\tau)(k > 0).$$

Second, the functions

$$L_n^k(t, \tau, \varphi) = [p_{n+2}^k(\tau)P_n^k(t) - P_{n+2}^k(t)p_n^k(\tau)]e^{ik\varphi},$$
$$\tilde{L}_n^k(t, \tau, \varphi) = [q_n^k(\tau)P_{n+2}^k(t) - P_n^k(t)q_{n+2}^k(\tau)]e^{ik\varphi}, \tag{3}$$

satisfy the biharmonic equation.

These functions were introduced and investigated in [5]. In addition to them, the functions t and τ are also biharmonic, which can be verified directly.

We fix that of cofocal spheroids corresponding to **r**. The quantity D^{-1} is an analytical function of t and φ at fixed τ and is expanded in orthogonal functions $Y_{nk} = P_n^k(t)e^{ik\varphi}$ by the general rule [1]. We obtain the series

$$D^{-1} = \sum_{n,k}\alpha_{nk}Y_{nk}.$$

Hereafter, the summation over n and k is performed in the ranges $n \geq 0$, $-n \geq k \geq n$. The coefficients α_{nk} are

conventionally represented by the integrals

$$\alpha_{nk} = \frac{1}{c_{nk}} \int_0^{2\pi} \int_{-1}^{1} D^{-1} Y_{nk}^* dt d\varphi,$$

where

$$c_{nk} = \int_0^{2\pi} \int_{-1}^{1} |Y_{nk}|^2 dt d\varphi = \frac{4\pi}{2n+1} \frac{(n+k)!}{(n-k)!}$$

(hereafter, asterisk means complex conjugation).

For clarity, we will use the expressions for an element of the spheroidal surface, $d\sigma = c^2\sqrt{(1+\tau^2)(t^2+\tau^2)}\, dt d\varphi$, and for an element of the normal in the neighborhood of the spheroid, $dn = c\sqrt{(t^2+\tau^2)/(1+\tau^2)}\, d\tau$, which follow from relation (2). Then,

$$\alpha_{nk} = \frac{1}{c_{nk} c^2 \sqrt{1+\tau^2}} \iint \frac{Y_{nk}^*}{\sqrt{t^2+\tau^2}} d\sigma, \qquad (4)$$

where integration is over the entire spheroid surface.

In this way, we come to the potential of an elementary layer with the density $Y_{nk}^*/\sqrt{t^2+\tau^2}$. This potential is locked for separately in the internal and external regions in the form $v_i = \beta_i p_n^k(\tau_1) P_n^k(t_1)\exp(-ik\varphi_1)$ and $v_e = \beta_e p_n^k(\tau_1) P_n^k(t_1)\exp(-ik\varphi_1)$, respectively, where β_i and β_e are still unknown numerical coefficients. To find them, we use standard sewing condition on the surface: the coincidence of β_i and β_e with the normal derivative having a jump that is equal to the density of the layer multiplied by the factor 4π.

We take advantage of the relation known in the theory of spherical functions (Wronski determinant)

$$q_n^k(\tau)\frac{dp_n^k(\tau)}{d\tau} - p_n^k(\tau)\frac{dq_n^k(\tau)}{d\tau} = \frac{(n+k)!}{(n-k)!}(1+\tau^2)^{-1}, \quad (5)$$

where the numerical coefficient is easily found from the asymptotical expressions obtained at $\tau \longrightarrow \infty$ for p_n^k and q_n^k.

At $\tau_1 = \tau$, the sewing condition for the potential has the form $\beta_i p_n^k(\tau) = \beta_e q_n^k(\tau)$. Then, by virtue of the relation

$$\frac{\partial}{\partial n} = \frac{1}{c}\sqrt{\frac{1+\tau^2}{\tau^2+t^2}}\frac{\partial}{\partial \tau}$$

which follows from formulas (2), the sewing together

of the normal derivatives $\partial v_i/\partial n_1$ and $\partial v_e/\partial n_1$ yields

$$\frac{1}{c}\sqrt{1+\tau^2}[\beta_e q_n'^k(\tau_1) - \beta_i p_n'^k(\tau_1)] = -4\pi.$$

Eventually, we obtain

$$\beta_i = 4\pi\frac{(n-k)!}{(n+k)!}\sqrt{1+\tau^2}q_n^k(\tau),$$

$$\beta_e = 4\pi\frac{(n-k)!}{(n+k)!}\sqrt{1+\tau^2}p_n^k(\tau)$$

with regard for formula (5).

Returning to formula (4), we arrive at

$$\alpha_{nk} = \frac{2n+1}{c}\left[\frac{(n-k)!}{(n+k)!}\right]^2 q_n^k(\tau)p_n^k(\tau_1)$$

$$\times P_n^k(t_1)\exp(-ik\varphi_1) \ (\tau_1 < \tau)$$

or

$$\alpha_{nk} = \frac{2n+1}{c}\left[\frac{(n-k)!}{(n+k)!}\right]^2 p_n^k(\tau)q_n^k(\tau_1)$$
$$\times P_n^k(t_1)\exp(-ik\varphi_1) \ (\tau_1 > \tau). \qquad (6)$$

Now, straightforward calculations in view of definitions (1) yield

$$D^{-1} = \frac{1}{c}\sum_{n,k}(2n+1)\left[\frac{(n-k)!}{(n+k)!}\right]^2 V_n^{*k}(\mathbf{r})W_n^k(\mathbf{r}_1)$$

or

$$D^{-1} = \frac{1}{c}\sum_{n,k}(2n+1)\left[\frac{(n-k)!}{(n+k)!}\right]^2 W_n^k(\mathbf{r})V_n^{*k}(\mathbf{r}_1), \quad (7)$$

the first formula in (7) being used at $\tau < \tau_1$, and the other, at $\tau > \tau_1$. The formulas for D^{-1} are given in [1]. They coincide with formulas (7) if we express the functions P and Q of imaginary arguments in terms of our p and q. In the limit $c \longrightarrow 0$, formulas (7) become well-known formulas for spherical functions [6].

Let us expand the quantity D in a way similar to that used above for D^{-1}; i.e.,

$$D = \sum_{n,k}\xi_{nk}(\mathbf{r}_1)Y_{nk}(t,\varphi),$$

where

$$\xi_{nk} = \frac{1}{c_{nk}c^2\sqrt{1+\tau^2}}\iint\frac{DY_{nk}^*}{\sqrt{t^2+\tau^2}}d\sigma.$$

Consider first the particular case $n = k = 0$, which, as will be demonstrated below, is associated with some computational difficulties. Then,

$$\xi_{00} = \frac{1}{c^2 \sqrt{1 + \tau^2}} \iint \frac{D}{\sqrt{t^2 + \tau^2}} d\sigma.$$

We will analyze the auxiliary function $\eta(\tau) = q_0(\tau) - q_0(\tau_*)$, where τ_* denotes the quantity τ corresponding to the spheroid selected, because integration over the spheroid surface $\tau = \tau_*$ should be replaced by integration over the volume at this point. The normal derivative is easily found:

$$\frac{\partial \eta}{\partial n} = \frac{1}{c\sqrt{(t^2 + \tau^2)/(1 + \tau^2)}} \frac{\partial \eta}{\partial \tau} = -\frac{1}{c\sqrt{(t^2 + \tau^2)(1 + \tau^2)}},$$

therefore,

$$\xi_{00} = -\frac{1}{c} \iint D \frac{\partial \eta}{\partial n} d\sigma.$$

If $\tau_1 < \tau_*$, we can apply the Green theorem [7] to the external space, more exactly, to the space between the spheroid $\tau = \tau_*$ and greater spheroid with $\tau = \hat{\tau}$, where $\hat{\tau}$ will then tend to infinity. In our notation,

$$\iint_{\tau = \hat{\tau}} \left(D \frac{\partial \eta}{\partial n} - \eta \frac{\partial D}{\partial n} \right) d\sigma - \iint_{\tau = \tau_*} D \frac{\partial \eta}{\partial n} d\sigma$$

$$= \iiint (D\Delta\eta - \eta\Delta D) d\mathbf{r} = -2 \int \frac{\eta}{D} d\mathbf{r}$$

(Δ is the Laplacian). Here, we at once have taken into account the fact that $\eta(\mathbf{r})$ is a harmonic function. In the preceding equalities, the volume integral is the potential at an internal point $\mathbf{r}_1$ that is induced by an inhomogeneous spheroid with cofocal layers of equal density. Such potentials are well known [8]. Partition into layers leads to the Stieltjes integral

$$\int \frac{\eta}{D} d\mathbf{r} = \int_{\tau_*}^{\hat{\tau}} \eta(\tau) dU(\tau),$$

where $U(\tau)$ is the potential at a fixed internal point $\mathbf{r}_1$ that is induced by a homogeneous ellipsoid of unit density bounded by the surface $\tau = \text{const}$. Integration by parts yields

$$\int \frac{\eta}{D} d\mathbf{r} = \eta(\hat{\tau}) U(\hat{\tau}) + \int_{\tau_*}^{\hat{\tau}} \frac{U(\tau)}{1 + \tau^2} d\tau.$$

The potential $U(\tau)$ is defined by the known expression

$$U = \pi a'^2 c' \int_0^\infty \left(1 - \frac{R_1^2}{a'^2 + s} - \frac{z_1^2}{c'^2 + s} \right)$$
$$\times \frac{ds}{(a'^2 + s)\sqrt{c'^2 + s}},$$

where the cylindrical coordinates $R = \sqrt{x_1^2 + x_2^2}$ and $z = x_3$ refer to the point $\mathbf{r}_1$ and a' and c' are the semiaxes of the spheroid corresponding to the parameter τ; i.e., $a' = c\sqrt{1 + \tau^2}$ and $c' = c\tau$.

Taking the integrals over s, we obtain

$$U = \pi c^2 \tau (1 + \tau^2) \left[2\arctan\frac{1}{\tau} - \frac{R_1^2}{c^2} \left(\arctan\frac{1}{\tau} - \frac{\tau}{1 + \tau^2} \right) \right.$$
$$\left. - 2\frac{z_1^2}{c^2} \left(\frac{1}{\tau} - \arctan\frac{1}{\tau} \right) \right],$$

$$\int \frac{U(\tau)}{1 + \tau^2} d\tau = \frac{\pi}{2} \left\{ c^2 \left[(1 + \tau^2)^2 \arctan\frac{1}{\tau} + \tau + \frac{\tau^3}{3} \right] \right.$$

$$- \frac{R_1^2}{2} \left[(1 + \tau^2)^2 \arctan\frac{1}{\tau} + \tau - \tau^3 \right]$$

$$\left. - z_1^2 \left[3\tau + \tau^3 - (1 + \tau^2)^2 \arctan\frac{1}{\tau} \right] \right\}.$$

On the other hand, the surface integral is represented in the form

$$\iint_{\tau = \hat{\tau}} \left(D\frac{\partial \eta}{\partial n} - \eta\frac{\partial D}{\partial n} \right) d\sigma = \iint \left[-\frac{D}{c\sqrt{(t^2 + \hat{\tau}^2)(1 + \hat{\tau}^2)}} \right.$$
$$\left. - \frac{\partial D}{\partial n} \left(\arctan\frac{1}{\hat{\tau}} - \arctan\frac{1}{\tau_*} \right) \right] d\sigma.$$

In the spheroidal coordinates,

$$D^2 = c^2 [2 + \tau^2 + \tau_1^2 - t^2 - t_1^2 - 2\tau t \tau_1 t_1$$
$$- 2\cos(\varphi - \varphi_1)\sqrt{(1 + \tau^2)(1 + \tau_1^2)(1 - t^2)(1 - t_1^2)}],$$

at large τ we have

$$D = c[\tau - t\tau_1 t_1 - \cos(\varphi - \varphi_1)$$
$$\times \sqrt{(1 + \tau_1^2)(1 - t^2)(1 - t_1^2)}] + O(\tau^{-1}).$$

After the substitution of this asymptotic expression into the surface integral, we use the limit $\hat{\tau}$ tends to ∞; then, $\hat{\tau}$ as an auxiliary quantity disappears in the formulas. The cylindrical coordinates R_1 and z_1 are expressed in terms of the spheroidal ones according to formulas (1). Asterisk at τ_* can now be omitted, and we eventually obtain

$$\xi_{00} = -2\pi c \left\{ \left[\frac{(1 + \tau_1^2)(1 - t_1^2)(\tau^2 - 1)}{2} \right. \right.$$

$$\left. - (1 + t_1^2 \tau_1^2)(1 + \tau^2) \right] \arctan\frac{1}{\tau}$$

$$\left. + \tau(t_1^2 \tau_1^2 - 1) - \frac{\tau}{2}(1 + \tau_1^2)(1 - \tau_1^2) \right\}$$

or, after the use of the particular expressions for the Legendre functions and for their combinations V, W, L, and $\tilde{L}$,

$$\xi_{00} = 2\pi c \left\{ \tau + \frac{4}{9}[1 + V_2^0(\mathbf{r}_1)][\tilde{L}_0^0(\mathbf{r}) + P_2(t)] \right.$$

$$\left. + \frac{4}{9} L_0^0(\mathbf{r}_1) W_0^0(\mathbf{r}) \right\}. \tag{8}$$

In this case, the difficulty arose since $\xi_{00}(\mathbf{r}_1)$ is an unbounded function. This is not the case for other ξ_{nk}, because the principal term in the asymptotics for D disappears immediately on integration and $\xi_{nk}(\mathbf{r}_1)$ proves to be bounded. This function can be found if the value of the Laplacian is known. Since

$$\Delta D = \Delta^{(1)} D = \frac{2}{D} \tag{9}$$

(superscript (1) denotes the coordinate of the point $\mathbf{r}_1$), we merely have

$$\Delta^{(1)} \xi_{nk} = 2\alpha_{nk},$$

where α_{nk} are defined by formulas (6).

It is sufficient to construct a bounded and everywhere continuous solution of this Poisson equation; then, by virtue of the Liouville theorem [9], it will coincide with the function $\xi_{nk}(\mathbf{r}_1)$ desired up to an additive constant (which proves to be zero while checking).

Note that we can use the formula (the proof is omitted)

$$\Delta \left\{ \frac{(n - k + 1)(n - k + 2)}{(2n + 3)^2} L_n^k \right.$$

$$\left. + \frac{(n + k - 1)(n + k)}{(2n - 1)^2} L_{n-2}^k \right\} = \frac{2(2n + 1)}{c^2} V_n^k \tag{10}$$

and construct exactly the same formula relating the functions $\tilde{L}$ and $-W$. In so doing, the functions L_{n-2}^k and $\tilde{L}_{n-2}^k$ should be defined carefully at $n = k$ or $n = k + 1$, because one has to check that the functions p and q appearing in L_{n-2}^k and $\tilde{L}_{n-2}^k$ obey standard recurrent formulas. It is easy to show that the functions p_n^k should be considered to be identically zeros at $n < k$ and the definition of the functions q_n^k should be completed as

$$q_{k-1}^k(\tau) = (2k - 2)!!(1 + \tau^2)^{-k/2},$$

$$q_{k-2}^k(\tau) = (2k - 2)!!\tau(1 + \tau^2)^{-k/2}. \tag{11}$$

The case $k = 0$ is an exception. Formulas (11) become invalid, we do not define the corresponding functions, and two of formulas (10) are replaced by

$$c^2 \Delta \left\{ -\frac{1}{9} \tilde{L}_0^0 + \frac{\tau}{2} \right\} = W_0^0,$$

$$c^2 \Delta \left\{ -\frac{1}{25} \tilde{L}_0^0 - \frac{t}{6} \right\} = W_1^0. \tag{12}$$

Therefore, except for another singular case $n = 1$ and $k = 0$, one can write immediately

$$\xi_{nk}(\mathbf{r}_1) = c \left[\frac{(n-k)!}{(n+k)!} \right]^2 q_n^k(\tau) \left[\frac{(n-k+1)(n-k+2)}{(2n+3)^2} \right.$$

$$\times L_n^k(\mathbf{r}_1) + \frac{(n+k-1)(n+k)}{(2n-1)^2} L_{n-2}^k(\mathbf{r}_1) \right] + \tilde{\zeta}(\mathbf{r}_1)$$

at $\tau > \tau_1$ and

$$\xi_{nk}(\mathbf{r}_1) = -c \left[\frac{(n-k)!}{(n+k)!} \right]^2 p_n^k(\tau) \left[\frac{(n-k+1)(n-k+2)}{(2n+3)^2} \right.$$

$$\times \tilde{L}_n^k(\mathbf{r}_1) + \frac{(n+k-1)(n+k)}{(2n-1)^2} \tilde{L}_{n-2}^k(\mathbf{r}_1) \right] + \tilde{\zeta}(\mathbf{r}_1) \tag{13}$$

at $\tau < \tau_1$. Here, the harmonic functions ζ and $\tilde{\zeta}$ are defined from the sewing conditions for the functions ξ_{nk} and their normal derivatives on the spheroid $\tau = \text{const}$ (this is equivalent to comparing the derivatives with

respect to τ_1 at the point $\tau = \tau_1$). As a result of analysis, which is clarified below, we arrive at the desired expression, namely,

$$
\zeta(\mathbf{r}_1) = \left\{ \frac{(n-k+2)!(n-k)!}{(2n+3)^2[(n+k)!]^2} P_n^k(t_1) p_n^k(\tau_1) q_{n+2}^k(\tau) \right.
$$

$$
- \frac{[(n-k)!]^2}{(2n-1)^2(n+k-2)!(n+k)!} P_n^k(t_1) p_n^k(\tau_1) q_{n-2}^k(\tau)
$$

$$
+ \frac{[(n-k+2)!]^2}{(2n+3)^2(n+k)!(n+k+2)!} P_{n+2}^k(t_1) p_{n+2}^k(\tau_1) q_{n+2}^k(\tau)
$$

$$
+ \frac{(n-k)!(n-k-2)!}{(2n-1)^2[(n+k-2)!]^2} P_{n-2}^k(t_1) p_{n-2}^k(\tau_1) q_{n-2}^k(\tau)
$$

$$
\left. + \gamma_{nk} P_n^k(t_1) p_n^k(\tau_1) q_n^k(\tau) \right\} c \exp(-ik\varphi_1),
$$

where

$$
\gamma_{nk} = -\frac{4(4k^2-1)(2n+1)}{(2n-1)^2(2n+3)^2} \left[\frac{(n-k)!}{(n+k)!} \right]^2, \qquad (14)
$$

and $\tilde{\zeta}$ is expressed in the same way but the symbols p and q change places everywhere. The verification is based on several identities for the Legendre functions, which are lacking in the reference books and call for special derivation.

First, we introduce the function $M_n^k = q_{n+2}^k p_n^k - q_n^k p_{n+2}^k$. One can easily show (from the definition of the functions q) that $M_n^k(\tau)$ is a first-degree polynomial odd with respect to τ. Therefore, it is sufficient to determine one coefficient using the asymptotics for p and q at large τ. The calculation yields

$$
M_n^k = -(2n+3) \frac{(n+k)!}{(n-k+2)!} \tau. \qquad (15)
$$

Second, we use the combination $\Omega_n^k = q_{n+2}'^k p_n^k + q_n'^k p_{n+2}^k - q_{n+2}^k p_n'^k - q_n^k p_{n+2}'^k$. Differentiating it and replacing the resulting second derivatives by the well-known differential equations for p_n^k and q_n^k [1], we obtain

$$
\Omega_n'^k = -\frac{2\tau}{1+\tau^2} \Omega_n^k + \frac{2(2n+3)}{1+\tau^2} M_n^k, \qquad (16)
$$

where the function M_n^k is defined by formula (15).

We treat formula (15) as a differential equation for Ω_n^k. It is easy to integrate:

$$
(1+\tau^2)\Omega_n^k = -\frac{(2n+3)^2(n+k)!}{(n-k+2)!}(\tau^2 + C).
$$

The constant of integration C can be determined, e.g., by the substitution $\tau = i$ We obtain

$$
\Omega_n^k = -\frac{(2n+3)^2(n+k)!}{(n-k+2)!}
$$

$$
+ \left[\frac{(n+k)!}{(n-k)!} + \frac{(n+k+2)!}{(n-k+2)!} \right] \frac{1}{1+\tau^2}.
$$

Along with $\Omega_n^k(\tau)$ and $M_n'^k(\tau)$, we also employ differential relation (5) while comparing the derivatives with respect to τ on both sides.

The above relations suffice to validate the choice of the functions ζ and $\tilde{\zeta}$ (it is not necessary to add a constant, because the expression for $\xi_{nk}(\mathbf{r}_1)$ shows the correct asymptotical behavior, vanishing at $r_1 \longrightarrow \infty$ at least on the average over the sphere). According to formulas (12), in the particular case $n = 1$ and $k = 0$, $-t/6$ must appear in the second formula of (13) in place of the second term in brackets.

It remains to collect all terms in the expansion for D. In this case, for each k, the pairwise grouping of the terms having the subscript n other than 2 seems to be the most appropriate. The singular cases $n = 0$ and $n = 1$ at $k = 0$ give additional terms with separated τ and t. As a result, we have

$$
D = c \sum_{k=-\infty}^{\infty} \sum_{n=|k|}^{\infty} \left[\frac{(n-k)!}{(n+k)!} \right]^2 \left[-\frac{(n-k+1)(n-k+2)}{(2n+3)^2} \right.
$$

$$
\times \tilde{L}_n^k(\mathbf{r}) V_n^{*k}(\mathbf{r}_1) - \frac{(n+k-1)(n+k)}{(2n-1)^2} \tilde{L}_{n-2}^k(\mathbf{r}) V_n^{*k}(\mathbf{r}_1)
$$

$$
+ \frac{(n-k+1)(n-k+2)}{(2n+3)^2} L_n^{*k}(\mathbf{r}_1) W_n^k(\mathbf{r}) \qquad (17)
$$

$$
+ \frac{(n+k-1)(n+k)}{(2n-1)^2} L_{n-2}^{*k}(\mathbf{r}_1) W_n^k(\mathbf{r})
$$

$$
\left. + \gamma_{nk} V_n^{*k}(\mathbf{r}_1) W_n^k(\mathbf{r}) \right] + c(\tau - tt_1\tau_1)
$$

in the range $\tau \geq \tau_1$; otherwise, $\mathbf{r}$ and $\mathbf{r}_1$ as well as the respective coordinates, change roles. The terms with L_{n-2}^k and $\tilde{L}_{n-2}^k$ are omitted at $|k| \geq n - 1$ [5].

Thus, formula (17) with specific γ_{nk} defined by relations (14) is fully proved. For testing, one can, e.g., consider the case when both points are located on the polar axis; i.e., $t = t_1 = 1$, $r = c\tau$, and $r_1 = c\tau_1$. If we put $r \geq r_1$ for definiteness, the simple rearrangement of the terms

on the right-hand side of Eq. (17) (only the terms with $k = 0$ remain, and all but the isolated term $c|\tau - \tau_1|$ cancel out) immediately yields the correct expression for the distance: $D = r - r_1$. In another test, r and r_1 tend to infinity simultaneously and proportionally with the angular coordinates retained. Then, we have the asymptotes $r \approx c\tau$ and $t \approx \cos\theta$ (recall that θ is the polar angle); similar asymptotes hold for the other point. From formula (17), we derive the relation

$$D = \sum_{k=-\infty}^{\infty} \sum_{n=|k|}^{\infty} \frac{(n-k)!}{(n+k)!} \left[\frac{r_1^{n+2}}{(2n+3)r^{n+1}} - \frac{r_1^n}{(2n-1)r^{n-1}} \right]$$
$$\times P_n^k(\cos\theta) P_n^k(\cos\theta_1) \exp[ik(\varphi - \varphi_1)],$$

which also directly follows from the theory of spherical functions. For this purpose, one multiplies the well-known series (see, for example, [1, p. 145])

$$D^{-1} = \frac{1}{\sqrt{r^2 + r_1^2 - 2rr_1 \cos v}} = \sum_{s=0}^{\infty} \frac{r_1^s}{r^{s+1}} P_s(\cos v)$$

by $D^2 = r_1^2 + r^2 - 2rr_1\cos v$ and applies the recurrent formula $(n - k + 1)P_{n+1}^k(t) - (2n + 1)tP_n^k(t) + (n + k)P_{n-1}^k(t) = 0$.

CONCLUSION

The importance of expansions (7) and (17) is that they give a solution of the Poisson equation and the equation

$$\Delta\Delta u(\mathbf{r}) = 4\pi f(\mathbf{r}), \tag{18}$$

respectively, in a certain coordinate system.

Indeed, as is known [10],

$$u(\mathbf{r}) = \frac{1}{2} \iiint Df(\mathbf{r}_1) d\mathbf{r}_1 \tag{19}$$

is a particular solution of Eq. (18). If the function f is different from zero in a finite region of the space, function (18) proves to be biharmonic in the entire remaining region and integral (19) with D in the form of expansion (17) yields an expression for this biharmonic function in the form of a series in some standard set of functions. At the same time, this validates the completeness of the set of the biharmonic functions V_n^k and L_n^k inside a specific compression of the spheroid S. Indeed, the function v, which is harmonic inside the spheroid, can always be represented by the potential of an elementary layer

$$v(\mathbf{r}) = \int_S \frac{\mu(\mathbf{r}_1)}{D} d\sigma_1,$$

where $d\sigma_1$ is an element of the surface S and μ is some surface density.

For an arbitrary sufficiently smooth biharmonic (four-times differentiable) function ψ we identify v with $\Delta\psi$ inside the spheroid. Then, for the function

$$\psi_1(\mathbf{r}) = \frac{1}{2} \int_S D\mu(\mathbf{r}_1) d\sigma_1$$

we obtain $\Delta\psi_1 = v$. Hence, the initial function ψ is represented by the sum

$$\psi = \frac{1}{2} \int_S D\mu(\mathbf{r}_1) d\sigma_1 + \chi, \tag{20}$$

where χ is some harmonic function.

However, $\chi(\mathbf{r})$ can always be expanded in the set of functions V [1]. According to our basic formula (17), the expansion of D represents the integral in formula (20) as the superposition of V, L, t, and τ, so that the completeness of this set in the above-mentioned class of sufficiently smooth biharmonic functions inside the spheroid is proved. The completeness of the set of the functions W, $\tilde{L}$, t, and τ outside the oblate spheroid in the class of sufficiently smooth biharmonic functions bounded at infinity is proved in a similar way.

Recall that sometimes the character of a problem requires just the spheroidal coordinates to be used. In particular, such conditions appear in a natural way if the function f is most conveniently expressed in the spheroidal coordinates. Physically, such a situation can be associated with the specific role of some reference spheroid or, in the limiting case, a circle of finite radius in a particular problem. The example of astrophysical applications is given in [11]. More complex examples may appear in the dynamics of self-gravitating fluids with regard for viscosity. The technique developed in this study can find wide application, because Eq. (18) and similar ones are often encountered in problems of the theory of elasticity, the theory of plasticity, and hydrodynamics [12, 13]. The specific role of some circle is also common in engineering problems of the equilibrium of a three-dimensional elastic medium [14].

REFERENCES

1. E. W. Hobson, *The Theory of Spherical and Ellipsoidal Harmonics* (Cambridge Univ. Press, Cambridge, 1931; Inostrannaya Literatura, Moscow, 1952).

2. V. A. Antonov, E. I. Timoshkova, and K. V. Kholshevnikov, *An Introduction to the Theory of Newtonian Potential* (Nauka, Moscow, 1988).

3. F. G. Tricomi, *Lezioni sulle equazioni a devivate parziali* (Torino, 1954; Inostrannaya Literatura, Moscow, 1957).

4. Yu. N. Rabotnov, *Mechanics of Deformable Solid* (Nauka, Moscow, 1988).

5. A. S. Baranov, Zh. Vychisl. Mat. Mat. Fiz. **37**, 395 (1997).

6. L. N. Sretenskiĭ, *The Theory of Newtonian Potential* (Gostekhteorizdat, Moscow, 1946).

7. H. Jeffreys and B. Swirles, *Methods of Mathematical Physics* (Cambridge Univ. Press, Cambridge, 1966; Mir, Moscow, 1969), Vol. 1.

8. S. Chandrasekhar, *Ellipsoidal Figures of Equilibrium* (Yale University Press, New Haven, Conn., 1969; Mir, Moscow, 1973).

9. A. F. Timan and V. M. Trofimov, *An Introduction to the Theory of Harmonic Functions* (Nauka, Moscow, 1968).

10. J. Boussinesq, *Application des Potentiels à l'Etude de l'Equilibre et du Mouvement des solides élastiques* (A. Blanchard, Paris, 1969).

11. S. Chandrasekhar and N. R. Lebovitz, Astrophys. J. **136**, 1037 (1962).

12. N. I. Bezukhov, *Foundations of the Theory of Elasticity, Plasticity and Creep* (Vysshaya Shkola, Moscow, 1968).

13. G. K. Batchelor, *An Introduction to Fluid Dynamics* (Cambridge Univ. Press, Cambridge, 1967; Mir, Moscow, 1973).

14. A. E. Andreĭkiv, *Three-Dimensional Problems of Theory of Cracks* (Naukova Dumka, Kiev, 1982).

Translated by D. Zhukhovitskiĭ

GASES
AND LIQUIDS

Investigation of a Free Convective Current Using Holographic Interferometry

N. M. Ganzherli, I. A. Maurer, and D. F. Chernykh

Ioffe Physicotechnical Institute, Russian Academy of Sciences,
Politekhnicheskaya ul. 26, St. Petersburg, 194021 Russia
e-mail: nina@holo.ioffe.rssi.ru
Received April 9, 2001; in final form, June 1, 2001

Abstract—By the method of holographic interferometry in real time, an investigation is carried out of the development of free convective currents in an enclosed cavity filled with different liquids exposed to heat fluxes from either linear or point sources. Characteristic times of the processes are measured. Temperature fields in the current are plotted. © 2002 MAIK "Nauka/Interperiodica".

INTRODUCTION

Nonstationary heat transfer processes in gases or liquids are complicated physical phenomena accompanied by the development of currents with varying temporal and spatial characteristics. Mathematical simulation of such processes requires the solution of a set of nonlinear differential equations with boundary conditions, also in the form of differential equations. The objective of simulation is to obtain patterns of varying temperature and velocity fields in the flow. An experimental study allows the mathematical model of the phenomenon to be made more precise and adequate to the physical process [1]. Probing nonstationary flows with the help of a dense grid of probes meets with difficulties because the probes interfere with the flow. Besides, the finite response time of the probes restricts the possibilities of studying fast nonstationary processes. Therefore, for studies of heat transfer processes, noncontact optical methods [2, 3] are of special interest—among them, interferometric and holographic techniques, which substantially expand capabilities of optical experiments in visualization of phase inhomogeneities [4, 5].

When light passes through an inhomogeneous medium, optical path differences arise caused by variations in the refractive index along the light path. These variations cause a change in the light wave phase $\Delta\varphi(x, y)$, which is defined by the expression

$$\Delta\varphi(x, y) = \frac{2\pi}{\lambda}\int_l [n(x, y, z) - n_0]dl, \qquad (1)$$

where x and y are the coordinates of a point in the plane of observation; λ is the light wavelength; n_0 is the refractive index of a medium in its undisturbed state at the initial time instant; $n(x, y, z)$ is the distribution of the refractive index in the medium at the moment of record-ing of the interferogram; and l is the light ray path through the medium.

The light wave phase incursions change the interference conditions in the observation plane and cause variations in the field illumination. Using measured $\Delta\varphi(x, y)$ one can evaluate the distribution of the refractive index $n(x, y, z)$ and such parameters of the medium related with the refractive index as the density, temperature, etc.

EXPERIMENTAL

As an example of a nonstationary process, we consider development of a quasi-free convective current in an enclosed cavity exposed to a localized heat source. Of importance to the development of the theory of currents in liquids is the study of buoyancy, which, in a number of situations, can appreciably influence the heat transfer intensity. Models of some convective processes that form currents due to buoyancy can be sorted, according to [6], into two vast groups: buoyant currents and thermal bubbles. In both cases, the motion in a medium is caused by the force of gravity because of lower liquid density near the heat source. Fundamentals of the theory of buoyant currents are presented in [7, 8].

Experimental investigations of heat transfer processes of a similar kind were carried out by a number of authors [9–11], for example, with the help of a grating interferometer [10, 11].

We studied the development features of a quasi-free convective flow in an enclosed cavity for a number of simple liquids with the Prandtl number Pr in a range from 4.6 to 34. The basic physical characteristics of the liquids used were taken from [12, 13].

As a heat source we used a standard MLT resistor (with the resistance $R = 430\ \Omega$) 0.5 cm in length and 0.18 cm in diameter. The amount of heat was controlled by the current passing through the resistor and varied from 0.02 to 1.7 W. The heat source was placed inside

a metal cavity with two side walls of glass for interferometric-holographic diagnostics. The dimensions of the cavity and the observation windows were $10 \times 30 \times 50$ mm and 30×50 mm, respectively.

Two series of experiments were carried out with different positions of the resistor at the cavity bottom. In the first case, the resistor was located along the illumination direction, which made the heat flux more uniform across the liquid layer thickness and the resistor could be assumed close to linear. In the second series of experiments, the heat source was housed inside a nozzle with an outlet diameter of about 1 mm. In this case, the resistor could be considered a pointsource.

The development of a free convective current was studied with the use of a holographic interferometer. A feature of the optical scheme of the interferometer was a ground glass placed immediately behind the cavity (Fig. 1) [14]. In this case, the ground screen represented the localization plane of the resulting interference pattern, which simplified recording of the interferograms; however, it imposed a restriction on the character of variations of the refractive index in the medium under study acceptable for analysis.

Rigorous analysis of the changes in the refractive index on the basis of interferograms could be made for plane objects and for objects in which the refractive index depended only on coordinates x and y and did not change along z axis coinciding with the direction of illumination of the object. In particular, the use of a nearly linear heat source allowed us to evaluate changes in the refractive index $\Delta n = n(x, y) - n_0$ using a simplified formula

$$\Delta \varphi(x, y) = \frac{2\pi l}{\lambda}[n(x, y) - n_0], \qquad (2)$$

or, changing from the phase incursion $\Delta \varphi(x, y)$ to the optical path difference expressed in terms of the number of fringes $k(x, y)$ [4, 5], by the formula

$$n(x, y) = \frac{k(x, y)\lambda}{l} + n_0. \qquad (3)$$

A convective current caused by a point heat source is an axisymmetric object, the axis of which is parallel to the current direction. In this case, expression (1) takes the form of Abel's integral equation. To solve this equation we used a step approximation method proposed by Pierce [15].

Information on the development of a free convective current obtained by the method of finite-fringe holographic interferometry was supplemented with that derived from infinite-fringe interferograms [4, 5]. Finite-fringe interferograms allowed us to determine the sign of the refractive index change, and infinite-fringe interferograms gave directly isothermic lines in the flow field.

The method of holographic interferometry in real time gives a possibility to follow the entire process of development and stabilization of a convective current.
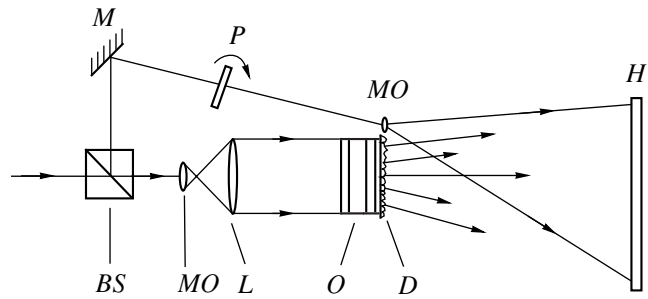
Fig. 1. Optical diagram of a holographic interferometer. *BS* beam splitter; *MO* microscope objective; *L* lens; *M* mirror; *P* plane-parallel plate; *O* object; *D* diffuser; *H* hologram.
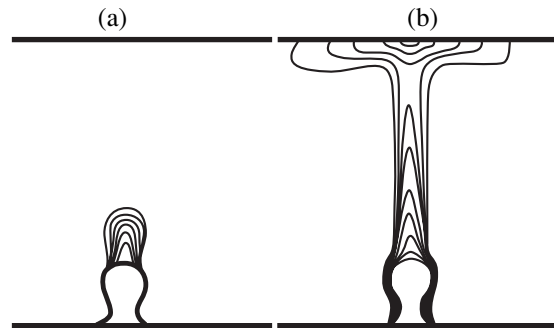

Fig. 2. Temperature fields for two development stages of a free convective current in ethyl alcohol. (a) Formation and detachment of a thermal bubble; (b) formation of a free convective current.

Recording of the interferograms was implemented with the help of filming equipment, a TV camera and a photographic camera. The filming equipment could record the process at speeds from 3 to 16 frames per second and determine the temporal characteristics of the process with good accuracy, namely, the time of heating (due to heat conduction) until detachment of a thermal bubble τ_1, the time elapsed before the convective current reached the top to the cavity, and the time τ_2 in which a quasi-stationary convection regime is established (all times were measured from the start of heating).

The temperature was measured with three thermocouples: one directly above the heat source, another at the top of the cavity where the current impinged, and the third in a region undisturbed by the heated liquid flow. Data from the thermocouples were used to correct the temperatures calculated from interferometric measurements with the use of the well-known dependence of the refractive index of the liquids on temperature for a light wavelength of 0.63 μm.

RESULTS OF THE EXPERIMENT

First, consider in detail the results of experiments on a free convective current caused by a linear heat source. From an analysis of infinite-fringe interferograms, we
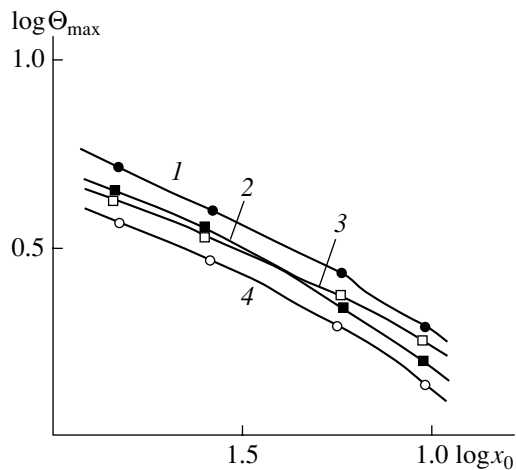
Θ/Θ_{max}
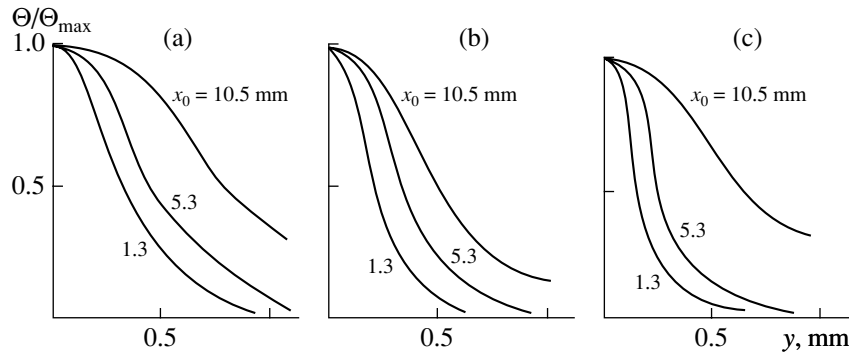
Fig. 3. Reduced temperature profiles in a free convective current at a heat source power of 0.23 W. Three cross sections at different heights x_0 above the heat source.
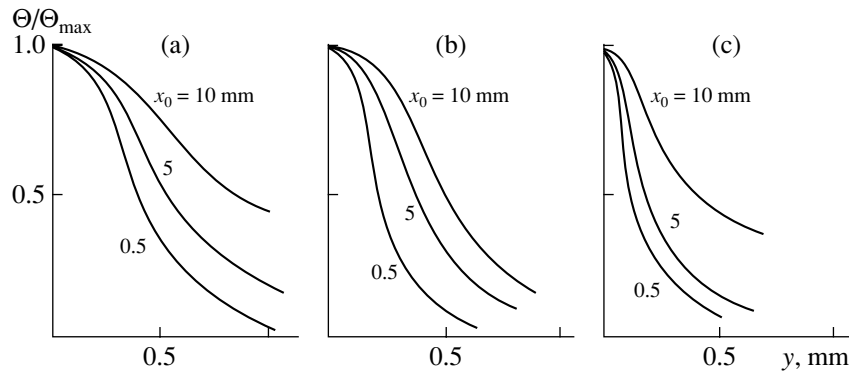
Fig. 4. Reduced temperature profiles in a free convective current in acetone at different heat source powers. Three cross sections at different heights x_0 above the heat source.

obtained the temperature fields for different instants of time as the free convective current developed in an enclosed cavity filled with liquid. As an example, temperature fields obtained at different moments of time in a cavity filled with ethyl alcohol at a heat source power of 0.23 W are presented in Fig. 2. Figure 2a depicts the

Fig. 5. Lapse of the temperature Θ_{max} along the symmetry axis of a free convective current with x_0 in (*1*) ethyl alcohol, (*2*) isopropyl alcohol, (*3*) acetone, and (*4*) distilled water.

formation of a thermal bubble (3 s after start of heating), and Fig. 2b corresponds to an encounter of the quasi-free convective current with the top of the cavity (10 s after start of heating). The temperature difference $\Delta\Theta$ between adjacent isotherms is a characteristic of the particular liquid and depends on the temperature factor of the refractive index. The temperature increment between isotherms in Fig. 2 is 0.395°C.

Analysis of the interferograms allowed us to obtain reduced temperature profiles in free convective currents that developed in the liquids under study at different heat fluxes. Figure 3 shows reduced temperature profiles in a free convective current along the coordinate y in horizontal cross sections at various heights x_0 above the 0.23-W heat source in (a) acetone, (b) ethyl alcohol, and (c) turpentine.

Figure 4 shows reduced temperature profiles in a free convective current in acetone at different heat source powers: (a) 0.02, (b) 0.08, and (c) 0.23 W.

In Fig. 5, a profile is shown of the maximum excess (with respect to the liquid temperature in the thermally undisturbed area) temperature Θ_{max} along the symmetry axis of a free convective current with height x_0 above the heat source. In the area of the free convective current where the influence of the heat source, as well as
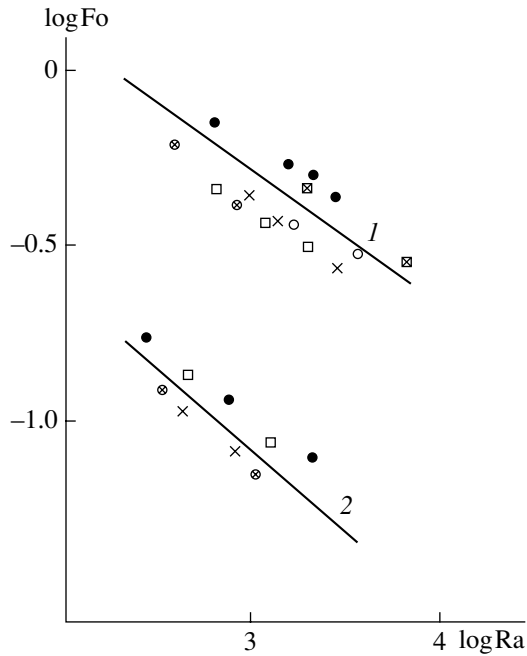
Fig. 6. Relationship between logFo and logRa for two development regimes of a free convective stream. (*1*) Convection regime; (*2*) heat conduction regime. ⊗ Water; × isopropyl alcohol, ● ethyl alcohol, ○ carbon tetrachloride; □ turpentine; ⊠ acetone.

the effect of the current colliding with the cavity top, is weak, variation of Θ_{max} in the current is roughly in proportion to $x_0^{-0.5}$.

A relationship between Fourier and Rayleigh's criteria of similarity, Fo and Ra, calculated for the cases considered, is of interest from the theoretical point of view. Fourier's number characterizes the rate of forma-

tion of a temperature field at given heating conditions and has the form Fo = $\alpha\tau/l^2$, where α is the temperature conductivity, τ is the characteristic time of the process, and l is the characteristic size of the object. Fourier's number expresses the ratio of the temperature distribution along the current axis to the temperature change with time at each point of the axis. Rayleigh's number Ra = $gl^3\beta_0\Theta/v\alpha$ characterizes the relationship between the heat flux in a medium due to the buoyancy and free convection, on the one hand, and two diffusion processes interfering with the motion and stabilizing it, namely, the internal friction due to viscosity and the heat conduction, on the other hand [8]. Here, g is the gravitational acceleration, l is the characteristic size of the cavity, β_0 is the coefficient of volumetric thermal expansion, Θ is the temperature excess, v is the coefficient of kinematic viscosity, and α is the temperature conductivity.

The dependence of Fourier's number on Rayleigh's number can be experimentally obtained by measuring the characteristic development times of convective current and the corresponding temperature increments. Functions Fo = f(Ra) were found for the characteristic heating time by conduction and the corresponding temperature increase in the thermal bubble, as well as for the characteristic time of formation of the free convection that developed and the corresponding temperature increase. The characteristic times for different liquids and different heat fluxes were determined to an accuracy of 0.1 s. The temperature changes were obtained from the interferograms and compared with the thermocouple measurement. The table contains the time τ_1 of conduction heating until detachment of the thermal bubble and the corresponding bubble temperature Θ_1, as well as the time τ_2 for formation of the quasi-stationary convection regime and the corresponding tempera-
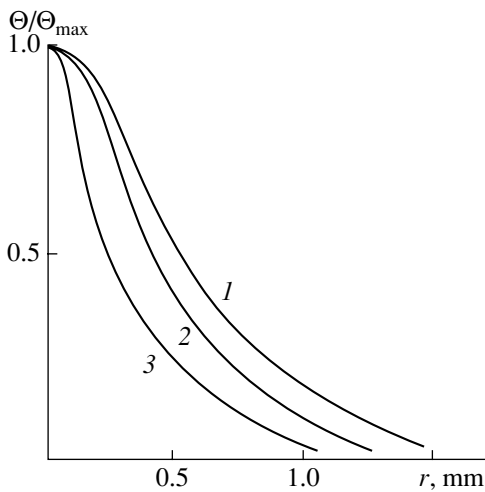
Fig. 7. Radial distributions of the dimensionless temperature changes in ethyl alcohol at the height $x_0 = 10$ mm above the nozzle. The heat flux is produced by a point source of different powers: (*1*) 0.23; (*2*) 0.52; (*3*) 0.67 W.
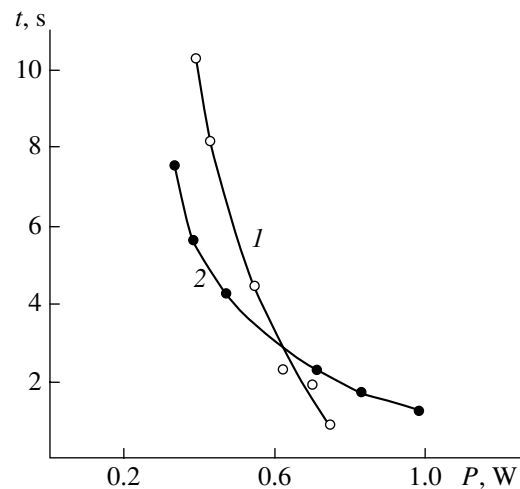
Fig. 8. Time before detachment of the thermal bubble versus heat source power in (*1*) isopropyl alcohol and (*2*) ethyl alcohol.

Times of formation of the thermal bubble, times of establishing the convection regime, and the corresponding temperature increments in the current at different heat fluxes

Liquid	0.02 W				0.08 W			
	τ_1, s	Θ_1, °C	τ_2, s	Θ_2, °C	τ_1, s	Θ_1, °C	τ_2, s	Θ_2, °C
Water (7.5)								
Ethyl alcohol (16.8)	5	1.2	10	2.4	3	2.4	9	5.15
Isopropyl alcohol (34)	6	1.1	10	2.7	4.2	4.2	8	5.0
Acetone (4.2)			12	2.2			6	4.6
Carbon tetrachloride (4.6)	3.3	1.3	10	2.1	2		8	5.0
Turpentine (21.2)			9	3.15			7	6.0
	0.23 W				0.48 W			
Water (7.5)	3.5	6.5			2.8	8.05		
Ethyl alcohol (16.8)	2.7		8	9.1	2		7	11.8
Isopropyl alcohol (34)	3		8		2		6	
Acetone (4.2)			5	6.6			3.5	6
Carbon tetrachloride (4.6)	1		6		1		4	
Turpentine (21.2)			5	6.5			4	10

Note: The Prandtl number is given in parentheses.

ture increase Θ_2 in the current for various liquids with different Prandtl numbers.

The plot of logFo as a function of logRa is given in Fig. 6. The points in the plane logFo, logRa were approximated by linear functions using the least-squares method. The slope of curve *1* corresponds to Fo ~ Ra$^{-0.21}$ which agrees with theoretical prediction for a free convective current in an enclosed cavity [8].

In the case of a free convective current produced by a point heat source, using changes of the refractive index Δn_i determined by Pierce's method of step approximation, radial distributions of the refractive index in the convective current for a number of liquids were plotted. The calculations were performed for several cross sections above the nozzle. Figure 7 demonstrates radial distributions of the temperature increases in ethyl alcohol at a height of $x_0 = 10$ mm above the nozzle at different heat source powers.

In Fig. 8, variation of the time until detachment *t* of the thermal bubble with the increase in heat source power *P* are shown for isopropyl alcohol (curve *1*) and ethyl alcohol (curve *2*).

CONCLUSION

The results of the investigation of development of a free convective current in an enclosed cavity provide evidence of the efficiency of holographic interferometry in problems of heat transfer in transparent media.

The reduced temperature profiles and the current development characteristics obtained are in agreement with the theory. The developed interferometric-holographic technique and the vast experimental data obtained with this technique can be of interest to specialists in the field of free convection.

REFERENCES

1. D. N. Popov, *Nonstationary Hydromechanical Processes* (Mashinostroenie, Moscow, 1982).
2. L. A. Vasil'ev, *Schlieren Methods* (Nauka, Moscow, 1968; Israel Program for Scientific Translations, Jerusalem, 1971).
3. W. Hauf and U. Grigull, *Optical Methods in Heat Transfer* (Academic, New York, 1970; Mir, Moscow, 1973).
4. Yu. I. Ostrovskiĭ, M. M. Butusov, and G. V. Ostrovskaya, *Holographic Interferometry* (Nauka, Moscow, 1977).
5. A. K. Beketova, A. F. Belozerov, A. N. Berezkin, *et al.*, *Holographic Interferometry of Phase Objects* (Nauka, Leningrad, 1979).

6. J. S. Turner, *Buoyancy Effects in Fluids* (Cambridge Univ. Press, Cambridge, 1973; Mir, Moscow, 1977).

7. Ya. B. Zel'dovich, Zh. Éksp. Teor. Fiz. **7**, 1463 (1937).

8. O. G. Martynenko and Yu. A. Sokovishin, in *Free-Convective Heat Exchange: Handbook* (Minsk, 1982), p. 162.

9. K. S. Mustafin, V. A. Seleznev, and E. I. Shtyrkov, Opt. Spektrosk. **22**, 319 (1967).

10. Yu. I. Lyakhov, Prikl. Mekh. Tekh. Fiz., No. 2, 169 (1970).

11. V. L. Zimin and Yu. N. Lyakhov, Prikl. Mekh. Tekh. Fiz., No. 3, 159 (1970).

12. N. B. Vargaftik, *Tables of the Thermophysical Properties of Liquids and Gases* (Nauka, Moscow, 1972; Halsted Press, New York, 1975).

13. *Tables of the Physical Quantities: Handbook*, Ed. by I. K. Kikoin (Atomizdat, Moscow, 1976).

14. A. L. Barannikov, N. M. Ganzherli, S. B. Gurevich, *et al.*, Izv. Akad. Nauk SSSR, Ser. Fiz. **49**, 711 (1985).

15. W. D. Pierce, in *Receiving and Studying of High Temperature Plasma* (Inostrannaya Literatura, Moscow, 1962).

Translated by N. Mende

The Possibility of Using a Low-Tritium-Content D–T Mixture Compressed with a High-Power Current Pulse

M. V. Fedulov
Received July 27, 2001

Abstract—The possibility of the propagation of a thermonuclear detonation wave in a deuterium pinch with a small (~0.5%) tritium admixture is considered. The required pinch parameters and the necessary value of the current produced by a high-power pulsed current generator are estimated. © *2002 MAIK "Nauka/Interperiodica"*.

INTRODUCTION

Progress in the development of high-power pulsed current generators with progressively increasing currents gives hope that, in the foreseeable future, it will be possible to achieve the parameters required for the ignition and maintenance of a thermonuclear detonation wave with an acceptable energy gain ratio. In earlier calculations of the propagation of a thermonuclear detonation wave [1–10], a 50/50 D–T mixture was considered. At the same time, it is interesting to consider a mixture with a tritium concentration corresponding to the complete tritium reproduction. Since the ratio between the rates of the reactions $d + d \longrightarrow p + t$ (which will be designated by S_p) and $d + t \longrightarrow n + \text{He}_4$ (S_t) in the temperature range 20–40 keV varies approximately from 0.005 to 0.01 [11], it reasonable to consider a D–T mixture with a tritium concentration of 0.5 to 1%. To avoid misunderstanding, hereinafter, by the reaction rate is meant the averaged product of the cross section by the relative particle velocity $\langle \sigma v \rangle$. Below, we will derive estimating relationships between the pinch parameters required for the detonation wave to propagate in a low-tritium-content mixture (LTCM), provided that the necessary temperature is attained in a certain region of the pinch.

EVALUATION OF THE PRODUCT OF THE DENSITY BY THE RADIUS FOR A FUEL WITH A LOW TRITIUM CONTENT

Calculations showed [5–10] that, for the detonation wave to propagate through a compressed column of a 50/50 D–T mixture, the sufficient value of the product of the mixture mass density ρ_c by the column radius r_c is about 0.25–0.5 g/cm². Let us denote this product by λ_r and estimate its value for an LTCM. Since, for the following analysis, it is necessary to know the qualitative relation between $\lambda_r = \rho_c r_c$ and the tritium concentration c_t, we will reproduce a brief derivation of the dependence of λ_r on the temperature T in the forepart of the burning zone (FBZ), which determines the heating of the cold plasma ahead of the thermonuclear detonation wave. For a low tritium content, it is the FBZ where the contribution from the reaction $d + \text{He}_3 \longrightarrow p + \text{He}_4$ with a released energy of 18.3 MeV can be neglected, because the He₃ accumulation period $\tau_g = 1/S_g n_d$ (where S_g is the reaction rate) is long compared to the ratio of the FBZ length δ_z to the detonation wave velocity $v_b \cong 10^8 T^{1/2}$ cm/s (here and below, the temperature is in keV, the particle energy is in MeV, the current J in MA, and other quantities are expressed in CGS units, unless otherwise specified). The long duration of the He₃ accumulation period means that

$$\lambda_z = \rho \delta_z \ll 3 \times 10^{16} T^{1/2}/S_g(T) \text{ g/cm}^2. \quad (1)$$

The left-hand side of inequality (1) is on the order of λ_r; i.e., as will be shown below, it is equal to several g/cm². At the same time, the right-hand side decreases monotonically with temperature in the range 20–40 keV, reaching approximately 55 g/cm² at $T = 40$ keV. Note that, even for $T = 40$ keV, τ_g is equal to ~100/ρ ns, so that, for ρ > 100 g/cm³, τ_g does not exceed the pinch duration. Hence, the contribution from the reaction $d + \text{He}_3 \longrightarrow p + \text{He}_4$ must be taken into account when calculating the total energy yield.

We assume that the temperature is uniform throughout the FBZ and the ratio of the average (more precisely, effective) density ρ to ρ_c in this region does not vary with time. We assume also that the energy loss per unit time through unit area of the FBZ because of the radial expansion and the fuel heating ahead of the wave front is equal to $k_0(T)\rho$, where the coefficient k_0 describes the dependence of the energy loss in the FBZ on T. Assuming that the ratio of the pinch radius r_p to the effective radius of the expanding burning zone (~2V/S, where V and S are the volume and the surface area of the FBZ) is constant (more precisely, independent of c_t), we obtain that the value of λ_r, accurate to a constant factor, is determined by the expression

$$\lambda_r = \text{const} k_0(T)\rho^2 \Big/ \Big(\sum S_{ij} n_i n_j \varepsilon_{ij} - W_r\Big), \quad (2)$$

where n_i is the density of nuclei of species i; S_{ij} are the reaction rates; ε_{ij} is the fraction of the reaction energy left by fast particles in the FBZ; and W_r is the bremsstrahlung power per unit volume, which is proportional to ρ^2.

Expression (2) depends only on T, and $a = \text{const}$ can be determined from the previous calculations for a 50/50 mixture. Since energy losses are determined primarily by the pressure and the expansion velocity, we can assume that k_0 is proportional to $T^{3/2}$. This allows us to estimate the dependence of λ_r on T for different tritium concentrations.

Below, we will assume that the minimum of the curve $\lambda_r(T)$ (at $T = T_m$) for a 50/50 mixture corresponds to 0.5 g/cm². Calculations were performed under the assumption that all the energy of an α-particle produced in the d–t reaction remains in the FBZ or is expended on the fuel heating ahead of the wave front, whereas a neutron escapes from the plasma without losing its energy. Figure 1 shows the dependence of λ_r on T calculated under the additional assumption that the medium is transparent to bremsstrahlung. We note that the left branch of the dependence of λ_r on T is unstable: as the temperature is slightly increased, it continues increasing up to T_m; in contrast, as the temperature T is decreased, the burning terminates.

To calculate the dependence of λ_r on T at low tritium concentrations, we should know the values of ε_{ij} both for d–t and d–d reactions. We should also keep in mind that (as will be seen from the results of calculations presented below) the minimum value of λ_r should increase by at least one order of magnitude as compared to a 50/50 mixture. Let us show that, for any slowing-down length of charged particles due to collisions with field particles, the full energy of charged particles remains in the FBZ due to magnetization. Indeed, for the maximum ratio of the Larmor radius r_L of a fast charged particle to the pinch radius r_p, we have

$$r_L/r_p = 0.7(AE)^{1/2}/J, \qquad (3)$$

where A and E are the atomic weight and the particle energy (in MeV) and J is the pinch current.

For LTCM, we are dealing with currents of at least several tens of MA; hence, the ratio r_L/r_p should be small. The total collision cross section σ_n of a neutron with $E = 2.45$ MeV produced in the d–d reaction (in deuterium) is about 6 b [12]; i.e., the ratio of its mean free path to r_p is equal to

$$I_n/r_p = 3.3/\sigma_n\lambda_r \cong 0.5/\lambda_r. \qquad (4)$$

Hence, at λ_r values of several g/cm², we can assume that all the energy of a neutron produced in the reaction $d + d \longrightarrow \text{He}_3 + n$ is transferred to the plasma. More complicated and important is the problem of the slowing down of neutrons produced in the d–t reaction ($E = 14.1$ MeV). According to [12], at $E > 4$ MeV, the value
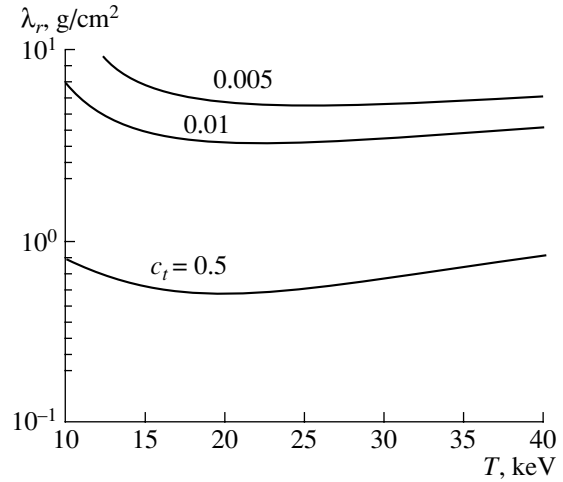
Fig. 1. Dependence of $\lambda_r = \rho_c r_c$ on the temperature for different tritium concentrations c_t.

of σ_n drops rapidly as E increases; consequently, the probability of the first neutron collision can be relatively low. At the same time, a neutron colliding with a deuteron loses, on average, almost one-half of its energy, so that the probability of subsequent collisions increases sharply. For a rough upper estimate, we can assume that the energy fraction left in the pinch plasma by such a d–t neutron is close to the probability of its first collision.

In addition to the case of $c_t = 0.5$, Fig. 1 also shows the dependence of λ_r on T for two other tritium concentrations. Calculations were performed under the assumption that the collisional cross section of a neutron produced in the d–t reaction is equal to 1 b. The results of calculations show that, as the tritium concentration decreases by two orders of magnitude, λ_r increases by more than one order of magnitude, whereas the value of T_m increases insignificantly (from 18 to 24 keV). In this case, the ratio between the d–d and d–t reaction rates increases only from 0.010 to 0.012 and, although c_t decreases, the d–t reaction again makes the main contribution to the heating of the FBZ, largely due to the contribution from a neutron. For a cross section equal to 1 b, the ratio of the mean free path to the pinch radius is equal to $\sim 3/\lambda_r$; i.e., r_p is nearly twice as large as the mean free path. Perhaps more detailed calculations will show that the contribution from a neutron produced in the d–t reaction is somewhat overestimated. We note, however, that, for σ_n equal to 0.5 b, the results of calculations change insignificantly (the minimum value of λ increased by only 14%).

It should also be noted that, as the tritium concentration decreases, the role of bremsstrahlung losses increases. Thus, for $c_t = 0.5$, bremsstrahlung losses amount to only 4% of the total energy lost by α-particles in the FBZ, whereas, for $c_t = 0.005$, this fraction
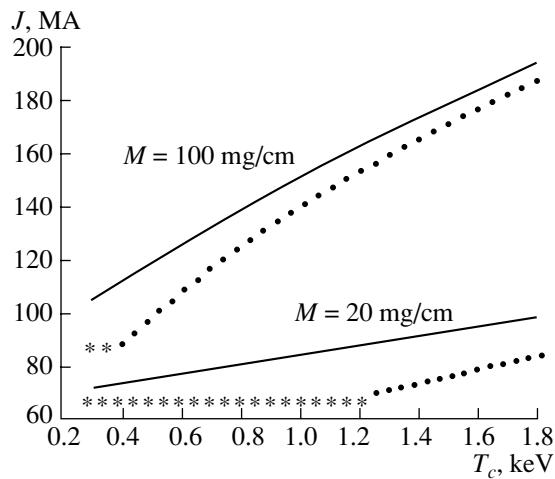
Fig. 2. Current required for confining the fuel heated up to the temperature T_c for $\lambda_r = 6$ g/cm^2: circles correspond to the ideal-gas model, asterisks correspond to the generate electron gas, and solid curves show the results obtained with the TFP model.

increases to one-fourth. Although, in the latter case, the FBZ is longer, it remains transparent to bremsstrahlung, because the ratio of r_p to the mean free path l_r in deuterium is equal to

$$r_p/l_r \cong 0.1\lambda_r^2/r_p T^{7/2}, \qquad (5)$$

so that, for $\lambda_r = 6$ g/cm^2 and $T = 20$ keV, we obtain $r_p/l_r \cong 10^{-4}/r_p$.

CURRENT VALUE AND ENERGY YIELD

Let us now determine the relation between the mass of a D–T mixture per unit length of the plasma column and the current J required for creating an equilibrium pinch with appropriate parameters. Keeping in view tritium reproduction or, in other words, assuming that the d–t reaction is provided by the tritium produced in the reaction $d + d \longrightarrow t + p$, we should have a rather high degree of burnout—no less than 10%. The utilization of six deuterons (with the contribution from the reaction $d + \text{He}_3 \longrightarrow p + \text{He}_4$ taken into account) results in an energy release of 43.25 MeV [13]. Hence, for 10% burnout, about $35M$ GJ per centimeter of plasma column is released, where M is the mass per unit length in g/cm. In this case, it is necessary to compress the plasma down to

$$r_p = M/\pi\lambda_r \text{ cm} \qquad (6)$$

and, accordingly, increase the density up to $\rho_c = \pi\lambda_r^2/M$ g/cm^3. Taking the value λ_r equal to 6 g/cm^2, we obtain that, even for $M = 0.1$ g/cm, the pinch diameter should be equal to $\sim$100 μm and the density should be $\rho_c \cong 1$ g/cm^3.

At such strong plasma compression, the plasma temperature T_c can be very high. Then, we can use the gas equation to determine the current J required for maintaining the equilibrium. Assuming that the temperature is spatially uniform, we can use the well-known formula

$$J = 440(MT_c)^{1/2} \text{ MA}. \qquad (7)$$

For example, if, during compression, a fuel with a mass per unit length of $M = 0.1$ g/cm is heated to only $T_c \cong 0.5$ keV, then a current as high as $J \cong 100$ mA will be required for the pinch confinement. On the other hand, we cannot decrease the value of M with the purpose of decreasing the current and the total released energy in a shot, because it is necessary to compress the fuel to a very small radius r_p.

Note that, when compressing solid deuterium (with a mass density of $\rho_0 = 0.16$ g/cm^3), the initial radius is at least larger than $r_0 = (M/\pi\rho_0)^{1/2}$. This means that compression down to $r = r_p$ requires an increase in the inductance by a value more than

$$L_* = 2\ln(r_0/r_p) = 2\ln((\pi/\rho_0 M)^{1/2}\lambda_r)$$
$$\cong 2\ln(4.4\lambda_r/M^{1/2}) \text{ nH/cm} \qquad (8)$$

(for $\lambda_r = 6$ g/cm^2 and $M = 0.1$ g/cm, we have $L_* \cong 9$ nH/cm).

In this case, the increase in the magnetic energy due to compression, $E_H = L_*J^2/2$ kJ/cm, substantially exceeds the pinch thermal energy $E_r = 0.75J^2$ kJ/cm.

The possibility of decreasing J and E_T by decreasing T_c is limited by the degeneration of the electron gas. Let us consider a limiting case of T_c (expressed in energy units) smaller than the Fermi energy ε_F (for deuterium, we have $\varepsilon_F = 16\rho_c^{2/3}$ eV). If we take into account the pressure of the degenerate electron gas, then, for deuterium, we obtain

$$P = 3.16\rho_c^{5/3} \text{ Mbar}. \qquad (9)$$

For the current value required for maintaining the equilibrium, we obtain

$$J \cong 18M^{5/6}/r_p^{2/3} \cong 39\lambda_r^{2/3}M^{1/6} \cong 130M^{1/6} \text{ MA}. \qquad (10)$$

Hence, for an indefinitely low temperature, the required value of the current exceeds the right-hand side of Eq. (10) (for $M = 0.1$ g/cm, we have $J > 87$ MA). Note that, for a 50/50 D–T mixture (for $\lambda_r = 0.5$ g/cm^2), the minimum current is equal to $25M^{1/6}$ MA (for $M = 0.1$ g/cm, $J > 17$ MA). If, as a reference value at the minimum of the curve $\lambda_r(T)$ for a 50/50 D–T mixture, we take not 0.5 g/cm^2, but one-half of this value [i.e., we halve const $k_0(t)$ in formula (2)], then the calculated value of $\lambda_r(T)$ for a fuel with $c_t = 0.005$ will also

decrease by nearly one-half (the value of λ_r has little effect on the fraction of the energy lost by neutrons produced in the d–t reaction). Hence, at $c_t = 0.005$, the minimum value of λ_r will decrease to ~3 g/cm^2. Under this assumption and with $M = 0.1$ g/cm, the minimum current for a 50/50 mixture at a low temperature will decrease to 10 MA; for a fuel with $c_t = 0.005$, this value will decrease to 55 MA.

Asymptotic expressions (7) and (10) are quite sufficient for the rough evaluation of the required current value. Computations of the dependence of the pressure on the temperature and density with the help of a TFP model [14] demonstrated that the required current value did not increase substantially (Fig. 2).

CONCLUSION

Estimates show that, with a current of 100–200 MA, it is possible to produce a thermonuclear detonation wave in a fuel consisting almost completely of deuterium. The most complicated problem is, apparently, to provide the compression of the fuel with the smallest possible mass to the required size.

ACKNOWLEDGMENTS

REFERENCES

1. A. L. Fuller and R. A. Gross, Phys. Fluids **11**, 534 (1968).

2. M. S. Chu, Phys. Fluids **15**, 13 (1972).

3. A. F. Nastoyashchiĭ, At. Énerg. **32**, 45 (1972).

4. A. F. Nastoyashchiĭ and L. P. Shevchenko, At. Énerg. **32**, 451 (1972).

5. V. V. Cernukha and M. V. Fedulov, in *Proceedings of the 8th European Conference on Controlled Fusion and Plasma Physics, Prague, 1977*, Vol. 11, p. 170.

6. M. V. Fedulov and V. V. Chernukha, Preprint No. 2939 (Kurchatov Institute of Atomic Energy, Moscow, 1978).

7. E. N. Avrorin *et al.*, Fiz. Plazmy **10**, 514 (1984) [Sov. J. Plasma Phys. **10**, 298 (1984)].

8. V. V. Vikhrev and V. V. Ivanov, Preprint No. 4161/6 (Kurchatov Institute of Atomic Energy, Moscow, 1984).

9. V. V. Vikhrev and V. V. Ivanov, Dokl. Akad. Nauk SSSR **282**, 1106 (1985) [Sov. Phys. Dokl. **30**, 492 (1985)].

10. V. V. Ivanov, M. V. Fedulov, and V. V. Chernukha, Preprint No. 4572/8 (Kurchatov Institute of Atomic Energy, Moscow, 1988).

11. B. N. Kozlov, At. Énerg. **12**, 238 (1962).

12. *Handbook of Physical Quantities*, Ed. by I. S. Grigoriev and E. Z. Meilikhov (Énergoatomizdat, Moscow, 1991; CRC Press, Boca Raton, 1997).

13. J. J. Duderstadt and G. A. Moses, *Inertial Confinement Fusion* (Wiley, New York, 1982; Énergoatomizdat, Moscow, 1984).

14. N. N. Kalitkin and L. V. Kuz'mina, Preprint (Institute of Applied Mathematics, USSR Acad. Sci., Moscow, 1975).

Translated by N. Larionova

=========== **SOLIDS** ===========

Crack Propagation upon Dynamic Failure
of Polymethylmethacrylate

S. A. Atroshenko, S. I. Krivosheev, and A. Yu. Petrov

Research Institute of Mathematics and Mechanics, St. Petersburg State University,
Universitetskaya nab. 7/9, St. Petersburg, 198904 Russia
e-mail: satroshe@atr.ipme.ru
Received January 29, 2001; in final form, July 9, 2001

Abstract—Crack propagation in polymethylmethacrylate shock-stressed by a pulsed magnetic field is experimentally investigated. The failure type as a function of the distance from the crack tip is analyzed, and the failure diagram is constructed. The surface failure energy is estimated. The tilt angles of mesocracks are correlated to the failure type. © *2002 MAIK "Nauka/Interperiodica"*.

INTRODUCTION

When subjected to high tensile stresses, glassy polymers may become brittle and fail. Brittle failure is characterized by uniform strains up to the instant of failure, which proceeds by fast crack propagation across the specimen, i.e., in the plane normal to the tensile force applied. In this case, the cross-sectional area in the failure plane does not change significantly, and the residual longitudinal strains in both pieces of the specimen are absent. The failure type depends on experimental conditions. The brittle failure of viscoelastic materials takes place at relatively low temperatures and small exposure times or at high strain rates.

EXPERIMENTAL MATERIALS
AND METHODS

An advantage of glassy polymers, such as polymethylmethacrylate (PMMA), is transparency, which makes it possible to observe crack propagation and control the process.

PMMA was used to determine the failure type upon dynamically stressing the specimen by a pulsed magnetic field. As in [1], the stressing conditions provided the uniform distribution of the pressure along the crack edge. The tests were performed at threshold stress pulses.

We tested organic glass specimens with $c_1 = 1970$ m/s, $c_2 = 1130$ m/s, and $K_{Ic} = 1.47$ MPa m$^{1/2}$, where c_1 and c_2 are the longitudinal and transverse velocities of acoustic waves, respectively, and K_{Ic} is the limit stress intensity factor under static stressing.

The stressing conditions and the specimen geometry are shown in Fig. 1. The specimens were prepared as follows. A 3-mm-wide groove was made in a PMMA slab 10 mm wide and 100 mm long. At the end of the groove, a 3-mm-deep sawcut of width about 0.3 mm

was made. We studied the propagation of cracks from the end of the sawcut when the specimen was subjected to a pulse of pressure uniformly distributed over the groove. The pressure pulse was generated by a 10-mm-wide plane conducting line made of 250-μm-thick copper foil. A current pulser was discharged in the oscillatory mode with an oscillation period $T = 5.5$–6.0 μs and the decay time constant $T_1 = 1$–4 μs. The current pulse amplitude was varied from 150 to 300 kA, which corresponded to a stress amplitude in the specimen from 140 to 320 MPa. The current was measured using a Rogowski loop having a passive RC integrator with an integration time constant of 68 μs and recorded with an S9-8 oscilloscope.

If the capacitor bank is discharged in the oscillatory mode to the plane copper lines whose minimal cross section fits the maximum allowable action integral for copper and if the ratio of the line width to the spacing $b/h \geq 10$, the pressure distribution can be considered uniform [2]. Then, under the assumption that the decay
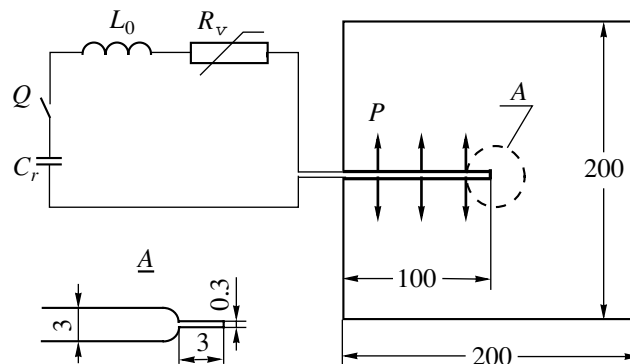
Fig. 1. Stressing setup and test specimen. *P*, load; C_r and L_0, capacitance and self-inductance of the current pulser, respectively; *Q*, switch. Sizes are given in millimeters.

is not too high, the pressure can be determined by the formula

$$P(t) = P_0 \exp\left(-\frac{2t}{T_1}\right) \sin^2\left(2\pi\frac{t}{T}\right). \qquad (1)$$

At the time instant $t_m = \arctan(2\pi T_1/T)T/2/\pi$, the pressure reaches the maximal value

$$P_m = P_0 \exp(-2t_m/T_1)/(1 + T^2/T_1^2/4/\pi^2).$$

Figure 2 shows the optical scheme for the visualization of crack propagation under pulsed stressing. An SFR-2 streak camera was used as a photographic recorder. An IFP-120 flash lamp was employed as a source of light. A part of specimen *7* is cut by adjustable slit *5* and is recorded by film *1*. The image is transferred to the film path with rotating mirror *2*. The film shows how the illuminance varies when the moving crack shuts off the optical path. The boundary between the regions of different illuminance (different optical density of the image) is the interface between the cracked and intact materials. To synchronize the crack propagation with the instant of stressing, the radiation from the discharge gap is transferred to the field of the camera with a pipeline.

Ten specimens were tested. The oscillation period was $T = 5.6$ μs, the decay time constant was $T_1 = 4.2$ μs, and the pressure pulse amplitude P_m was varied from 140 to 320 MPa. Typical results of processing the streak photographs (crack propagation dynamics) are given in Fig. 3.

RESULTS AND DISCUSSION

1. Failure under Pulsed Stressing

Note the stepwise character of the crack propagation. The crack travels a distance of 9–12 mm until it stops for the first time. Initially, the crack moves with a velocity of 100 to 500 m/s according to the stressing conditions. If the stress exceeds the threshold value, the initial portion of the distance to the first stop (several millimeters) is traveled with a velocity of 420–450 m/s. The rest of this distance is traveled with a much lower velocity, about 100 m/s.

The time instant the crack stops correlates with the time it takes for the stress wave to doubly travel the distance from the crack edge to the boundary of the specimen. Tests with the smaller-size specimens confirm this observation. The stop can be explained by the arrival of the waves reflected from the specimen boundary at the tip crack.

Processing the streak photographs, we constructed the dependence of the crack origination instant on the applied stress amplitude. To compare the dependence thus obtained with model data, we considered a semi-infinite crack with the normally stressed edges as a mathematical model. The solution of this problem yields the actual experimentally observed stress pattern
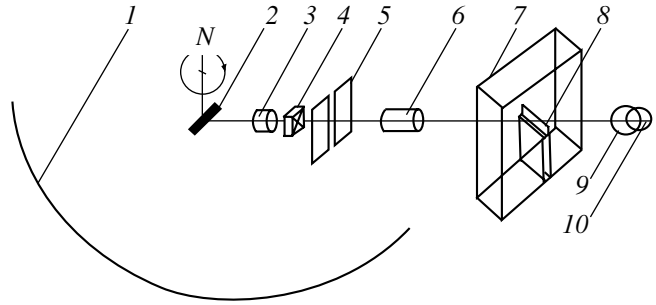
Fig. 2. Optical detection scheme. *1*, photographic film; *2*, mirror; *3* and *6*, objective lenses; *4*, shutter; *5*, adjustable slit; *7*, specimen; *8*, running crack; *9*, IFP-120 flash lamp; and *10*, pipeline.
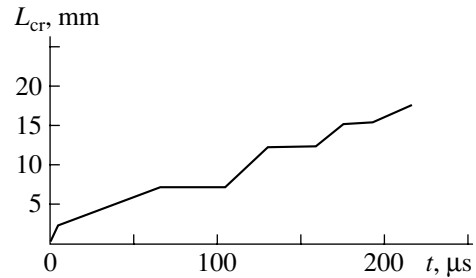

Fig. 3. Typical L_{cr} vs. time dependence obtained by processing streak photographs.

prior to the arrival of the waves reflected from the model boundaries. For the specimens used in our experiments, the time to arrival was about 100 μs. In our problem, the stress intensity factor has the form [3]

$$K_I(t) = \int_0^t \frac{\alpha P(s)}{\sqrt{t-s}} ds, \qquad (2)$$

where $\alpha = 2c_2\sqrt{c_1^2 - c_2^2}/(c_1\sqrt{\pi c_1})$.

Since in the experiments we varied only the stress amplitude $P(t)$, the value of the intensity factor remained the same up to multiplier and reached the maximum within $t \approx 1.7$ μs after the start of stressing.

The structural-temporal criterion [4] allows us to calculate the dependence of the failure initiation instant on the applied stress amplitude. In our case, this criterion has the form [5]
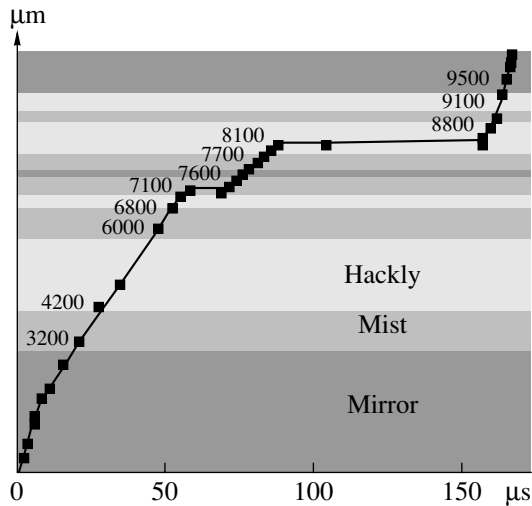
$$\frac{1}{\tau} \int_{t-r}^t K_I(t')dt' \le K_{Ic}, \qquad (3)$$

where τ is the incubation time (intrinsic material property).

For PMMA, $\tau = 32$ μs, as follows from experiments [1]. The calculation shows that the least failing amplitude in our experiments is $P_m = 94.7$ MPa. Figure 4 depicts the calculated dependence of the applied stress

P_m, MPa
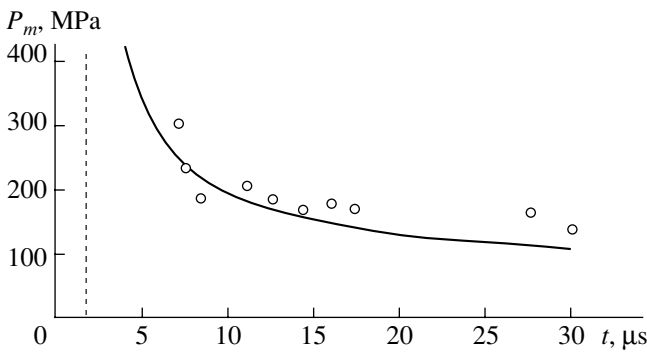
Fig. 4. Amplitude of the stress pulse applied vs. time to failure.

μm

Fig. 5. Sequence of fracture types vs. stressing time and distance to notch. Figures indicate the distance traveled by the crack.

amplitude on the failure initiation instant. Circles denote data points. The vertical dotted line ($t \approx 1.7$ μs) indicates the time when the stress intensity factor reaches the maximum. The time in Fig. 4 is counted from the stress application time. The test data suggest that the crack starts to move within some time after the local force field at the tip, the stress intensity factor, has reached the maximum, that is, in the descending branch of the curve. Such a phenomenon is called failure delay. Failure delay has been observed in experiments on

Table 1. Energy of cup fracture

Crack length L_{cr}, mm	Cup size d, μm	Crack velocity V, m/s	Fracture energy E, J/cup
5.0	32	92.4	17×10^{-18}
17.5	18	50.696	5.3×10^{-18}
25.5	13	194.25	2.8×10^{-18}

scabbing [6] and spalling [7]. The failure delay is the time between the instant the local force field reaches a maximum and the instant of failure initiation. This phenomenon cannot be explained in terms of the concept of the critical stress intensity factor but finds an explanation within the structural-temporal approach. In Fig. 4, the failure delay is the distance between the curve or a data point to the dotted line.

2. Surface Energy to Failure

The surface energy to failure is a measure of the fracture strength of a material. It is a more general index of the material strength properties than the strength itself, because it does not depend on local characteristics of the material or on the type of mechanical action. The discrepancy between the estimated and experimental values of the surface energy to failure and also the appearance of the fracture surface lead us to conclude that, when subjected to local stresses at the crack tip, the material exhibits the plastic behavior although its macroscopic properties are typical of a brittle substance. It is of interest to compare the behavior of metals and polymers. At room temperature, the differences between the estimated and experimental values of the surface energy to failure for PMMA and steel are close to each other [8], since both materials show similar local responds. In both cases, the energy dissipation mechanism involves a combination of shear stresses and diffusion; however, the type of diffusion and the parameters of the energy-dissipating field in these materials greatly differ. In steel, the energy is dissipated because of the dislocation motion in the metal lattice; in the polymer, the dissipation is due to the motion of dislocation segments in the Van der Waals force field generated by surrounding molecules. Inelastic processes taking place at the crack tip are a result of a high local stress concentration, to judge from the value of the specific surface energy. In this respect, the behavior of glassy polymers is akin to that of metals.
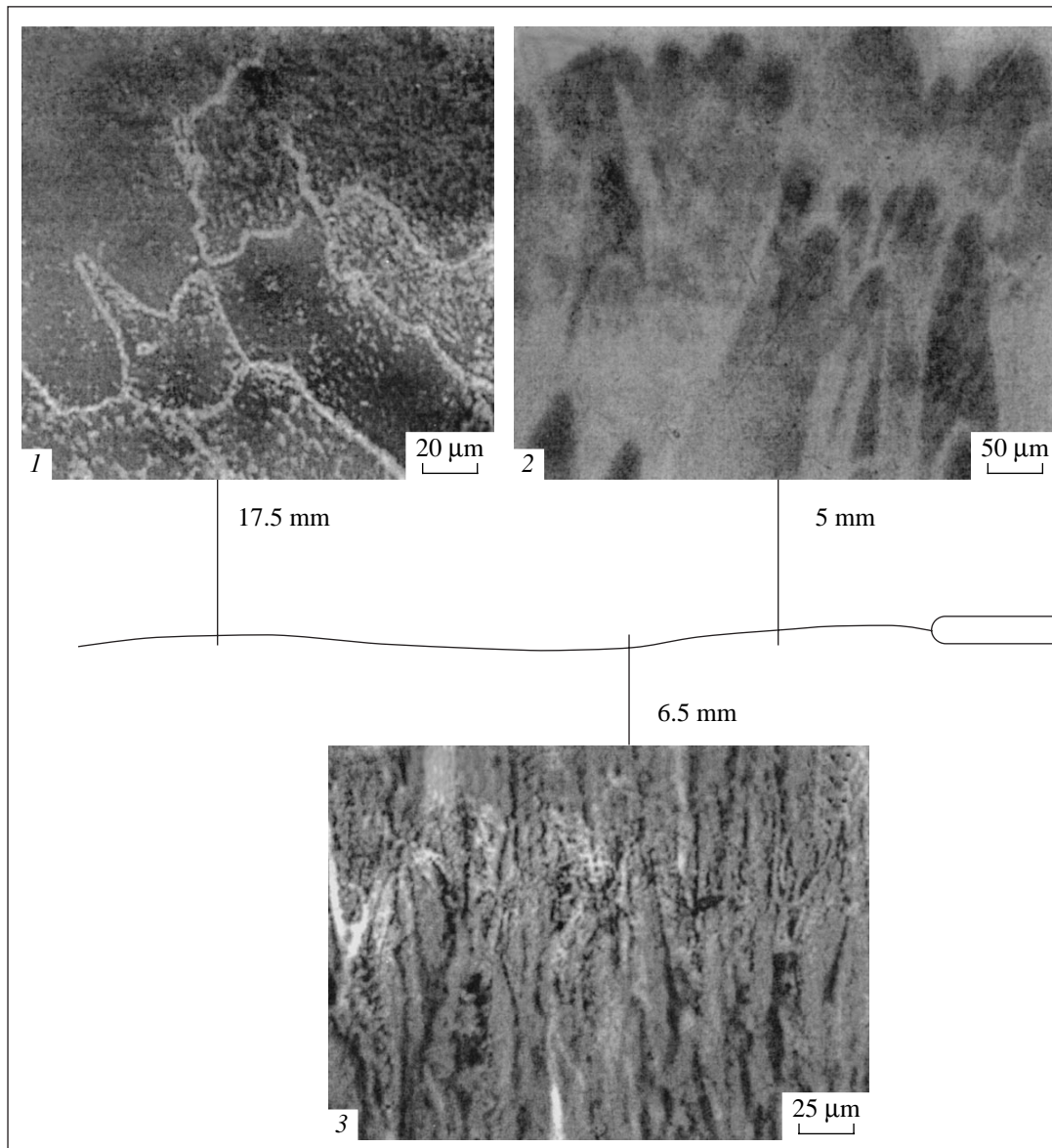
Processes responsible for the failure of solids can be judged from the fracture surface. Analyzing it, one can clarify a failure mechanism and theoretically investigate the strength properties of a material.

If, under tension, the crack propagates from the fracturing source slowly, the initial region of propagation is smooth and offers a high reflectivity. This region is called a smooth (mirror) zone or a zone with "silver" cracks—thin layers of a highly strained partially exfoliated material that can specularly reflect light [9]. Next to the mirror zone, or the zone of slow crack propagation, is a transition zone, which is characterized by the increased fracture surface roughness. In notched specimens, the transition zone is larger than in those without a notch and features irregular lines extended along the crack propagation direction. The following is the zone containing a number of geometric figures, largely parabolas and hyperbolas [10]. They arise when a secondary crack originates from a defect in front of the

Fig. 6. Typical fracture surfaces vs. distance traveled by the crack (*1*, mirror; *2*, mist; and *3*, hackly).

propagating primary crack. The secondary crack propagates with a constant speed as a set of concentric circles "radiating" from many local defects, which are activated by the propagating stress wave. When crossing the primary crack, the secondary crack forms parabolas [11] if the rate of propagation of the primary and secondary cracks are the same. The defect from which the secondary crack originates is the focus of the parabola. A slight difference between the levels of the propagating stress waves renders these figures visible and leads to the substantial absorption of the energy. The appearance of the fracture surface reflects basic deformation processes on which the fracture energy is spent. Outside this zone, the stress level and the crack propagation velocity increase, and high-rate (explosive) fail-

ure takes place. In this case, starlike structures arise on the fracture surface [12], and the cracks propagate more or less uniformly. Under high dynamic overstressing, many of the defects act simultaneously and the material undergoes brittle fracture irrespective of its behavior under static stressing. It has been found [13] that the rate of PMMA destruction (crack propagation velocity) grows with distance to the fracturing source and lies between 400 and 700 m/s, depending on the static stress at the instant of failure (the crack velocity increases with the stress value). Note also that the destruction rate and the appearance of the fracture surface do not strictly correlate: the latter depends also on the stressing rate and history.
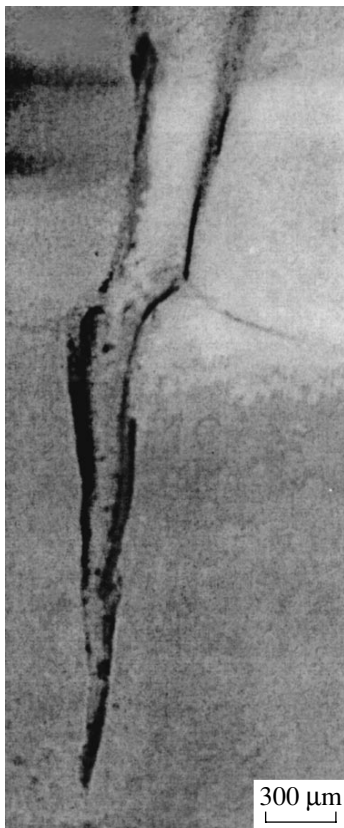
Fig. 7. Wedge crack in PMMA.
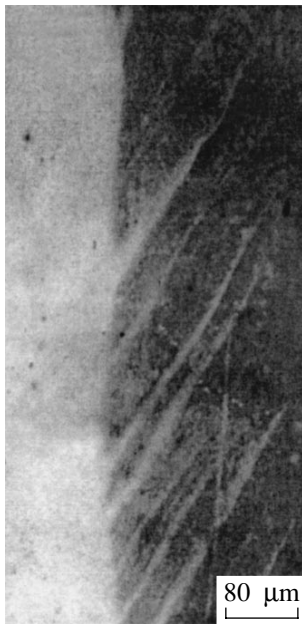
Fig. 8. Bifurcation of mesocracks in PMMA.

During the destruction, the major portion of the energy is spent on plastic deformation around fracturing sources. These are usually the foci of the parabolas where substantial plastic strains are observed. The rougher is the fracture surface, the higher energy is spent on crack propagation. The energy absorption mechanisms can be judged by the surface appearance. When a defect is initiated, the major part of the energy is absorbed because of plastic deformation. As has been mentioned above, the fracture surface is smooth in the mirror, or cleavage, zone, where the crack velocity is slow. In this zone, the energy needed for sustaining the crack propagation is small.

The next region, the region of increased surface roughness (mist region), exhibits randomly arranged irregular triangles with a great spread in their sizes. Outside this region, the PMMA fracture surface becomes similar to the surface of metals at plastic (hackly or cup) fracture. A great number of parabolic cups requires a great energy for the crack propagation. A growth of the energy absorbed by the fracture surface increases the cup size, so that the degree of plastic fracture declines.

The parabolic features in PMMA resemble the regions of cup fracture in high-viscosity steel. In the steel, material shear is seen near the cup crater. Upon cracking, parabolic surface features (cups), which look like those on the PMMA surface, appear. Parabolas reflect the conventional type of PMMA failure [13]. The mechanism of parabola formation sheds light on the whole process of failure: crack initiation; intermittent evolution of cracking; further initiation of cracks at defect sites; and the occurrence of secondary cracks and their relation to the position of the primary crack front (the secondary and primary cracks usually propagate at different levels).

3. Failure Type

We analyzed the character of PMMA failure under dynamic stressing and constructed the diagram (Fig. 5) that shows how the failure type (vertical axis) varies with the crack length (counted from the notch top) and pulse duration (horizontal axis). As the crack propagates from the notch, the type of failure changes in the order: cleavage, quasi-cleavage, and cup fracture (in terms of metal science), which correspond to the smooth region, mist region, and parabolic (cup) region, respectively. As the crack propagates further, these basic types of failure alternate. The images of the fracture surface and the change-over from one fracture type to another are shown in Fig. 6. As the distance to the notch top increases, the cup size decreases from 32 to 13 μm, according to Fig. 6. Knowing the cup diameter and the PMMA specific surface fracture energy ($\gamma = 2.1 \times 10^2$ J/mm^2 [8]), one can calculate the fracture energy per middle cup. This energy drops from 17×10^{-8} to 2.8×10^{-8} J as the crack propagates inward to the specimen (Table 1).

It should be noted that a small crack propagating into the specimen normally to the fracture surface was detected at the notch end on the fracture surface. More-

Table 2. Tilts of mesocracks

Crack length L_{cr}, mm	Mesocrack tilt, deg	Crack velocity V, m/s
1	32	271.6
3	33	119.8
6.5	29	70.58
7.1	19	58.95
7.6	28	62.27
27	26	126.79

over, a wedge crack originated at the cups was detected at a distance of about 10 mm from the notch (Fig. 7). At a distance of about 4.2 mm, this crack bifurcates (Fig. 7): two branches normal to the main crack propagate in the direction coincident with that of the main crack front. This crack causes cup (hackly) fracture.

At the same time, mesocracks bifurcate from the main crack (Fig. 8); however, they differ from those described in [5–10]. Their tilt varies nonmonotonically with distance to the notch end (Table 2). In addition, the mesocracks are initiated in the smooth and rough regions. They are absent on the surfaces of cup (parabolic) fracture.

As for the crack propagation velocity, it also varies nonmonotonically with distance to the notch, unlike data reported in [8].

CONCLUSIONS

The method to visualize the brittle fracture of cracked PMMA subjected to dynamic stressing (microsecond pulses of threshold amplitude) is developed.

The following features of the fracture were revealed.

(i) Crack propagation delays relative to the time instant the stress applied reaches the threshold value.

(ii) Crack length is proportional to the amplitude of above-threshold stress pulses.

(iii) Fracture types alternate until the crack stops first.

(iv) Mesocracks are initiated only in the smooth and rough (mist) regions. Their tilt to the direction of the main crack propagation was found to be independent of the crack velocity and distance to the notch.

The fracture type-pulse duration-notch distance diagram was constructed.

The surface fracture energy per cup was shown to decrease from 17×10^{-8} to 2.8×10^{-8} J/cup, this energy being related to the characteristic crack size.

REFERENCES

1. S. I. Krivosheev, Preprint No. 142, IPMash RAN (Institute of Problems in Machine Science, Russian Academy of Sciences, St. Petersburg, 1997).

2. G. A. Shneerson, *Fields and Transient Processes in Ultrahigh Current Equipment* (Énergoatomizdat, Moscow, 1991).

3. G. P. Cherepanov, *Mechanics of Brittle Fracture* (Nauka, Moscow, 1974; McGraw-Hill, New York, 1979).

4. N. F. Morozov, Yu. V. Petrov, and A. A. Utkin, Izv. Akad. Nauk SSSR, Mekh. Tverd. Tela, No. 5, 180 (1988).

5. Yu. V. Petrov and A. A. Utkin, in *Mechanics of Fracture: Theory and Experiment* (St. Petersb. Gos. Univ., St. Petersburg, 1995), p. 94.

6. N. A. Zlatin, G. S. Pugachev, S. M. Mochalov, and A. M. Bragov, Fiz. Tverd. Tela (Leningrad) **17**, 2599 (1975) [Sov. Phys. Solid State **17**, 1730 (1975)].

7. D. A. Shockey, D. C. Erlich, J. F. Kalthoff, and H. Homma, Eng. Fract. Mech. **23**, 311 (1986).

8. J. P. Berry, J. Polym. Sci., No. 50, 107 (1961).

9. V. I. Aleshin and E. V. Kuvshinskiĭ, Mekh. Polim., No. 6, 989 (1978).

10. J. Fineberg and M. Marder, Phys. Rep. **313**, 1 (1999).

11. C. E. Felner, *Theoretical and Applied Mechanics Report*, No. 224 (University Illiois, 1962).

12. J. A. Kies, A. M. Sallivan, and G. R. Irwin, J. Appl. Phys. **21**, 716 (1950).

13. *Fracture Processes in Polymeric Solids* (International Publishers, New York, 1970).

Translated by V. Isaakyan

$=\!=\!=\!=\!=\!=\!=\!=\!=$ **SOLIDS** $=\!=\!=\!=\!=\!=\!=\!=\!=$

Charged Defect Diagnostics from Luminescence Spectra

L. P. Ginzburg and A. P. Zhilinskiĭ

Moscow Technical University of Communication and Informatics, Moscow, 111024 Russia
Received November 25, 1999; in final form, April 20, 2001

Abstract—A method allowing the detection of an absolutely rigid Coulomb gap formed by localized charged defects from the low-frequency part of the luminescence spectrum is proposed. Its efficiency is demonstrated with luminescence in amorphous semiconductors and optical fibers, as well as with triboluminescence. © *2002 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The subsystem of randomly distributed charged defects (CDs) defines many of the physical properties of a material. One reason is that this subsystem gives rise to diagonal disorder, which is responsible for such effects as the Anderson transition in the inverse layer of MIS structures [1], the position of the mobility gap in v-SiO$_2$ [2, 3], the concentration dependence of the photoconductivity threshold in semiconductors [4], the scattering of carriers [5], etc.

However, the presence of CDs is difficult to predict. The point is that, under the equilibrium condition, such defects arise only if the occupied energy level of a negative center is below the empty level of a positive one. But even if there is an indirect evidence for the presence of CDs (for example, so-called U^- centers in chalcogenide glassy semiconductors (CGSs)), mechanisms behind their formation remain unclear and are discussed within different models (see, e. g., [6] and Ref. therein). Therefore, the direct experimental detection of CDs seems to be of great importance.

It has been demonstrated with CGSs [7] that CDs must affect the low-frequency region of the photoluminescence (PL) spectrum. This effect can be understood from the fact that the spectral density of states in the CD subsystem is characterized by the Coulomb gap (CG)

$$g(E) \propto (E - \delta)^2, \qquad (1)$$

where

$$\delta \simeq \hbar\omega_0/2 \qquad (2)$$

and $\hbar\omega_0$ coincides with the PL threshold. Since $\omega_0 > 0$, this CG can be called "absolutely rigid." This means that the 2δ interval about the Fermi level remains "forbidden" [8].[1] The expression for the electrical conductivity near the gap edge acquires the factor $\exp(-2\delta/k_0 T)$. That is why this gap cannot be detected

by the methods usually applied to reveal a soft gap, in particular, by low-temperature experiments on hopping conduction or tunneling spectroscopy.

On the other hand, according to [7], Eq. (1) involves two features, which can be detected by accurately measuring the low-frequency tail of the PL spectrum. The first one is the presence of the interval

$$\hbar\omega_0 < \hbar\omega \le \hbar\omega_m, \qquad (3)$$

within which the value

$$\xi(\hbar\omega) = \frac{\Delta I(\hbar\omega)}{(\hbar\omega)^2}, \qquad (4)$$

($\Delta I(\hbar\omega)$ is the luminescence intensity) obeys the rule

$$\xi(\hbar\omega) = A - B\hbar\omega + C(\hbar\omega)^2, \qquad (5)$$

where A, B, and C are positive constants.

The second feature is that the coefficients A, B, and C relate as

$$AC \simeq 0.25 B^2. \qquad (6)$$

As was mentioned above, CGSs provide an example of materials where the presence of U^- centers has been established [6, 9]. The analysis of all the PL spectra ever observed in CGSs has shown [7] that, in most cases, relationships (5) and (6) hold. The approach developed in [7] was extended to v-SiO$_2$ [3], in which a significant amount of O$_3^+$ and O$^-$ centers is surmised by several authors (see, e. g., [10]). The result was in favor of this supposition.

The aforesaid leads one to the question of to what degree the technique of PL spectrum analysis developed in [7] can be applied for CD diagnostics. In other words, whether one can conclusively establish the presence of CDs (or even estimate their concentration [3, 11]) based only on a conventional "bell-like" curve $\Delta I(\hbar\omega)$ or more precisely, its carefully measured low-

---

[1] From this point on, the abbreviation CG implies an absolutely rigid gap.

frequency part where conditions (5) and (6) are fulfilled in interval (3).

Our purpose is to show that, with the exception of some specific situations, this sufficiency is justified. In Section 1, we demonstrate that, for the density of states given by the general series

$$g(E) = \sum_{n=1}^{\infty} a_n (E - \delta)^n, \qquad (7)$$

meeting conditions (5) and (6) within interval (3) leaves (apart from some specific cases) only the term with $n = 2$ in (7). Since no other model leads to a similar dependence in the system of localized states (see, e. g., [12, 13]), this proves the presence of CDs. Section 2 considers examples illustrating how the application of conditions (5) and (6) to a short low-frequency portion of seemingly identical PL spectra enables one to prove or disprove the presence of the CD subsystem.

1. SUFFICIENCY OF CONDITIONS (5) AND (6) FOR CD DIAGNOSTICS

According to [7], the general formula for $\Delta I(\hbar\omega)$ is

$$\Delta I(\hbar\omega) \propto (\hbar\omega)^2 \int_{\delta}^{\infty} dE F(E) g(E) g(E - \hbar\omega), \qquad (8)$$

where $F(E)$ is the amount of localized radiative states with an energy E that may be occupied by an electron that absorbed a photon.

Let interval (3), where the value $\xi(\hbar\omega)$ defined by (4) follows law (5), be detected. The assumption that only the $\delta - E_m$ interval contributes to (3) enables us to replace the upper limit of integration in (8) by E_m. Then, before being substituted into (8), series (7) should be reduced to the sum

$$g(E) = \alpha_1 (E - \delta) + \alpha_2 (E - \delta)^2. \qquad (9)$$

Only under condition (9) does not the expansion of $g(E - \hbar\omega)$ in powers of $\hbar\omega$ contain terms with powers greater than two. Because of the proportional relationship between the quantities in (8), we can put $\alpha_2 = 1$ without loss in generality. Then, the density of states becomes

$$g(E) = \beta(E - \delta) + (E - \delta)^2. \qquad (10)$$

Substituting (10) into the expression

$$\xi(\hbar\omega) = \int_{\delta}^{E_m} dE F(E) g(E) g(E - \hbar\omega) \qquad (11)$$

using the equality

$$(E - \delta - \hbar\omega)^2 = (E - \delta)^2 - 2(E - \delta)\mu\omega + (\hbar\omega)^2, (12)$$

and taking into account (5), we obtain

$$A \propto \int_0^{\Omega_m} dU F(U + \delta)(\beta^2 U^2 + 2\beta U^3 + U^4), \qquad (13)$$

$$B \propto \int_0^{\Omega_m} dU F(U + \delta)(\beta^2 U + 3\beta U^2 + 2U^3), \qquad (14)$$

$$C \propto \int_0^{\Omega_m} dU F(U + \delta)(\beta U + U^2), \qquad (15)$$

where

$$\Omega_m = E_m - \delta. \qquad (16)$$

Following [7], we take advantage of the expansion

$$F(U + \delta) = \sum_{n=1}^{\infty} a_n(\delta) U^n, \qquad (17)$$

where the fact that $F(\delta) = 0$ is taken into consideration. Substituting (17) into (13)–(15) and integrating yields

$$A \propto A(\beta) + a, \quad B \propto B(\beta) + b, \quad C \propto C(\beta) + c, \quad (18)$$

where

$$A(\beta) = \beta^2 \gamma(2) \Omega_m^3 + 2\beta\gamma(3)\Omega_m^4, \quad a = \gamma(4)\Omega_m^5, (19)$$

$$B(\beta) = \beta^2 \gamma(1)\Omega_m^2 + 3\beta\gamma(2)\Omega_m^3, \quad b = 2\gamma(3)\Omega_m^4, (20)$$

$$C(\beta) = \beta\gamma(1)\Omega_m^2, \quad c = \gamma(2)\Omega_m^3 \qquad (21)$$

and the designation

$$\gamma(p) = \sum_{n=1}^{\infty} a_n \frac{\Omega_m^n}{n + p + 1} \qquad (22)$$

is used.

Combining Eqs. (18), we find

$$\frac{AC}{B^2} = \frac{A(\beta)C(\beta) + aC(\beta) + cA(\beta) + ac}{B^2(\beta) + 2B(\beta)b + b^2} \frac{b^2}{ac}. \qquad (23)$$

As was shown in [7], at $p \geq 4$, the expression

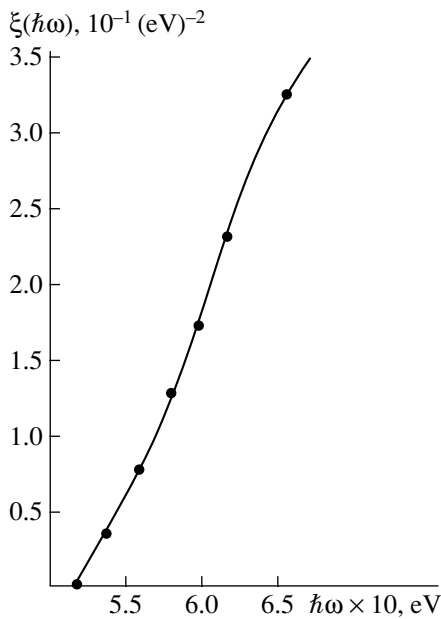$$\frac{\gamma(p)\gamma(p-2)}{\gamma(p-1)} \simeq 1 \qquad (24)$$
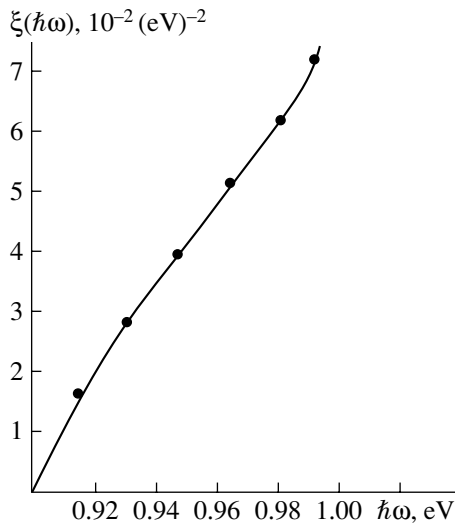
is fairly accurate.

Therefore, in view of (19)–(21),

$$ac \simeq 0.25 b^2. \qquad (25)$$

Let us assume that both conditions (5) and (6) are met. In this case, Eqs. (23) and (25) suggest that

$$4[A(\beta)C(\beta) + aC(\beta) + cA(\beta)] \simeq B^2(\beta) + 2B(\beta)b. (26)$$

Fig. 1. $\xi(\hbar\omega)$ function. Reference points are obtained from data reported in [14, Fig. 3].

Fig. 2. $\xi(\hbar\omega)$ function. Reference points are obtained from data reported in [15, Fig. 2D].

The substitution of (19)–(21) into (26) and appropriate rearrangements give

$$\beta(\beta^3 + r\beta^2 + s\beta + t) \simeq 0, \qquad (27)$$

where

$$r = 2\frac{\gamma(2)}{\gamma(1)}\Omega_m, \quad s = \frac{\gamma(2)}{\gamma(1)}\left(5\frac{\gamma(2)}{\gamma(1)} - 4\frac{\gamma(3)}{\gamma(2)}\right)\Omega_m^2,$$

$$t = 4\frac{\gamma(3)}{\gamma(1)}\left(\frac{\gamma(2)}{\gamma(1)} - \frac{\gamma(4)}{\gamma(3)}\right)\Omega_m^3. \qquad (28)$$

One of the roots of (27) is evident. It corresponds to the condition

$$\beta \simeq 0. \qquad (29)$$

From Eq. (28), it follows that other real positive root (if any) must explicitly depend on Ω_m defined by (16). According to (10), β is the rate of change of the density of states at the point $E = \delta$. Therefore, the use of any root other than (29) entails the situation when β depends not only on the behavior of $g(E)$ in the vicinity of $E = \delta$ but also on the energy E_m, which is separated from δ by a finite interval. This seems to be a rather specific case. If it is out of consideration, only root (29) remains. As was mentioned above, it indicates the presence of the CG.

2. EXAMPLES

Amorphous Semiconductors

We will compare two PL spectra taken from amorphous semiconductors [14, Fig. 3; 15, Fig. 2D]. These spectra were chosen because of their apparent similarity. Both start from the zero value of $\Delta I(\hbar\omega)$, peak at close energies (1.05 and 1.33 eV), and have comparable half-widths (0.46 and 0.29 eV). However, as is seen from Figs. 1 and 2, the low-frequency behavior of the $\xi(\hbar\omega)$ functions defined by (4) is entirely different. The curve in Fig. 1 [14] is well fitted by (15). This can be checked by plotting the curve

$$\eta(\hbar\omega) = \frac{d\xi(\hbar\omega)}{d(\hbar\omega)}, \qquad (30)$$

which, under condition (5), must have the linear part

$$\eta(\hbar\omega) = 2C\hbar\omega - B. \qquad (31)$$

For testing, we used the least-squares method. Constructing tangents to the curve in Fig. 1, we obtained a set of closely spaced (0.005 eV apart) empirical points approximating the $\eta(\hbar\omega)$ function. The root-mean-square deviation for the first ten of them turned out to equal 6.7%. This value indicates that these points can actually be approximated with a straight line. The coefficients of the line were determined by solving the corresponding normal equations: $C = 19.11$ (eV)$^{-4}$ and $B = 8.38$ (eV)$^{-3}$. The straight line plotted with these coefficients (Fig. 3) is in good agreement with dependences (30) and (31).

Substituting C and B derived above into Eq. (5) and using the curve in Fig. 1, one finds that $A \simeq 1.78$ (eV)$^{-2}$. Hence, it follows that $AC \simeq 0.24B^2$, which is consistent with condition (6).

The behavior of the curve in Fig. 2 indicates that, conditions (5) and (6) fail in the case [15]. In [14], a-As$_2$S$_3$, which belongs to the CGS class, was studied. As was mentioned in the Introduction, semiconductors of these type contain charged U^- defects. The PL spec-
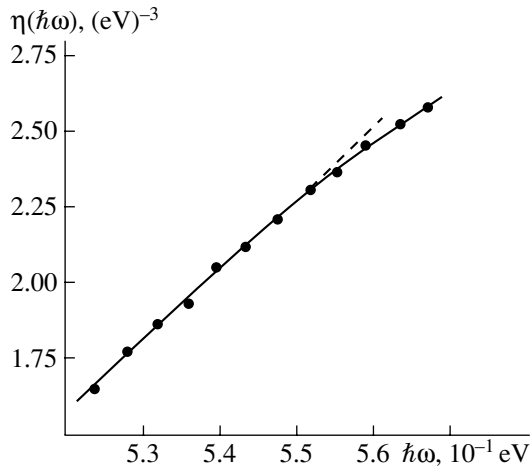
$\eta(\hbar\omega)$, (eV)$^{-3}$

Fig. 3. $\eta(\hbar\omega)$ function constructed with the tangents to the curve in Fig. 1.

tra from a-Si, where the most typical defects are neutral [16] dangling bonds, were considered in [15].

Note that the fundamental difference in the character of the defects was revealed by studying the PL low-frequency tails, comprising no more than 10% of the spectrum half-width.

The same result was obtained for other six PL spectra from CGSs [7]. In [7], a-Si samples prepared under different conditions were also studied. Their PL spectrum was shifted towards lower frequencies. The corresponding $\xi(\hbar\omega)$ function is presented in [7, Fig. 6] and looks like that shown in Fig. 2.

Red Glow in Optical Fibers

Red glow (RG) observed in 3D samples of g-SiO$_2$ near 1.9 eV is usually attributed to nonbridging oxygen atoms, which form the $\equiv$Si–O$\uparrow$ structure (the lines depict bonds; the arrow, an unpaired electron (spin)). However, the appearance of such defects during drawing low-OH optical fibers has not been completely understood. In [11], a process accounting for all attendant factors (an excess of oxygen and the lack of E' centers) was proposed in the form

$$2(\equiv Si–O–O–Si\equiv)$$
$$\longrightarrow (\equiv Si–O\uparrow O^+–Si\equiv) + (\equiv Si–O^-\uparrow O–Si\equiv). \tag{32}$$
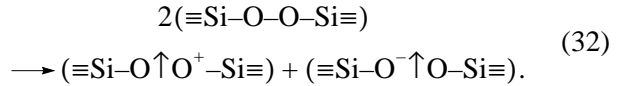
A specific feature of Eq. (32) is that nonbrinding atoms are incorporated into oppositely charged complexes; thus, it is impossible to exclude *a priori* the competition between the transitions within the CG. Analysis of the data reported in [17, 18] confirmed the fulfillment of conditions (5) and (6) [11]. Special attention should be paid to the spectra presented in [18, Fig. 4]. These spectra were obtained by two different ways. In the first case, measurements were performed at the end of a 300-m-long optical fiber, so that transmission losses had to be taken into account (*e* measurement). The other spectrum was detected in the immediate proximity to the side face of the fiber (*s* measurement). Being generally similar, the associated curves exhibit a small yet visible difference in the low-frequency region [18, Fig. 4]. The $\xi(\hbar\omega)$ and $\eta(\hbar\omega)$ functions plotted with the use of [18, Fig. 4] and data in [11] are shown in Figs. 4 and 5, respectively. The difference in the low-frequency regions of the RG curves is in the different arrangements of the linear parts of the $\eta(\hbar\omega)$ functions
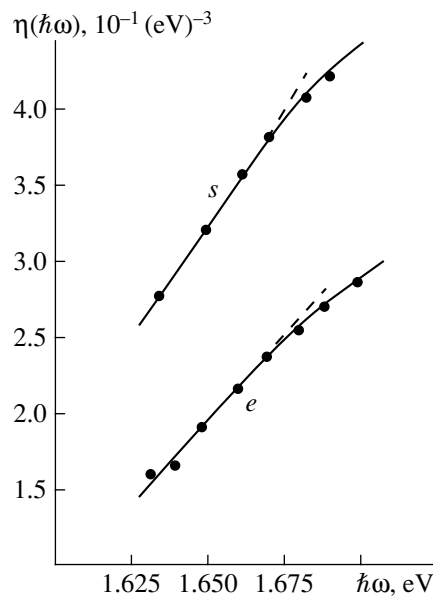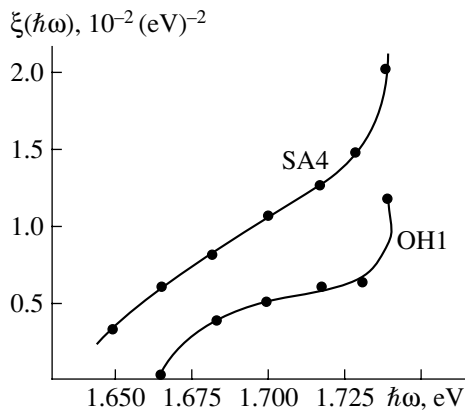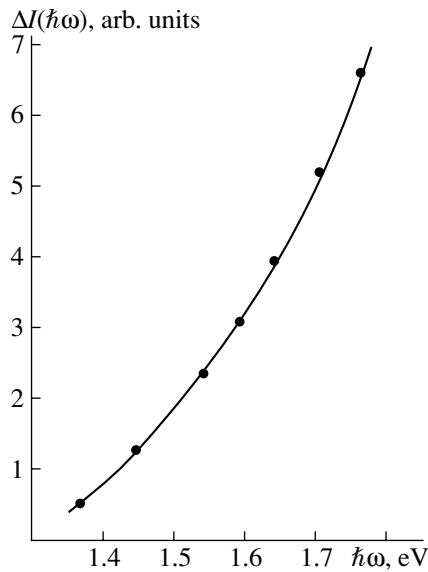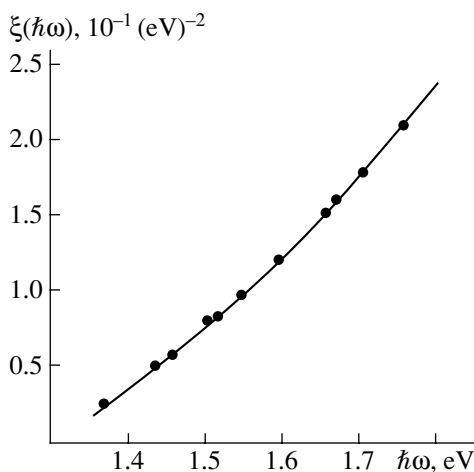
Fig. 4. $\xi(\hbar\omega)$ functions. Reference points were obtained from the data presented in [18, Fig. 4]. The curves are taken from [11].

Fig. 5. $\eta(\hbar\omega)$ functions constructed with the tangents to the curve in Fig. 4. The curves are taken from [11].
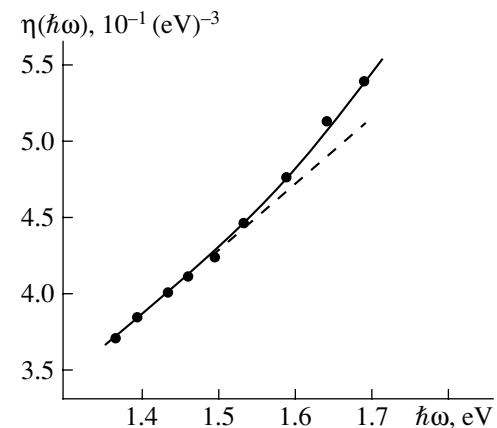
$\xi(\hbar\omega)$, 10^{-2} $(eV)^{-2}$

Fig. 6. $\xi(\hbar\omega)$ functions. Reference points were obtained from the data presented in [19, Fig. 12]. SA4 and OH1 indicate low-OH samples.

$\Delta I(\hbar\omega)$, arb. units

Fig. 7. Averaged low-frequency dependence $\Delta I(\hbar\omega)$ according to [20].

$\xi(\hbar\omega)$, 10^{-1} $(eV)^{-2}$

Fig. 8. $\xi(\hbar\omega)$ function. Reference points were obtained from Fig. 7.

defined by (30) and (31). Nevertheless, as was mentioned above, Eqs. (5) and (6) hold in both cases. This example illustrates the high reliability of the technique for CD diagnostics, since it proves to be insensitive to small variations in the shape of the low-frequency region of the PL spectrum.

Figure 6 shows the $\xi(\hbar\omega)$ functions constructed from the RG spectra presented in [19, Fig. 12]. These curves markedly deviate from law (5). At first glance, this looks surprising, because dehydrated optical fibers used in [19] were similar to those studied in [17, 18]. However, the former were subjected to γ irradiation, which is known to generate Compton electrons; thereby, the following processes seem to be plausible:

$$O_1^- + \hbar\omega_1 \longrightarrow \uparrow O_1 + \hbar\omega_2 + e^-, \qquad (33)$$

$$\equiv Si-O^- + \hbar\omega_1 \longrightarrow \equiv Si-O\uparrow + \hbar\omega_2 + e^-. \qquad (34)$$
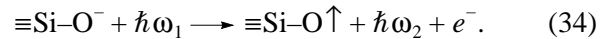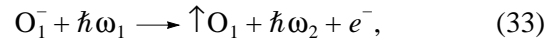
Both processes give rise to nonbrindging oxygen atoms, and process (34) destroys the CG formed by the charged complexes.

Triboluminescence

Triboluminescence (TL) is luminescence induced by mechanical action. This phenomenon still lacks adequate explanation. Our interest in TL in the context of this study is dictated by the following. In nonirradiated optical fibers, red glow was attributed [11] to transitions between charged defects. But unlike other types of glow, such as that from U^- centers in CGSs, it displays no detectable Stokes shift. Hence, one could suggest a correlation between process (32), which results from fiber drawing, and TL due to fiber breakage, which latter is centered near 1.97 eV [20]. Difficulties emerge when we analyze the TL spectrum presented in [20, Fig. 2b], since it is described by a polygonal line with the spread of $\Delta I(\hbar\omega)$ comparable to the value of $\Delta I(\hbar\omega)$ in the low-frequency region. Visually, however,

$\eta(\hbar\omega)$, 10^{-1} $(eV)^{-3}$

Fig. 9. $\eta(\hbar\omega)$ function constructed with the tangents to the curve in Fig. 8.

the spread is highly symmetric about some center line. Then, when applied to this line, or technique for CG detection may give a valid result. The low-frequency region of this line is shown in Fig. 7. The dependence $\xi(\hbar\omega)$ derived from the curve in Fig. 7 and the corresponding $\eta(\hbar\omega)$ function are plotted in Figs. 8 and 9, respectively. At first glance, they show evidence of the CG. Counting in favor of this are both the shape of the curve in Fig. 8 and the presence of the linear part in the curve in Fig. 9 (the variation of the first-order differences between the first five points obtained with the tangents to Fig. 8 amounted to 9.2%). However, the application of Eqs. (5) and (31) to Figs. 8 and 9 yields $B = 0.21$ (eV)$^{-3}$, $C = 0.21$ (eV)$^{-4}$, and $A \simeq -0.08$ (eV)$^{-2}$. These values are is consistent with Eq. (6); hence, the TL under study does not relate to the CG. Note that this result reflects the well-know situation when the peaks of TL and PL are coincident but the low-frequency regions are considerably different [21].

REFERENCES

1. B. I. Shklovskiĭ and A. L. Efros, *Electronic Properties of Doped Semiconductors* (Nauka, Moscow, 1979; Springer-Verlag, New York, 1984).

2. A. Hartstein, Z. A. Weinberg, and D. Y. DiMaria, in *Physics of SiO₂ and Its Interfaces*, Ed. by S. T. Pantelides (Pergamon, New York, 1978), p. 51.

3. L. P. Ginzburg, J. Non-Cryst. Solids **171**, 164 (1994).

4. L. P. Ginzburg, Fiz. Tekh. Poluprovodn. (Leningrad) **23**, 1629 (1989) [Sov. Phys. Semicond. **23**, 1008 (1989)].

5. G. Lucovsky, Philos. Mag. B **39**, 531 (1979).

6. K. D. Tséndin, in *Electron Phenomena in Glassy Chalcogenide Semiconductors*, Ed. by K. D. Tséndin (Nauka, St. Petersburg, 1996).

7. L. P. Ginzburg, J. Non-Cryst. Solids **171**, 172 (1994).

8. R. Chicon, M. Ortuño, and M. Pollak, Phys. Rev. B **37**, 10520 (1988).

9. M. Kastner and H. Fritsche, Philos. Mag. B **37**, 199 (1978).

10. G. Lukovsky, Philos. Mag. B **41**, 457 (1980).

11. L. P. Ginzburg, Fiz. Khim. Stekla **25** (2), 47 (1999).

12. J. M. Ziman, *Models of Disorder: The Theoretical Physics of Homogeneously Disordered Systems* (Cambridge Univ. Press, Cambridge, 1979; Mir, Moscow, 1982).

13. J. H. Davies, P. A. Lee, and T. M. Rice, Phys. Rev. B **29**, 4260 (1984).

14. F. Mollot, J. Chernogora, and C. Benoit à la Gillaume, Phys. Status Solidi A **21**, 281 (1974).

15. I. G. Austin, T. S. Nashashibi, T. M. Searle, and W. E. Spear, J. Non-Cryst. Solids **32**, 373 (1979).

16. B. L. Gel'fand and B. I. Shklovskiĭ, in *Physics of Compound Semiconductors. Proceedings of the IX Winter School on Physics of Semiconductors, Leningrad, 1979*, p. 5.

17. P. Kaiser, J. Opt. Soc. Am. **64**, 475 (1974).

18. Y. Hibino, H. Hanafusa, and S. Sakaguchi, Appl. Phys. Lett. **47**, 1157 (1985).

19. S. Munekuni, T. Yamanaka, Y. Shimogauchi, *et al.*, J. Appl. Phys. **68**, 1212 (1990).

20. A. J. Smiel and T. A. Fisher, Appl. Phys. Lett. **41**, 324 (1982).

21. J. I. Zink, Acc. Chem. Res. **11** (8), 289 (1978).

Translated by A. Sidorova-Biryukova

====================== **SOLIDS** ======================

Giant Piezoelectric Anisotropy in Sodium Niobate with a Composite-Like Structure

L. A. Reznichenko, L. A. Shilkina, A. V. Turik, and S. I. Dudkina

Research Institute of Physics, Rostov State University,
pr. Stachki 194, Rostov-on-Don, 344090 Russia
e-mail: klevtsov@ip.rsu.ru
Received May 18, 2001

Abstract—On the basis of X-ray diffraction and electrophysical data for stoichiometric and nonstoichiometric sodium niobate, it is concluded that a mechanism responsible for the infinite ratio of the electromechanical coupling factors for the thickness and plane modes may be related to the composite-like structure of the material. It is shown that such a structure may result from the presence of ordered extended voids (arising at the joints between perfect structure blocks), which behave as microcracks. Stresses generated at the joints may modify the domain structure and lead to the anisotropic distribution of polarized clusters in the ferroelectric phase. This also may enhance the piezoelectric anisotropy. © 2002 MAIK "Nauka/Interperiodica".

In [1], we have discovered the piezoelectric effect in polarized stoichiometric and nonstoichiometric sodium niobate ($NaNbO_3$) ceramics in the temperature range of 20–300°C. This effect is accompanied by the piezoresonant dispersion of the permittivity. It was noted that this circumstance, together with the observation of dielectric hysteresis loops [2], indicates that polar microdomains, or ferroelectric clusters, form in the *Pbma* antiferroelectric rhombic phase of $NaNbO_3$ subjected to a polarizing electric field. For the stoichiometric composition, the values of the piezoelectric moduli d_{33} measured at room temperature ranged from 25 to 50 pC/N. The values of the electromechanical coupling coefficients for the thickness mode were found to be $K_t = 0.25$–0.31; those of the mechanical quality factor, $Q_{Mm} = 2$–7. All the compositions had very small planar electromechanical coupling coefficients ($K_p \Rightarrow 0$) and piezoelectric moduli ($d_{31} \Rightarrow 0$) and, hence, very large anisotropies of the electromechanical coupling coefficients, $K_t/K_p \longrightarrow \infty$, and piezomoduli, $d_{33}/|d_{31}| \longrightarrow \infty$.

Earlier, a similar situation was investigated [3, 4] in several $PbTiO_3$-based ferroelectric ceramics. Therefore, it was supposed [1] that the reason for the large piezoelectric anisotropy in $NaNbO_3$ and $PbTiO_3$ is the same: small values and low anisotropy of the permittivity tensor components in single-domain metastable ferroelectric $NaNbO_3$ crystals, as well as the specific domain structure and distribution of clusters in the ferroelectric phase.

In this study, we hypothesize that the very large piezoelectric anisotropy in $NaNbO_3$ is related to its composite-like structure. The main distinctions between composite-like materials and ordinary ceramics are [5] rather high K_t when plane modes are close to

zero or even vanish and extremely small (down to several unities) values of the mechanical quality factor Q_{Mm}. Since such a combination of parameters is characteristic of some $NaNbO_3$-based ceramics, one can expect that their structure is also composite-like.

It is likely that the mechanisms of forming the composite-like structure in $NaNbO_3$- and $PbTiO_3$-based ceramics is the same, i.e., the development of the anisotropic structure of extended defects (like microcracks [6–9]). However, since high-resolution optical and electron microscopy methods have failed to visualize regularly arranged microcracks in $NaNbO_3$ (unlike $PbTiO_3$), one may assume that other extended defects form voids similar to microcracks. The origination of such voids can be related to the specific crystal-chemical structure of the object, namely, its block structure due to developing shear microstructures [10–12]. Their existence in $NaNbO_3$, as well as in more complex niobium oxides and their solid solutions (as demonstrated with the method of X-ray powder diffraction [13–15]), is explained by the fact that Nb_2O_5, the thermally stablest compound of those participating in the synthesis of niobate materials, is prone to void formation [16, 17].

A characteristic feature of the shear microstructures is the elimination of anion vacancies, arising when the degree of oxidation of Nb slightly changes $(Nb^{5+}_{1-x}Nb^{4+}_x)_2 O^{2-}_{5-x} \square_x$ ($\square$ is an anion vacancy), upon shearing one part of the structure relative to the other along a certain crystallographic direction. In Nb_2O_5, shear develops, as a rule, in two or three almost orthogonal directions, giving rise to rectangular infinitely long $m \times n$ columns of oxygen octahedrons or $m \times n \times p$ ($p = 1, 2$) blocks [16, 17], respectively. At sites where the

columns and blocks join together, voids (channels) form, which may contain both octa- and tetrahedral positions [16]. The existence of the latter was demonstrated by X-ray emission spectroscopy [18]. Here, Nb atoms, retaining the octahedral configuration, occupy two levels ($Z = 0$ and $Z = 1/2$), being displaced by half-diagonal of the octahedron in neighboring blocks. Twelve of 14 known Nb_2O_5 modifications have the block structure. The stablest high-temperature polymorphic modification, α-Nb_2O_5, is obtained from the ReO_3-type structure by alternating the $(20\bar{9})$ and (601) planes. The resulting block structure has the form $3 \times 4 \times 1 + 3 \times 5 \times \infty$, the columns $3 \times 5 \times \infty$ being parallel to the c axis of the Nb_2O_5 unit cell.

In $NaNbO_3$, we found three systems of crystallographic shear planes that are also parallel to the c axis ([010]) of the crystal. They form two types of the cross sections of the $m \times n$ octahedrons. One refers to the $m \times n \times p$ blocks; the other, to the $m \times n \times \infty$ columns of octahedrons. The direction of displacement of Nb atoms from their perfect sites, which was determined in the single crystals, suggests that the (102) and (201) planes are crystallographic shear planes and p in the formula $m \times n \times p$ equals two.

The alternation of the crystallographic shear planes must cause satellite maxima to appear in the X-ray diffraction pattern near the basic reflections. The spacing between the satellites and the basic reflections is inversely proportional to the wavelength (λ) of a disturbance caused by crystallographic shear, deforming the ideal lattice, i.e., to the spacing between the crystallographic shear planes [19, 20]. The X-ray diffraction patterns from the samples under investigation contain two (C_1 and C_2) (Fig. 1a) or three (C_1, C_2, and C_3) (Fig. 1b) satellites corresponding to the spacing between the crystallographic shear planes (70–80 Å (C_1), 60–70 Å (C_2), and 40 Å (C_3)) and indicating the formation of the columns and/or perfect blocks. The appearance of the satellites only on the side of smaller angles θ seems to reflect the correlation of shift and substitution disorders. The infinitely long columns regularly arranged along the [010] direction join via ordered channels with vacant or filled octahedral positions.
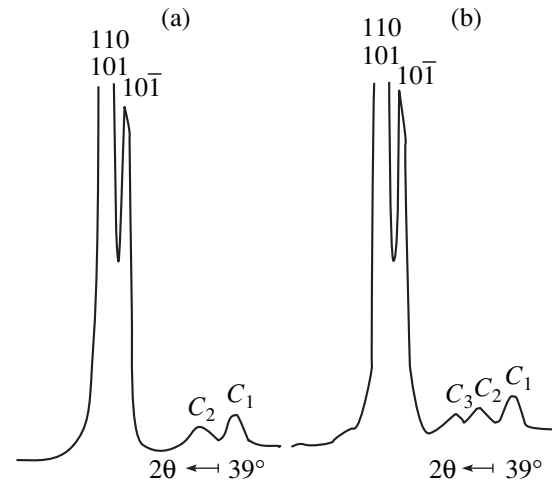
Fig. 1. Fragments of the X-ray diffraction patterns for the $NaNbO_3$ ceramics including the (100) and $(10\bar{1})$ reflections and (a) two (C_1, C_2) and (b) three (C_1, C_2, C_3) satellites.

$NaNbO_3$ samples of stoichiometric and nonstoichiometric compositions were obtained by the conventional ceramic technology (solid-phase synthesis at 800–850°C, followed by pressure-free sintering at 1220–1240°C). We used extra-pure grade and commercial-grade Nb_2O_5. Powders synthesized were (or were not) granulated by the two-stage grinding of pressed blocks first through a sieve with a mesh of 0.7–0.9 mm and after four-hour keeping in a closed vessel, through a sieve with a mesh of 0.25–0.30 mm. Such a procedure significantly increases the density of the granulated material and (after its sintering) ceramic samples.

The table lists the density ρ; wavelength λ; the number N of oxygen octahedrons between the crystallographic shear planes; factors K_t and K_p; and I_{rel}, which is the ratio of the (430) reflection intensity for impurity compound $NaNb_3O_8$ to the (100) reflection intensity for $NaNbO_3$ of stoichiometric and nonstoichiometric composition.

It is seen that the less dense samples with a small amount of the $NaNb_3O_8$ intermediate phase has the infinite anisotropy of properties. The finite value of the

Structure parameters and the electrophysical properties of stoichiometric and nonstoichiometric sodium niobate

Composition	ρ, ρ/cm^3	K_t	K_p	C_1		C_2		I_{rel}
				λ, Å	N	λ, Å	N	
$NaNbO_3$ extra-pure grade granulated	4.31	0.31	0.10	69.4	12.6	42.12	7.6	0
$NaNbO_3$ extra-pure grade nongranulated	3.1	0.31	0	79.3	14.4	40.9	7.4	8
$NaNbO_3$ commercial-grade granulated	3.23	0.26	0	67.5	12.5	41.38	7.5	10
$Na_{0.88}NbO_3$ commercial-grade nongranulated	3.09	0.25	0	68.5	12.4	41.02	7.44	10
$Na_{0.86}NbO_3$ commercial-grade nongranulated	3.40	0.25	0	67.9	12.6	41.38	7.50	10

ratio K_t/K_p is typical of high-density and pure NaNbO$_3$ obtained from the granulated powders. The reason may be as follows.

It is well known [21] that, in perovskites, shear causes not only anion but also cube–octahedral *A* sites to disappear. As a result, NaNbO$_3$ always contains an excess amount of Na atoms (4–6 at.%). If the sample preparation conditions provide the maximum possible solubility of the components (granulation at the stage of synthesis, sintering by hot pressing, etc.), these free atoms can occupy regular and irregular vacant sites in the NaNbO$_3$ structure, including those inside the inter-columnar and interblock channels. Then, since evidence for superlattice ordering of Na is lacking, it is reasonable to assume that its distribution in these channels agrees with the model of random filling, as is the case for Ge atoms in the GeO$_2$: 9Nb$_2$O$_5$ compound [22]. Sodium niobate crystallizing under these conditions is pure and has a high density. When Na atoms form a side impurity compound rather than fill the voids, these channels seem to behave as directional microcracks. High stresses arising at the sites of crack nucleation abruptly increase the internal friction and, as a consequence, decrease Q_{Mm}. At the same time, the voids suppress the plane mode, thereby increasing the ratio K_t/K_p. The modification of the domain structure and the anisotropic distribution of polarized ferroelectric clusters, which is induced by the stresses, may also be factors contributing to the piezoelectric anisotropy. The validity of this assumption is supported by the infinite piezoelectric anisotropy observed in high-porosity PbTiO$_3$-based ceramics, having a large number of microcracks and, consequently, a low strength [6–9].

Thus, the additional reason for giant piezoelectric anisotropy in some types of stoichiometric and nonstoichiometric NaNbO$_3$ is its composite-like structure caused by the presence of ordered extended defects (voids), which arise where the columns and perfect blocks join and behave as microcracks.

Note that since the shear microstructures "inherited" from rutile TiO$_2$ [16] also take place [23] in PbTiO$_3$-based ceramics and related Ti-containing materials, this factor may contribute to the anisotropy of the piezoelectric coefficients in these materials.

ACKNOWLEDGMENTS

REFERENCES

1. L. A. Reznitchenko, A. V. Turik, E. M. Kuznetsova, and V. P. Sakhnenko, J. Phys.: Condens. Matter **13**, 3875 (2001).

2. R. H. Dungan and R. D. Golding, J. Am. Ceram. Soc. **47** (2), 73 (1964).

3. A. V. Turik and V. Yu. Topolov, J. Phys. D **30**, 1541 (1997).

4. A. V. Turik, Ferroelectrics **222**, 33 (1999).

5. W. A. Smith and A. A. Shaulov, Ferroelectrics **87**, 309 (1988).

6. D. I. Makar'ev, A. N. Klevtsov, V. A. Servuli, and L. A. Shilkina, in *Proceedings of the International Conference "Fundamental Problems of Piezoelectric Instrument Making" ("Piezotechnology-99"), Rostov-on-Don, 1999*, Vol. 1, p. 44.

7. E. A. Dul'kin, L. I. Grebenkina, D. I. Makar'ev, *et al.*, Pis'ma Zh. Tekh. Fiz. **25** (22), 21 (1999) [Tech. Phys. Lett. **25**, 894 (1999)].

8. D. I. Makar'ev, A. N. Klevtsov, and V. A. Servuli, Electronic Journal *Issledovano v Rossii*, No. 57 (1999), http://Zhurnal/mipt/rssi/ru/articles/1999/058.pdf.

9. D. I. Makar'ev, A. N. Klevtsov, and V. A. Servuli, in *Proceedings of the XV All-Russia Conference on Physics of Ferroelectrics (VKS-XV), Rostov-on-Don, 1999*, p. 292.

10. A. D. Wadsley, Rev. Pure Appl. Chem. **5**, 165 (1955).

11. A. D. Wadsley, in *Non-Stoichiometric Compounds* (Academic, London, 1964), p. 98.

12. J. S. Anderson, in *Surface Defect Properties of Solids* (Chemical Society, London, 1972), Vol. 1, p. 5.

13. L. A. Reznichenko, O. N. Razumovskaya, and L. A. Shilkina, in *Proceedings of the International Conference "Piezotechnology-99," Obninsk, 1997*, p. 191.

14. L. A. Shilkina, L. A. Reznichenko, and O. N. Razumovskaya, in *Proceedings of the 8th International Symposium on Physics of Ferroelectric Semiconductors (JMES-8), Rostov-on-Don, 1998*, Vol. 7, p. 192.

15. L. A. Reznichenko, L. A. Shilkina, and O. N. Razumovskaya, in *Proceedings of the XV All-Russia Conference on Physics of Ferroelectrics (VKS-XV), Rostov-on-Don, 1999*, p. 294.

16. C. N. R. Rao and J. Gopalakrishnan, *New Directions in Solid State Chemistry* (Cambridge Univ. Press, Cambridge, 1986; Nauka, Novosibirsk, 1990).

17. F. Fairbrother, *The Chemistry of Niobium and Tantalum* (Elsevier, Amsterdam, 1967; Khimiya, Moscow, 1972).

18. L. A. Reznichenko, L. A. Shilkina, S. V. Titov, *et al.*, in *Proceedings of the International Symposium "Ordering in Minerals and Alloys" (OMA-2000), Rostov-on-Don, 2000*, p. 111.

19. W. A. Wooster, *Diffuse X-ray Reflections from Crystals* (Clarendon, Oxford, 1962; Inostrannaya Literatura, Moscow, 1963).

20. A. Guinier, *Théorie et Technique de la Radiocristallographie* (Dunod, Paris, 1956; Fizmatgiz, Moscow, 1961).

21. A. G. Petrenko and V. V. Prisedskiĭ, *Structure Defects in Ferroelectrics* (Uchebno-metodicheskiĭ Kabinet po Vysshemu Obrazovaniyu pri Minvuze USSR, Kiev, 1989).

22. A. Ann, A. McConnell, J. S. Anderson, and C. N. R. Rao, Spectrochim. Acta A **32**, 1067 (1976).

23. L. A. Reznichenko, L. A. Shilkina, S. V. Titov, and O. N. Razumovskaya, in *Proceedings of the International Symposium "Ordering in Minerals and Alloys" (OMA-2000), Rostov-on-Don, 2000*, p. 127.

Translated by Yu. Vishnyakov

<div align="center">

SOLID-STATE
ELECTRONICS

</div>

Multislot Microwave Transmission Lines Fabricated on an Insulating Substrate Covered with a Ferroelectric Film

I. G. Mironenko and A. A. Ivanov

St. Petersburg State Electrotechnical University, St. Petersburg, 197376 Russia

e-mail: mit@eltech.ru, MironencoIG@ramber.ru, iva@solaris.ru

Received January 25, 2001

Abstract—A rigorous electrodynamic model is used to calculate the dispersion characteristics of multislot microwave transmission lines fabricated on a layered structure containing a ferroelectric film. The attenuation due to the finite conductivity of the metal electrodes is found. © 2002 MAIK "Nauka/Interperiodica".

INTRODUCTION

A structure consisting of a ferroelectric film applied on a dielectric substrate underlies slot and coplanar microwave transmission lines [1–3]. These lines can be used to construct devices with electrically controlled characteristics basically because the permittivity ε of the ferroelectric film can vary with an applied electric field. The permittivity of most ferroelectric materials changes by a factor of 1.5–2 when the electric field intensity is 2–3 kV/mm. Therefore, for the voltage controlling the line to be low, the slots must be sufficiently narrow. However, narrow slots introduce losses, because the conductivity of the electrodes is finite [1–3]. This contradiction is reconciled in a planar waveguide structure which we will further refer to as a multislot microwave transmission line. In this line, narrow electrodes located inside of a wide slot on the surface of the ferroelectric film serve simultaneously as a waveguide, transmitting the electromagnetic energy, and as control electrodes, which produce the control electric field of a desired intensity in the narrow interelectrode gaps over the entire wide slot.

The purpose of this paper is to analyze the dispersion and attenuation characteristics of such multislot transmission structures.

STATEMENT OF THE PROBLEM AND VALIDATION OF THE ELECTRODYNAMIC MODEL

Figure 1 illustrates the cross section of the transmission line under study. We will deal with a rigorous electrodynamic model. Though relatively intricate, it accurately describes the real structure. The fundamental mode of this structure is a hybrid wave ($E_z \neq 0$ and $H_z \neq 0$). It can be described in terms of two scalar potentials, for example, the electric and magnetic Hertz

vectors aligned with one of the coordinate axes (z or y). It is preferred to align the Hertz vectors with the y axis, because they produce the *LE* and *LM* modes, which turn out to be related through the electrode current at the final stage of the analysis when the boundary conditions are imposed in the plane $y = 0$. In the absence of the electrodes, the structure illustrated in Fig. 1 is a layered inhomogeneous waveguide sandwiched in conducting planes. A natural basis for this waveguide consists of the *LE* and *LM* waves, which are related through the electrode current in the hybrid mode.

As is known, to analyze the dispersion and attenuation characteristics in such planar waveguides, one should solve integral equations for the current distribution over the electrodes or for the electric field distribution over the slots. An efficient procedure for solving the equations is the Galerkin method, which provides as high an accuracy as other ways of solving integral equations. Therefore, we use the Galerkin method at the final stage of the analysis.
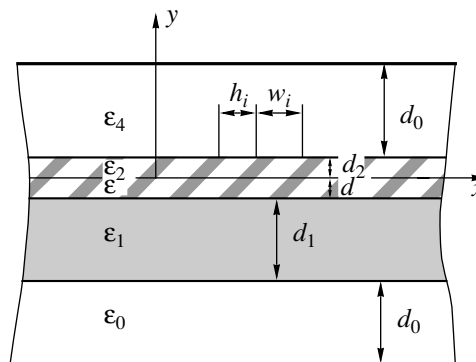

Fig. 1. Cross section of the multislot planar structure.

BASIC STAGES OF THE SOLUTION PROCEDURE

(1) Formulas for the scalar potentials. Let us specify the electric and magnetic potentials as

$$\mathbf{A} = \mathbf{e}_y \varphi(x, y) \exp(-j(\gamma z - \omega t)),$$

$$\mathbf{F} = \mathbf{e}_y \Psi(x, y) \exp(-j(\gamma z - \omega t)),$$

where γ is the propagation constant.

Both potentials satisfy the wave equation in any domain of the cross section. Consider the Fourier transforms of the potentials:

$$\begin{Bmatrix} \bar{\varphi}(y, s) \\ \bar{\Psi}(y, s) \end{Bmatrix} = \int\limits_{-\infty}^{+\infty} \begin{Bmatrix} \varphi(x, y) \\ \Psi(x, y) \end{Bmatrix} \exp(-jsx) dx.$$

The Fourier transforms $\bar{\varphi}(y, s)$ and $\bar{\Psi}(y, s)$ satisfy the homogeneous equation

$$(\partial^2/\partial y^2 + \alpha_i^2) \begin{Bmatrix} \bar{\varphi}_i(y, s) \\ \bar{\Psi}_i(y, s) \end{Bmatrix} = 0 \qquad (1)$$

in any domain of the cross section. In Eq. (1), $\alpha_i^2 = (\gamma^2 + s^2 - k^2 \varepsilon_i)$, $k^2 = \omega^2 \varepsilon_0 \mu_0$, and i is the index of the corresponding domain in Fig. 1.

The solution to Eq. (1) in the ith domain has the form

$$\bar{\varphi}_i(y, s) = A_i(s) \sinh(\alpha_i y) + B_i(s) \cosh(\alpha_i y),$$

$$\bar{\Psi}_i(y, s) = C_i(s) \sinh(\alpha_i y) + D_i(s) \cosh(\alpha_i y). \qquad (2)$$

Clearly, the boundary conditions for the Fourier transforms are the same as those for the electromagnetic potentials: (i) $\partial \bar{\varphi}(y, s)/\partial y = 0$ and $\bar{\Psi}(y, s) = 0$ on the perfectly conducting planes and (ii) the Fourier transforms and their normal derivatives $(\partial \bar{\varphi}(y, s)/\partial y)(1/\varepsilon)$ and $\partial \bar{\psi}(y, s)/dy$ are continuous at the boundaries of the insulating layers. Using (2) and the boundary conditions, the scalar functions $\bar{\varphi}(y, s)$ and $\bar{\psi}(y, s)$ can be found up to constant factors. Let us represent these functions in different domains of the cross section as

$$y \geq 0: \bar{\varphi}^{(+)}(y, s) = B^m(s) A^{(+)}(y, s);$$

$$\bar{\Psi}^{(+)}(y, s) = B^e(s) F^{(+)}(y, s);$$

$$y \leq 0: \bar{\varphi}^{(-)}(y, s) = B_1^m(s) A^{(-)}(y, s); \qquad (3)$$

$$\bar{\Psi}^{(-)}(y, s) = B_1^e(s) F^{(-)}(y, s),$$

where $B^m(s)$, $\ldots B_1^e(s)$ are arbitrary functions and $A^{(+)}(y, s)$, $F^{(+)}(y, s)$, $A^{(-)}(y, s)$, and $F^{(-)}(y, s)$ are the known functions presented in the Appendix.

Relationships (3) thus determine the Fourier transforms of the scalar potentials up to arbitrary factors.

(2) Derivation of the integral equations. To derive the integral equations, we use the boundary conditions for the Fourier transforms of the tangential components of the electric and magnetic field intensities in the plane $y = 0$. Let us write the relationships for the Fourier transforms of the field components that directly follow from the Maxwell equations:

$$\bar{H}_x(y, s) = j\gamma\bar{\varphi}(y, s) + (s/\omega\mu_0)(\partial\bar{\Psi}(y, s)/\partial y),$$

$$\bar{E}_x(y, s) = (s/\omega\varepsilon_0\varepsilon)(\partial\bar{\varphi}(y, s)/\partial y) - j\gamma\bar{\Psi}(y, s),$$

$$\bar{H}_z(y, s) = js\bar{\varphi}(y, s) - (\gamma/\omega\mu_0)(\partial\bar{\Psi}(y, s)/\partial y), \qquad (4)$$

$$\bar{E}_z(y, s) = -(\gamma/\omega\varepsilon_0\varepsilon)(\partial\bar{\varphi}(y, s)/\partial y) - js\bar{\Psi}(y, s).$$

The tangential components E_x and E_z, as well as their Fourier transforms, are continuous within the slots:

$$\bar{E}_z^{(+)}(0, s) = \bar{E}_z^{(-)}(0, s),$$

$$\bar{E}_x^{(+)}(0, s) = \bar{E}_x^{(-)}(0, s). \qquad (5)$$

In other words, relationship (5) must be satisfied in the plane $y = 0$. Writing Eqs. (5) in view of (3) and (4) and eliminating two unknown functions, we obtain

$$B_1^m(s) = B^m(\varepsilon/\varepsilon_2)([A^{(+)}(0, s)]/[A^{(-)}(0, s)]),$$

$$B_1^e(s) = B^e(s)([F^{(+)}(0, s)]/[F^{(-)}(0, s)]), \qquad (6)$$

where $[A^{\pm}(0, s)] = \partial A^{\pm}(y, s)/\partial y|_{y=0}$.

Using relationships (3), (4), and (6), $\bar{E}_x$ and $\bar{E}_z$ in the plane $y = 0$ can be written as

$$\bar{E}_x(y, s) = (s/\omega\varepsilon_0\varepsilon_2)(B^m(s)/[A^{(+)}(0, s)]) \\ - j\gamma B^e F^{(+)}(y, s), \qquad (7)$$

$$\bar{E}_z(y, s) = -(\gamma/\omega\varepsilon_0\varepsilon_2)(B^m(s)/[A^{(+)}(0, s)]) \\ - js B^e F^{(+)}(y, s). \qquad (8)$$

These relationships for the Fourier transforms of the electric field components in the plane $y = 0$ are independent of the electrode geometry.

As is known, the narrower the domain of definition of unknown functions, the better the convergence of the Galerkin method. In our structure, these domains are slots between the electrodes. Therefore, we will derive integral equations for the fields in the slots. When the conductivity (σ) of the electrode metal is finite, the boundary conditions on the electrodes are written in terms of the electrode currents. Assume that the electrodes are nontransparent for the field, though infinitesimally thin. In this case, the tangential magnetic field experiences a discontinuity equal to the surface current density j_s on the electrodes; i.e., the boundary condition

on the conducting electrodes (the plane $y = 0$) can be written as

$$[\mathbf{e}_y(\overline{\mathbf{H}}^{(+)}(x, 0) - \overline{\mathbf{H}}^{(-)}(x, 0))] = \mathbf{j}_s, \qquad (9)$$

where the brackets mean the vector product.

We introduce the surface impedance of the electrodes in the form $Z_s = (1 + j)\sqrt{\omega\mu_0/2\sigma}$ and write expression (9) in the equivalent form in terms of the field on the electrodes

$$\dot{\mathbf{E}}(x, 0) = Z_s[\mathbf{e}_y(\dot{\mathbf{H}}^{(+)}(x, 0) - \dot{\mathbf{H}}^{(-)}(x, 0))]. \qquad (10)$$

With (10), the Fourier transform of the component $\dot{E}_x$ of the field in the plane $y = 0$ can be represented as

$$\overline{E}_x(0, s) = \int_{-\infty}^{+\infty} \dot{E}_x(x, 0)\exp(-jsx)dx$$

$$= Z_s \int_{-\infty}^{\text{slot edge}} (\dot{H}_z^{(+)}(x, 0) - \dot{H}_z^{(-)}(x, 0))\exp(-jsx)dx$$

$$+ \sum_{\text{over slots}} \int \dot{E}_x(x, 0)\exp(-jsx)dx \qquad (11)$$

$$+ \sum_{\text{over electrodes}} \int (\dot{H}_z^{(+)}(x, 0) - \dot{H}_z^{(-)}(x, 0))\exp(-jsx)dx$$

$$+ Z_s \int_{\text{slot edge}}^{+\infty} (\dot{H}_z^{(+)}(x, 0) - \dot{H}_z^{(-)}(x, 0))\exp(-jsx)dx.$$

Since $\dot{H}_z$ is continuous in the slots, formula (11) can be recast as

$$\overline{E}_x(0, s)$$

$$= Z_s(\dot{H}_z^{(+)}(0, s) - \dot{H}_z^{(-)}(0, s)) + \sum_{\text{over slots}} q_i(s). \qquad (12)$$

A similar expression can be obtained for $\overline{E}_z(0, s)$:

$$\overline{E}_z(0, s)$$

$$= Z_s(\dot{H}_x^{(+)}(0, s) - \dot{H}_x^{(-)}(0, s)) + \sum_{\text{over slots}} g_1(s). \qquad (13)$$

In these formulas,

$$q_i(s) = \int_{i\text{th slot}} E_x(x, 0)\exp(-jsx)dx,$$

$$g_i(s) = \int_{i\text{th slot}} E_z(x, 0)\exp(-jsx)dx.$$

The left-hand side of formulas (12) and (13) is known from expressions (7) and (8). The Fourier transforms for the components of the magnetic field are found from (3) and (4). Therefore, in view of (12) and (13), the remaining two unknown coefficients B_1^m and B_1^e can be expressed through the field distribution in the slot as

$$B^m = \omega\varepsilon_0\varepsilon_2\left(\left(-\gamma\sum_i g_i(s)\right.\right.$$

$$\left.\left. + s\sum_i q_i(s)\right)\right/((\gamma^2 + s^2)[A^{(+)}(0, s)]\xi^m)\right),$$

$$\qquad (14)$$

$$B^e = j\left(\left(\gamma\sum_i q_i(s)\right.\right.$$

$$\left.\left. + s\sum_i g_i(s)\right)\right/((\gamma^2 + s^2)[F^{(+)}(0, s)]\xi^e)\right),$$

where

$$\xi^e = 1 + (Z_s/120\pi k)G(s),$$

$$\xi^m = 1 - (Z_s 120\pi)k\varepsilon_2 P(s),$$

$$G(s) = ([F^{(+)}(0, s)]/F^{(+)}(0, s))$$
$$- ([F^{(-)}(0, s)]/F^{(-)}(0, s)),$$

$$P(s) = (A^{(+)}(0, s)/[A^{(+)}(0, s)])$$
$$- (\varepsilon/\varepsilon_2)([A^{(-)}(0, s)]/A^{(-)}(0, s)).$$

The condition that the Fourier transforms of the tangential magnetic field components be continuous in the slot,

$$H_x^{(+)}(0, s) = H_x^{(-)}(0, s),$$

$$H_z^{(+)}(0, s) = H_z^{(-)}(0, s),$$

yields the equations

$$j\gamma[A^{(+)}(0, s)]P(s)B^m$$

$$+ (s\omega\mu_0)F^{(+)}(0, s)G(s)B^e = 0,$$

$$js[A^{(+)}(0, s)]P(s)B^m$$

$$+ (\gamma/\omega\mu_0)F^{(+)}(0, s)G(s)B^e = 0. \qquad (15)$$
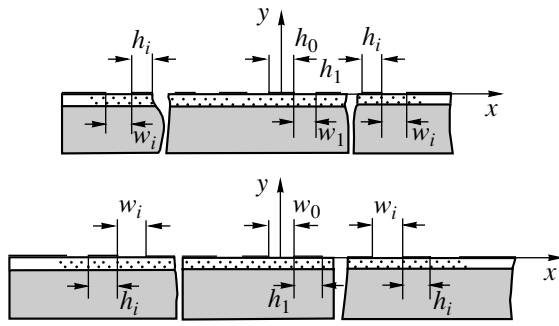
Fig. 2. Cross section of multislot planar structures with an even and odd number of slots.

With (14), Eq. (15) can be reduced, after some algebra manipulation, to

$$f_{11}(s, \dot{\gamma}) \sum_i q_i(s) + f_{12}(s, \dot{\gamma}) \sum_i g_i(s) = 0,$$

$$\dot{\gamma} s f_{21}(s, \dot{\gamma}) \sum_i q_i(s) + f_{22}(s, \dot{\gamma}) \sum_i g_i(s) = 0, \tag{16}$$

where

$$f_{11}(s, \dot{\gamma})$$
$$= (k^2 s^2 P(s) \xi^e - \dot{\gamma}^2 G(s) \xi^m)/((\dot{\gamma}^2 + s^2) \xi^e \xi^m),$$

$$f_{12}(s, \dot{\gamma}) = f_{21}(s, \dot{\gamma})$$
$$= -(k^2 P(s) \xi^e + G(s) \xi^m)/((\dot{\gamma}^2 + s^2) \xi^e \xi^m),$$

$$f_{22}(s, \dot{\gamma}) = (k^2 \dot{\gamma}^2 P(s) \xi^e - s^2 G(s) \xi^e)/((\dot{\gamma}^2 + s^2) \xi^e \xi^m),$$

$$\dot{\gamma} = \gamma' - j\gamma''.$$

Equations (16) constitute a system of integral equations for the electric field distribution in the slots.

(3) Solution of the integral equations. The technique for solving integral equations is well-known. The electric field components $\dot{E}_x$ and $\dot{E}_z$ in the slots are expanded into Chebyshev polynomials of the first and second kind with the weighting functions meeting the Meixner edge condition [4, 5].

Let us calculate the integrals $q_i(s)$ and $g_i(s)$ in Eq. (16). We assume that the slots and the electrodes, which are arranged symmetrically about the axis of symmetry (magnetic axis for an even number of slots and electric for an odd number of slots) have equal dimensions (Fig. 2). In this case, the components E_x and E_z of the fundamental mode in the symmetrically arranged slots are equal. Let x_{0i} be the center of the ith

slot. Then,

$$q_i(s) = 2 \int_{-w_i/2}^{w_i/2} E_{x, i}(x, 0) \cos(x_{0, i}, s) \exp(-jsx) dx,$$

$$\tag{17}$$

$$g_i(s) = 2 \int_{-w_i/2}^{w_i/2} E_{z, i}(x, 0) \cos(x_{0, i}, s) \exp(-jsx) dx.$$

As is known, the integrals of the Chebyshev polynomials are calculated exactly [5]. In view of this, (17) can be written as

$$q_i(s) = 2 \sum_{n = 0, 1} (-1)^n a_n^{(i)} J_{2n}(sw_1/2) \cos(x_{0, i}, s),$$

$$g_i(s) = 2j \sum_{m = 1} (-1)^{m + 1} b_m^{(i)} 2m \tag{18}$$

$$\times (J_{2m}(sw_{i/2})/sw_i/2) \cos(x_{0, i}, s),$$

where $J_v(z)$ is the Bessel function and $a_n^{(i)}$ and $b_m^{(i)}$ are the coefficients of $E_x(x, 0)$ and $E_z(x, 0)$ expansion in the Chebyshev polynomials for the ith slot.

Substituting (18) into Eq. (16) yields

$$\sum_{i = 1}^N \sum_{n = 0, 1}^\infty (-1)^n a_n^{(i)} J_{2n}(sw_i/2) f_{11}(s, \dot{\gamma}) \cos(x_{0, i}, s)$$

$$+ 2j\dot{\gamma} \sum_{i = 1}^N \sum_{m = 1, 2}^\infty (-1)^{m + 1} 2m b_m^{(i)} (J_{2m}(sw_{i/2}) w_i)$$

$$\times f_{12}(s, \dot{\gamma}) \cos(x_{0, i}, s) = 0,$$

$$\tag{19}$$

$$\dot{\gamma} s \sum_{i = 1}^N \sum_{n = 0, 1}^\infty (-1)^n a_n^{(i)} J_{2n}(sw_i/2) f_{21}(s, \dot{\gamma}) \cos(x_{0, i}, s)$$

$$+ j \sum_{i = 1}^N \sum_{m = 1, 2}^\infty (-1)^{m + 1} 2m b_m^{(i)} (J_{2m}(sw_i/2)/(sw_i/2))$$

$$\times f_{22}(s, \dot{\gamma}) \cos(x_{0, i}, s) = 0,$$

where N is the number of slots on either side of the plane of symmetry.

To find the unknown coefficients $a_n^{(i)}$ and $b_m^{(i)}$ and the propagation constant $\dot{\gamma}$, we multiply the first and second equations in (19) by $(-1)^p J_{2n}(s\omega_i/2) \cos(x_{0, t}, s)$ and $j(-1)^{q + 1} 2q(J_{2n}(s\omega_t/2)/s\omega_t/2) \cos(x_{0, t}, s)$, respectively, and integrate the results over the interval $[0, \infty]$
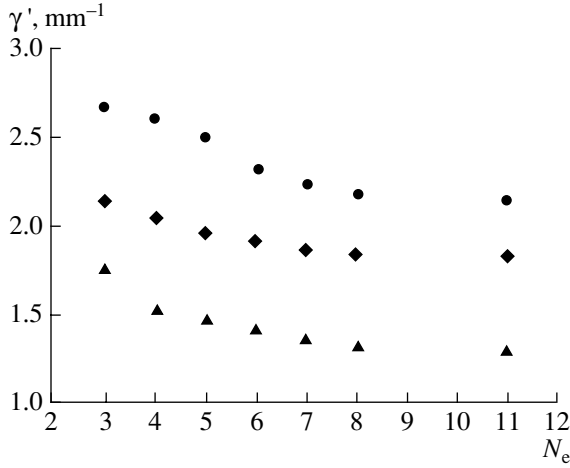
Fig. 3. Propagation constant versus number of electrodes N_e at $\varepsilon = $ (▲) 500, (◆) 1500, and (●) 2500.
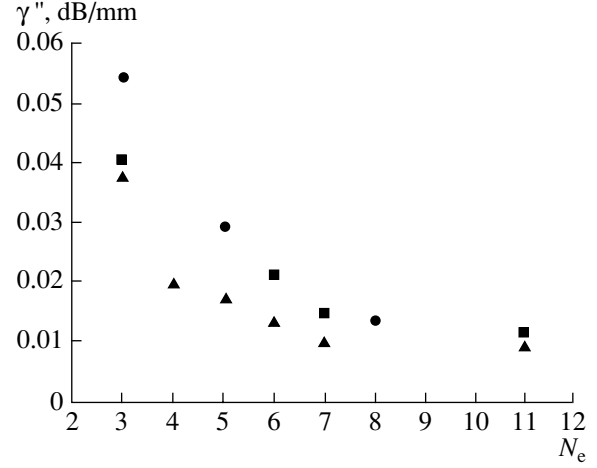
Fig. 4. Attenuation constant versus number of electrodes N_e at the same ε as in Fig. 3.

to obtain

$$\sum_{i=1}^{N}\sum_{n=0,1}^{\infty} a_n^{(i)} K_{p,n,i,t}^{(11)}(\dot{\gamma}) + \sum_{i=1}^{N}\sum_{m=1,2}^{\infty} b_m^{(i)} K_{p,m,i,t}^{(12)}(\dot{\gamma}) = 0,$$

$$\sum_{i=1}^{N}\sum_{n=0,1}^{\infty} a_n^{(i)} K_{q,n,i,t}^{(21)}(\dot{\gamma}) + \sum_{i=1}^{N}\sum_{m=1,2}^{\infty} b_m^{(i)} K_{p,m,i,t}^{(22)}(\dot{\gamma}) = 0,$$

(20)

where

$$K_{p,n,i,t}^{(11)}(\dot{\gamma}) = (-1)^{n+p}\int_0^\infty J_{2p}(sw_t/2)f_{11}(s,\dot{\gamma})$$

$$\times J_{2p}(sw_t/2)\cos(x_{0,i},s)\cos(x_{0,t},s)ds,$$

$$K_{p,m,i,t}^{(12)}(\dot{\gamma}) = j(4\gamma_m/w_i)(-1)^{m+p+1}(I^{12}),$$

$$I^{12} = \int_0^\infty J_{2p}(sw_t/2)f_{12}(s,\dot{\gamma})J_{2m}(sw_i/2)$$

$$\times \cos(x_{0,i},s)\cos(x_{0,t},s)ds,$$

$$K_{q,n,i,t}^{(21)}(\dot{\gamma}) = K_{p,m,i,t}^{(12)}(\dot{\gamma}),$$

$$K_{q,m,i,t}^{(22)}(\dot{\gamma}) = (-1)^{m+q+2}mqI^{22},$$

$$I^{22} = \int_0^\infty [J_{2q}(sw_t/2)f_{22}(s,\dot{\gamma})J_{2m}(sw_t/2)$$

$$\times \cos(x_{0,i},s)\cos(x_{0,t},s)/s^2 w_i w_t]ds.$$

The condition that the determinant of system of equations (20) be zero yields an equation for $\dot{\gamma}$.

NUMERICAL RESULTS

Figures 3 and 4 show the real and imaginary parts of $\dot{\gamma}$ that were calculated for the structure with $w = d = 0.05$ mm, $d_1 = 0.34$ mm, $\varepsilon_1 = 9.5$, $\varepsilon_2 = 1$, and $\sigma = 5 \times 10^7 \, (\Omega \, \text{mm})^{-1}$ at 30 GHz. The screening effect is neglected. A ferroelectric film can be characterized by the product εd. In the ranges $\varepsilon \le 2500$ and $d \le 5 \times 10^{-3}$ mm, the calculation error due to the use of this combined parameter is negligibly small.

The dependence of the propagation constant γ' and attenuation constant γ'' of the multislot transmission line versus number of electrodes (Figs. 3, 4) corroborates the starting idea of the paper: a multislot planar transmission line fabricated on a dielectric substrate covered with a ferroelectric film transmits the electromagnetic field like a slot line. The attenuations due to ohmic losses in the electrodes (in our calculations, $\sigma = 5 \times 10^7 \, (\Omega \, \text{mm})^{-1}$ of multislot and slot lines are close to each other when the slot sizes are comparable. In particular, for a single-slot line based on our structure, $\dot{\gamma} = 1.8 - j2 \times 10^{-3}$ mm^{-1} at $(\varepsilon d) = 1.75$ mm and $w = 0.5$ mm, while for a multislot line with six $d = 0.05$-mm-wide electrodes spaced at 0.05 mm intervals, $\dot{\gamma} = 1.9 - j2 \times 10^{-3}$ mm^{-1} at the same parameters of the insulating structure.

Thus, the multislot planar transmission line fabricated on a dielectric substrate covered by a ferroelectric film can be used for designing microwave devices with electrically tunable parameters.

APPENDIX

Relationships for calculating the scalar potentials are as follows. In the domain $y > 0$,

$$A^{(+)}(\gamma, s) = B^{(+)}\sinh(\alpha_2 y) + C^{(+)}\cosh(\alpha_2 y),$$

where

$$B^{(+)} = \tanh(\alpha_2 d_2) - (\alpha_0/\alpha_2)\varepsilon_2 \tanh(\alpha_0 d_0),$$

$$C^{(+)} = (\alpha_0/\alpha_2)\varepsilon_2 \tanh(\alpha_2 d_2)\tanh(\alpha_0 d_0) - 1,$$

$$F^{(+)}(\gamma, s) = B_1^{(+)}\sinh(a_2 y) + C_1^{(+)}\cosh(\alpha_2 y),$$

$$B_1^{(+)} = (\alpha_0/\alpha_2) - \tanh(\alpha_2 d_2)\tanh(\alpha_0 d_0),$$

$$C_1^{(+)} = \tanh(\alpha_0 d_0) - (\alpha_0/\alpha_2)\tanh(\alpha_2 d_2).$$

In the domain $y < 0$,

$$A^{(+)}(y, s) = [B^{(-)}((\alpha_1\varepsilon/\alpha\varepsilon_1) - \tanh(\alpha_1 d)\tanh(\alpha d))$$

$$+ C^{(-)}(\tanh(\alpha d) - (\alpha_1\varepsilon/\alpha\varepsilon_1)\tanh(\alpha_1 d))]\sinh(\alpha y)$$

$$+ [B^{(-)}(\alpha_1\varepsilon/\alpha\varepsilon_1)\tanh((\alpha d) - \tanh(\alpha_1 d))$$

$$+ C^{(-)}(1 - (\alpha_1\varepsilon/\alpha\varepsilon_1)\tanh(\alpha_1 d)\tanh(\alpha d))]\cosh(\alpha y),$$

where

$$B^{(-)} = \tanh(\alpha_1 d_1') + ((\alpha_0\varepsilon_1/\alpha_1)\tanh(\alpha_0 d_0)),$$

$$C^{(-)} = (\alpha_0\varepsilon_1/\alpha_1)\tanh(\alpha_0 d_0)\tanh(\alpha_1 d_1') + 1,$$

$$F^{(-)}(y, s) = [B_1^{(-)}((\alpha_1/\alpha) - \tanh(\alpha_1 d)\tanh(\alpha d))$$

$$+ C_1^{(-)}(\tanh(\alpha d) - (\alpha_1/\alpha)\tanh(\alpha_1 d))]\sinh(\alpha y)$$

$$+ [B_1^{(-)}((\alpha_1/\alpha)\tanh(\alpha d) - \tanh(\alpha_1 d))$$

$$+ C_1^{(-)}(1 - (\alpha_1/\alpha)\tanh(\alpha_1 d)\tanh(\alpha d))]\cosh(\alpha y),$$

$$B_1^{(-)} = \tanh(\alpha_1 d_1')\tanh(\alpha_0 d_0) + (\alpha_0/\alpha_1),$$

$$C_1^{(-)} = (\alpha_0/\alpha_1)\tanh(\alpha_1 d_1') + \tanh(\alpha_0 d_0).$$

Here,

$$\alpha_0 = (\gamma^2 + s^2 - k^2)^{1/2}, \quad \alpha_1 = (\gamma^2 + s^2 - k^2\varepsilon_1)^{1/2},$$

$$\alpha_2 = (\gamma^2 + s^2 - k^2\varepsilon_2)^{1/2}, \quad \alpha_1 = (\gamma^2 + s^2 - k^2\varepsilon)^{1/2},$$

$$d_1' = d_1 + d.$$

REFERENCES

1. O. G. Vendik, I. S. Danilov, and S. P. Zubko, Zh. Tekh. Fiz. **67** (9), 94 (1997) [Tech. Phys. **42**, 1068 (1997)].

2. I. G. Mironenko and A. A. Ivanov, Pis'ma Zh. Tekh. Fiz. **27** (13), 16 (2001) [Tech. Phys. Lett. **27**, 536 (2001)].

3. I. G. Mironenko and A. A. Ivanov, Pis'ma Zh. Tekh. Fiz. (in press).

4. Integrated Ferroelectrics International Journal **22**, 1420 (1998).

5. K. C. Gupta, R. Garg, I. Bahl, and P. Bhartia, in *Microstrip Lines and Slot Lines* (Artech, Boston, 1996), p. 535.

Translated by A. Khzmalyan

SOLID-STATE ELECTRONICS

Electroluminescence Kinetics in Thin-Film ZnS-Based Devices at Ultralow Frequencies

N. T. Gurin, A. V. Shlyapin, and O. Yu. Sabitov

Ul'yanovsk State University, Ul'yanovsk, 432700 Russia
e-mail: soy@sv.uven.ru
Received March 13, 2001

Abstract—The variation of the instantaneous luminance from thin-film electroluminescent structures excited by alternating-sign sawtooth voltage pulses with a pulse rate of 0.1–2.0 Hz is studied. From the solution of the kinetic equation, the time dependences of the instantaneous luminance and the internal quantum yield are obtained. The time dependences of the luminance and the current exhibit two, "fast" and "slow," rising portions, which correlate with various specific regions in the field and charge dependences of the luminance and other electrophysical parameters. The results are explained by the formation of space charges in the phosphor layer, causing a decrease in its effective thickness, and by a change in the luminescence mechanism. © 2002 MAIK "Nauka/Interperiodica".

The voltage-luminance (*V–L*) characteristic, or the dependence of the mean (apparent) luminance on the alternating-sign exciting voltage amplitude, is a basic characteristic of thin-film electroluminescent devices (TF ELDs) [1]. The mean luminance, in this case, is the result of time averaging over instantaneous luminance pulses, so-called luminance waves. The shape of these waves depends on that of the exciting voltage. From the decay time constant of the instantaneous luminance, the lifetime of excited luminescence centers, for example, Mn^{2+}, was estimated at 1.3–1.5 ms in ZnS at low Mn concentrations and was found to decrease with increasing Mn concentration [1, 2]. However, available data for the *V–L* curve shapes cannot shed light on physical processes responsible for the EL kinetics and do not allow the detailed description of TF ELD properties. The point is that most studies have been performed for the exciting voltage frequencies higher than 50 Hz, at which neighboring waves overlap. Another reason is the use of a rectangular pulse or sinusoidal exciting voltage, which makes the analysis of physical processes very difficult. However, the application of a sawtooth voltage with a frequency of lower than 50 Hz (in this case, all characteristic luminescence times are smaller than a one quarter of the exciting voltage cycle) makes the in-depth analysis of EL in thin-film devices possible [2].

In this work, we study the time variation of the TF ELD instantaneous luminance when the structure is excited by sawtooth voltage pulses with a low repetition rate. The time variation of the luminance is correlated with those of the current and the field in the phosphor layer, with the field and charge dependences of the instantaneous luminance, and also with the current-voltage and voltage-charge characteristics of the phosphor.

Experimental TF ELDs were metal–insulator–semiconductor–insulator–metal (MISIM) structures. Here, M refers to the lower transparent 0.2-μm-thick SnO_2-based electrode applied on the glass substrate and to the upper nontransparent thin-film (0.15 μm thick) Al electrode of diameter 1.5 mm; S, to the 0.54-μm-thick ZnS : Mn (13 wt %) electroluminescent layer; and I, to 0.15-μm-thick $ZrO_2 : Y_2O_3$ (13 wt %) insulating layers. The phosphor layer was applied by thermal evaporation in a quasi-closed chamber at a substrate temperature of 250°C, followed by annealing at 250°C for 1 h; the nontransparent electrode, by thermal evaporation; and the insulating layers, by electron-beam evaporation.

We took the dependences of the instantaneous luminance L and the current I_e through the TF structure on time t when the device was excited by alternating-sign sawtooth voltage pulses from a G6-34 generator equipped with a driver amplifier and an external G5-89 trigger generator. The pulse amplitude was 160 V at a nonlinearity coefficient of no more than 2%. In the "single-shot" excitation mode, a train of two triangular pulses with a repetition rate of 2 Hz is applied. The time T_s between single-shot excitations was varied from 1 to 100 s. In the continuous excitation mode, the pulse repetition rate was 0.1, 0.2, 0.5, and 2 Hz. The current I_e was measured with a 100 Ω- to 10 kΩ-resistor series-connected to the EL structure. The voltage drop across the resistor was equal to or less than 1 V. The instantaneous luminance value was measured with an FÉU-84-3 photoelectric multiplier. The time dependences of the exciting voltage, current through the structure, and instantaneous luminance were recorded with an S9-16 two-channel storage oscilloscope connected to a PC through the associated interface. This system can measure (and memorize) 2048 points within a discretiza-

tion period with an accuracy of better than 2%. The mathematical and graphical data processing were carried out with the application programs Maple V Release4 Version 4.00b and GRAPHER Version 1.06 (from the 2-D Graphing System).

For given variations of the exciting voltage $V(t)$ and current $I_e(t)$ in the external circuit, the mean field $F_p(t)$ in the phosphor layer is given by [3]

$$F_p(t) = \frac{1}{d_p}\left[V(t) - \frac{1}{C_i}\int_0^t I_e(t)dt\right] + F_{pol}, \qquad (1)$$

where d_p is the phosphor layer and C_i is the capacitance of the capacitor made of two TF ELD insulating layers.

The field F_{pol} includes the field F_{pi} due to the polarization charge Q_{pi} accumulated at the phosphor-insulator interface and the field F_{ps} due to the space charge in the phosphor layer. Generally, the latter may have several components associated with different space charges and differently oriented relative to F_{pi}. The field F_{pol} decays in the interval between two active states of the device, in which the current I_p passes through the phosphor. Because of the space charges, this current has the reactive component and can be assigned to a conventional semiconductor device that has a phosphor layer sandwiched in insulating layers. For the initial geometric capacitance of this phosphor layer C_p, its thickness d_p, and the voltage $V_p(t) = F_p(t)d_p$ across it [3, 4], the current through the phosphor, in view of (1), is given by

$$I_p(t) = I_e(t)\frac{C_i + C_p}{C_i} - C_p\frac{dV(t)}{dt}$$
$$= C_i\frac{dV(t)}{dt} - (C_i + C_p)d_p\frac{dF_p(t)}{dt}. \qquad (2)$$

The charge transferred through the phosphor in the absence of losses due to recombination and charge trapping equals the charge accumulated at the interfaces during the active state and is related to the charge Q_e passing in the external circuit as [2, 3]

$$Q_p(t) = \frac{C_i + C_p}{C_i}[Q_e(t) - C_eV(t)] + Q_{pol}, \qquad (3)$$

where

$$Q_e(t) = \int_0^t I_e(t)dt \qquad (4)$$

and Q_{pol} is the residual polarization charge.

The values of Q_{pol} and F_{pol}, as well as the true positions of the $Q_p(t)$ and $F_p(t)$ curves relative to the abscissa axis, were found as in [3] with an accuracy of ±2%. The dependences calculated were obtained with formulas (1)–(4) at $C_i = 986$ pF and $C_p = 250$ pF with allowance for the voltage drop across the resistor,

which was subtracted from $V(t)$. The values of C_i and C_p were determined starting from the TF ELD total capacitance $C_e = 200$ pF measured with an E7-14 impedance meter and from TF ELD geometry.

The experimental curves $L(t)$ and $I_e(t)$ (Fig. 1), as well as the calculated curves $I_p(t)$, $F_p(t)$, $L(Q_p)$, $I_p(F_p)$, and $Q_p(F_p)$ (Fig. 2), have the following distinctive features: (1) The weak asymmetry of the current pulses $I_p(t)$ in the continuous excitation mode when the positive and negative voltage half-waves are applied to the upper electrode (+Al and –Al conditions, respectively), as was observed in [3, 4], the asymmetry diminishing with increasing frequency f; and considerable asymmetry of the luminance waves $L(t)$, which also diminishes with increasing frequency. At the frequencies $f = 0.1$ and 2 Hz, the amplitude of the luminance wave under the +Al conditions is, respectively, 3.4 times and ~1.5 times greater than under the –Al conditions (Fig. 1). (2) The appearance of an extra peak of the current I_p, as was also observed in [3, 4], in the single-shot excitation mode during the first half-cycle under the −Al conditions, the amplitude of this peak growing with T_s, and the appearance of the associated peak of the luminance L (Figs. 2a, 2e), which is explained by the increased probability of emitting center excitation (the probability is proportional to the I_p current density [2, 4]. (3) The presence of two distinct, "fast" and "slow," rising portions in the curves $L(t)$ and $I_p(t)$. Point *1* is the separating point where the rate of rise of L and I_p changes the sign (Figs. 2a, 2e). (4) The similar shapes of the $L(t)$ and $I_p(t)$ curves except for the case +Al, $f = 2$ Hz (Figs. 2a, 2e). (5) The proximity of point *1'*, where the rise in L and I_p markedly slows down (Figs. 2a, 2e), to the point where the S-shaped portions appear in the current-voltage, or $I_p(F_p)$, characteristic of the phosphor layer (Fig. 2d) [3–5], in the curves $L(F_p)$ (Fig. 2c), and in the curves $Q_p(F_p)$ (Fig. 2h). (6) The nearly linear early portion of the curve $L(I_p)$, except for the very beginning, where the curve is superlinear, and the decrease in the slope of this curve (Fig. 2b) when passing over point *1* and especially over point *1'*. (7) The presence of two, fast and slow, rising portions in the curve $L(Q_p)$, which are separated by point *1* (Fig. 2g), as was also observed in [2].

To explain the above observations, we will consider the EL kinetics. As is known [1, 6], the instantaneous luminance $L(t)$ of TF ELDs depends on the excited emitting centers $N^*(t)$ as

$$L(t) = \frac{A\eta_{int}d_pN^*(t)}{\tau^*} = \frac{AN_1P_rd_pN^*(t)}{\tau^*}, \qquad (5)$$

where $A = K_0f_\lambda h\nu/\pi$ is the constant determined in the approximation of emission monochromatism and diffusely scattering surface, K_0 is the coefficient of emission extraction from the device, f_λ is the luminous efficiency, h is the Planck constant, ν is the emission fre-
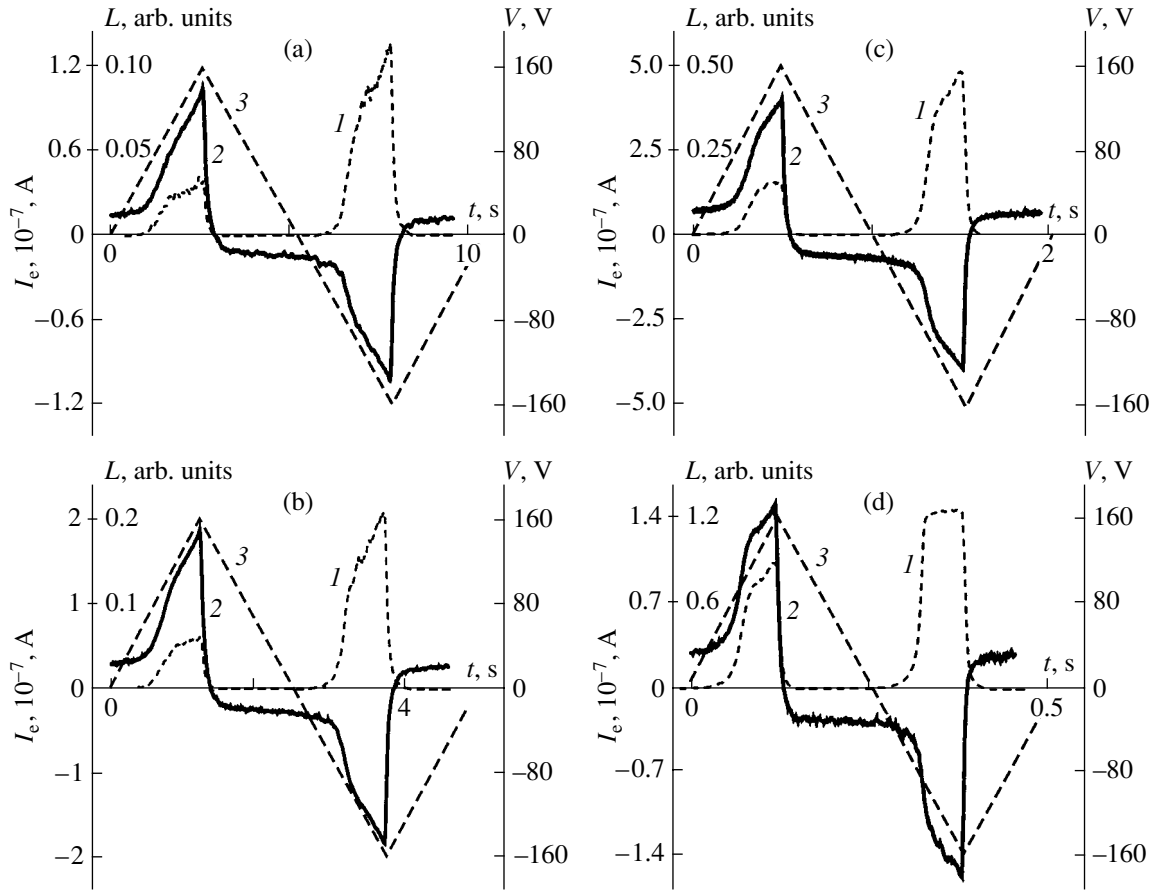
Fig. 1. (*1*) Instantaneous luminance, (*2*) current, and (*3*) exciting voltage vs. time for the continuous excitation. The exciting voltage frequency $f =$ (a) 0.1, (b) 0.2, (c) 0.5, and (d) 2.0 Hz. The first half-cycle corresponds to the conditions (–Al).

quency, $\eta_{int} = N_1 P_r$ is the intrinsic quantum yield, $N_1 = d_p \sigma N$ [1] is the number of emitting centers excited by one electron passing through the phosphor, σ is the cross section of impact excitation of emitting centers, N is the density of the emitting centers, $P_r = \tau^*/\tau_r$ is the probability of radiative relaxation for an emitting center, τ^* is the lifetime of an excited emitting center, and τ_r is the relaxation time constant of excited centers due to radiative transitions to the ground state.

Under the assumption of the impact excitation of emitting centers, $N^*(t)$ is found by solving the kinetic equation [6, 7]

$$\frac{dN^*(t)}{dt} = \alpha(t)[N - N^*(t)] - \frac{N^*(t)}{\tau^*}, \qquad (6)$$

where

$$\alpha(t) = \frac{\sigma I_p(t)}{q S_e} \qquad (7)$$

is the probability of emitting center excitation per unit time, S_e is the TF ELD surface, and q is the electron charge.

If σ is constant, α is independent of N^*, and τ^* is independent of t, the solution of Eq. (6) is [8]

$$N^*(t) = e^{-\left[\int \alpha(t)dt + \frac{1}{\tau^*}\int dt\right]}$$
$$\times [\int \alpha(t)N e^{\int \alpha(t)dt + \frac{1}{\tau^*}\int dt} dt + C], \qquad (8)$$

where C is a constant of integration.

The maximal value of α corresponds to the greatest current in our frequency range, $I_{p2} \approx 1.6 \times 10^{-6}$ A at a frequency of 2 Hz at point *2* (Fig. 2e). At $\sigma \approx 2 \times 10^{-16}$ cm^2 and $S_e = 2$ mm^2 [1, 6], the value of α determined from (7) is $\alpha = 10^{-1}$ s^{-1}. For $\tau^* \approx 1.5 \times 10^{-3}$ s [2], $1/\tau^* \approx 667$ s^{-1}. Hence, the condition $\alpha(t) \ll 1/\tau^*$ is valid throughout the frequency range. Taking into account that $I_p(t) = 0$ and $N^*(t) = 0$ at $t = 0$, so that the constant of integration $C = 0$, and that condition (7) is met, the solution (8) of Eq. (6) takes the form

$$N^*(t) = \frac{\sigma N}{q S_e} e^{-t/\tau^*} \int I_p(t) e^{1/\tau^*} dt. \qquad (9)$$
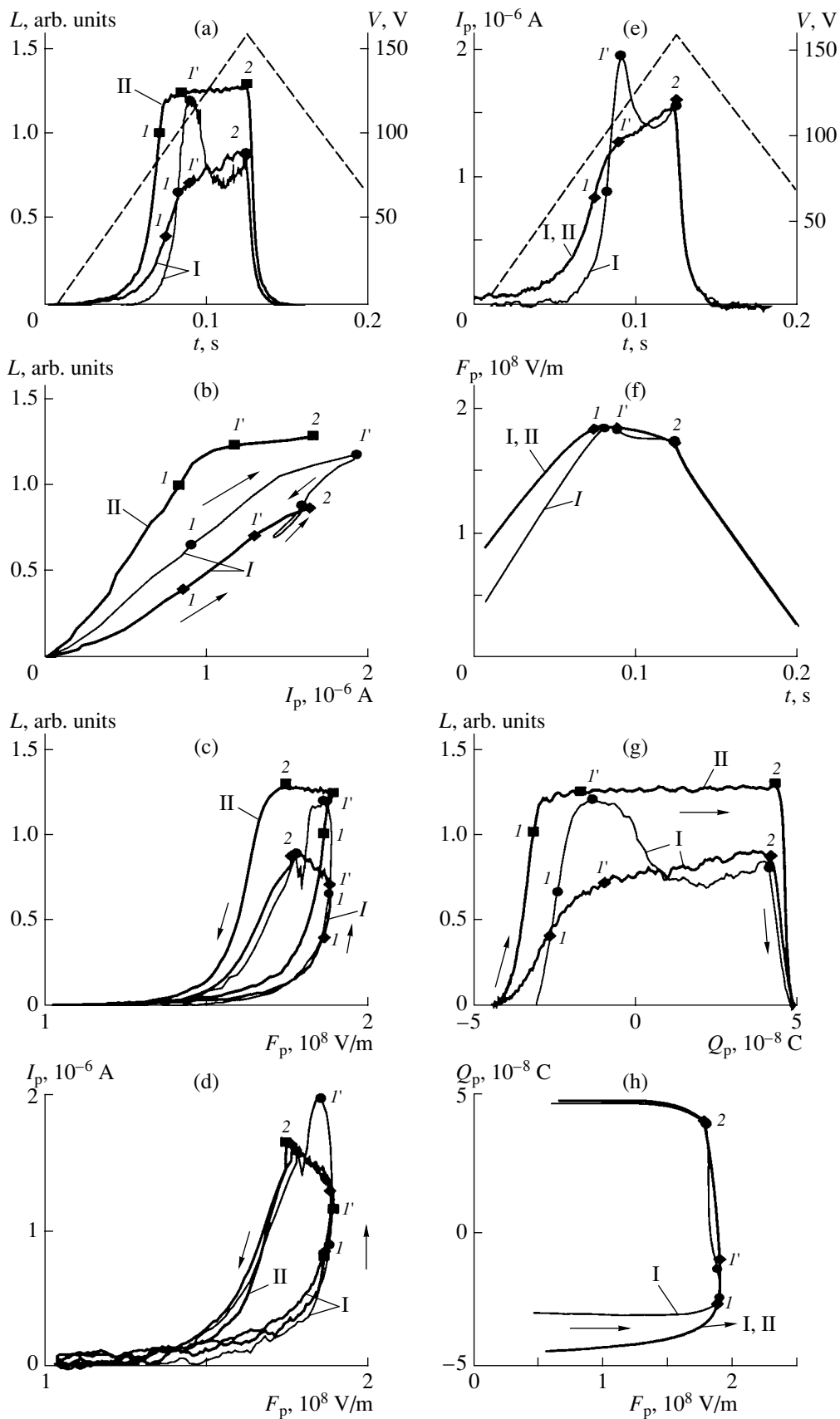
Fig. 2. TF ELD characteristics for $f = 2$ Hz. Dashed lines, $V(t)$; thick continuous lines, continuous excitation; thin continuous lines, single-shot excitation with $T_s = 100$ s. I, conditions (–Al); II, conditions (+Al).

In the fast-rising portion of $I_p(t)$ (Fig. 2e), the curve $I_p(t)$ under the continuous excitation mode is approximated as

$$I_p(t) = I_{p1}e^{\frac{t-t_1}{\tau_1}}, \qquad (10)$$

where τ_1 is the rise time constant of the current $I_p(t)$ and t_1 is the time instant corresponding to point 1.

Solution (9) is reduced to

$$N^*(t) = \frac{\sigma N}{qS_e}\frac{\tau_1\tau^*}{\tau_1+\tau^*}I_{p1}e^{\frac{t-t_1}{\tau_1}} = \frac{\sigma N}{qS_e}\frac{\tau_1\tau^*}{\tau_1+\tau^*}I_{p1}(t) \quad (11)$$

and the dependence $L(t)$ in view of (5) has the form

$$L(t) = \frac{Ad_p^2\sigma^2N^2P_r}{qS_e}\frac{\tau_1}{\tau_1+\tau^*}I_p(t)$$
$$= \frac{A\eta_{int}^2}{qS_eP_r}\frac{\tau_1}{\tau_1+\tau^*}I_p(t). \qquad (12)$$

For $\tau_1 \gg \tau^*$,

$$L(t) = \frac{A\eta_{int}^2}{qS_eP_r}I_p(t). \qquad (13)$$

In the slowly rising portion of the curve $I_p(t)$ (Fig. 2e), it is approximated (under the continuous excitation mode) by

$$I_p(t) = I_{p1}(1 - e^{-t/\tau_2} + e^{t/\tau_3}), \qquad (14)$$

where τ_2 and τ_3 are the rise time constants, respectively, before and after the point where the rate of rise of the current $I_p(t)$ changes the sign in the slowly rising portion (Fig. 2e).

Then, the solution (8) of Eq. (6) takes the form

$$N^*(t) = \frac{\sigma N\tau^*}{qS_e}I_{p1}\left[1 - \frac{\tau_2}{\tau_2-\tau^*}e^{-t/\tau_2} + \frac{\tau_3}{\tau_3+\tau^*}e^{t/\tau_3}\right.$$
$$\left. + \left(\frac{\tau_1}{\tau_1+\tau^*} - 1 + \frac{\tau_2}{\tau_2-\tau^*} - \frac{\tau_3}{\tau_3+\tau^*}\right)e^{-t/\tau^*}\right] \qquad (15)$$

and

$$L(t) = \frac{A\eta_{int}^2}{qS_eP_r}I_{p1}\left[1 - \frac{\tau_2}{\tau_2-\tau^*}e^{-t/\tau_2} + \frac{\tau_3}{\tau_3+\tau^*}e^{t/\tau_3}\right.$$
$$\left. + \left(\frac{\tau_1}{\tau_1+\tau^*} - 1 + \frac{\tau_2}{\tau_2-\tau^*} - \frac{\tau_3}{\tau_3+\tau^*}\right)e^{-t/\tau^*}\right]. \qquad (16)$$

For $\tau_2 \gg \tau^*$ and $\tau_1 = \tau_3$ (which holds in our frequency range, as will be shown later),

$$L(t) = \frac{A\eta_{int}^2}{qS_eP_r}I_{p1}\left(1 - e^{-t/\tau_2} + \frac{\tau_3}{\tau_3+\tau^*}e^{t/\tau_3}\right). \qquad (17)$$

For $\tau_1, \tau_2, \tau_3 \gg \tau^*$,

$$L(t) = \frac{A\eta_{int}^2}{qS_eP_r}I_p(t). \qquad (18)$$

In the descending portion, the curve $I_p(t)$ is approximated as

$$I_p(t) = \frac{I_{p2}}{2}(e^{-t/\tau_4} + e^{-t/\tau_5}), \qquad (19)$$

where τ_4 and τ_5 are the decay time constants of $I_p(t)$.

In this case, the solution (8) of Eq. (6) is

$$N^*(t) = \frac{\sigma N\tau^*}{qS_e}\left\{\frac{I_{p2}}{2}\left(\frac{\tau_4}{\tau_4-\tau^*}e^{-t/\tau_4} + \frac{\tau_5}{\tau_5-\tau^*}e^{-t/\tau_5}\right)\right.$$
$$+ \left[I_{p1}\left(-\frac{\tau_2}{\tau_2-\tau^*}e^{-t_2/\tau_2} + \frac{\tau_3}{\tau_3+\tau^*}e^{t_2/\tau_3}\right.\right.$$
$$\left.+ \frac{\tau_1}{\tau_1+\tau^*} + \frac{\tau_2}{\tau_2-\tau^*} - \frac{\tau_3}{\tau_3+\tau^*}\right) \qquad (20)$$
$$\left.\left. - \frac{I_{p2}}{2}\left(\frac{\tau_4}{\tau_4-\tau^*} + \frac{\tau_5}{\tau_5-\tau^*}\right)\right]e^{-t/\tau^*}\right\},$$

where $t_2 = 1/4f$ is the time instant corresponding to point 2. The dependence $L(t)$ is given by

$$L(t) = \frac{A\eta_{int}^2}{qS_eP_r}\left\{\frac{I_{p2}}{2}\left(\frac{\tau_4}{\tau_4-\tau^*}e^{-t/\tau_4} + \frac{\tau_5}{\tau_5-\tau^*}e^{-t/\tau_5}\right)\right.$$
$$+ \left[I_{p1}\left(-\frac{\tau_2}{\tau_2-\tau^*}e^{-t_2/\tau_2} + \frac{\tau_3}{\tau_3+\tau^*}e^{t_2/\tau_3}\right.\right.$$
$$\left.+ \frac{\tau_1}{\tau_1+\tau^*} + \frac{\tau_2}{\tau_2-\tau^*} - \frac{\tau_3}{\tau_3+\tau^*}\right) \qquad (21)$$
$$\left.\left. - \frac{I_{p2}}{2}\left(\frac{\tau_4}{\tau_4-\tau^*} + \frac{\tau_5}{\tau_5-\tau^*}\right)\right]e^{-t/\tau^*}\right\}.$$
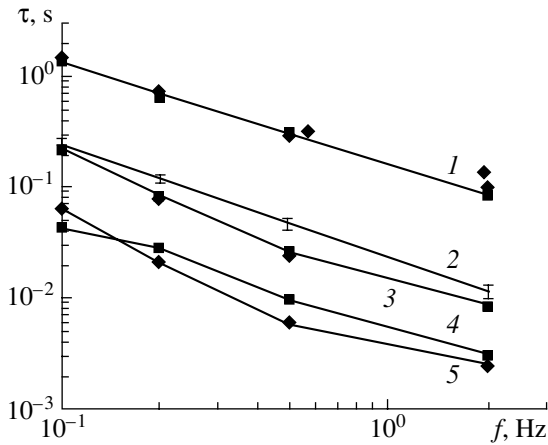
For $\tau_2 \gg \tau^*$ and $\tau_1 = \tau_3$,

$$L(t) = \frac{A\eta_{int}^2}{qS_eP_r}\left\{\frac{I_{p2}}{2}\left(\frac{\tau_4}{\tau_4-\tau^*}e^{-t/\tau_4} + \frac{\tau_5}{\tau_5-\tau^*}e^{-t/\tau_5}\right)\right.$$
$$+ \left[I_{p1}\left(1 - e^{-t_2/\tau_2} + \frac{\tau_3}{\tau_3-\tau^*}e^{t_2/\tau_3}\right)\right. \qquad (22)$$
$$\left.\left. - \frac{I_{p2}}{2}\left(\frac{\tau_4}{\tau_4-\tau^*} + \frac{\tau_5}{\tau_5-\tau^*}\right)\right]e^{-t/\tau^*}\right\}.$$

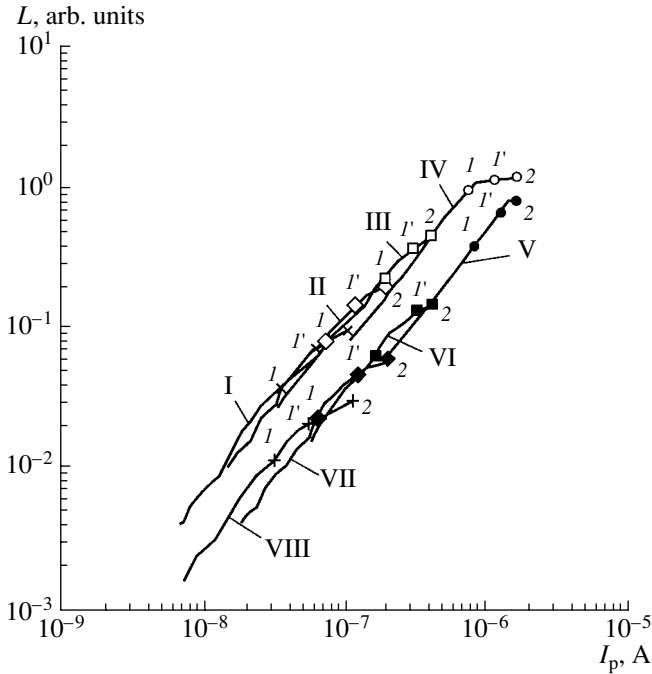If $\tau_3, \tau_4, \tau_5 \gg \tau^*$, then, in view of (14),

$$L(t) \approx \frac{A\eta_{int}^2}{qS_eP_r}\frac{I_{p2}}{2}(e^{-t/\tau_4} + e^{-t/\tau_5}) = \frac{A\eta_{int}^2}{qS_eP_r}I_{p2}(t). \quad (23)$$

The table lists the time constants determined from the experimental and calculated data and characterizing the EL kinetics. For the conditions ($+Al$, $f = 2$ Hz), the

Fig. 3. $\tau_1-\tau_5$ vs. frequency. (*1*) τ_2, (*2*) τ_1 and τ_3 for the conditions (–Al) and (+Al), (*3*) τ_5, (*4*) τ_4 (–Al), and (*5*) τ_4 (+Al). (*1, 3*) ■ (+Al) and ◆ (–Al).

Fig. 4. $L(I_p)$ for the continuous excitation. $f =$ (I, VIII) 0.1, (II, VII) 0.2, (III, VI) 0.5, and (IV, V) 2. (I–IV) +Al and (V–VIII) –Al; ■ (–Al) and ◆ (+Al).

values of τ_2 and τ_3 are given for the curve $I_p(t)$, while for the curve $L(t)$, $\tau_2 \approx 2$ ms and $\tau_3 \approx 560$ ms. Here, the function decays starting from the additional portion with a time constant τ of 10–40 ms, which is 3 ms wide. The descending portion with a similar value of τ, $\tau \approx 14$ ms, is also typical of higher excitation frequencies, $f = 10$ and 50 Hz [2]. It is observed at $L \geq 1.1$ as well.

From the table, it follows that the time constants τ_1 and τ_3 are coincident with each other at all the frequencies within the overall (measurement + calculation + approximation) accuracy ($\pm 10\%$). This may indicate

that the processes responsible for the corresponding portions of the curves $I_p(t)$ and $L(t)$ are similar in nature. The values of τ_1, τ_2, and τ_3 linearly decrease with rising frequency with nearly the same slopes (Fig. 3). The time constants τ_4 and τ_5 apparently characterize the capture of free carriers by bulk and surface traps. The detailed explanation of the curves $I_p(t)$, as well as the of frequency dependences of τ_1, τ_2, τ_3, τ_4, and τ_5, calls for special investigation into charge transfer in the phosphor layer.

The results obtained can be interpreted as follows. As follows from Figs. 2b and 4, for all the frequencies, the dependences $L(I_p)$ have the same slope of the initial portion within the measurement accuracy, in accordance with formulas (12) and (13). In the second portion, for all the frequencies, the dependence of L on I_p weakens, which means that the proportionality coefficient in formulas (17) and (18) for $L(I_p)$ decreases (compared to the initial portion). The luminances L_1 and L_2 and the currents I_{p1} and I_{p2} at points *1* and *2* (Figs. 2a, 2e) linearly depend on frequency f (Fig. 5), while the field F_p varies with frequency only slightly (Figs. 2b, 2d). This can be explained by the nearly linear dependence of I_p on the rate of change of the voltage [see (2)] in this case.

According to (5), (13), and (18), the change in the slope of the $L(I_p)$ curves in passing to the second portion (Figs. 2b, 4) can be explained by a decrease in the cross section σ of impact excitation; in the probability P_r of radiative transitions; or in the phosphor effective thickness d_p, within which the impact excitation of emitting centers takes place.

The reasons for such a variation of σ or P_r can be (i) a change in the Mn^{2+} ion excitation mechanism (from direct impact excitation to resonance excitation involving neighboring atoms or ZnS lattice defects [9]; this seems to be hardly probable, because the excitation level at ultralow frequencies is low for all the conditions except (+Al, 2 Hz); (ii) the excitation of emitting complexes (along with ordinary centers) that combine Mn^{2+} ions and ZnS lattice defects, such as sulfur vacancies [10], and have other impact excitation cross sections σ, relaxation times τ^*, and radiative relaxation probabilities P_r. However, these complexes form at elevated Mn concentrations (2.2 wt %), i.e., when the phosphor layer is deposited at a substrate temperature of 200°C without subsequent annealing, and their room-temperature EL spectra do not depend on the exciting voltage [10].

The effective thickness d_p may decrease because of space charges left at the anode and the cathode after the preceding half-cycle. Also, d_p shrinks because of the occurrence and expansion of space charge regions when deep donors ionize at the anode. This decreases the field at the anode and gives rise to the negative differential resistance region in the current–voltage char-
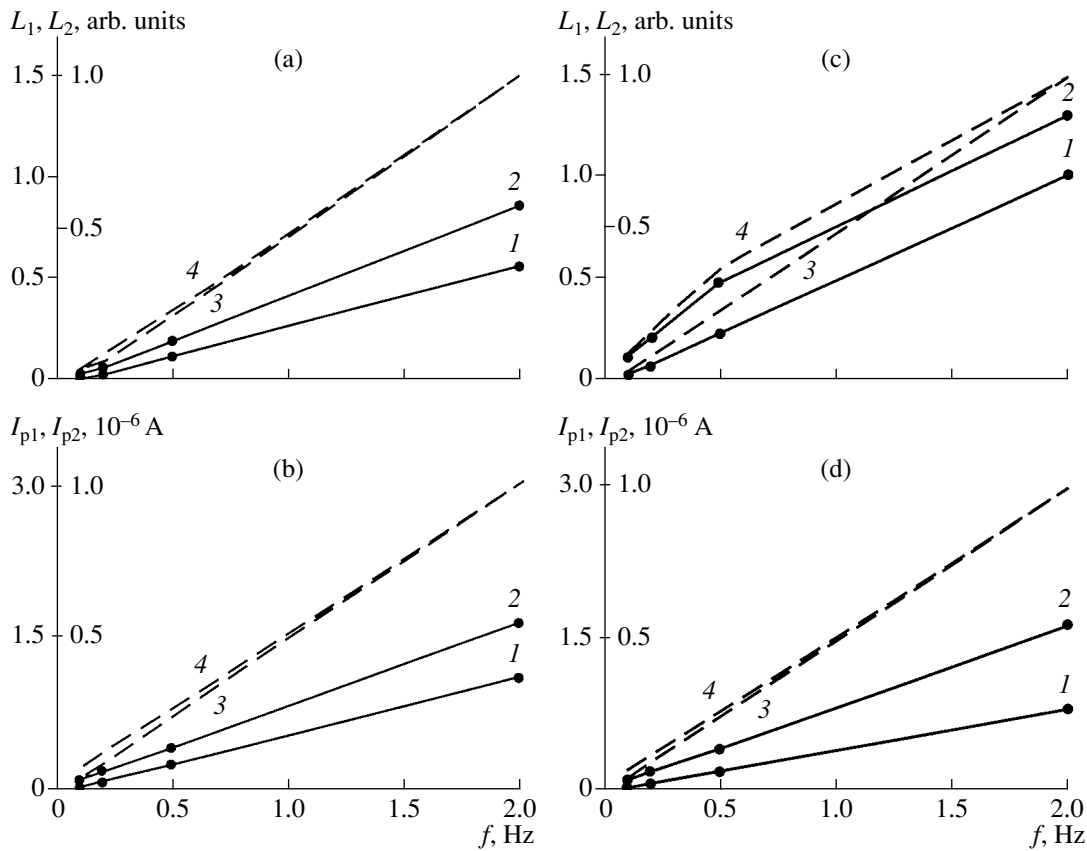
Fig. 5. (a, c) Luminances L_1, L_2 and (b, d) currents I_{p1}, I_{p2} vs. frequency f. (*1*) L_1, I_{p1}; (*2*) L_2, I_{p2}; (*3*) L_1, I_{p1} and (*4*) L_2, I_{p2} normalized to the associated values at $f = 2$ Hz.

acteristic (Fig. 2d). Finally, the effective thickness decreases because of the formation of space charges when deep acceptors ionize at the cathode in the single-shot excitation mode. In this case, the cathode field decreases [3]. As a result, the free electron energy at the anode may be insufficient for the ionization of Mn^{2+} emitting centers. The fact that τ_1 and τ_3 equal each other supports our previous supposition. Namely, once the deep donors have ionized, the field of the resulting space charge at the anode has expanded, and the mean field in the phosphor has decreased (Fig. 2f), the tunnel emission current from the cathode is first limited in the second portion. Subsequently, as the external field grows, this current begins to rise again (in the second portion) with the same time constant ($\tau_1 = \tau_3$). In this situation, the effect of the residual space charges and their fields is appreciably reduced in the single-shot excitation mode with $T_s = 100$ s (Fig. 2b) [3] and the effective thickness d_p is the largest.

At the same time, under the conditions (+Al, $f = 2$ Hz), the slope of the curve $L(I_p)$ in the second portion decreases much sharper (Fig. 2b), this decrease being observed for higher levels of L (close to those at higher frequencies, 10 and 50 Hz). This can be explained by

Table

f, Hz	Conditions	τ_1, ms	τ_2, ms	τ_3, ms	τ_4, ms	τ_5, ms
0.1	−Al	246	1440	230	44	220
	+Al	240	1420	220	62	236
0.2	−Al	127	580	155	29	88
	+Al	118	750	120	21	84
0.5	−Al	57	330	55	10	28
	+Al	47.5	320	45	4	24
2	−Al	14	93	11	3.3	9.1
	+Al	13.5	108	12	2.6	9.8
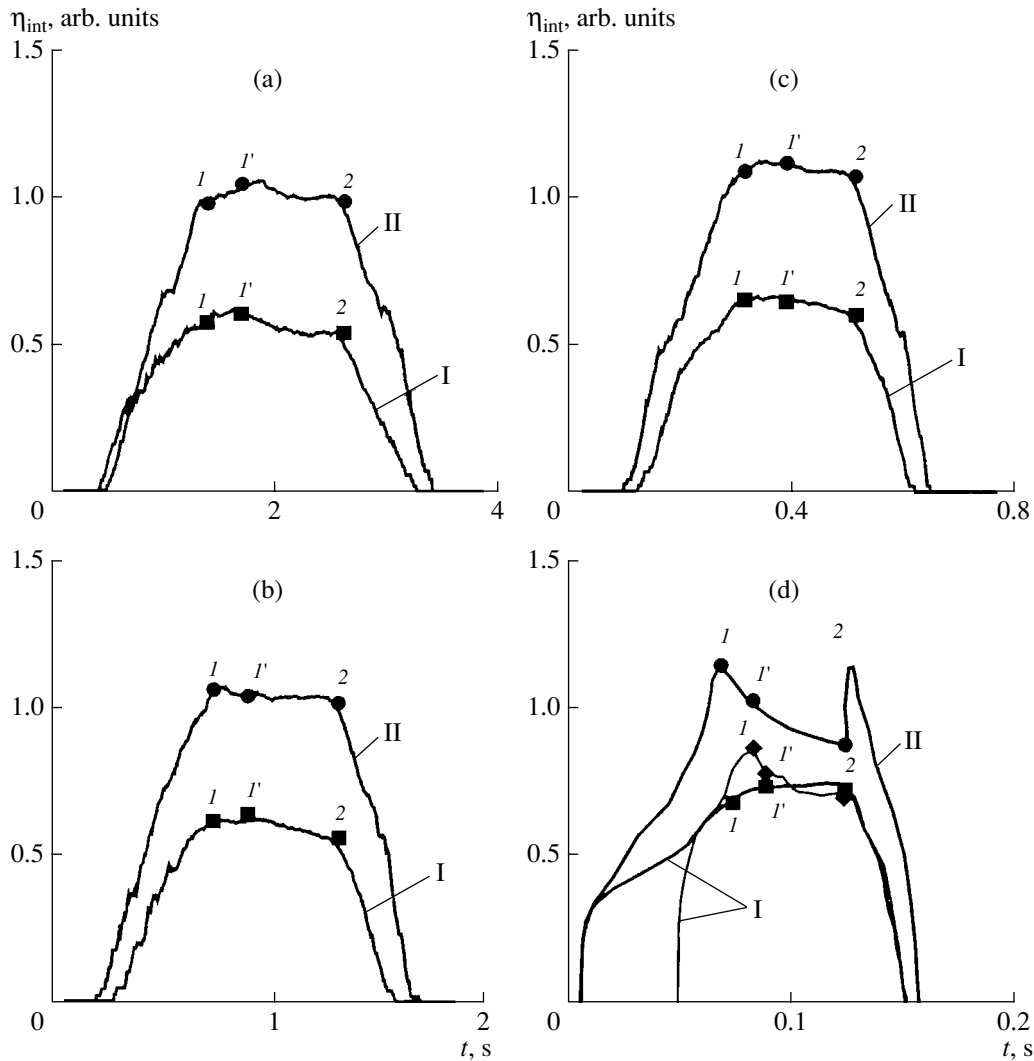
Fig. 6. $\eta_{int}(t)$ for $f =$ (a) 0.1, (b) 0.2, (c) 0.5, and (d) 2 Hz. Thick continuous lines, continuous excitation; thin continuous lines, single-shot excitation. I, –Al; II, +Al.

the resonant interaction of excited Mn^{2+} ions with each other and/or with surrounding atoms or ZnS lattice defects [2].

The equal slopes of the normalized dependences of L_1, L_2, I_{p1}, and I_{p2} on f in all the cases except for the conditions (+Al, $f = 2$ Hz) for the curve $L(t)$ (Figs. 5a–5c) suggest that the same mechanism is responsible for the rise in I_p in the initial (first) portion and at the end of the second one (Fig. 2f), and also validate the use of simplified expressions (13) and (18) for finding the dependence $L(t)$.

The increased slope of the curves $L(I_p)$ in the first portion for the single-shot excitation and also when the frequency f declines in the continuous excitation mode (Figs. 2b, 4) is explained by the large relaxation times of the space charges at both electrodes, the decrease in the fields of these charges [3], and the increase in the field in the bulk with the mean field F_p in the phosphor

remaining nearly the same (Fig. 2c). The last effect causes the energy of electrons to increase and the ionization efficiency N_1 to rise. The superlinear initial region in the curve $L(I_p)$ (Fig. 2b) is explained by the same reasons.

Using expressions (5), (12), (13), (16)–(18), and (21)–(23), we can estimate the instantaneous values of the internal quantum yield, η_{int}, and the external quantum yield, $\eta_{ext} = K_0\eta_{int}$, both are most important parameters of TF ELDs. From formulas (5), (13), (18), and (23), the instantaneous value $\eta_{int}(t)$ for the entire curves $L(t)$ and $I_p(t)$ is

$$\eta_{int}(t) = \sqrt{\frac{qS_e}{A}\frac{L(t)}{I_p(t)}}. \qquad (24)$$

The curves $\eta_{int}(t)$ constructed with the data tabulated and formulas (13), (16)–(18), and (21)–(24) (Fig. 6)

sharply rise in the first portion of $L(t)$ and then smoothly decline after point $1'$ in the second portion for $f = 0.1$, 0.2, and 0.5 Hz. Under the conditions ($-$Al, $f = 2$ Hz), η_{int} slightly increases after point $1'$. Note that the decline in the second portion is much greater for the single-shot excitation, although the absolute values of η_{int} at the beginning of the second portion exceed those for the continuous excitation (Fig. 6d). As the frequency f increases, the maximal values of η_{int} slightly increase (Fig. 6). Under the conditions ($+$Al, $f = 2$ Hz) (Fig. 6d), $\eta_{int}(t)$ sharply drops in the second portion and the additional peak appears in the descending branches of $L(t)$ and $I_p(t)$. Such behavior of $\eta_{int}(t)$ can be explained as follows. In the general form, η_{int} and the physical parameters of the phosphor layer are related as

$$\eta_{int}(t) = N_1(t)P_r(t) = \sigma(t)N(x, t)d_P(t)P_r(t), \quad (25)$$

where $N(x, t)$ is the distribution of the emitting centers across the phosphor layer.

The fast rise in η_{int} in the first portion up to point 1 (Fig. 6) is associated with the growth of the mean field $F_p(t)$ in the phosphor (Fig. 2f), the energy of accelerated electrons, and the number $N_1(t)$ of emitting centers excited by an electron. In the interval between points 1 and $1'$, η_{int} varies insignificantly, since the parameters of Mn^{2+} ion excitation remain fairly constant and the mean field $F_p(t)$ varies only slightly (Fig. 2f). From point 1 to point 2, the decline in $\eta_{int}(t)$ can be explained, according to (25), by a decrease in the effective value of $d_p(t)$ due to the field redistribution in the phosphor. In the interval between the voltage pulses in the single-shot excitation, the space charges considerably relax; therefore, upon the subsequent ionization of deep centers, the scattering of the carrier energy by these centers is more intense. As a result, η_{int} in the second portion sharply drops (Fig. 6d). The extra peak appearing in the descending portion under the conditions ($+$Al, $f = 2$ Hz) (Fig. 6d) is due to the slower decay of the luminance $L(t)$ compared with the current $I_p(t)$. This is because the curve $L(t)$ has the extra decay portion with τ of about several milliseconds, which is presumably related to another mechanism of Mn^{2+} center excitation and, accordingly, with the changed parameters $N_1(t)$, $\sigma(t)$, $N(x, t)$, and $P_r(t)$ in formula (25). For the same reason, in this case, η_{int} in the second portion drops sharper than under the conditions ($+$Al, $f = 0.1$, 0.2, 0.5 Hz).

The asymmetry of $L(t)$ and $\eta_{int}(t)$ (Figs. 1, 2a, 6) at various polarities of the excitation voltage [conditions ($-$Al) and ($+$Al)] stems from the nonuniform distribution of Mn^{2+} centers across the phosphor: it is higher at the upper electrode, hence, the higher values of $L(t)$ and $\eta_{int}(t)$ under the conditions ($+$Al) [3, 4].

It should be noted that the decay of $\eta_{int}(t)$ down to zero (Fig. 6) occurs in parallel with the same decrease in $I_p(t)$ (Fig. 2e). The mean field F_p in the phosphor, while being reduced roughly to 10^8 V/m (Fig. 2f), nev-

ertheless provides the ionization of the Mn^{2+} emitting centers. This means that kinetic equation (6), where the generation term $\alpha(t)[N - N^*(t)]$ is present, must be solved for the descending branch of $L(t)$.

Thus, our study of the instantaneous luminance and other physical characteristics of TF ELDs showed the presence of fast and slow portions in the time dependences of the instantaneous luminance and current through the phosphor layer. The former is characterized by the nearly linear current dependence of the luminance and by a fast rise in the internal quantum yield presumably because of an increase in the number of emitting centers excited by an electron passing through the phosphor layer when the mean field in this layer grows. In the slow portion, the current and the luminance grow with a lower rate and the current dependence of the luminance weakens, which is accompanied by a decrease in the internal quantum yield. It appears that, at relatively low excitation levels (currents I_p), the phosphor effective thickness, within which the impact excitation of Mn^{2+} emitting centers takes place, changes. The excitation is associated with the ionization of deep donors and acceptors. The ionized donors and acceptors, along with the field of residual space charges produced during the preceding half-cycle of the exciting voltage, diminish the local fields at the electrodes. When the current I_p exceeds some critical level, the direct impact excitation of Mn^{2+} centers is likely to change to the resonant interaction of these centers with each other and/or with neighboring atoms or ZnS lattice defects.

REFERENCES

1. *Electroluminescent Sources of Light*, Ed. by I. K. Vereshchagin (Énergoatomizdat, Moscow, 1990).

2. N. T. Gurin, O. Yu. Sabitov, A. V. Shlyapin, and A. V. Yudenkov, Pis'ma Zh. Tekh. Fiz. **27** (4), 12 (2001) [Tech. Phys. Lett. **27**, 138 (2001)].

3. N. T. Gurin, O. Yu. Sabitov, and A. V. Shlyapin, Zh. Tekh. Fiz. **71** (8), 48 (2001) [Tech. Phys. **46**, 977 (2001)].

4. N. T. Gurin, A. V. Shlyapin, and O. Yu. Sabitov, Zh. Tekh. Fiz. **71** (3), 72 (2001) [Tech. Phys. **46**, 342 (2001)].

5. N. T. Gurin, A. V. Shlyapin, and O. Yu. Sabitov, in *Proceedings of the International Conference "Optics of Semiconductors," Ul'yanovsk, 2000*, p. 80.

6. N. T. Gurin and O. Yu. Sabitov, Zh. Tekh. Fiz. **69** (5), 65 (1999) [Tech. Phys. **44**, 537 (1999)].

7. N. A. Vlasenko, Yu. V. Kopytko, and V. S. Pekar, Phys. Status Solidi A **81**, 661 (1984).

8. G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers* (McGraw-Hill, New York, 1968; Nauka, Moscow, 1974).

9. A. A. Zhigal'skiĭ, E. V. Nefedtsev, and P. E. Troyan, Izv. Vyssh. Uchebn. Zaved., Fiz., No. 2, 37 (1995).

10. A. N. Georgobiani, A. N. Gruzintsev, Xu Xurong, and Lou Zidong, Neorg. Mater. **35**, 1429 (1999).

Translated by V. Isaakyan

OPTICS,
QUANTUM ELECTRONICS

Combined Photoreflectance/Photoluminescence Studies of the Stability of Semiconductor Surface Passivation

R. V. Kuz'menko*, A. V. Ganzha*, É. P. Domashevskaya*, and P. V. Ryasnoĭ**

* Voronezh State University,
Universitetskaya pl. 1, Voronezh, 394693 Russia
e-mail: phssd2@main.vsu.ru (Kusmenko)
** Voronezh State Academy of Architecture and Civil Engineering,
Voronezh, 394006 Russia
Received March 30, 2001

Abstract—Combined photoreflectance/photoluminescence measurements are proposed for the study of the stability of semiconductor surface passivation. With such an approach, laser action on both charged and recombination-active electron states can be investigated. The efficiency of the method proposed is demonstrated with GaAs substrates passivated by selenium. © *2002 MAIK "Nauka/Interperiodica".*

The stationary laser irradiation of the semiconductor surface by photons with an energy exceeding the energy E_0 of the direct optical transition activates desorption, chemisorption, and reoxidation processes, as well as the generation of defects on the surface and in the near-surface area. The mechanisms of the corresponding reactions for the naturally oxidized surfaces of III–V semiconductors have been discussed in details in [1–6]. However, the associated models consider only changes in the integral intensity of luminescence. The works cited above studied only the effect of laser radiation on recombination-active electron states. The combination of luminescence excitation spectroscopy with photoreflectance measurements based on the periodic electromodulation of reflected light by the near-surface area of a semiconductor (the irradiation of the surface by a laser light with a photon energy higher than the band gap leads to the modulation of the surface electric self-field) makes it possible to gain additional information on the surface electric field [7, 8]. Thus, the combination of photoreflectance and photoluminescence measurements allows one to study the effect of laser irradiation on both recombination-active and charged states at the surface. An additional advantage of such an approach is the possibility of making measurements with the same setup described in [9].

In this work, our method is demonstrated by studying the stability of the GaAs surface passivated by Se. Samples used were *n*-GaAs(100) substrates annealed in a Se or Se + As vapor at the temperature 660–680 K for 5–45 min. The annealing was performed at a Se vapor pressure ranging from 0.1 to 1 Pa. The processing of a GaAs substrate in a chalcogen-containing medium at these temperatures results in the heterovalent substitution reaction and the formation of the Ga_2Se_3 surface pseudomorphic layer. It has been demonstrated [10–12]

that the Ga_2Se_3/GaAs interface exhibits a lower density of surface states in comparison with the naturally oxidized surface. However, the stability of this phenomenon has not been studied.

All the measurements were carried out in air at room temperature with the setup described in [9]. A modulated laser beam (He–Ne laser, $\lambda = 632.8$ nm) is focused on the surface area measuring 0.1×0.1 mm, and the E_0 spectrum of GaAs photoreflectance is recorded (the pseudomorphic Ga_2Se_3 layer is transparent at 632.8 nm). It is known that high-power laser irradiation may induce photostimulated reactions near the GaAs surface. Therefore, the laser power density was taken such that the integral intensity of luminescence remained nearly unchanged in the course of the measurements. After recording the first photoreflectance spectrum, we increased the power density, stopped the modulation of the laser beam, and recorded the time dependence of the photoluminescence (PL) intensity $I^{PL}(t)$. After a time depending on the PL signal curve shape, the experiment was interrupted, the irradiation power density was decreased, and the next measurement of the photoreflectance spectrum was performed. Each time after recording the spectrum, the laser intensity was increased to the previous level in such a way as to avoid discontinuities in the curve of the integral photoluminescence intensity (IPLI), and the curve $I^{PL}(t)$ was taken again.

We empirically found that the IPLI signal from our samples varies insignificantly for several hours when the power density is $L \sim 1$ mW/cm². On the contrary, the signal markedly changes within several seconds at $L > 100$ W/cm². Therefore, we used the above values for the measurements of the photoreflectance and IPLI, respectively. The times of recording the photoreflec-

tance spectrum and the curve $I^{PL}(t)$ were 600 and 1500 s, respectively.

Figure 1 shows a typical result of the combined measurements for our samples. The decay (upper) curve, obtained at $L \sim 2500$ W/cm^2, represents the first type of the time dependence of the PL intensity. The ten-fold decrease in the power density affects the initial part of the dependence: the increase in the signal to a certain maximum level is followed by the decay of the first type (the lower curve in Fig. 1). A further decrease in the laser intensity causes an extension of the rising portion, so that only the slow growth of the IPLI is observed during the measurements at $L \sim 1$ W/cm^2.

The simulation of the $I^{PL}(t)$ dependences showed that the curve of the first type is well described by a single damped exponential. The curve of the second type can be approximated by the sum of rising and damped exponentials:

$$I^{PL}(t) = I^{PL}(t = \infty)\left[C + A\exp\left(-\frac{t}{t_1}\right) - B\exp\left(-\frac{t}{t_2}\right)\right],$$

where t_1 and t_2 are the characteristic time constants of the laser-induced processes.

The photoreflectance spectra have a mid-field shape [13]. In the mid-field case, the spectrum consists of a central peak in the vicinity of the energy of the transition and high-energy Franz–Keldysh oscillations (Fig. 2). The theory of photoreflection predicts that, when the photoreflectance spectra are measured near the transition E_0 for a three-dimensional critical point, the Franz–Keldysh oscillations can be approximated by the asymptotic formula [14, 15]

$$\frac{\Delta R}{R}(\hbar\omega) \approx \mathrm{Re}\left(\frac{1}{\hbar\omega - E_0}\frac{1}{(\hbar\omega)^2}\exp\left(-\frac{\sqrt{\hbar\omega - E_0}\,\Gamma}{(\hbar\Omega)^{3/2}}\right)\right.$$
$$\left. \times \cos\left[\frac{2}{3}\left(\frac{\hbar\omega - E_0}{\hbar\Omega}\right)^{3/2} + \Theta_0\right]\right).$$

Here, $\hbar\Omega$ is the electrooptical energy

$$\hbar\Omega = \left[\frac{e^2 F^2 \hbar^2}{8\mu_\parallel}\right]^{1/3},$$

where $\mu_\parallel$ is the reduced effective electron–hole mass in the direction of the electric field and e is the elementary charge, which exceeds the phenomenological energy Γ of the spectral broadening of the transition.

Thus, one can determine the value of the electrooptical energy and, consequently, the surface electric field intensity (if the value of $\hbar\Omega$ is known) from the period of the Franz–Keldysh oscillations. Knowing the surface electric field intensity and the equilibrium concentration n of charge carriers, one can obtain the surface potential φ and the density of charged surface states Q_{ss}
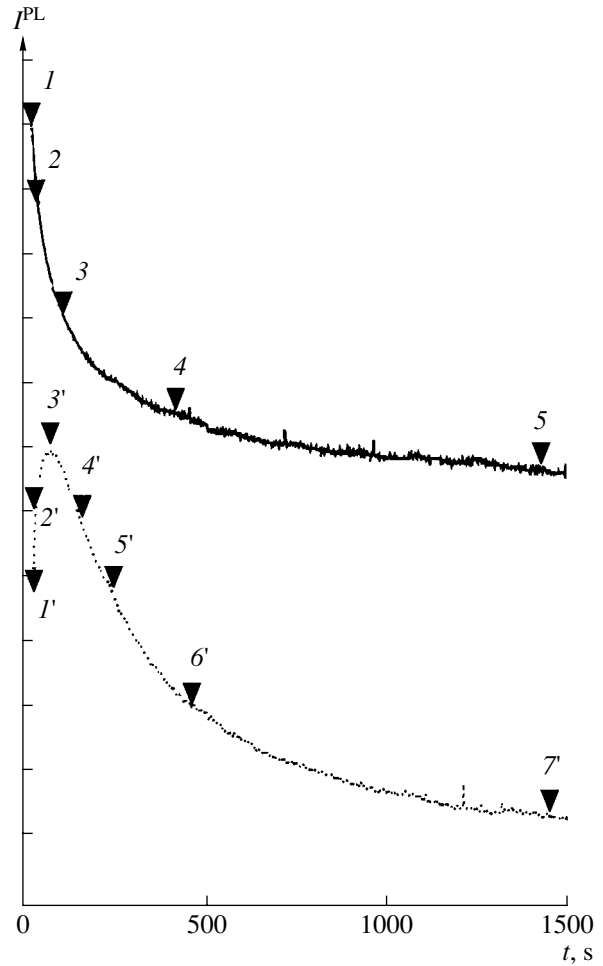

Fig. 1. Results of the combined photoreflectance/photoluminescence measurements at $L = 2500$ (upper curve) and 250 W/cm^2 (lower curve). The electric field intensities are (*1*) 3.01×10^6, (*2*) 2.90×10^6, (*3*) 2.89×10^6, (*4*) 2.89×10^6, (*5*) 2.89×10^6, (*1'*) 3.12×10^6, (*2'*) 3.04×10^6, (*3'*) 2.99×10^6, (*4'*) 2.97×10^6, (*5'*) 2.95×10^6, (*6'*) 2.95×10^6, and (*7'*) 2.95×10^6 V/m.

by the formula [7]

$$\varphi = \frac{\varepsilon\varepsilon_0 F^2}{2en} = \frac{Q_{ss}^2}{2\varepsilon\varepsilon_0 en}.$$

The quantitative analysis of the photoreflectance spectra taken prior to the photostimulated reactions yields the following mean values of the parameters: $F \approx 3 \times 10^6$ V/m, $e\varphi \approx 0.3$ eV, and $Q_{ss}/e \approx 2 \times 10^{11}$ cm^{-2}. Figure 1 shows the results of the photoreflectance measurements. The initial portion of the upper curve exhibits the decrease in the surface electric field, which, however, saturates in a short time. It is clear from the lower curve that the decrease in the electric field intensity is related to the rising exponential. Thus, the photoreflectance measurements unambiguously confirm that the
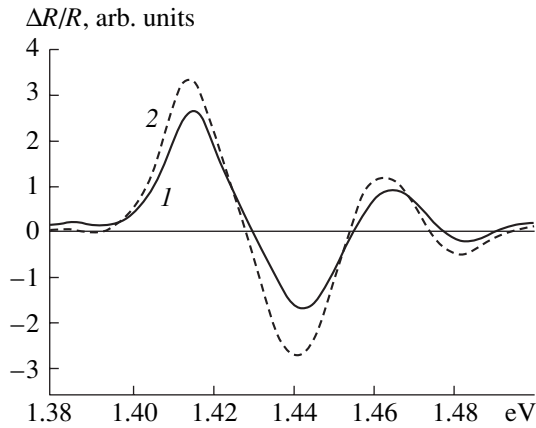
Fig. 2. Experimental E_0 spectra of the photoreflectance for GaAs samples passivated by Se. The difference in the Franz–Keldysh oscillation period implies the difference in the surface electric field intensities: (*1*) $F_1 = 3.19 \times 10^6$ V/m and (*2*) $F_2 = 2.91 \times 10^6$ V/m.

first type of the time dependence of the IPLI also has the growing component with the small time constant.

Our dependences of the IPLI can be interpreted within the model assuming that photoinduced chemical reactions on the GaAs surface and the generation of "non-radiation-induced" defects in the near-surface area proceed simultaneously. The laser-induced generation of non-radiation-induced defects in the near-surface area of GaAs has been reported in [1–6].

We assume that the surface of the samples studied is inhomogeneous and has the areas of both stable and incomplete passivation. In the former areas, the heterovalent substitution reaction is completed, and they can be assigned the minimum density of states. In the latter, the density of states is increased, since the intrinsic states of the semiconductor surface add up with Se-induced states. The existence of the areas with different surface state densities was confirmed by taking the high-resolution (10×10 μm) photoreflectance spectra from various sites of the passivated substrates. It was demonstrated that the surface electric field intensity varies over the passivated surface, unlike the unprocessed substrate.

The laser-induced transformation of the areas of incomplete passivation into the areas of stable passivation decreases the rate of surface recombination and enhances the luminescence. On the contrary, the generation of the non-radiation-induced defects in the near-surface area causes the degradation of the signal. The fact that these defects appear in the near-surface area,

rather than on the surface of the semiconductor, follows from the constant value of the surface electric field intensity in the decay portion of the curve, whereas the incomplete-to-stable passivation transition is characterized by the variable surface field. Based on this fact, one can assume that the states due to this transition must either appear inside the band gap near the band edges or eliminate the states at the midgap.

Our study of the stability of the selenium-passivated GaAs surface showed that combined photoreflectance/photoluminescence measurements are an effective way for investigating the electronic properties of the semiconductor surface. The generation of non-radiation-induced defects in the near-surface area of the semiconductor is responsible for the decay of the photoluminescence intensity. The photostimulated reactions on the passivated surface enhance the photoluminescence signal.

REFERENCES

1. M. Y. A. Raja, S. R. J. Brueck, M. Osinski, and J. McInerney, Appl. Phys. Lett. **52**, 625 (1988).
2. T. Suzuki and M. Ogawa, Appl. Phys. Lett. **31**, 473 (1977).
3. N. M. Haegel and A. Winnacker, Appl. Phys. A **A42**, 233 (1987).
4. D. Guidotti, E. Hasan, H.-J. Hovel, and M. Albert, Appl. Phys. Lett. **50**, 912 (1987).
5. J. Ogawa, K. Tamamura, K. Akimoto, and Y. Mory, Appl. Phys. Lett. **51**, 1949 (1987).
6. D. Guidotti, E. Hasan, H.-J. Hovel, and M. Albert, Nuovo Cimento D **11**, 583 (1989).
7. R. B. Darling, Phys. Rev. B **43**, 4071 (1991).
8. N. Bottka, D. K. Gaskill, R. S. Sillmon, *et al.*, J. Electron. Mater. **17**, 161 (1988).
9. J. Schreiber, S. Hildebrandt, W. Kircher, and T. Richter, Mater. Sci. Eng. B **9**, 31 (1991).
10. B. I. Sysoev, V. F. Antyukhin, V. D. Strygin, and V. N. Morgunov, Zh. Tekh. Fiz. **56**, 913 (1986) [Sov. Phys. Tech. Phys. **31**, 554 (1986)].
11. B. I. Sysoev, N. N. Bezryadin, G. I. Kotov, and V. D. Strygin, Fiz. Tekh. Poluprovodn. (St. Petersburg) **27**, 131 (1993) [Semiconductors **27**, 69 (1993)].
12. N. N. Bezryadin, É. P. Domashevskaya, I. N. Arsent'ev, *et al.*, Fiz. Tekh. Poluprovodn. (St. Petersburg) **33**, 719 (1999) [Semiconductors **33**, 665 (1999)].
13. D. E. Aspnes, Surf. Sci. **37**, 418 (1973).
14. D. E. Aspnes, Phys. Rev. B **10**, 4228 (1974).
15. D. E. Aspnes and A. A. Studna, Phys. Rev. B **7**, 4605 (1973).

Translated by A. Chikishev

═══════════════ **RADIOPHYSICS** ═══════════════

Broadband Emission from a Relativistic Electron Bunch in a Semi-Infinite Waveguide

V. A. Balakirev, I. N. Onishchenko, D. Yu. Sidorenko, and G. V. Sotnikov

Kharkov Institute for Physics and Technology, National Science Center,
Akademicheskaya ul. 1, Kharkov, 61108 Ukraine
e-mail: sotnikov@kipt.kharkov.ua
Received March 27, 2001; in final form, July 6, 2001

Abstract—A study is made of the excitation of a transition radiation pulse during the injection of a charged particle bunch through the end metal wall into a semi-infinite cylindrical waveguide. Exact analytic expressions for the fields of a thin ring-shaped bunch are obtained in terms of the Lommel functions of two variables. The energy efficiency, power, and spectrum of radiation emitted from a finite-size charged bunch in a vacuum waveguide are calculated numerically with allowance for the multimode nature of the excited field. It is shown that, under certain conditions, the bunch can generate a short, high-intensity electromagnetic pulse with a broad frequency spectrum. The effect of various parameters of the charged bunch–waveguide system (such as the bunch current, bunch duration, and waveguide radius) on the generation efficiency of a transition radiation pulse is investigated. © 2002 MAIK "Nauka/Interperiodica".

INTRODUCTION

At present, much attention is focused on the problem of the generation of short intense electromagnetic pulses such that the width of their frequency spectra is comparable to their mean frequency. Such pulses are called "ultrawide-band pulses" (UWBPs) [1, 2]. The current interest in the problem of pulsed emission of high-power electromagnetic signals stems primarily from their application in ultrawide-band radio location. UWBPs have also found use in, e.g., solving ecological problems, detection of objects under the Earth's surface, geologic exploration of the marine shelf, and the goal of ensuring flight and road safety [1–6].

Traditional methods for generating short UWBPs are based on the use of ultrawide-band antennas (TEM-horns, spiral and biconical antennas, and others [4]). However, UWBPs can also be generated by intense pulsed relativistic electron beams (IREBs). Short (and ultrashort) IREBs with durations of 1–100 ps, energies of 0.5–1 MeV, and peak currents of 1–100 kA can be either produced by converting a continuous IREB into a sequence of electron pulses (modulated beams) [7] or generated directly by high-current devices (explosive-emission diodes).

The most efficient way of generating electromagnetic UWBPs is to exploit nonresonant (impact) mechanisms for the production of IREBs. In this case, the duration of the excited UWBP is determined by the duration of a pulsed IREB and the power of the emitted UWBP is determined by the beam current and energy. The mechanism for the impact excitation of a UWBP by a relativistic electron beam during the charging of a whip antenna was studied experimentally by Gaponenko *et al.* [8]. Balakirev *et al.* [9] reported the results of experimental investigations on the emission of UWBPs by a TEM-horn antenna during its impact excitation by an IREB. High-power UWBPs can be generated not only in direct processes of the excitation of ultrawide-band antennas by IREBs but also due to the effect of coherent transition radiation [10].

The transition radiation from an individual charged particle is one of the fundamental elementary radiation processes [11]. This effect occurs when a charged particle moves through an electrically nonuniform medium. The emission of radiation by a charge propagating in such a medium can be interpreted as follows. The spatial electromagnetic field distribution of a moving charged particle is determined by both its velocity and the dielectric permittivity (magnetic permeability) of the medium through which it propagates. Even in the case of uniform straight-line motion in an electrically nonuniform medium, the particle emits radiation, resulting in the spatial redistribution of its electromagnetic field. The situation with a particle undergoing accelerated motion in a uniform medium is similar; the only difference is that the electromagnetic field of the particle is redistributed because of the change in the particle velocity. The transition radiation effect can be greatly enhanced by using charged particle bunches [12]. The radiation fields of the particles in a bunch whose dimensions are substantially smaller than the wavelength of the emitted radiation are coherent; as a result, the radiation from a bunch of N particles is N^2 times more intense than that from one particle. Because of this effect, electron bunches provide an efficient means of generating intense electromagnetic transition radiation.

Here, we theoretically investigate the excitation of ultrawide-band transition radiation during the injection of a short-pulse IREB into a semi-infinite circular-cross-section waveguide whose end wall (through which the beam is injected) is short-circuited by a conducting diaphragm. In such a waveguide, a distinguishing feature of the transition radiation is a large dispersion of electromagnetic waves, so that the shape of a UWBP of transition radiation propagating in a waveguide will deform permanently.

THE FIELD OF A THIN-WALLED RING-SHAPED BUNCH

We consider a semi-infinite ($0 \le z < \infty$) cylindrical metal waveguide of radius b that is filled with a homogeneous dielectric with permittivity ε. The waveguide end ($z = 0$) is short-circuited by a metal wall transparent to relativistic electrons. An axisymmetric monoenergetic electron bunch is injected through the metal wall; the bunch then moves with a constant velocity $v_0 < c/\sqrt{\varepsilon}$ along the symmetry axis of the waveguide (the z axis). We neglect the influence of a narrow vacuum drift channel on the electrodynamics of the waveguide system and, in order to simplify the calculations, assume that the dielectric fills the entire waveguide. Under these assumptions, we will obtain an exact solution in the important case of a vacuum waveguide ($\varepsilon = 1$).

We start by determining the field of an infinitely short and infinitely thin charged ring with the charge density

$$\rho = -\frac{eN}{2\pi r_0 v_0}\delta(r - r_0)\delta(t - t_0 - z/v_0),$$

where e is the charge of an electron, N is the number of electrons in the ring, v_0 is the ring velocity, r_0 is the ring radius, and t_0 is the time at which the ring enters the waveguide.

Solving Fourier-transformed Maxwell's equations with allowance for the boundary condition $E_r = 0$ at the end metal wall, we obtain the following expression for the radial electric field of an axisymmetric E-wave:

$$E_r(t, r, z, t_0, r_0) = \frac{2Ne}{\pi b\varepsilon v_0}$$

$$\times \sum_n \frac{\omega_{0n}^2 J_0(\lambda_n r_0/b) J_1(\lambda_n r/b)}{\lambda_n J_1^2(\lambda_n)}\{I_{2n} - I_{1n}\}, \tag{1}$$

$$I_{1n} = \int_{-\infty}^{+\infty} d\omega \frac{\exp[-i\omega t + i\omega(t_0 + z/v_0)]}{(\omega - i\omega_{0n})(\omega + i\omega_{0n})}, \tag{2}$$

$$I_{2n} = \int_{-\infty}^{+\infty} d\omega \frac{\exp(-i\omega\tau + i\xi\sqrt{\omega^2 - \alpha_n^2})}{(\omega - i\omega_{0n})(\omega + i\omega_{0n})}, \tag{3}$$

where $\tau = t - t_0$, $\xi = z\sqrt{\varepsilon}/c$, $\omega_{0n} = \lambda_n v_0/(b\sqrt{1 - \varepsilon v_0^2/c^2})$, $\alpha_n = \lambda_n c/(b\sqrt{\varepsilon})$, and λ_n is the nth root of the Bessel function J_0.

Integral (2) describes the Coulomb field of a charge moving in an infinite waveguide and is equal to

$$I_{1n} = \begin{cases} \dfrac{\pi}{\omega_{0n}}\exp\left[-\omega_{0n}\left(t - t_0 - \dfrac{z}{v_0}\right)\right] & \text{for } t - t_0 - \dfrac{z}{v_0} \ge 0 \\[2ex] \dfrac{\pi}{\omega_{0n}}\exp\left[\omega_{0n}\left(t - t_0 - \dfrac{z}{v_0}\right)\right] & \text{for } t - t_0 - \dfrac{z}{v_0} < 0. \end{cases} \tag{4}$$

Integral (3) corresponds to free oscillations of a cylindrical waveguide and describes the transition radiation. The analytic solution with integral (3) was first obtained by Denisov [13] in studying the propagation of an electromagnetic signal in an ionized gas. In what follows, we will use the method proposed in [13].

The function $\sqrt{\omega^2 - \alpha_n^2}$ has branch points $\omega = \pm\alpha_n$. We draw a cut along the closed interval $(-\alpha_n, \alpha_n)$ in the complex plane ω. Integral (3) should be taken along a contour lying above the cut because, in this case, the total field (1) vanishes for $t < t_0$.

For $t - t_0 - z\sqrt{\varepsilon}/c < 0$, integral (3) is given by the residue of the integrand at the pole $\omega = i\omega_{0n}$:

$$I_{2n} = \frac{\pi}{\omega_{0n}}\exp\left[\omega_{0n}\left(t - t_0 - \frac{z}{v_0}\right)\right]. \tag{5}$$

For $t - t_0 - z\sqrt{\varepsilon}/c > 0$, we can apply the residue theorem to obtain

$$I_{2n} = I_{el} + \frac{\pi}{\omega_{0n}}\exp\left[\omega_{0n}\left(t - t_0 - \frac{z}{v_0}\right)\right]. \tag{6}$$

Here, the integral I_{el} of the integrand in (3) is taken along a closed contour (in the negative direction) that envelopes both the poles, $\omega = \pm i\omega_{0n}$, and the cut $(-\alpha_n, \alpha_n)$. As the integration contour, we choose an ellipse C_ω^- whose foci are the branch points $\omega = \pm\alpha_n$. Let us appropriately transform the integral I_{el}. First, we make the replacement $p = -i\omega$. Second, we switch to the new variable $\alpha_n\zeta = \sqrt{p^2 + \alpha_n^2} - p$. Third, we make another replacement $\zeta = -\beta\omega$, where $\beta = \sqrt{(\tau - \xi)/(\tau + \beta)}$. As a result, the integral I_{el} becomes

$$I_{el} = \frac{i}{2\omega_{0n}}\int_{C_\omega^+} d\omega \exp\left[\frac{\alpha_n}{2}\sqrt{\tau^2 - \xi^2}\left(\omega - \frac{1}{\omega}\right)\right]$$

$$\times \left\{\frac{1}{\omega - \omega_1} - \frac{1}{\omega - \omega_2} - \frac{1}{\omega - \omega_3} + \frac{1}{\omega - \omega_4}\right\}, \tag{7}$$

where $\omega_{1,2} = \beta^{-1}\sqrt{(c \mp \sqrt{\varepsilon}v_0)(c \pm \sqrt{\varepsilon}v_0)^{-1}}$ and $\omega_{3,4} = -\omega_{1,2}$.

These transformations convert the elliptic integration contour C_ω^- into a circular contour C_ω^+, which is centered at the point $\omega = 0$ in the complex plane ω and encloses no singular points except $\omega = 0$. Along the contour C_ω^+, the integration is carried out in the positive direction and we can use the expansions

$$-\frac{1}{\omega - \omega_j} = \frac{1}{\omega_j}\sum_{k=0}^{\infty}\left(\frac{\omega}{\omega_j}\right)^k \quad (8)$$

because the contour does not enclose the poles $\omega = \omega_j$. We can also use the relationship [14]

$$\frac{1}{2\pi i}\int_{C_\omega^+} d\omega\, \omega^k \exp\left[\frac{x}{2}\left(\omega - \frac{1}{\omega}\right)\right] = (-1)^{k+1}J_{k+1}(x). \quad (9)$$

We substitute the series expansions (8) into integral (7), change the order of integration and summation, and take into account relationship (9). As a result, we obtain

$$I_{el} = \frac{2\pi i}{\omega_{0n}}\sum_{m=0}^{\infty}(-1)^m \quad (10)$$
$$\times [(ir_2)^{1+2m} - (ir_1)^{1+2m}]J_{1+2m}(y_n),$$

where $y_n = \alpha_n\sqrt{\tau^2 - \xi^2}$ and $r_{1,2} = \omega_{2,1}^{-1}$.

According to [15], the nth-order Lommel function $U_n(q, x)$ of two variables is defined as

$$U_n(q, x) = \sum_{m=0}^{\infty}(-1)^m\left(\frac{q}{x}\right)^{n+2m}J_{n+2m}(x). \quad (11)$$

For $t - t_0 - z\sqrt{\varepsilon}/c > 0$, expressions (10) and (11) put integral (6) into the form

$$I_{2n} = \frac{2\pi}{\omega_{0n}}\left\{iU_1(ir_2 y_n, y_n) - iU_1(ir_1 y_n, y_n)\right.$$
$$\left. + \frac{1}{2}\exp\left[\omega_{0n}\left(t - t_0 - \frac{z}{v_0}\right)\right]\right\}. \quad (12)$$

The properties of the Lommel functions [15] imply that

$$iU_1(iry, y)$$
$$= \begin{cases} -\sum_{m=0}^{\infty} r^{2m+1}J_{2m+1}(y) & \text{for} \quad |r| \le 1 \\ -\sinh\left(\frac{ry}{2} - \frac{y}{2r}\right) - \sum_{m=0}^{\infty}\frac{J_{2m+1}(y)}{r^{2m+1}} & \text{for} \quad |r| > 1. \end{cases} \quad (13)$$

Finally, using expressions (4), (5), and (12) and taking into account relationships (13), we arrive at the following expression for the total electric field (1), which is generated by a thin ring-shaped electron bunch propagating with constant velocity in a semi-infinite waveguide and is represented as a linear superposition of the Coulomb field of a moving charge, E_r^{coul}, and the transition radiation field E_r^{trans}:

$$E_r(t, r, z, t_0, r_0)$$
$$= E_r^{\text{coul}}(t, r, z, t_0, r_0) + E_r^{\text{trans}}(t, r, z, t_0, r_0),$$

$$E_r^{\text{coul}}(t, r, z, t_0, r_0) = -\frac{2Ne}{b^2\varepsilon\sqrt{1 - \varepsilon v_0^2/c^2}}$$
$$\times \sum_n \frac{J_0(\lambda_n r_0/b)J_1(\lambda_n r/b)}{J_1^2(\lambda_n)}$$
$$\times \left\{\epsilon_1 \exp\left[\omega_{0n}\left(t - t_0 - \frac{z}{v_0}\right)\right]\right.$$
$$\left. + \epsilon_2 \exp\left[-\omega_{0n}\left(t - t_0 - \frac{z}{v_0}\right)\right]\right\}, \quad (14)$$

$$E_r^{\text{trans}}(t, r, z, t_0, r_0) = \frac{4Ne}{b^2\varepsilon\sqrt{1 - \varepsilon v_0^2/c^2}}$$
$$\times \sum_n \frac{J_0(\lambda_n r_0/b)J_1(\lambda_n r/b)}{J_1^2(\lambda_n)}$$
$$\times \left\{\epsilon_1\left[\sum_{m=0}^{\infty}(r_1^{2m+1} - r_2^{2m+1})J_{2m+1}(y_n)\right.\right.$$
$$\left. + \frac{1}{2}\exp\left[\omega_{0n}\left(t - t_0 - \frac{z}{v_0}\right)\right]\right]$$
$$+ \epsilon_2\left[\sum_{m=0}^{\infty}\left(r_1^{2m+1} - \frac{1}{r_2^{2m+1}}\right)J_{2m+1}(y_n)\right.$$
$$\left.\left. + \frac{1}{2}\exp\left[-\omega_{0n}\left(t - t_0 - \frac{z}{v_0}\right)\right]\right]\right\}. \quad (15)$$

Here, we use the notation $v_{pr} = c/\sqrt{\varepsilon}$ and introduce the functions

$$\epsilon_1 = \begin{cases} 1 & \text{for } t - t_0 - \frac{z}{v_0} \le 0 \text{ and } t - t_0 - \frac{z}{v_{pr}} \ge 0 \\ 0 & \text{for } t - t_0 - \frac{z}{v_0} > 0 \text{ or } t - t_0 - \frac{z}{v_{pr}} < 0, \end{cases}$$

$$\epsilon_2 = \begin{cases} 1 \ \text{ for } \ t - t_0 - \dfrac{z}{v_0} > 0 \\[2mm] 0 \ \text{ for } \ t - t_0 - \dfrac{z}{v_0} \leq 0. \end{cases}$$

The expression for the magnetic field generated by a thin ring-shaped bunch can be obtained in an analogous way (in which it is also represented as a linear superposition of the magnetic field associated with the bunch current, H_φ^{curnt}, and the transition radiation field H_φ^{trans}):

$$H_\varphi(t, r, z, t_0, r_0)$$

$$= H_\varphi^{\text{curnt}}(t, r, z, t_0, r_0) + H_\varphi^{\text{trans}}(t, r, z, t_0, r_0),$$

$$H_\varphi^{\text{curnt}}(t, r, z, t_0, r_0) = \frac{\varepsilon v_0}{c} E_r^{\text{coul}}(t, r, z, t_0, r_0), \quad (16)$$

$$H_\varphi^{\text{trans}}(t, r, z, t_0, r_0) = \frac{4 N e v_0}{b^2 c \sqrt{1 - \varepsilon v_0^2/c^2}}$$

$$\times \sum_n \frac{J_0(\lambda_n r_0/b) J_1(\lambda_n r/b)}{J_1^2(\lambda_n)}$$
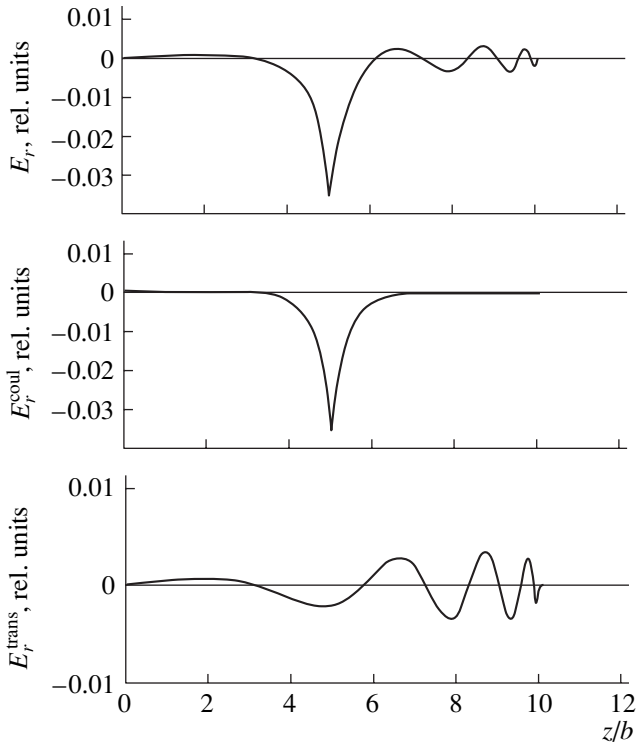
Fig. 1. First harmonic of the field of a bunch in the form of a thin disc of radius $a/b = 0.125$. The observation time is $tc/b = 10$, and the radius of the observation region is $r_{\text{obs}}/b = 0.25$. The remaining parameter values are $t_0 = 0$, $\varepsilon = 1.0$, and $v_0/c = 0.5$.

$$\times \left\{ \epsilon_1 \left[-\sum_{m=0}^{\infty} (r_1^{2m+1} + r_2^{2m+1}) J_{2m+1}(y_n) \right. \right.$$

$$\left. + \frac{1}{2} \exp\left[\omega_{0n}\left(t - t_0 - \frac{z}{v_0} \right) \right] \right] \tag{17}$$

$$+ \epsilon_2 \left[-\sum_{m=0}^{\infty} \left(r_1^{2m+1} + \frac{1}{r_2^{2m+1}} \right) J_{2m+1}(y_n) \right.$$

$$\left. \left. + \frac{1}{2} \exp\left[-\omega_{0n}\left(t - t_0 - \frac{z}{v_0} \right) \right] \right] \right\}.$$

The quasistatic field components (14) and (16) and the transition radiation field components (15) and (17) are defined (and are nonzero) in the region $z \leq (t - t_0) v_{pr}$. The quantity v_{pr} introduced above is merely the highest velocity of the propagation of electromagnetic perturbations in a waveguide. This indicates that, for $t > t_0$, neither Coulomb nor transition radiation fields enter the region to the right of the point $z_{pr} = (t - t_0) v_{pr}$. The fastest and highest-frequency component of the electromagnetic signal is the precursor, which propagates with the velocity v_{pr} [16]. Since the bunch propagation velocity is $v_0 < v_{pr}$, the field overtakes the bunch. A qualitative pattern of the propagation of a transition radiation pulse is illustrated in Fig. 1, which shows the longitudinal profiles of the first harmonics of the total radial electric field and its Coulomb and transition radiation components. The bunch generates an ultrawideband transition radiation pulse, whose shortest-wavelength and highest-frequency components are the components of the precursor. The oscillation amplitude in the pulse decreases toward the precursor, so that the total field vanishes in the cross section $z = z_{pr}$.

NUMERICAL RESULTS

For simulations, we chose an electron bunch with the current density distribution

$$j_z(r_0, t_0) = j_0 J_0\left(\lambda_1 \frac{r_0}{a} \right) \exp\left[-4\left(\frac{2 t_0}{T_b} - 1 \right)^2 \right], \tag{18}$$

where j_0 is the peak current density, $0 \leq t_0 \leq T_b$, T_b is the bunch duration, $0 \leq r_0 \leq a$, and a is the bunch radius.

Analogous current density distributions are typical of many experiments. Such a bunch can be modeled by a number of macroparticles in the form of thin rings with corresponding radii, charges, and longitudinal coordinates. The field generated by the bunch can be represented as a sum of the fields of all constituent macroparticles.

The qualitative pattern of the propagation of a transition radiation pulse (see Fig. 1) was obtained by taking into account only the first harmonic of the total
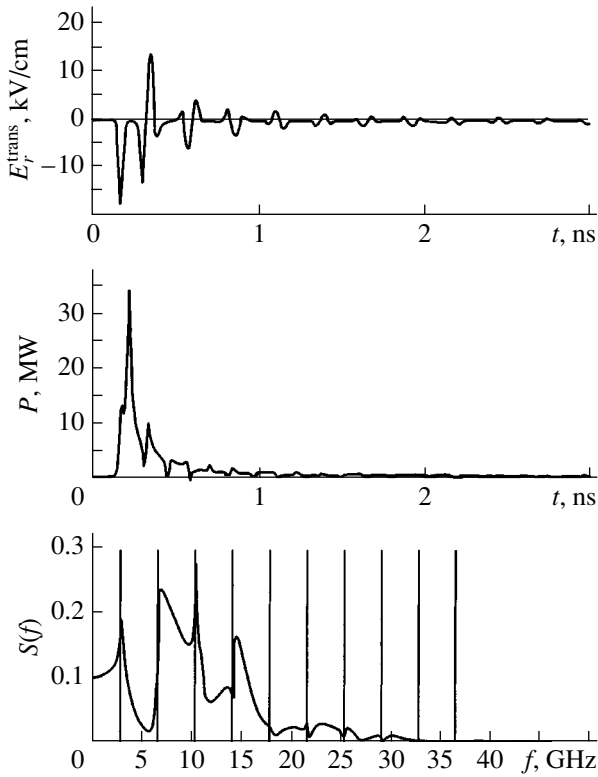
Fig. 2. Parameters of the transition radiation signal generated by a Gaussian bunch. The vertical lines are the critical frequencies f_n. The signal is calculated at the observation point with the coordinates $z_{obs}/b = 1$ and $r_{obs}/b = 0.25$ for $b = 4.0$ cm, $\gamma = 2.29$, $a = 0.5$ cm, $L_b = 2.0$ cm, and $I_{max} = 1.9455$ kA.
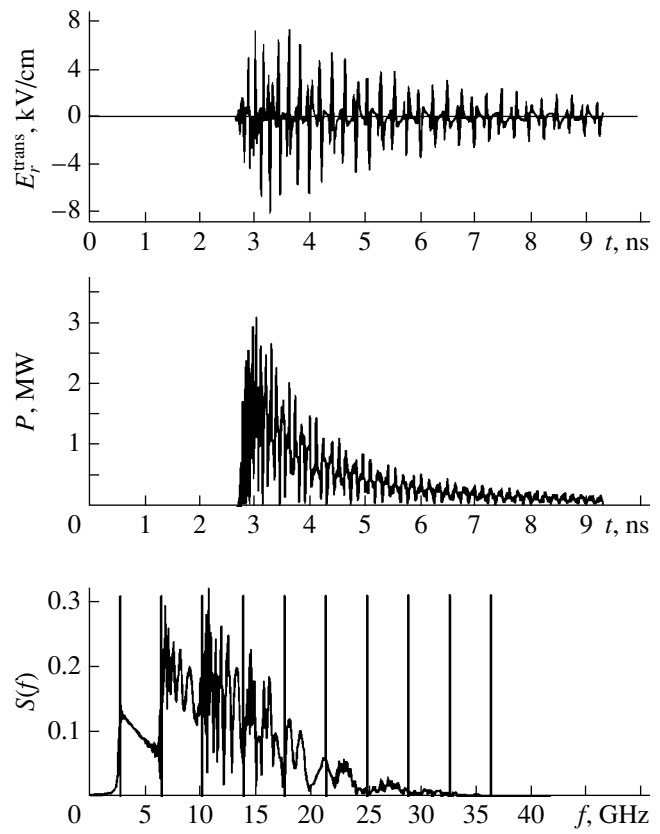

Fig. 3. The same as in Fig. 2 but at $z_{obs}/b = 20$.

radial field. However, in reality, a charged bunch in a cylindrical waveguide generates the full set of radial field harmonics. This results in the narrowing of the field maxima and the appearance of characteristic peaks whose amplitudes are much larger than the amplitude of the first radial field harmonic [17]. The numerical results presented in this section were obtained for the first ten field harmonics. For the above parameters of the waveguide and of the bunch, computations with a larger number of harmonics revealed no significant changes in the shape of the field profile. We simulated a vacuum waveguide ($\varepsilon = 1$) in which case the radial inhomogeneity of the dielectric filling and radiation losses in the dielectric are absent.

In what follows, we restrict ourselves to analyzing the transition radiation component of the field. We calculated only the measurable radiation parameters. For this reason, we first investigated the time evolutions of the field at certain spatial points. Second, we computed the time evolutions of the transition radiation power. Finally, we determined the frequency spectra $S(f)$ of the generated electromagnetic pulses (here, f is the linear frequency). The efficiency with which the oscillations are generated was characterized by the following quantities (in percentages of their maximum values):

the energy efficiency η, defined as the ratio of the radiation energy to the kinetic energy of the bunch electrons, and the power-conversion efficiency χ, defined as the ratio of the maximum radiation power to the maximum power of the bunch current. The efficiencies η and χ so defined are directly proportional to the bunch charge. The parameters of the bunch–waveguide system were chosen so that the maximum bunch current was lower than the vacuum limiting current [18].

Let us consider transition radiation signals near the waveguide end and far from it. The first case is illustrated in Fig. 2. We can see that an electromagnetic pulse near the waveguide end is characterized by a large field amplitude (15 kV/cm), high maximum power (about 33 MW), and short duration (less than 1 ns). The spectrum of this signal is broadband and has sharp narrow peaks at frequencies close to the critical frequencies $f_n = \lambda_n c/(2\pi b \sqrt{\varepsilon})$ of the waveguide. The presence of the low-frequency ($f < f_1$) part of the spectrum can be explained as being due to the fact that the transition radiation signal has propagated only a short distance and, therefore, did not have enough time to form completely. Far from the waveguide end (see Fig. 3), the field amplitude is smaller than 8 kV/cm, the maximum
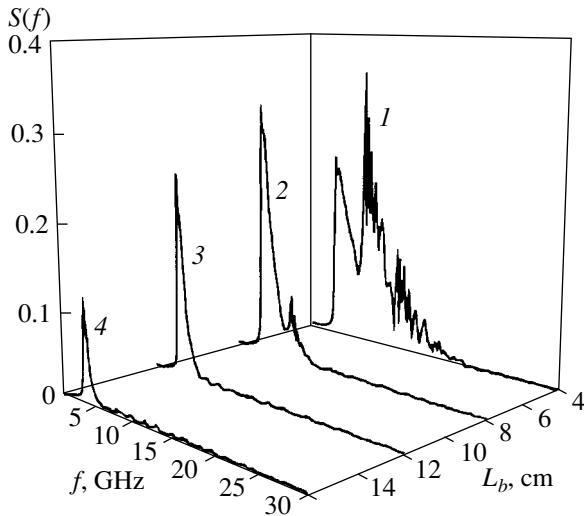
Fig. 4. Dependence of the shape of the transition radiation spectrum on the bunch length at the point with the coordinates $z_{obs}/b = 20$ and $r_{obs}/b = 0.25$ for $b = 4.0$ cm, $\gamma = 2.29$, $a = 0.5$ cm, and $I_{max} = 1.9455$ kA.
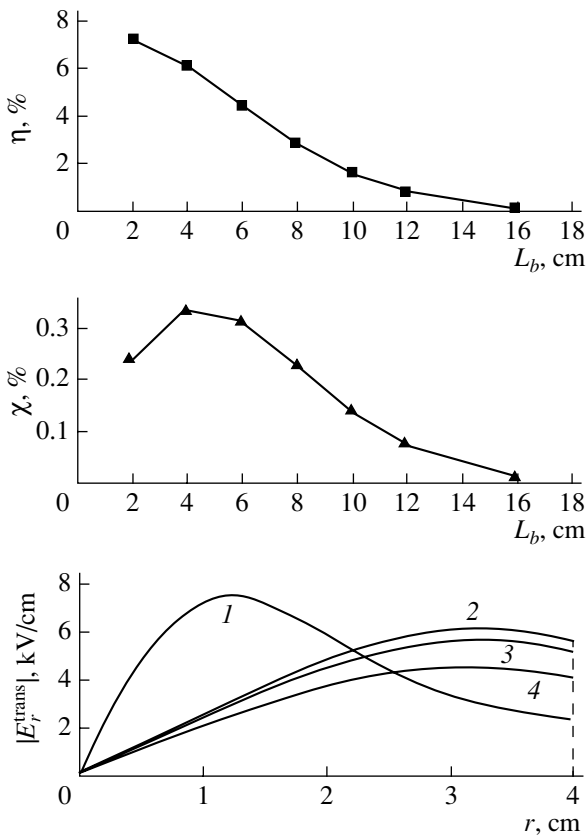
Fig. 5. Dependence of η and χ on the bunch length and the radial structure of the field strength at the time at which the radiation power is maximum. The parameters of the bunch–waveguide system are the same as in Fig. 4.

radiation power is about 3 MW (which is one order of magnitude lower than that in the previous case), and the pulse duration is significantly longer (more than 5 ns). The spectrum peaks at frequencies somewhat higher

than the corresponding critical frequencies f_n and contains no low-frequency part (below the frequency f_1). We can see that the gaps between the neighboring critical frequencies disappeared and that the entire spectrum is somewhat broader. Such changes in the shape of the transition radiation signal may be explained as follows. When a charged bunch enters the waveguide, it excites all possible natural waveguide modes simultaneously with different amplitudes. The dispersion properties of the waveguide are such that the natural modes propagate at different velocities. As a result, the initially short wave packet spreads out continuously as it moves away from the waveguide end. Figure 3 shows that the frequency of the recorded oscillations decreases with time because higher frequency oscillations propagate at higher velocities and are the first to reach the observation point. The radiation energy remains constant; consequently, as the pulse duration increases, both the radiation power and the field strength decrease.

Figure 4 shows how the shape of the transition radiation spectrum changes as the length L_b of a charged bunch increases. The total bunch charge was adjusted to be proportional to the bunch length in such a way that the maximum current in bunches of different lengths was the same. The spectrum width is seen to decrease significantly as the bunch length L_b increases. In a waveguide with the radius $b = 4$ cm, a bunch with the length $L_b = 4$ cm efficiently excites at least three waveguide modes (curve *1*) and a bunch with the length $L_b = 8$ cm excites two modes (curve *2*). In contrast, a bunch with the length $L_b = 12$ cm (curve *3*), as well as a bunch with the length $L_b = 16$ cm (curve *4*), excites only one mode; moreover, the amplitude of the mode excited by the longer of these two bunches is smaller. This dependence of the shape of the radiation spectrum on the bunch length can be explained as a consequence of the current density distribution (18), which was chosen above for an electron bunch. Figure 5 displays the dependence of the efficiencies η and χ on the bunch length L_b and the corresponding radial structure of the transition radiation field. The energy efficiency η is seen to decrease as L_b increases. This is explained by the smaller number of the excited waveguide modes and higher bunch kinetic energy, which is proportional to the number of bunch electrons. The power-conversion efficiency χ has a maximum at $L_b = 4$ cm. In order to explain this maximum, we turn to the field structure. As the bunch length L_b increases from 2 cm (curve *1*) to 4 cm (curve *2*), the spectral content of the generated signal changes and the field maximum is displaced toward the side wall of the waveguide. As a result, the transition radiation power increases. With a further increase in L_b under such conditions that the first field harmonic is dominant (curves *3* and *4* for $L_b = 6$ and 8 cm, respectively), the radial profile of the field does not change qualitatively but the maximum field amplitude and, accordingly, the radiation power decrease.

It is evident from Figs. 4 and 5 that the generation of transition radiation in the waveguide is inefficient when the waveguide diameter is smaller than or comparable to the length of the charged bunch. Let us consider how the waveguide radius influences the generation efficiency of the transition radiation. The dependence of η and χ on the waveguide radius b and the radial structure of the transition radiation field that are shown in Fig. 6 were calculated for a bunch with fixed parameters. The maximum bunch current was lower than the vacuum limiting current in a waveguide with the largest radius, and the dimensionless longitudinal coordinate of the observation point was $z_{obs}/b = $ const. With increasing waveguide radius, the energy efficiency η increases monotonically and tends to saturate at large radii. On the other hand, as the waveguide radius increases, the current of the bunch that can be injected into the waveguide decreases. As b increases, the power-conversion efficiency χ, first increases to its maximum value then gradually decreases. The radial structure of the transition radiation field at the time at which the radiation power is maximum changes as follows. In waveguides with radii 4 and 8 cm (curves *1*, *2*), the field is maximum near the side wall; the radiation power is higher in a waveguide with the larger radius. In waveguides with radii 12 and 16 cm (curves *3*, *4*), a 10-cm-long bunch excites a multimode signal and the field maximum is displaced toward the waveguide axis. The maximum field amplitude and, accordingly, the radiation power are both lower. Note that curves *3* and *4* are qualitatively similar in shape to the radial profile of the transition radiation field generated by a 2-cm-long bunch in a waveguide with a radius of 4 cm (curve *1* in Fig. 5). Figure 7 illustrates how the shape of the radiation spectrum changes as the waveguide radius increases. We can see that, in a narrow waveguide (with radius $b = 4$ cm), the bunch excites an essentially single-mode signal (curve *1*). In a waveguide with radius $b = 8$ cm, the bunch excites two waveguide modes (curve *2*), and, in a waveguide with radius $b = 12$ cm, it excites four modes (curve *3*). In a waveguide with radius $b = 16$ cm, at least five modes are excited (curve *4*). The left boundary of the radiation spectrum is displaced toward the lower-frequency range, the spectrum width remains essentially unchanged, and the critical frequencies of the waveguide with the largest radius are spaced most closely.

CONCLUSION

When a charged bunch enters a semi-infinite cylindrical waveguide, it generates transition radiation. If the Cherenkov resonance condition is not satisfied, the excited electromagnetic field is a superposition of the quasistatic field of a moving charge and the transition radiation field. The fastest component of the field (the precursor) propagates with the velocity $c/\sqrt{\varepsilon}$, which is higher than the bunch velocity.
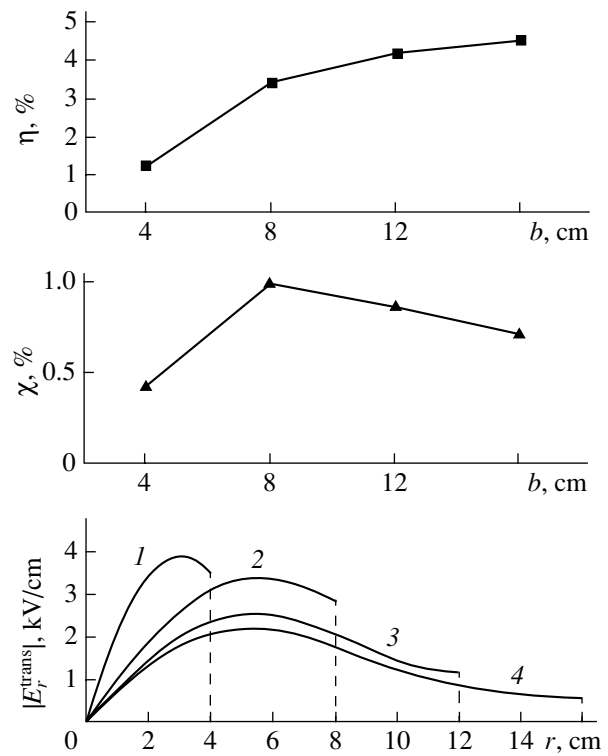
Fig. 6. Dependence of η and χ on the waveguide radius and the radial structure of the field strength at the time at which the radiation power is maximum. The curves were calculated at the waveguide cross section $z_{obs}/b = 3$ for $\gamma = 2.29$, $a = 0.5$ cm, $L_b = 10.0$ cm, and $I_{max} = 0.973$ kA.
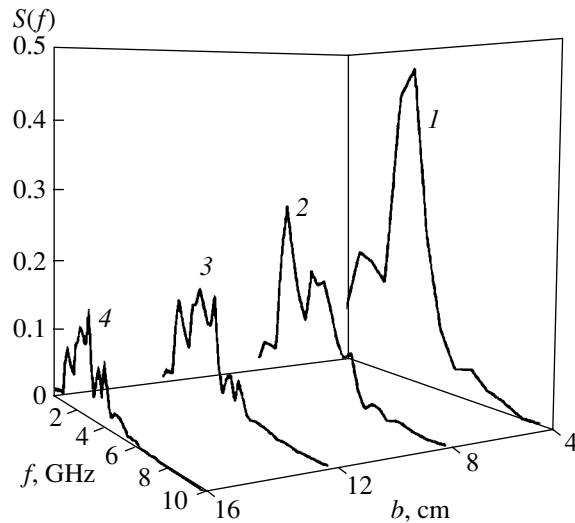

Fig. 7. Dependence of the shape of the transition radiation spectrum on the waveguide radius. The parameters of the bunch–waveguide system are the same as in Fig. 6.

The spectrum of the transition radiation signal contains waveguide's eigen oscillations with all possible frequencies and wavenumbers. Because of the dispersion properties of the waveguide under consideration, these oscillations propagate at vastly different veloci-

ties. The longitudinal profile of the transition radiation field calculated at a certain time shows that the wavelengths of oscillations are the shortest in the "head" of the wave packet and become progressively longer toward the waveguide end (toward the "tail" of the packet). From the head to the tail of the packet, the spatial envelope of the signal first increases from zero to maximum then decreases gradually. The time evolution of the transition radiation field calculated at a sufficiently large distance from the waveguide end shows that the frequency of the recorded oscillations decreases with time. By analogy with the spatial envelope, the temporal envelope of the signal first reaches a maximum then decreases.

The spectrum of the transition radiation signal is broadband and contains peaks corresponding to several harmonics of the radial electric field. The wide and asymmetric peaks that are clearly distinguished in the spectrum occur at frequencies somewhat higher than the corresponding critical frequencies of the waveguide.

The transition radiation signal calculated near the waveguide entrance is characterized by high maximum power and short duration. At a sufficiently large distance from the waveguide entrance, the maximum power of the signal is lower and the duration of the signal is significantly longer.

The longer the bunch, the smaller the number of the excited field harmonics. A sufficiently long bunch excites an essentially single-mode signal, in which case the generation efficiency of a transition radiation pulse sharply decreases. The efficiency with which a UWBP is generated by a long-duration bunch can be raised by using a waveguide whose radius is larger than the bunch length. In this way, however, the bunch current will be low because it will be restricted by the vacuum limiting current in the waveguide.

In order to achieve the greatest efficiency of the generation of UWBPs, it is necessary to use short electron bunches with currents close to the limiting current at which the bunches can be transported through a vacuum waveguide.

REFERENCES

1. H. F. Harmuth, *Nonsinusoidal Waves for Radar and Radio Communication* (Academic, New York, 1981; Radio i Svyaz', Moscow, 1985).

2. L. Yu. Astanin and A. A. Kostylev, *Foundations of Super-broadband Measurements* (Radio i Svyaz', Moscow, 1989).

3. L. Yu. Astanin and A. A. Kostylev, Zarubezhn. Radioélektron., No. 9, 3 (1981).

4. N. V. Zernov and G. V. Merkulov, Zarubezhn. Radioélektron., No. 9, 85 (1981).

5. V. G. Stroitelev, Zarubezhn. Radioélektron., No. 9, 95 (1981).

6. S. P. Pan'ko, Zarubezhn. Radioélektron., No. 9, 106 (1981).

7. M. Friedman, V. Serlin, Y. Y. Lay, and J. Krall, in *Proceedings of the 8th International Conference on High-Power Particle Beams, Novosibirsk, 1990*, Vol. 1, p. 53.

8. M. I. Gaponenko, V. I. Kurilko, S. M. Latinskiǐ, *et al.*, Vopr. At. Nauki Tekh., Ser. Yad.-Fiz. Issled. **11** (4, 5), 151 (1997).

9. V. A. Balakirev, M. I. Gaponenko, A. M. Gorban', *et al.*, Vopr. At. Nauki Tekh., Ser. Fiz. Plazmy, No. 3, 118 (2000).

10. V. A. Balakirev and G. L. Sidel'nikov, Zh. Tekh. Fiz. **69** (10), 90 (1999) [Tech. Phys. **44**, 1209 (1999)].

11. V. L. Ginzburg and I. M. Frank, Zh. Éksp. Teor. Fiz. **16**, 15 (1946).

12. V. L. Ginzburg and V. N. Tsytovich, *Transient Radiation and Transient Scattering* (Nauka, Moscow, 1984).

13. N. G. Denisov, Zh. Éksp. Teor. Fiz. **21**, 1354 (1951).

14. M. A. Lavrent'ev and B. V. Shabat, *Methods of the Theory of Functions of a Complex Variable* (Nauka, Moscow, 1973).

15. G. N. Watson, *Treatise on the Theory of Bessel Functions* (Cambridge Univ. Press, Cambridge, 1945; Inostrannaya Literatura, Moscow, 1949).

16. L. Brillouin, Ann. Phys. (Leipzig) **44** (10), 203 (1914).

17. T.-B. Zhang, J. L. Hirshfield, T. C. Marshall, and B. Hafizi, Phys. Rev. E **56**, 4647 (1997).

18. A. A. Rukhadze, L. S. Bogdankevich, S. E. Rosinskiǐ, and V. G. Rukhlin, *Physics of High-Current Relativistic Electron Beams* (Atomizdat, Moscow, 1980).

Translated by O. Khadin

SURFACES,
ELECTRON AND ION EMISSION

Visualization of Nonuniform Magnetic Fields by Garnet Ferrite Films

V. G. Vishnevskiĭ, R. M. Mikherskiĭ, and S. V. Dubinko

*Domen Design Office, Vernadsky National University,
Studencheskaya ul. 12, Simferopol, 95001 Ukraine*

e-mail: domain@home.cris.net

Received March 5, 2001; in final form, July 20, 2001

Abstract—The magnetization distribution in epitaxial garnet ferrite films subjected to the harmonic field of a signal from a magnetic tape is studied. By means of image processing techniques and diffraction spectrum analysis, it is shown that the distortion of the signal shape grows with increasing amplitude and period of the field. © 2002 MAIK "Nauka/Interperiodica".

Epitaxial garnet ferrite films (EGFFs) are used as effective magnetooptic transducers, which are needed, in particular, for phonoscopic examination in criminalistics [1]. They help to visualize magnetically recorded data.

In some cases, EGFFs with easy-plane anisotropy are promising, since they, unlike the films with perpendicular anisotropy, are locally magnetized by rotating the moment rather than by domain wall displacement. In this respect, the question of to what extent the external field and the magnetization distribution induced in an EGFF coincide becomes of prime interest.

In [2], it has been shown theoretically that the distortions of the magnetization distribution that is induced by a harmonic field depend on the amplitude and period of this field. The easy-magnetization axis lay in the film plane and was parallel to the shear component of the external field. The aim of this work is to experimentally check the theoretical findings obtained in [2].

In the experiments, we used the diffraction of light by a phase grating induced in the EGFF by the sinusoidal field of a magnetic tape. It is assumed that the spectrum of a perfect sinusoidal grating contains mostly the first orders of diffraction. In addition, we studied the distribution of the intensity of the grating image observed in a polarization microscope. In the latter case, digital image processing techniques were employed.

$(Bi,Lu,Ca)_3(Fe,Ga)_5O_{12}$ and $(Bi,Tm)_3(Fe,Ga)_5O_{12}$ EGFFs with easy-plane anisotropy were grown on $Gd_3Ga_5O_{12}(111)$ substrates. The films were covered by a TiN reflective layer and brought into contact with the magnetic tape. The tape contained sinusoidal signals with a discrete set of frequencies and an amplitude gradation within 20 dB. The linear distortion of the records
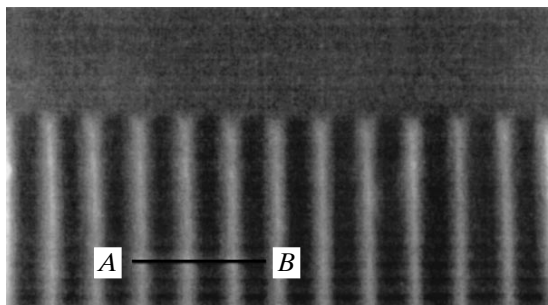

Fig. 1. Magnetization structure in the EGFF with easy-plane anisotropy. The structure is induced by the sinusoidal field of the magnetic tape with the period $d = 105$ μm.
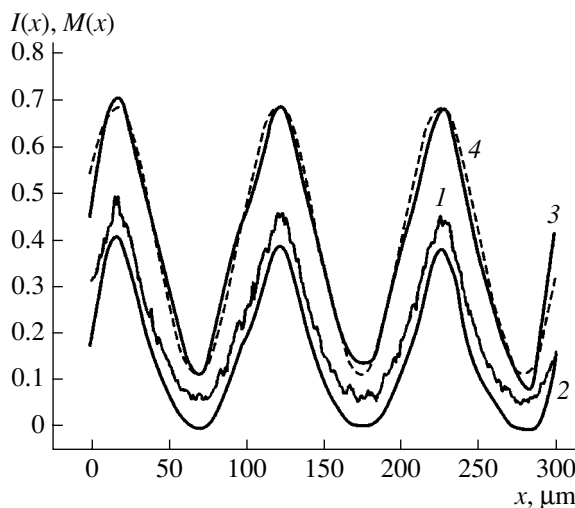
Fig. 2. Results of the digital image processing of image *1* taken from the segment *AB*. *1*, intensity profile; *2*, after filtering higher frequencies in *1*; *3*, magnetization distribution reconstructed from *2*; and *4*, cosinusoid
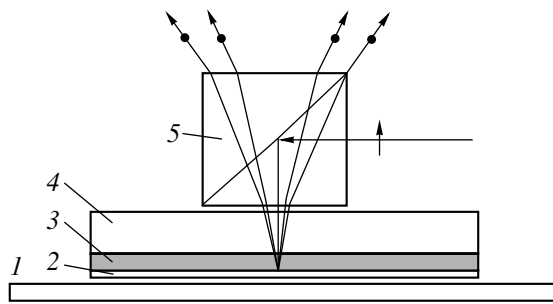
Fig. 3. Geometry of the magnetooptic diffraction experiments.
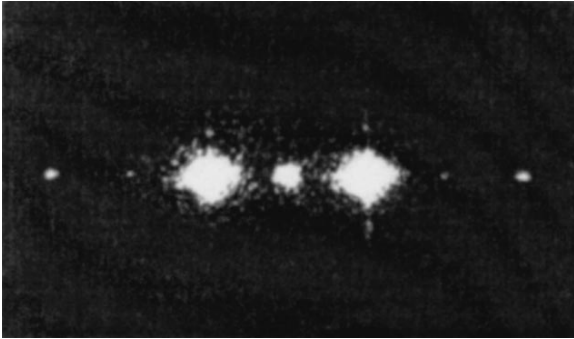
Fig. 4. Diffraction of light by the phase grating induced in the EGFF by the sinusoidal field of the magnetic tape.

was no greater than 1.5%. The thicknesses of the EGFF and the TiN coating were 3.5 and 0.2 μm, respectively. The saturation magnetization of the films was $4\pi M_s \leq$ 23 900 A/m and the specific Faraday rotation, $\theta = 0.9°-$ $1.1° \, \mu m^{-1}$. The EGFF either had a spontaneous single-domain structure or was brought to the saturation state by a uniform field applied along the easy magnetization axis (EMA). The magnetization structure induced in the EGFF by the tape field is shown in Fig. 1 (as it is seen in the polarization microscope).

Figure 2 demonstrates the digital processing of this image. Curve *1* is the intensity profile along segment *AB*, curve *2* is the result of filtering high frequencies of curve *1*, and curve *3* is the magnetization distribution

reconstructed from curve *2* according to the expression

$$M(x) \sim \frac{\varphi + \arcsin(\sqrt{I(x)/I_0})}{k\theta_0}. \qquad (1)$$

Here, $I(x)$ is the intensity distribution along the coordinate x, I_0 is the intensity of light incident on the polarizer-EGFF-analyzer system, φ is the angle of deflection of the polarizer axes from the crossed state, $\theta_0 = \theta h$ is the total Faraday rotation, $k = H_0(H_A + 4\pi M_s)^{-1} < 1$, H_0 is the external field amplitude, H_A is the axis of anisotropy, and $M(x)$ is the magnetization normal component distribution along the x axis.

Comparing curve *3* and cosinusoid *4*, we can judge the presence of distortions when the film images the signal shape. Diffraction spectrum analysis provides more accurate estimates of these distortions.

A beam from a helium-neon laser passes through the side face of a polarizing prism-cube and strikes the EGFF; beams (except for the zero-order ones) diffracted by the induced magnetization distribution in the film leave the cube through its top side. The intensities of the diffraction orders are measured with a photoelectric multiplier.

The experimental geometry is shown in Fig. 3, where the polarization state of each of the beams is also indicated. Such a geometry excludes both the effect of the magnetization tangential component on the Faraday rotation and polarization effects due to inclined incidence of light [3]. The scattered fields of the active layer of magnetic tape *1* act on EGFF *3* through transparent layer *2*, producing a magnetic diffraction grating. Prism-cube *5* is arranged on substrate *4* via the immersion contact so as to provide the film rotation in the plane. During the experiments, the in-plane EMA of the film was oriented either parallel (||) or perpendicular ($\perp$) to the tangential component of the tape field. Preliminary control of the EMA position was accomplished by the inductive method. The actual diffraction spectrum for the orientation ($\perp$) is depicted in Fig. 4.

The spectrum has the well-defined symmetry (relative to the zero order), the energy maxima are at the orders ± 1, and the intensity of even (second) orders is negligible.

Table

Sample no.	d, μm	H, A/m	I_1, arb. units	$\dfrac{I_3}{I_1}$	$\dfrac{I_5}{I_1}$	$\dfrac{I_7}{I_1}$	$\sqrt{\dfrac{I_3 + I_5 + I_7}{I_1}}$	EMA orientation
1	100	11937	4.0	0.081	0.015	0.0025	0.3142	||
1	50	11937	2.25	0.044	0.009	0.0022	0.2357	||
2	100	11937	2.50	0.080	0.012	0.0028	0.3079	||
2	100	3979	1.50	0.047	0.011	0.0040	0.2620	||
2	100	11937	3.50	0.066	0.014	0.0035	0.2891	$\perp$
2	100	3979	1.00	0.030	0.005	0.0010	0.1897	$\perp$

The distortion of the actual grating relative to the sinusoidal one, like the nonlinear distortion level, was estimated by the test

$$L = \sqrt{\frac{I_3 + I_5 + I_7}{I_1}}, \qquad (2)$$

where I_1, I_3, I_5, and I_7 are the intensities of the first and higher orders, respectively.

The experimental data are summarized in the table.

It is easy to see that the experimental data agree well with the theoretical results; namely, the maximal angle between the magnetization and the film plane is the larger and the distortions of the induced grating are the higher, the larger the period d of the inducing field and the higher its amplitude. The level of the distortions depends on the applied field period stronger than on the field amplitude.

As follows from the table, when the films with easy-plane anisotropy are used for the visualization, the dis-tortions can be suppressed if the EMA is oriented nor-mally to the tangential component of the field applied. Such a geometry is harder to treat theoretically, since the moment has two rotations different in phase: in the EGFF plane and normally to it.

REFERENCES

1. S. V. Levy, Yu. S. Agalidi, and V. G. Vishnevskiĭ, Izv. Vyssh. Uchebn. Zaved., Radioélektron. **41** (7–8), 74 (1998).

2. R. M. Mikherskiĭ, S. V. Dubinko, V. I. Butrim, and V. G. Vishnevskiĭ, Uch. Zap. Tavrichesk. Nats. Univ. **12** (2), 156 (1999).

3. Yu. F. Vilesov, V. G. Vishnevskiĭ, and N. A. Groshenko, Pis'ma Zh. Tekh. Fiz. **15** (13), 14 (1989) [Sov. Tech. Phys. Lett. **15**, 500 (1989)].

Translated by V. Isaakyan

SURFACES,
ELECTRON AND ION EMISSION

Production and Properties of Metal–Carbon Composite Coatings with a Nanocrystalline Structure

S. A. Shiryaev, M. V. Atamanov, M. I. Guseva, Yu. V. Martynenko, A. V. Mitin, V. S. Mitin, and P. G. Moskovkin

Russian Research Center Kurchatov Institute, pl. Kurchatova 1, Moscow, 123182 Russia
e-mail: martyn@nfi.kiae.ru
Received May 4, 2001

Abstract—It is suggested to produce metal–carbon composite coatings by magnetron sputtering of mosaic cathodes, which are Group IV, V, and VI metals. The mosaic structure of the cathode elements are computer-optimized for each of the metals. Reflection electron diffraction studies show that the coatings have the amorphous or nanocrystalline structures, which are thermally stable. The coatings offer specific physical properties, in particular, low friction factor and high hardness. © *2002 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Modern methods for solid surface modification, such as ion implantation and magnetron sputtering, as well as their combination, allow one to alter the surface properties of structural materials in a wide range. Coatings made of transition metal (Groups IVA–VIA) carbides with the face-centered cubic (fcc) lattice are of particular interest, because this lattice is typical of metals characterized by high hardness, excellent high-temperature strength, and good corrosion resistance.

In this work, we synthesize the coatings of transition metal carbides (TiC, Ta_2C, Ta_3C_2, CrC, NbC, Mo_2C, and $(TiC)N(Mo_2C)N$) by co-sputtering of the mosaic cathodes. The metallic cathodes with graphite inserts were sputtered by Ar^+ ions or by an $Ar^+ + N^+$ mixture in the magnetron plasma.

Magnetron sputtering is widely used for the deposition of various coatings [1], in studying composite coatings. In the latter case, the cathode is usually made from an alloy of necessary composition, which is sputtered in the magnetron [2].

We suggest to apply composite coatings by using a mosaic cathode. The mosaic structure of the cathode (target) is tailored in such a way as to provide coatings of uniform composition.

THEORETICAL GROUNDS OF FILM DEPOSITION USING MAGNETRON SPUTTERING OF A MOSAIC CATHODE

If the cathode is inhomogeneous, the erosion of the different materials proceeds at a various rate due to a difference in the sputtering coefficients. As a result, the materials with a higher sputtering coefficient turn out to be farther from the plasma than those with a lesser sputtering coefficient. The ion current density on the cathode is equal to the number of ions incident on a unit area per unit time:

$$j = n v_T/4, \tag{1}$$

where v_T is the ion thermal velocity and n is the plasma density.

The ion current decreases at the recessed cathode parts, i.e., at the materials with the higher sputtering coefficient. Indeed, the plasma boundary is located at the level of the highest cathode point along the magnetic field line. Downward, i.e., toward the cathode, the plasma density n rapidly decreases, because plasma electrons are magnetized and be deflected from the plasma boundary by no more than the Larmor radius

$$r_L = m v_e/eB,$$

where m is the electron mass, v_e is the electron velocity, e is the electron charge, and B is the magnetic field.

The ions and the electrons cannot be spaced more than the Debye length apart. The Debye length is given by

$$r_D = (T/4\pi n e^2)^{1/2},$$

where T is the plasma temperature.

Therefore, the characteristic distance of plasma density decrease can be estimated as

$$r_0 = r_L + r_D.$$

As a result, the erosion rates of the different cathode materials become equal. In this case, the material surface moves due to erosion with a velocity

$$u = Yj/N, \tag{2}$$

where Y is the sputtering coefficient and N is the volume atom concentration.

Thus, assuming that the ion velocities do not depend on the distance from the cathode, we obtain the following condition for the erosion of two materials:

$$S_1 n(x_1)/N_1 = S_2 n(x_2)/N_2, \qquad (3)$$

where x_1 and x_2 are the surface coordinates of the first and second materials, respectively, along the axis normal to the surface.

Knowing the dependence $n(x)$, one can calculate from Eq. (3) the displacement $\Delta x = |x_1 - x_2|$ of the surfaces for the same erosion rates.

If the plasma-facing surfaces of different materials constituting the target used for the deposition of multicomponent coatings are coplanar, the ratio of the components being deposited will first be proportional to the product of the area of sputtering by the sputtering coefficient. As time passes, the erosion process becomes steady-state, the erosion rates of all the components become equal, and the ratio of the components is proportional to the area of sputtering on the cathode. The transient period of the erosion process is

$$\tau = \Delta x/\Delta u, \qquad (4)$$

where $\Delta u = |u_1 - u_2|$ is the difference in the rates of surface sputtering.

Therefore, in practice, the cathode composition should be selected, starting from the equal erosion rates for all components. However, the coating deposited for the early period of time τ should be removed. Subsequently, when the values of the surface displacement Δx for the components with higher sputtering coefficients are accurately determined experimentally, these components can be recessed at once by Δx when manufacturing the cathode. In our conditions, $r_0 \approx 10^{-4}$ m and $\tau = \Delta x/\Delta u \approx 10^3$ s.

A computational code has been developed to simulate the percentage of the components in the film deposited and the film thickness. The code makes it possible to analyze the film deposition process when the targets of different geometry and composition are employed.

A mosaic target used in the simulation is composed of base material 1 and overlying inserts made of materials 2 and 3 (Fig. 1). However, our model applies to targets consisting of an arbitrary number of components.

To determine the percentage of the components in the coating, the contributions from each of the target inserts were calculated and then summed. The calculations, using several angular distribution functions for particles sputtered, showed that the distribution of particles deposited slightly depends on the distribution function; therefore, the distribution function in the form $f(\Theta) = \cos \Theta$ was used.

The distribution of the components in the film depends on the distance h between the target and the plane of deposition. Therefore, the calculations were performed for different values of h. The results are given in Figs. 2 and 3, where the distribution of one
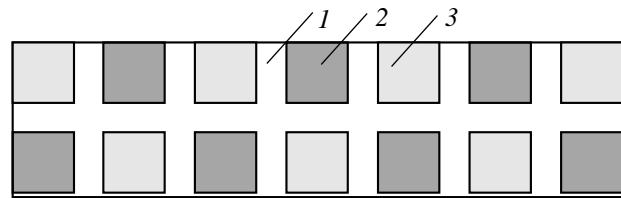

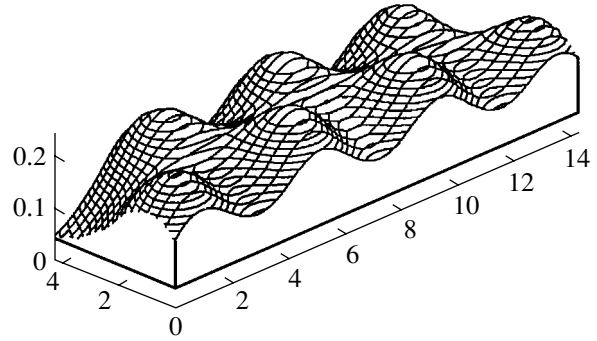**Fig. 1.** Geometry of the composite target used in simulation.

Fig. 2. Concentration of one component in the coating for the coating–cathode distance $h = 2$ cm.
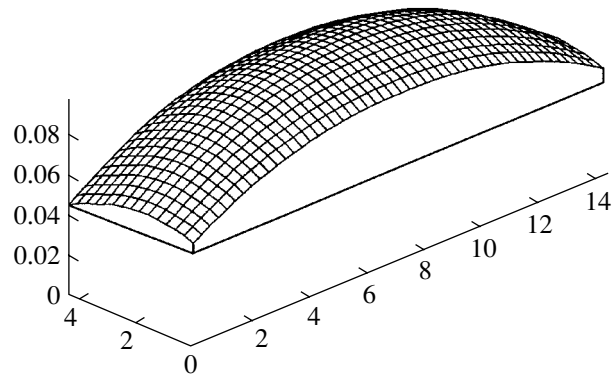

Fig. 3. The same for $h = 6$ cm.

component in the film is shown. It is seen that, at a small distance from the target, the components are deposited nonuniformly. The point is that, at a small distance from the substrate, the material of each of the islands is deposited directly underneath the insert. As the distance increases, the distribution becomes more uniform. At the same time, the amount of the material reaching the substrate decreases. Therefore, the exposure time should be increased to obtain a coating of a desired thickness. From geometric considerations, one can conclude that the material distribution in the film becomes uniform when the target–film distance R exceeds the insert spacing. At the same time, the density of the deposit is proportional to the ratio of the film area S to R^2. Hence, to minimize the deposition time, the substrate should be placed nearer to the target.

Concentration, at. %
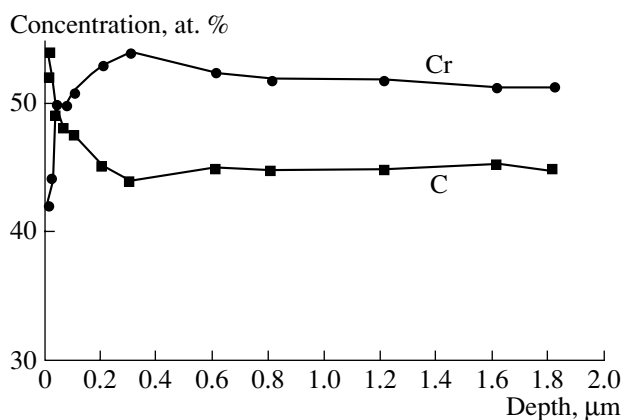
Fig. 4. Cr and C distributions in the coating. The chromium and graphite areas on the cathode relate as 50 : 50%.

Note also that the film thickness near the edges is smaller, which should be taken into account when arranging the substrates. The simulation showed the following: (1) the magnetron sputtering of the mosaic target provides the deposition of the homogeneous films; (2) to produce a film of uniform component distribution over the surface, the substrate surface should be placed at a distance far exceeding (about 3 times) the distance between the islands; (3) as the distance increases, the number of target atoms reaching the film decreases, so that the exposure time must be increased to obtain the film of a desired thickness.

EXPERIMENTAL

The experiments were carried out using a Vita ion accelerator equipped with a high-energy ion source, a plasma accelerator, and a magnetron. Metal–graphite compositions were used as the cathode. The metals were tantalum, molybdenum, niobium, vanadium, chromium, zirconium, titanium, and oxygen-free copper. The refractory metals were prepared by vacuum melting. We also used high-purity fine-grain graphite of

density 1.81–1.86×10^3 kg/m^3, which can be mechanically treated and polished.

To perform complex investigations, the coating was deposited onto different metals; glass; ceramics; and hard-alloy foils, plates, and beads with a polished surface. The TiC and TiCN, NbC, WC, CrC, MoC, and MoCN coatings were synthesized by sputtering the cathodes made of corresponding metals with graphite inclusions by Ar$^+$ ions or an Ar$^+$ + N$^+$ mixture.

Depending on the material to be sputtered, the cathode voltage was varied from 400 to 600 V; the current, in the range of 5–10 mA.

Before the deposition, the substrates were cleaned by a high-energy (300 eV) Ar$^+$ ion beam extracted from the plasma accelerator. After the deposition, the composition and the thickness of the coatings were studied with Rutherford backscattering.

The phase composition and the structure of the coatings were examined by high-energy electron diffraction with a transmission electron microscope and an electron diffractometer, X-ray diffraction, X-ray photoelectron spectroscopy, Auger electron spectroscopy, scanning electron microscopy, X-ray microspectral analysis with an SEM, and optical microscopy.

The coating thickness was estimated on the cleavage surface using a scanning electron microscope and a profilometer.

The coating adhesion to the substrate was tested by a REVETEST standard automatic tester available from the LSRH using a Rockwell C indenter. The loading rate was 100 N/min; the indenter velocity, 10 mm/min. The coating microhardness was determined using a PMT-3 hardness meter and an MI-7 microscope under a load of 0.2–2.0 N.

The wear resistance of the coatings was determined by rubbing against a loosely fixed abrasive (diamond paste on a thick paper) of granulometric composition $(10$–$15) \times 10^{-6}$ m on a friction machine at a constant load of 26 Pa. The coating thickness was found to be $(3$–$25) \times 10^{-6}$ m. The wear resistance was estimated
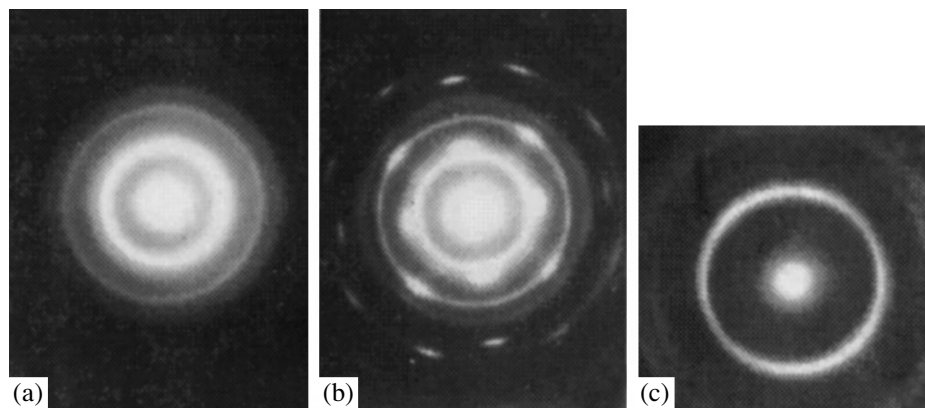
Fig. 5. Diffraction patterns for the (a) Ta$_2$C, (b) TaC, and (c) Ta$_3$C$_2$ in the Ta–C coating.

Composition and properties of coatings

Compound	Cathode composition, ratio Me : C over surface in %	Phase composition of coating (X-ray data), wt %	Lattice types and parameters, $\times 10^{-1}$ nm	Crystallite size, $\times 10^{-1}$ nm	Chemical composition, wt %	Microhardness H_V (0.2–1 H), GPa
(WG)	50W + 50C	$WC_{0.97} = 96.4 \pm 0.4$ $C = 3.6 \pm 0.4$	hcp $a = 2.907$ $b = 2.839$	30–40		16.00–27.00 (~24)*
(7GTa) (13TaG)	50Ta + 50C	$TaC_{0.97-0.1} = 97.6 \pm 0.4$ $C = 3.4 \pm 0.4$	fcc $a = 4.44$	60–80		15.30–27.00 (15.7)
(6MoG)	70Mo + 30C	$Mo_2C = 67.6 \pm 0.4$ $Mo = 32.4 \pm 0.4$ Closer to β-Mo_2C, high-temperature	hcp $a = 3.01$ $c = 4.74$	Texture 22–50	$Mo(tot) = 77.6 + 2.3$ $C(tot) = 2.3 \pm 0.3$ $Mo(fr) = 38.6 \pm 2.3$	13.00–30.00
(18MoG) (19MoG)	50Mo + 50C	$MoC_{0.95-1.0} = 95.6 \pm 0.4$ $C = 4.2$	fcc α-MoC $a = 4.27$	60–80	$Mo(tot) = 82.6 \pm 2.6$ $\quad\quad\quad 84.5 \pm 0.6$ $C(tot) = 15.5 \pm 0.6$ $\quad\quad\quad 16.2 \pm 0.1$ $C(fr) = 1.4 \pm 1.0$ $\quad\quad\quad 5.6 \pm 1.5$	18.00–31.00
(11NbC)	70Nb + 30C	Nb_2C	Hexagonal $a = 3.115$, $c = 4.948$	Texture 10–30	$Nb(tot) = 2.6 \pm 0.7$ $C(tot) = 6.07 \pm 1.5$	23.00–29.00 (24)
(NbG)	50Nb + 50C	$NbC_{0.95-1.0}$	fcc $a = 4.436$	Texture 20–50	$Nb(tot) = 84.3 \pm 3.0$ $C(tot) = 11.45 \pm 0.8$	18.90–23.00 (24)
(10GV)	50V + 50C	$VC_{0.85-0.88}$	fcc $a = 4.188$	15–50	$V(tot) = 84.3 \pm 3.0$ $C(tot) = 9.5 \pm 0.5$	13.00–26.00 (28)
(8GCr)	50Cr + 50C	$Cr_3C + Cr$ Closer to Cr_6C phase	Rhombic $a = 2.86$, $b = 5.58$, $c = 11.48$	80–120	$Cr(tot) = 84.3 \pm 2.8$ $C(tot) = 13.4 \pm 0.2$ $C(fr) = 0.59 \pm 0.05$	17.00–27.00
(2TiG)	50Ti + 50C	$TiC_{0.95-0.98} = 98.5 \pm 0.4$	Cubic, type V-1 4.330	Texture 111 30–50		21.50–37.40 (29.3)
(3TiG)	80Ti : 20C	$TiC_{0.5-0.6} = 68.6 \pm 0.4$ $Ti = 31.4 \pm 0.4$	Cubic, type V-1 4.334	80–120		11.00–22.00
(4ZrG)	50Zr : 50C	$ZrC_{0.9-0.95} = 96.5 \pm 0.4$	Cubic, type V-1 4.677	50–60		18.50–30.00 (28.6)
(5ZrG)	80Zr : 20C	$ZrC_{0.5-0.6} = 76.5 \pm 0.4$ $Zr = 23.5 \pm 0.4$	Cubic, type V-1 4.683	Texture 120–180		10.00–21.00

Note: tot, total content; fr, in free state; * refers to tabulated value.

from the rate of linear wear and friction track coloration.

The friction coefficient was determined using a sliding friction machine as follows: a coated rotating disk rubbed against an uncoated fixed ball under dry friction conditions in air.

The chemical composition of the coatings, i.e., the content of the metal and carbon in bound and free states, was determined by specially developed techniques with a Leco C-212 AH-7529 analyzer.

The thermal oxidation resistance in air and in a pure oxygen flow was tested by the method of thermogravimetric oxidation using a Setaram instrument (France).

The corrosion resistance of the coatings was estimated electrochemically using a PS4 potentiostat by plotting anode potentiodynamic polarization curves at a
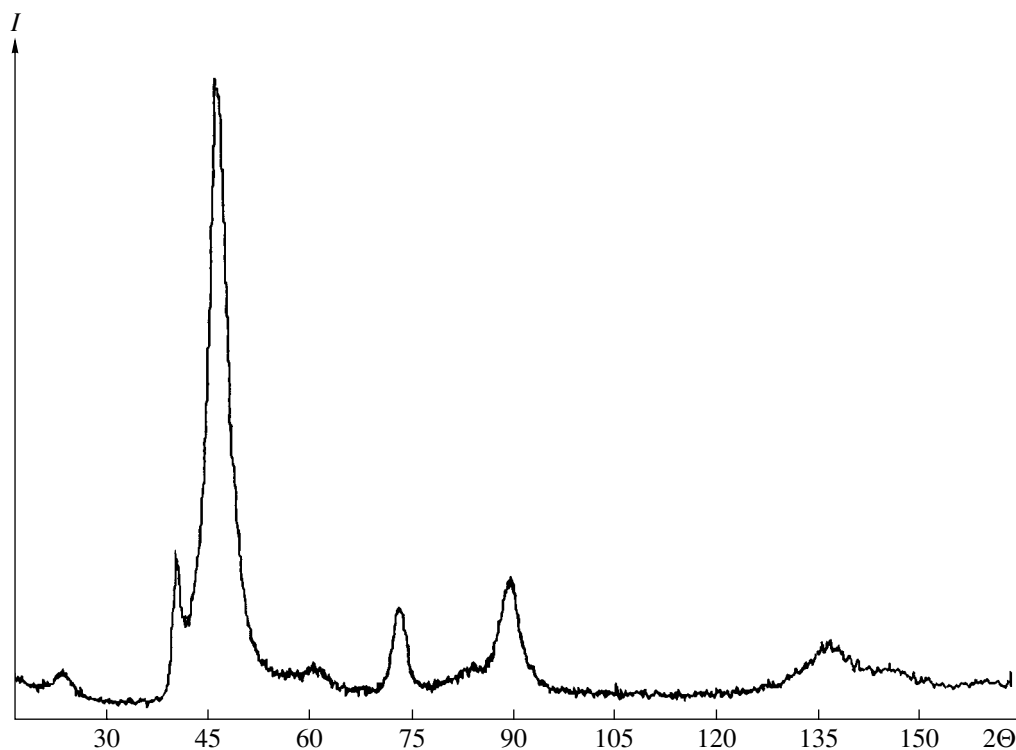
Fig. 6. X-ray reflections for the Mo–C coating. The phases are Mo$_2$C with a crystallite size of 4.8 nm, Mo$_4$C (5.1 nm), and a small amount of MoC (2.29 nm).

rate of 7.8 V/h in H$_2$SO$_4$ and HCl solutions at room temperature.

EXPERIMENTAL RESULTS

For discharge power densities from 5×10^5 to 11×10^5 W/m^2, we obtained smooth homogeneous condensates with metallic luster on different substrates. On the polished and glass substrates, the condensate had the specular surface, which is typical of refractory metal layers.

The investigation of the phase composition and the structure of the coating shows that when the mosaic cathode is sputtered, the metal and carbon atoms homogeneously mix to form carbides, if the metals interact with carbon, or pseudosolid carbon solutions in metals otherwise. All the Group IVA–VIA refractory metals studied in this work interact with carbon at temperatures no more than 473 K to form carbide phases, including high-temperature phases (see table). The formation of these phases is also confirmed by photoelectron energy spectrum analysis. The variation in the element composition of the coating across the depth is typical of almost all of the compositions. The Auger spectroscopy data for Cr–C, Ta–C, and Mo–C demonstrate that ion etching causes the C concentration to decrease and the metal concentration to increase (Fig. 4). Apparently, this results from graphite diffusion into the coating and its segregation at the surface.

The measurement of residual stress in the 6MoG, 19MoG, and 8GCr samples (see table) by the X-ray diffractometer equipped with a special ψ attachment shows that in the Mo phase, compressive stresses in the coating plane are ≈ 0.3 GN/m^2 (6MoG), while in MoC phases, they are 0.19 GN/m^2 (19MoG). In both cases, the coatings were deposited on carbon ball bearing steel.

The estimation of the adhesion for metal–carbon coatings as a whole (both qualitative and quantitative) shows that the highest adhesion (about the mean carbide tensile strength, 0.18–0.25 GPa; incomplete detachment: film islands remain on the substrate) is observed for the ductile substrates (such as copper, nickel, low-carbon and stainless steel). The adhesion was 1.5–2 times lower for the substrates made of the hard allow, high-speed and ball bearing steel, as well as of glass and ceramics. This is especially true for the coatings of a thickness $>5 \times 10^{-6}$ m.

The reflection and transmission electron microscopy studies of the coating surface confirm the presence of the ultrafine-grain nearly amorphous structure, as evidenced by the diffraction patterns (Fig. 5).

X-ray reflections for the Mo–C coating are shown in Fig. 6. Their identification shows that the coating is multiphase, containing Mo$_2$C with a crystallite diameter of 4.8 nm, Mo$_4$C (5.1 nm), and a small amount of
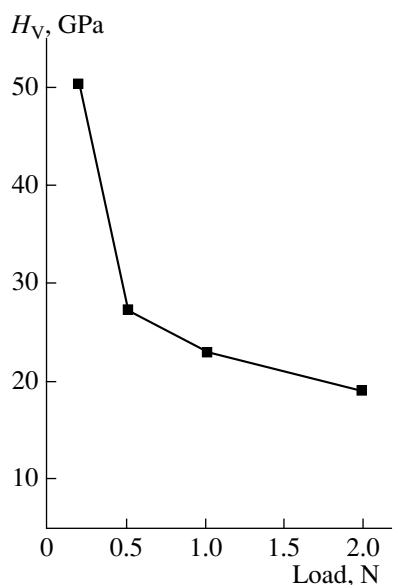
Fig. 7. Microhardness vs. load for the Mo–C coating.

MoC (2.29 nm). One can conclude that the molybdenum carbides have the nanocrystalline structure.

As the carbon content decreases and phases, such as Ta_2C, Nb_2C, and Mo_2C appear, i.e., when hcp MeC subcarbides form instead of the fcc MeC carbides (NaCl-like structure), the structure of coatings sharply changes. The lines 200, 220, and 111, which are typical of the fcc lattice with a parameter of 0.444 nm, are successively smeared. The ultrafine-grain nanocrystalline structure of the coatings provides high resistance to oxidation and corrosion. They also offer improved mechanical properties: hardness and wear resistance. The Mo–C coating is the hardest. For this coating, the microhardness as a function of load is shown in Fig. 7. The surface layer is superhard at a load of 0.2 N, $H_V = 50$ GPa.

The wear tests using a diamond paste showed the coatings with the composition $Me_2C + Me_{C_B}$ and $MeC + C_{C_B}$ offer the highest wear resistance.

Thus, it is shown that the co-sputtering of a metal and carbon provides their homogeneous mixing. The resulting deposits may contain both bound and free metal and carbon atoms. The coatings have the ultrafine-grain near-amorphous nanocrystalline structure, and this state is retained up to an annealing temperature of ≈ 1000 K.

In comparison with crystalline materials of this class, the metal–carbon coatings with the nanocrystalline structure offer higher corrosion resistance in a number of acids, higher oxidation resistance in air, lower friction coefficient, and higher wear resistance. The MoC coating is superhard ($H_V = 50$ GPa at 0.2 N).

ACKNOWLEDGMENTS

REFERENCES

1. P. J. Kelly and R. D. Arnell, Vacuum **56**, 159 (2000).
2. J. Musil and J. Vicek, in *Proceedings of the 5th Conference on Modification of Materials with Particle Beams and Plasma Flows, Tomsk, 2000*, p. 393.

Translated by M. Astrov

SURFACES,
ELECTRON AND ION EMISSION

Tip Field Emitters Coated with Fullerenes

T. A. Tumareva, G. G. Sominskiĭ, A. A. Efremov, and A. S. Polyakov

St. Petersburg State Technical University,
ul. Politekhnicheskaya 29, St. Petersburg, 195251 Russia
e-mail: tumareva@rphf.spbstu.ru, sominski@rphf.spbstu.ru
Received July 6, 2001

Abstract—Formation and characteristics of fullerene coatings on the surface of tungsten tip field emitters and emitters with ribbed crystals formed on their surface are studied. It is found that electric fields and heat treatments affect the structure and emission characteristics of the fullerene coatings. Methods of creating microprotrusions on the surface of the coatings that considerably enhance the electric field have been developed and tested, and emitters with a single microprotrusion yielding emission current densities up to 10^6–10^7 A/cm^2 are demonstrated. Emitters with fullerene coatings in the form of a distributed microcluster structure can yield currents in excess of 100 μA from a single micron-sized tip in a stationary regime. © *2002 MAIK "Nauka/Interperiodica".*

INTRODUCTION

One of the important unresolved tasks of vacuum electronics is the development of long-service-life field emitters for producing intense electron beams under technical vacuum conditions. It is known (see, for example, [1–3]) that coatings containing carbon are, as a rule, resistant to ion bombardments and the harmful effects of gaseous ambients and can be used as protective coatings for increasing the service life of field emitters. Carbon, as a readily available and inexpensive material, appears to be an obvious choice for such coatings. However, a protective coating should at the same time ensure high efficiency of the field emitter. Because of the large work function value of carbon materials (for example, for graphite $e\varphi \sim 4.7$ eV) there appears to be only practical means for obtaining high currents at low voltages with tip emitters coated with carbon: microprotrusions should be formed at the emitter surface to enhance the electric field. Standing in the way of the wide use of carbon are problems relating to preparation of this sort of structure and implementation of the process of depositing this high-temperature material. It might seem that good results for coatings with microstructures enhancing the electric field could be expected from the application of nanotubes [3, 4], but, so far, acceptable current characteristics for tip cathodes with nanotubes could not be obtained because a simple enough technology for forming ordered structures of nanotubes is not yet available. Our purpose is to study the possibility of using C_{60} fullerene molecules as a field emitter coating. This material has not been used earlier for such applications. Our presumption is that the large spherical C_{60} molecules are better suited for creating field-enhancing microstructures on the surface of refractory materials than nanotubes are. The very first data obtained [5, 6] mostly supported these expectations, and research was continued. This study summarizes the results of our research on fullerene coatings on the surface of tip field emitters of tungsten.

EXPERIMENTAL TECHNIQUE

The experiments were performed in a field emission projector. The fullerene coatings were deposited onto tip field emitters from a Knudsen cell [7]. The coating structure was determined from emission images of the tip apex observed on the projector screen at magnifications of 10^5–10^6. To reveal the effect of the coating on the performance of the field emitters, the current–voltage characteristics of the entire tip and of separate areas on its surface were measured. In addition, values of voltage U_I corresponding to certain fixed field emission currents I were monitored. Characteristic features of the U_I dependence of deposition time ("deposition curves") were used to determine the approximate coating thickness or identify the moments of time at which structural transformations occurred corresponding to the emergence (or disappearance) on the emitter surface of microprotrusions enhancing the electric field.

Images on the projector screen were registered with a sensitive TV camera, then digitized in a special controller and transferred to a PC for storage and subsequent processing. This system could reliably register images obtained with extremely low currents from the tip, down to $I \sim 0.01$–0.02 μA, at a rate of 24 images per second. The high sensitivity and speed of this image registration system afforded the possibility, where necessary, of minimizing the effect of the electric field on the coating structure during registration, even in a stationary regime. The effect of the electric fields on the emitter characteristics being registered could be further reduced using a regime of short (~4 μs) pulses at a low controllable repetition rate, $f \leq 400$ Hz. With reduced effect from electric fields during the measurements, the
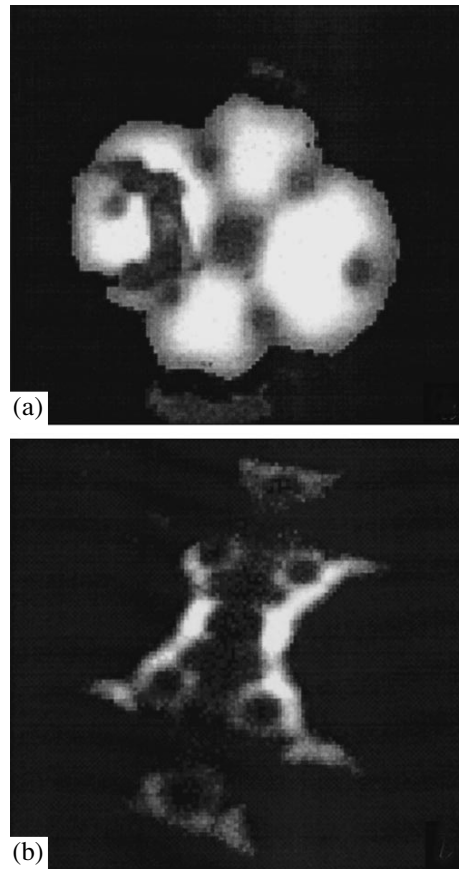
effect of the field specially applied for coating treatment could be better understood.

The effect on the field emitter parameters of interest of the quantity of deposited substance (an approximate coating thickness Θ of up to 50–100 monolayers), substrate temperature T (from room temperature up to 2800 K), and the electric fields created by application of voltage between the cathode and the screen (up to approximately 10^8 V/cm) was studied. Most of the experiments were carried out at a residual gas pressure in the experimental setup of ~ 10^{-10}–10^{-9} torr.

Before deposition of the fullerene coating, a single crystal was formed on the apex of a tungsten tip. Heat treatment at $\geq$950–1000 K of the deposited fullerene layers caused decomposition of the fullerene molecules and formation on the tip apex of a so-called ribbed crystal having a layer of tungsten carbide on its surface [8]. Transition from the ribbed crystal back to pure tungsten single crystal ("purification" of the emitter) was performed by heating the tip to temperatures just below the limiting temperature or by heating to lower temperatures after a prior considerable (by a few orders of magnitudes) increase of the pressure in the setup. Representative images of the tungsten tip and the ribbed crystal are shown in Fig. 1. The tungsten tip emitters studied had a tip apex diameter of approximately 0.3 to 2 μm. Formation of the fullerene coatings deposited both on the tungsten single crystal and the ribbed crystal was studied. Although the measurement technique used introduced a slight perturbation, the main results presented below were obtained in the stationary regime. Therefore, below unless otherwise indicated, the regime is assumed stationary.

Fig. 1. Images of emission from (a) tungsten crystal and (b) ribbed crystal. On the left side in frame (a) (and some subsequent emission images as well), a shadow of the field emitter holder can be seen.

EXPERIMENTAL RESULTS AND DISCUSSION

1. Characteristics of the fullerene coatings formed without applying a strong electric field. As found in the experiments, if a coating is deposited with no electric field about the tip, the layer of fullerene molecules formed both on pure tungsten crystals and ribbed crystals in early deposition stages is fairly uniform over the surface, without pronounced protrusions. As its thickness grows, the current to the screen at fixed voltage U diminishes because of the increasing work function of the surface and, for the same reason, the voltage U_I corresponding to the fixed current I increases. Further variations of U_I depend essentially on the registered magnitude of the field emission current I. This is illustrated in Fig. 2, where "deposition curves" $u_I(t)$ (dependences on deposition time t of the quantity u_t, which is a quotient of the voltage U_I measured at the time moment t and its initial value $U_I(0)$ at time $t = 0$) are compared. Curves 1 and 2 were obtained at currents 0.02 and 0.25 μA, respectively. Deposition characteristics for a ribbed crystal are shown also in Fig. 3, where $u_{0.02}(t)$ curves measured at the same current $I = 0.02$ μA

are compared for different fullerene deposition rates (different heating of the Knudsen cell).

Setting aside for the time being the reasons for changes of the deposition curves with emission current, let us make some inferences concerning characteristics of the fullerene coatings from the data obtained at $I = 0.02$ μA, i.e., under conditions of minimum influence of the electric fields involved in the measurement process. Under these conditions, u_I, after attaining a certain peak value u_{Im}, slowly diminishes with increasing coating thickness in the case of a ribbed crystal and stays practically unchanged for a long time for deposition on pure tungsten. $u_I(t)$ characteristics measured at different deposition rates are qualitatively similar and differ mainly in the time required for reaching u_{Im}. The time t_m diminishes with deposition rate. Usually the deposition rate is determined from deposition curves. The shape of deposition characteristics in Figs. 2 and 3 is typical of field emission systems where the work function increases with the coating thickness. The approach of the voltage u_I towards the maximum u_{Im} of the deposition curves might be related to formation of a monolayer coating. If this is the case, then from the charac-
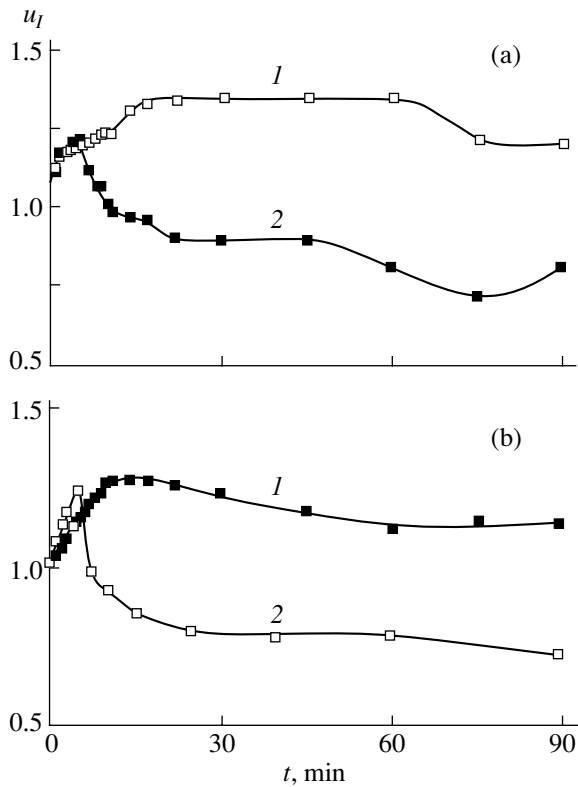
Fig. 2. Normalized deposition curves u_I for (a) pure tungsten and (b) ribbed crystal. Curves *1* and *2* were obtained at $I = 0.02$ and 0.25 μA, respectively.
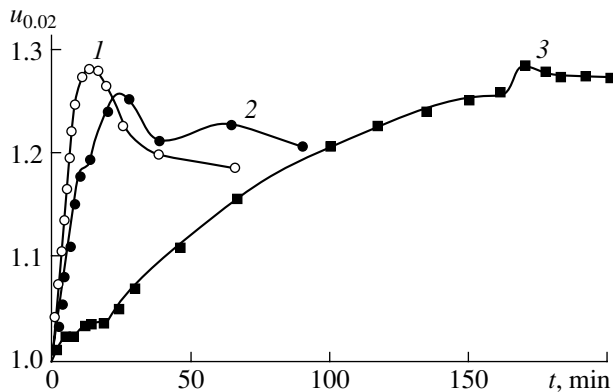
Fig. 3. Deposition curves $u_{0.02}(t)$ for ribbed crystal at $I = 0.02$ μA and different deposition rates of fullerene coatings (different temperatures of the Knudsen cell). (*1*) $T = 600$; (*2*) 580; and (*3*) 550 K.

teristics given in Fig. 3, it can be inferred that a monolayer coating is formed on a ribbed crystal in times of ~15 (curve *1*), ~30 (curve *2*), and ~175 min (curve *3*). For the same reasons, at the given deposition rate, a monolayer coating on pure tungsten is formed in a time of ~20 min (Fig. 2a)

Comparison of the $U_{0.02}$ values necessary for obtaining a specified fixed current from pure tungsten and

tungsten with a fullerene coating of a thickness of $\Theta \sim$ 1.5–2 monolayers (mL) yields the value of the work function $e\varphi$ of the fullerene coating. From the data obtained it follows that the work function equals approximately 5.5 eV.

In our opinion, estimates of the deposition rate and work function values are possible only under conditions where the coating is fairly uniform and there are no highly localized regions of enhanced electric field as a result of formation of 3D microstructures. Such conditions, even with the weakly perturbing measurement technique employed, could be obtained only for coatings of a thickness of $\Theta < 2$–4 mL. Figure 4 shows representative emission patterns for different stages of the fullerene coating deposition onto a tungsten tip. The experimental observations indicate that at coating thicknesses $\Theta \sim 0.2$–1.1 mL, only insignificant emission nonuniformities in the form of separate spots exist. Nonuniformities become more pronounced as the coating thickness increases; however, up to $\Theta \sim 2$–4 mL, their contribution to the total emission current from the tip is relatively small. For this reason, the initial distribution of emission from the tungsten tip is well discernible. Only on thick coatings bright emission spots are observed from emerging microprotrusions, which significantly enhance the electric field and obscure the image of the substrate.
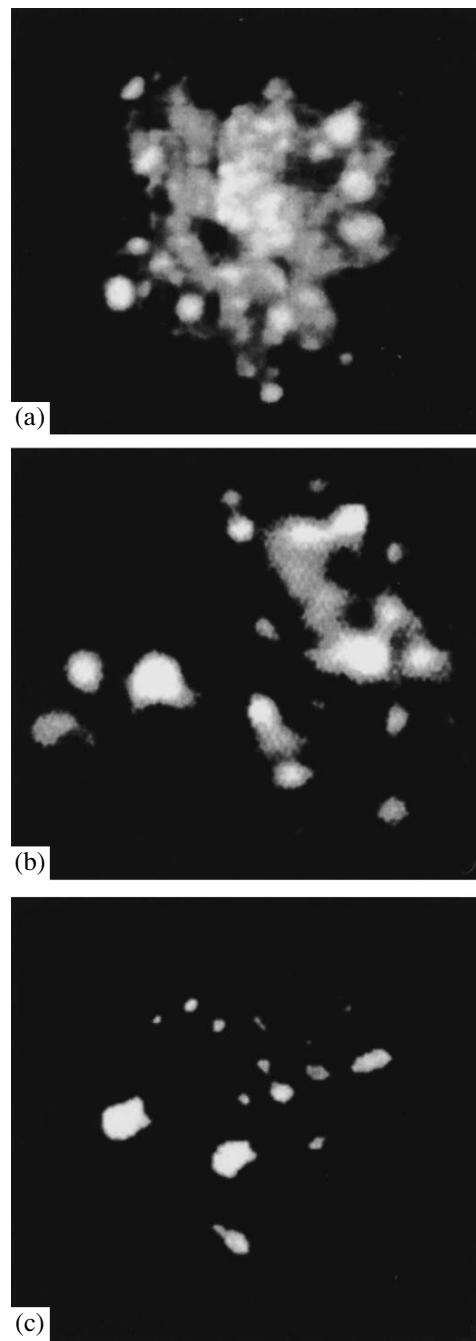
2. Effect of the electric field on characteristics of fullerene coatings. The above-mentioned formation of 3D microprotrusions on the fullerene coating can be caused by even those relatively weak perturbations from the electric field applied briefly for obtaining an emission image on the projector screen. This supposition is supported by considerable changes of the form of deposition curves with increasing registered current (compare characteristics *1* and *2* in Fig. 2). Probably, the image control at high emission currents (and voltages) causes reconstruction of the coating already in the course of the measurement process. In Fig. 5, a series of emission images is shown that were obtained at an emission current of 0.25 μA. During observations in this regime, microprotrusions on the tungsten tip surface contributing considerably to the emission arose at coating thicknesses as small as $\Theta \leq 0.5$ mL. At large thicknesses, the nonuniformities grow to such an extent that already at $\Theta = 1.7$ mL the substrate cannot be distinguished. The field enhancement at nonuniformities is what causes the abrupt decline of the deposition curves measured at the current of 0.25 μA compared with curves measured at 0.02 μA.

These data are not the only evidence of the effect of electric field. Special experiments were carried out in which the obtained coatings were treated in an electric field. The dynamics of the emission characteristics of a fullerene coating under applied electric field can be seen in Fig. 6. An emission system was formed in the measurement process of curve *1* in Fig. 2b after 90 min of the coating deposition on a ribbed crystal. At this
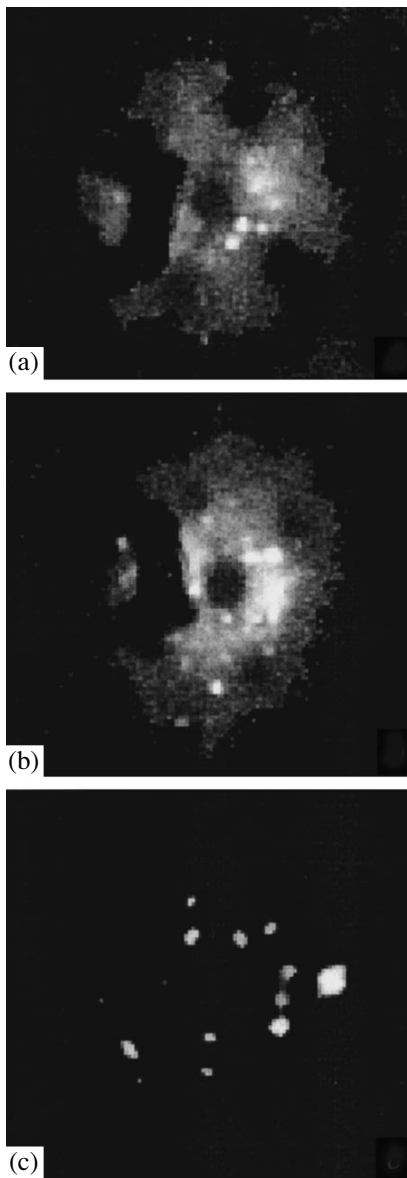
Fig. 4. Emission images of a tungsten tip registered at $I = 0.02$ µA; Θ: (a) 0.2; (b) 1.1; and (c) 5 mL.

**Fig. 5.** Emission images of the tungsten tip registered at $I = 0.25$ µA; Θ: (a) 0.5; (b) 1.7; and (c) 4.5 mL.

moment, the voltage was increased to $U_{0.25}$, a value corresponding to the current $I = 0.25$ µA. Figure 6 shows the dependence of the voltage quotient $u_{0.25}(t) = U_{0.25}(t)/U_{0.25}(0)$ on the time in the electric field starting from the moment ($t = 0$) when the voltage was applied. Rapid drop of this specific quantity is due to formation after voltage increase of the number of centers of intense emission on the cathode surface. In time, some of these centers break away from the tip (or disintegrate), which makes the $u_{0.25}(t)$ curve nonmonotonic. The microprotrusions on a ribbed crystal coating surface tend to accumulate near the ribs, i.e., in areas of initially enhanced field (Fig. 7a). Similar transformations of the deposition curves have been observed after

field treatment of fullerene coatings deposited on pure tungsten.

Summarizing the data obtained in studies of the effect of electric field on the structure of a fullerene coating, the mechanism of this effect appears to be as follows. In highly nonuniform electric fields, C_{60} molecules can be polarized and the molecular dipoles attracted to the regions of enhanced electric field emerging as a result of the roughness or microprotru-
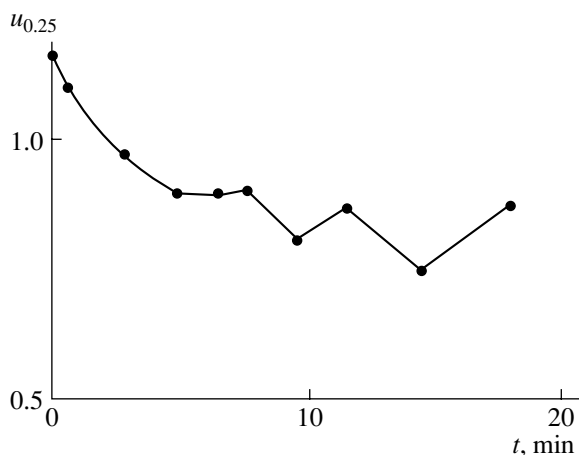
Fig. 6. Time variation of the quantity $u_{0.25}$ under applied electric field ("field treatment").

sions on the initial surface. Completion of small non-uniformities enhances the electric field still further, and the emission current from them increases. Efficient growth of microprotrusions on the emitting surface starts, according to our estimates, at electric field strengths of $E \sim (1–5) \times 10^7$ V/cm.

It should be noted that the formation of microprotrusions on the fullerene coating surface has been observed not only in the stationary regime but in pulsed measurements as well. Formation of these nonuniformities only occurred at much higher collected currents (and electric field strengths) than in the stationary regime, ≥ 10 μA. So far, not enough data has been accumulated on processes in the pulsed regime. Therefore, the temporal characteristics of the growth of microprotrusions cannot be clearly recognized as yet.

3. Formation of stable emitting structures on the surface of fullerene coatings. As mentioned above, field treatment of the fullerene coatings on tungsten can be used to create microformations enhancing the electric field at the surface and considerably reduce the voltages required for reaching a given current. However, emitters with nonuniformities of such type do not show stable emission even at currents $I \leq 1$ μA.

These studies demonstrated that fullerene microstructures displaying stable emission are possible to create only using ribbed crystals having a layer of tungsten carbide on the surface. It has been found that on thick enough coatings ($\Theta > 5$–10 mL), separate microprotrusions initially accumulating on the ribs (along [111] zonal lines) in an applied electric field combine and form one large intensively emitting center in the region of {023} planes close to the central (110) plane. Sometimes this effect takes place precisely during observations or while a Fowler–Nordheim characteristic is being measured. In Fig. 7, screen emission patterns are shown before (Fig. 7a) and after (Fig. 7b) the merging of separate microprotrusions into one emitting
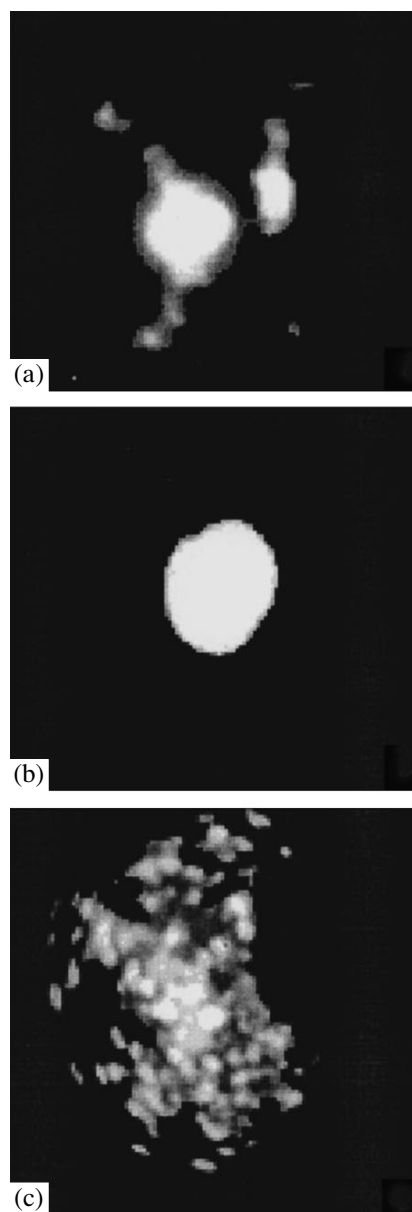
Fig. 7. Images of emission from the ribbed crystal with fullerene coatings of various types: (a) several microprotrusions on ribs; (b) a separate efficiently emitting center; and (c) a distributed microcluster structure providing emission currents of >100 μA.

center. Formation of such a center results in a considerable (up to 50%) reduction of U_I from its value for the initial ribbed crystal. Initially, the unstable spot stabilized after brief (1–5 min) heating at a temperature of about 950 K, which is close to the fullerene molecule decomposition temperature. Prolonged (≥ 20–30 min) heating at this temperature caused its destruction; the image returned to that of the original ribbed crystal. An emitting center treated at ~950 K yielded a current of ~10 μA. The estimated current density reached ~10^6–10^7 A/cm^2.

Substantially different coatings in regards to their structure and emission properties could be obtained on a ribbed crystal after a special treatment of its surface consisting in repeated cycles of deposition/thermal decomposition of fairly thick deposited layers ($\Theta > 5$–10 mL). After two to four cycles of the surface treatment under an electric field of a strength necessary for producing a current of a few microamperes, numerous (a few tens) 3D clusters formed. The cluster structure (Fig. 7c) produced in this way was stable under applied electric field, and in the stationary regime a micron-size tip could yield currents of more than 100 μA. Such structures are resistant to gaseous media. Exposure to vacuum for several days produced practically no changes in their properties. Measurements have shown that in pulsed regime currents from a field emitter with the fullerene coating could be increased severalfold.

Physical phenomena that take place during the procedure required for creating the above-described efficiently emitting distributed structures on the surface of a ribbed crystal is as of yet not quite clear. At this stage it can only be supposed that the multiple deposition/decomposition acts to increase the thickness of the transitional layer of tungsten carbide (or tungsten carbide with inclusions of carbon) at the base of the multicluster system being formed.

CONCLUSIONS

In summary, the main results of this work are as follows.

(i) The mechanism has been established by which fullerene coatings form on the surface of tungsten tip field emitters, and their characteristics have been determined. It was found that electric field affects the structure and the emission characteristics of the fullerene coating on the surface of tungsten tips and ribbed crystals formed by thermal decomposition of fullerene molecules on tungsten.

(ii) It has been found that thermal and field treatments produce isolated microprotrusions on the fullerene coatings, which considerably enhance the field and ensure stable field emission with high current densities of up to 10^6–10^7 A/cm^2 in the stationary regime.

(iii) The possibility has been demonstrated of fabricating emitters with fullerene coatings in the form of a distributed multicluster system yielding stable currents exceeding 100 μA from a single micron tip in the stationary regime.

REFERENCES

1. B. V. Bondarenko, V. A. Seliverstov, A. G. Shakhovskoĭ, and E. P. Sheshin, Radiotekh. Élektron. (Moscow) **32**, 199 (1987).

2. N. Chubun, N. Lazarev, E. Sheshin, and A. Suvorov, in *Abstracts of the 42th IFES, Madison, 1995.*

3. Yu. V. Gulyaev, Yu. A. Grigor'ev, N. I. Sinitsyn, *et al.*, in *Proceedings of All-Russia University Conference "Modern Problems of Electronics and Radiophysics of Ultra-High Frequensies"* (Kolledzh, Saratov, 1997), p. 90.

4. N. I. Sinitsyn, Yu. V. Gulyaev, G. Torgashov, *et al.*, Appl. Surf. Sci. **111**, 145 (1997).

5. G. G. Sominski, T. A. Tumareva, A. S. Polyakov, and K. K. Zabello, in *Proceedings of the International University Conference "Electronics and Radiophysics of Ultra-High Frequensies"* (Nestor, St. Petersburg, 1999), p. 327.

6. T. A. Tumareva and G. G. Sominskii, J. Commun. Technol. Electron., Suppl. 1 **45**, S110 (2000).

7. V. I. Karataev, Pis'ma Zh. Tekh. Fiz. **24** (5), 1 (1998) [Tech. Phys. Lett. **24**, 167 (1998)].

8. M. V. Loginov and V. N. Shrednik, Pis'ma Zh. Tekh. Fiz. **24** (11), 45 (1998) [Tech. Phys. Lett. **24**, 432 (1998)].

Translated by B. Kalinin

Formation on Field Emitters Coated with Fullerenes of Microformations Producing Ordered Emission Images

T. A. Tumareva, G. G. Sominskiĭ, and A. S. Polyakov

St. Petersburg State Technical University,
ul. Politekhnicheskaya 29, St. Petersburg, 195251 Russia
e-mail: tumareva@rphf.spbstu.ru, sominski@rphf.spbstu.ru
Received July 6, 2001

Abstract—Field emission projector studies of fullerene coatings deposited on tungsten tip field emitters reveal specific ordered patterns in the form of doublets, quadruplets, rings, disks, and other forms in the emitter images. The ways in which these types of ordered emission patterns arise and their relation to C_{60} microformations have been established. Possible causes of the emergence of the ordered emission images are analyzed on the basis of published data and experimental results obtained. A modification of the model of field emission from the surface of microformations taking into account internal reflection of the electronic waves from the formation boundaries has been proposed. © *2002 MAIK "Nauka/Interperiodica".*

INTRODUCTION

In the 1950s and 1960s, in studies of tip field emitters using a field emission projector, a number of researchers discovered an unusual phenomenon. After deposition onto an emitter of large organic molecules (such as copper phthalocyanine and anthraquinone), carbon clusters, and some other materials, ordered structures in the form of two- or four-petal figures ("doublets" and "quadruplets"), rings, disks and other forms appeared in the emission images (see, for example, [1–6]). The most detailed studies of the formation of such emission patterns, termed by E.W. Müller "molecular patterns," were performed on emitters with coatings of organic molecules. In particular, characteristic dimensions of the complexes of copper phthalocyanine molecules responsible for the molecular patterns were determined [2, 5, 6]. According to [2], 13 molecules are sufficient for the emergence of such pictures. Savchenko [6] estimated that for a cross-sectional dimension of ~16–22 Å, the copper phthalocyanine complexes have a height of ~70 Å; their long side is oriented along the electric field lines; and they include a considerably larger number of molecules, about 50. For other materials, including carbon, little reliable data have been accumulated. There are indications that carbon clusters responsible for producing such images are formed not on the field emitter surface but in the evaporator [6]. This is plausible but gives no information on the carbon cluster structure. Regrettably, information of their dimensions is also lacking.

Two theories have been developed to explain the nature of molecular pictures. In one of these (monomolecular) [2], their shape is related to the spatial and electronic structure of the large adsorbed molecules. In a waveguide model [3, 6] the same pictures are explained taking into account propagation of electronic waves (de Broglie waves) along a sort of waveguide formed by the molecular complex.

Authors of this work, in studies by the method of field emission projector coatings of tungsten tip emitters consisting of C_{60} fullerene molecules [7, 8], found that in this case emission images with ordered structure similar to those described in [1–6] can be observed on the projector screen. The results obtained made it possible to explain the process of emergence and characteristics of the complexes of C_{60} molecules producing the ordered emission images. On this basis, the formation model of the ordered field emission patterns from the emitter surface was elaborated. Below, the results of this study are presented.

MEASUREMENT TECHNIQUE AND EQUIPMENT

Detailed description of the design of a field emission projector and methods of depositing and studying the fullerene coatings on the field emitter surface have been given elsewhere [7, 8]. Therefore, we consider here only the details of the experimental technique and equipment to be taken into account for better understanding of the results of this work.

Emission images of the surface of tip field emitter magnified by 10^6 was obtained on the projector screen by applying voltage U between the screen and the cathode. Tungsten tip emitters had a tip apex diameter of about 0.3 to 2 μm. On the tip apex, a tungsten single crystal or a ribbed crystal with a coating of tungsten carbide on its surface was formed [9]. The coatings were deposited onto the tip by evaporation of C_{60} molecules from the Knudsen cell. Preparation of a ribbed
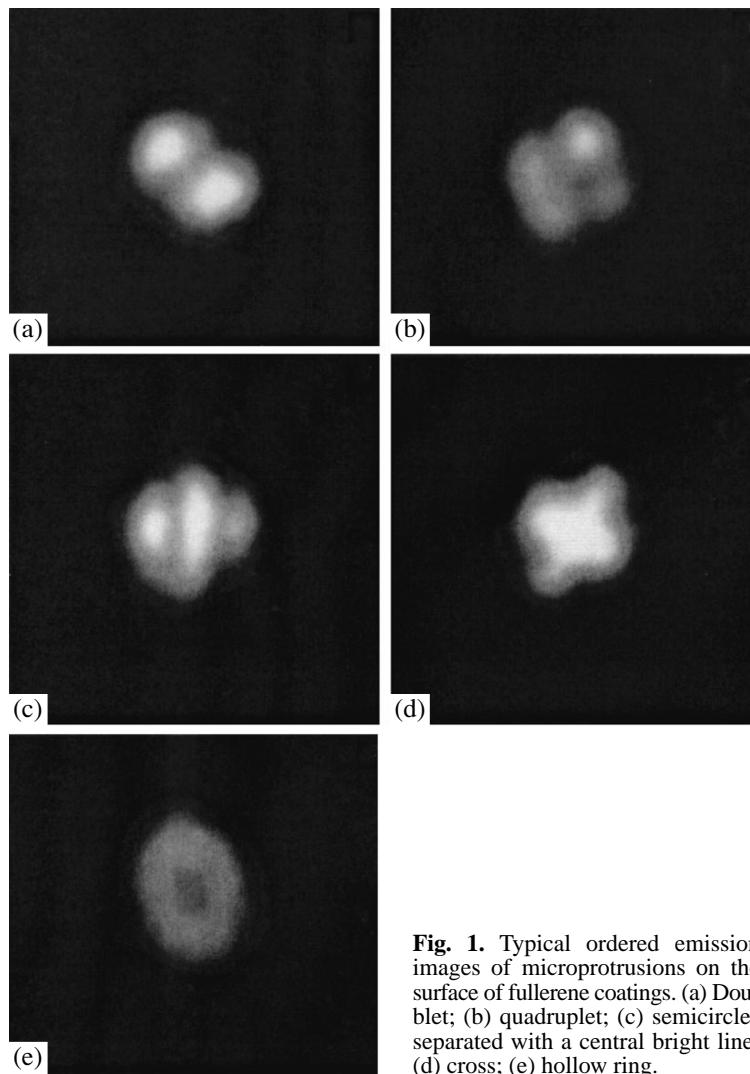
Fig. 1. Typical ordered emission images of microprotrusions on the surface of fullerene coatings. (a) Doublet; (b) quadruplet; (c) semicircles separated with a central bright line; (d) cross; (e) hollow ring.

crystal on the tungsten tip apex was achieved by thermal decomposition at temperatures $T \geq 950$–1000 K of fullerene coatings deposited earlier. The deposition rate of the fullerene coating and its thickness Θ were determined by a technique described in [8], from measured "deposition curves," which represented variations with time t of the voltage U_I necessary for obtaining a fixed current I from the tip. As a rule, the deposition of coatings was done in the absence of an electric field about the cathode. In order to determine the effect of electric fields, variations of the structure and emission characteristics of the fullerene coatings were then examined for different fixed currents from the cathode.

Most of the experiments were carried out at a residual gas pressure of $\sim 10^{-10}$–10^{-9} torr in the experimental setup.

EXPERIMENTAL RESULTS

Ordered patterns in the emission images from the tips similar to those described in [1–6] could be observed at fullerene coating thicknesses of $\Theta > 0.1$–0.2 monolayers (mL). Magnified images of typical ordered patterns are shown in Fig. 1. Figure 2 shows the general appearance of the emission images observed at different ribbed crystal coating thicknesses. In the general view of the field emitter emission image, the features of the formations of this sort are hard to distinguish; however, at moderate coating thicknesses (less than ~2–4 mL), these images carry information on their locations relative to the characteristic details of the substrate.

It was noted that the cross-sectional dimensions and the brightness of a spot of ordered structure on the projector screen sometimes varied during observation. A decrease in their brightness and dimensions can be related, for example, to reduction, as a result of ionic bombardment, of the number of fullerene molecules in a microformation on the tip surface, which produces a given image. The reverse process of increasing brightness and dimensions is probably caused by additional C_{60} molecules joining the structure. An increase in the
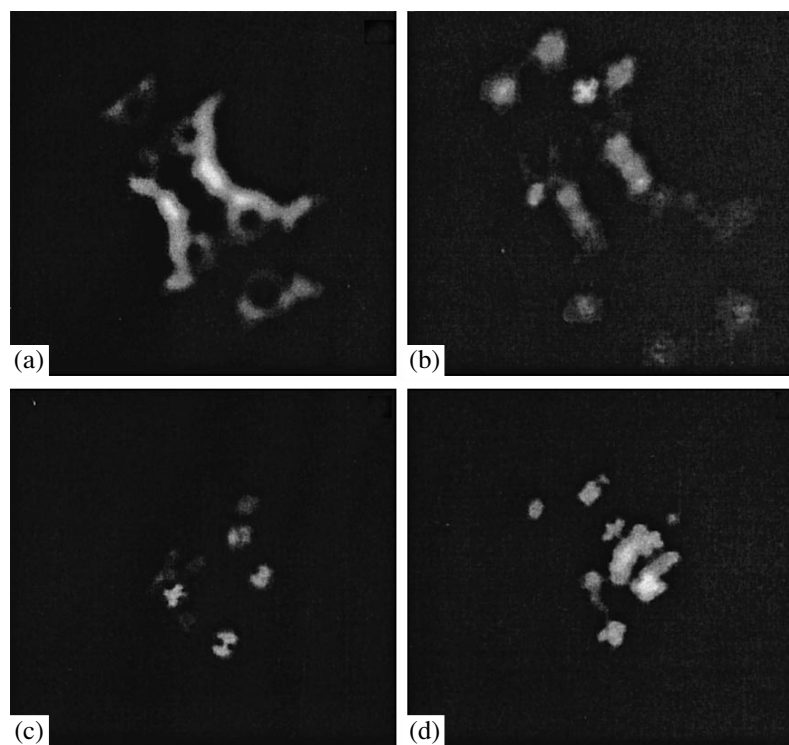
Fig. 2. Emission images of (a) an uncoated ribbed crystal and ribbed crystals with fullerene coatings of different thicknesses Θ: (b) 0.3; (c) 0.5; and (d) 3 mL.

dimensions of microformations on the emitter surface can be caused by the electric field. The mechanism of this process is supposedly as follows. Nonuniform electric fields existing near the emitter surface polarize the C_{60} molecules and displace them toward nearby surface nonuniformities, which enhance the field, that is, along the field gradient. This results in additional enhancement of the field about the microformation, increasing the brightness of its image on the screen. The supposed strong influence of nonuniform fields on the emergence of ordered microformations is supported by data obtained in studies of the fullerene coatings on ribbed crystals. In this case it is hard to provide a different explanation of the fact that at coating thicknesses up to about two monolayers, all major microformations on the emitter surface concentrate along the ribs, where the electric field is stronger (Fig. 2).

Strong electric field can also explain experimental observations of the sudden emergence and disappearance of the ordered structures. The former can be related to the grouping of fullerene molecules near a minor surface nonuniformity, and the latter with field evaporation of molecules or destruction of the microformations as a result of, for example, explosive emission. In some cases, it was possible to observe fast reconstruction of the existing pattern of spots on the projector screen when voltage U was varied. For example, repeated transitions could be observed from two-petal to four-petal structure and back.

In order to determine the formation mechanism of ordered spots on the projector screen, we thought it necessary to estimate the dimensions of the corresponding microformations on the tip surface. For this purpose, it was necessary to measure the current–voltage characteristics of the emitter as a whole and separate well-defined microformations. Using these data, current–voltage characteristics in the Fowler–Nordheim coordinates were plotted. The data obtained were used for determining the apex radius of the tip emitter, the cross section of the microformations D, and their height h. Figure 3 shows a typical Fowler–Nordheim plot of a microformation in the form of quadruplet. It was obtained by measuring currents directly from this microformation under conditions where the currents from the rest of the emitter surface were negligible.

The cross-sectional dimension of the microformation was calculated by Oostrom's method [10] and their height by the technique suggested by Gomer [11]. The characteristic cross-sectional dimension of the microformations producing doublets and quadruplets was found to be variable from 15 to 30 Å. The maximum height of this sort of microformations did not exceed their cross-sectional dimension even in those instances where these formations were seen as bright spots on the projector screen and the current from their surface exceeded by far the currents from other areas of the tip. The reason was probably that the height of considerably less bright ordered microformations discernible

against the background of the substrate image (for example, against the image of the tungsten single crystal) can only be less.

The data obtained on the characteristics of the fullerene coatings studied seem to provide grounds for making corrections to the existing models of formation of ordered emission patterns on the projector screen. Estimates that we made taking into account data on the dimensions of microformations giving ordered images and the information available in the literature on solid formations of fullerene molecules (see, for example, [12]) indicate that the largest microformations of this type include not more than a few tens of C_{60} molecules. Calculations have shown that the above-noted change of the emission spot structure taking place in electric field is accompanied by marked variations of the dimensions of initial microformations on the coating surface. So, a microformation producing an image in the form of doublet had a cross-sectional dimension of $D \approx 30$ Å and a height of $h \approx 15$ Å. Transformation into a quadruplet practically did not change its height but reduced the cross-sectional dimension down to 15 Å. An abrupt change of the image structure of microformations with variation of their dimensions probably indicates that the image depends not so much on the type of molecule as on the dimensions of microformation; i.e., a sort of dimensional effect takes place. It would seem that this is an argument in favor of the waveguide model of formation of ordered emission patterns. However, our data also provide arguments against the existing waveguide model [3, 6]. Anyway, the microformations in our experiments, of a height not exceeding (or even smaller) than its cross-sectional dimension, can hardly be considered waveguides. Even in [3, 6], where the height of copper phthalocyanine microformations, from which field emission was registered, was well in excess of the cross-sectional dimension, the emitting microformation could hardly be identified with a waveguide channeling de Broglie electronic waves without posing any obstacles (not reflecting). Indeed, the probability of electrons exiting into a vacuum under conditions typical for field emitters is considerably less than unity, which means that in constructing a model of the phenomenon, reflection of the electronic waves inside the solid-state system should be taken into account. Therefore, we consider that the resonator model is more adequate for describing the field emission from the surface of ordered microformations. In its essence, it is very close to the standard quantum-mechanical model describing allowed energy states of electrons in solid-state microstructures.

A microformation producing an ordered emission image can possibly be treated as a sort of 3D potential well (resonator) with walls that reflect electrons. In field emission studies, it is necessary to take into account decreasing by electric field the potential barrier at the outer microformation boundary relative the substrate. It is evident that electrons taking part in field
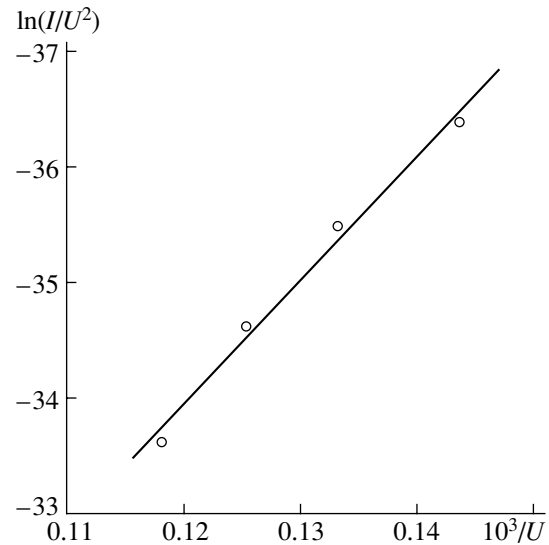
Fig. 3. Fowler–Nordheim plot of a microformation on the fullerene coating producing an emission image in the form of quadruplet (current I and voltage U are in A and V, respectively).

emission come mainly from the upper filled levels in the energy structure. It is generally assumed (see, for example, [3]) that the field emission from microformations made up of large organic molecules is determined by collective π-electrons. For explaining our data, it can be supposed that in the purely carbon fullerene microformations there exist such collective electrons. These electrons moving inside the microformation will be "bombarding" the boundary with a lower potential barrier as well. A fraction of these will tunnel into the vacuum. Electrons withdrawn from the microformation should be replenished by an influx of electrons from the substrate through the opposite boundary.

Spatial distribution of electrons exiting into a vacuum is obviously determined by the distribution of the amplitude of standing electronic waves in the potential well (resonator) considered, in particular, by characteristics of this distribution near the boundary with a reduced potential barrier. The shape of the distribution should vary in a discrete manner with the geometrical dimensions of the microformation because of changes in the oscillation modes in the resonator, which explains the dimensional effect noted above. This essentially qualitative model of the origin of ordered emission images must obviously be analyzed theoretically, but it appears attractive to us because it eliminates some drawbacks in the waveguide theory [3, 5] and allows us to understand the specifics of field emission from microformations on the surface of a field emitter.

CONSLUSION

In summary, note the most important results of this work. For the first time, microformations of C_{60}

fullerene molecules on the surface of a tip field emitter, which yield emission images with ordered structure, have been observed and studied. Major characteristics of these microformations were determined, including their dimensions and the transformations taking place under applied electric field and ion bombardment.

A "resonator model" has been proposed that qualitatively explains the mechanism of formation and reconstruction of the ordered patterns in the emission images from microformations on the surface of a field emitter.

REFERENCES

1. E. W. Müller, Z. Naturforsch. **27**, 290 (1955).

2. A. J. Melmed and E. W. Müller, J. Chem. Phys. **29**, 1937 (1958).

3. A. P. Komar and A. A. Komar, Zh. Tekh. Fiz. **31**, 231 (1961) [Sov. Phys. Tech. Phys. **6**, 166 (1961)].

4. V. A. Shishkin, Candidate's Dissertation (Inst. Élektrokhimii Akad. Nauk SSSR, Moscow, 1962).

5. A. P. Komar and V. P. Savchenko, Dokl. Akad. Nauk SSSR **158**, 1310 (1964) [Sov. Phys. Dokl. **9**, 917 (1965)].

6. V. P. Savchenko, Candidate's Dissertation (Fiz.-Tekh. Inst. im. A. F. Ioffe Akad. Nauk SSSR, Leningrad, 1966).

7. T. A. Tumareva and G. G. Sominskii, J. Commun. Technol. Electron., Suppl. 1 **45**, S110 (2000).

8. T. A. Tumareva, G. G. Sominskiĭ, A. A. Efremov, and A. S. Polyakov, Zh. Tekh. Fiz. **72** (2), 105 (2002) [Tech. Phys. **47**, 244 (2002)].

9. M. V. Loginov and V. N. Shrednik, Pis'ma Zh. Tekh. Fiz. **24** (11), 45 (1998) [Tech. Phys. Lett. **24**, 432 (1998)].

10. A. G. J. Oostrom, Philips Res. Rep., Suppl., No. 1, 1 (1966).

11. R. Gomer, J. Chem. Phys. **28**, 457 (1957).

12. A. V. Eletskiĭ, Usp. Fiz. Nauk **170**, 113 (2000).

Translated by B. Kalinin

EXPERIMENTAL INSTRUMENTS
AND TECHNIQUES

Nitrogen Oxide Conversion in Nitrogen–Oxygen Mixtures Excited by a Microsecond Electron Beam

Yu. N. Novoselov, G. V. Denisov, and R. M. Tkachenko

Institute of Electrophysics, Ural Division, Russian Academy of Sciences,
Yekaterinburg, 620016 Russia
e-mail: nov@iep.uran.ru
Received June 14, 2001

Abstract—The removal of nitrogen oxides from gas mixtures under the action of an ionizing microsecond pulsed electron beam is studied experimentally. Experiments were carried out with gas mixtures simulating flue gases of thermal power plants. The current density and pulse duration of the electron beam, as well as the concentrations of oxygen and the impurity to be removed, are shown to affect the cleaning characteristics. It is found that the nitrogen oxide conversion initiated by a pulsed electron beam cannot be described by models developed for steady-state electron beams. © 2002 MAIK "Nauka/Interperiodica".

INTRODUCTION

NO_x nitrogen oxides, which are in abundance in flue gases of thermal power plants, metallurgy, and some other industries, are very dangerous from the environmental standpoint. Removing them from industrial gaseous wastes is one of the most important environmental problems. One of the methods for removing NO_x is based on pretreating the flue gas by electron beams and/or electric discharges (see, e.g., [1]). According to theoretical concepts [1–3], the thermalized electrons of an electron beam or electric discharge in a wet flue gas stimulate the formation of the O, OH, O_2H, and other free radicals, which leads to the oxidation of NO and NO_2 to HNO_3 nitric acid. Adding ammonia results in reactions with nitric acid leading to the production of powdery NH_4NO_3 ammonium salt, which can be caught with different types of filters.

A cleaning technique based on steady-state electron accelerators was approved under industrial conditions [1, 4, 5]. It was shown that, with this method, the NO_x concentration can be reduced by more than 80%. However, the high energy cost of the removal of an impurity by means of a steady-state electron beam did not allow widespread use of this method. Depending on specific conditions, the energy cost ranges from 20 to 50 eV per molecule (eV/mol), which results in a high power consumption and makes the method noncompetitive. Similar problems occur when removing SO_2 sulfur oxide: the use of steady-state electron beams enables a high decontamination factor, but the energy cost is unacceptably high.

In [6, 7], it was shown that microsecond pulsed electron beams can be efficiently used to remove sulfur oxides. In this case, at the optimum current density and pulse duration of the electron beam, a chain mechanism

of SO_2 oxidation occurs in the gas to be cleaned [8]. The main reactions responsible for this mechanism are reactions involving charged and excited particles. With this method, an energy cost as low as ~1 eV/mol [9] and a decontamination factor of more than 90% were attained. Such a high removal efficiency is related to the repeated involvement of thermalized beam electrons in attachment reactions resulting in the production of negative molecular oxygen ions. By varying the beam current density, the optimum concentration of active particles can be obtained, whereas varying the beam pulse duration allows one to rule out competitive processes [7]. In [10, 11], it was experimentally shown that the use of pulsed electron beams to remove nitrogen oxides also reduces the energy cost. In this study, the conversion of NO_x nitrogen oxides in flue gas mixtures irradiated by a microsecond electron beam is studied experimentally.

EXPERIMENTAL SETUP

In experiments, a facility based on a pulsed electron accelerator with a plasma cathode was used. The accelerator design was similar to that described in [12]. The accelerator produced a radially divergent electron beam with a cross-section area of 1.44 m^2 behind the output foil. The beam electron energy was 280–300 keV. The beam current density j and the FWHM pulse duration τ could be varied within $(0.2–1.2) \times 10^{-3}$ A/cm^2 and 32–90 μs, respectively. The electron beam was injected into a 0.17-m^3 gas chamber, bounded by a grounded output foil of the accelerator and a metal cylinder, to which a 9-μF capacitor charged to a maximum voltage of 20 kV could be connected. The distance between the chamber walls was 10 cm. The energy dissipated in the gas gap was determined as $W = j\tau D$, where D is the beam
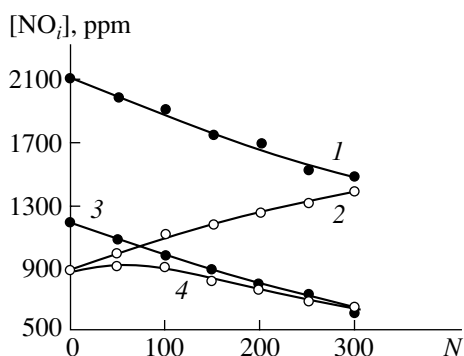
Fig. 1. Concentrations of (*1*) NO_x, (*2*) NO_2 [calculated by reaction (1)], (*3*) NO_2, and (*4*) NO nitrogen oxides vs. number of irradiation pulses N for $j = 0.3 \times 10^{-3}$ A/cm^2 and $\tau = 42$ μs.
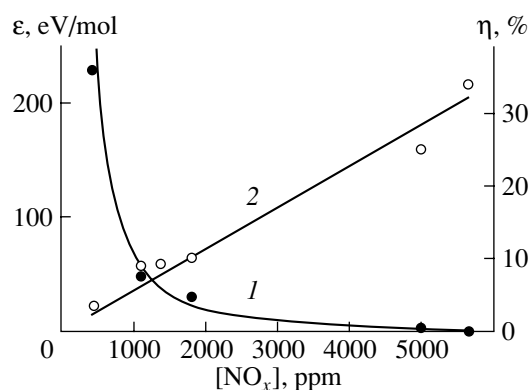
Fig. 2. Dependences of (*1*) the energy cost ε and (*2*) the decontamination factor η on the initial NO_x concentration in the $N_2 : O_2 : H_2O \approx 87 : 10 : 3$ mixture.

energy absorbed by the gas per unit length. The product $j\tau$ was determined by integrating the current and voltage waveforms, and D was measured with the help of TsDP-F-2 film detectors by a routine technique [13].

Experiments were carried out with $N_2 : O_2 \approx 90 : 10$ mixtures with an admixture of NO_x. A N_2/NO model mixture with a controlled component content was prepared in a separate mixer. Just before the beginning of the experiment, this mixture was puffed into the gas chamber, which had already been filled with oxygen. The mixture stayed in the chamber for several hours; during this period, the density of a nitrogen monoxide impurity decreased by 15–20% due to the oxidation reaction

$$2NO + O_2 \longrightarrow 2NO_2. \quad (1)$$
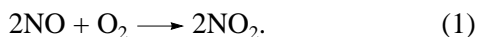
While the mixture stayed in the chamber, the NO concentration was measured at regular intervals. After reaching the dynamic equilibrium among the components of the mixture in the chamber (this mixture composition was assumed to be "initial"), the mixture was exposed to the electron beam. Thus, we excluded the influence of natural oxidation of NO to NO_2 both in the

chamber volume and on the chamber wall during the measurements.

Different methods were used to monitor the gas mixture composition. A Tsvet-500 gas chromatograph was used to determine the nitrogen and oxygen concentrations. The concentrations of oxygen and NO_x nitrogen oxide were measured with the help of a TESTO-350 gas analyzer. In addition, a conductometric method [14] with cryogenic separation of NO from NO_2 was used to monitor the NO and NO_x contents in the mixture. The [NO] concentration was determined after the cryogenic separation of NO from NO_2, whereas the $[NO_x] = [NO] + [NO_2]$ concentration was measured without separation. The concentration of nitrogen oxide was found from the calibration graph of the [NO] concentration versus the conductance of the absorbing solution, which was measured in a separate experiment for reference NO/N_2 gas mixtures with the NO concentration ranging from 200 to 5500 ppm. The statistical error in measuring the concentrations of the mixture components, including nitrogen oxides, was no higher than 0.03.

The gas mixture was irradiated by single pulses. Each experiment was performed with a series of 300–500 pulses. The nitrogen oxide content was checked after every 50 pulses. In experiments, the number of the removed impurity molecules, the decontamination factor $\eta = \Delta[NO_x]/[NO_x]_0$ (in %), and the energy cost per molecule removed $\varepsilon = WN/e\Delta[NO_x]$ (in eV/mol) were determined. Here, $[NO_x]_0$ and $\Delta[NO_x]$ (in cm^{-3}) are the initial impurity density and its change after an exposure to a series of pulses, respectively; W (in J/cm^3) is the specific energy deposited in the gas in one electron beam pulse; N is the number of pulses in the series; and e (in C) is the electron charge. The total error in measuring ε (with allowance for the error in determining the deposited energy W) was no higher than 0.3.

EXPERIMENTAL RESULTS

Preliminary experiments were carried out with N_2/O_2 mixtures in the absence of NO_x. In this case, whatever the electron beam parameters, no nitrogen oxides were formed. In the presence of NO_x, we always observed an almost linear decrease in the NO and NO_x concentration with increasing the number of the electron beam irradiation pulses N. The higher the initial impurity concentration $[NO_x]_0$, the higher the slope of the $[NO_x] = f(N)$ linear dependence [10]. These results indicate that (i) the conversion of nitrogen oxides depends on their initial concentration and (ii) the higher the initial impurity concentration $[NO_x]_0$, the larger the number of molecules removed per irradiation pulse.

In accordance with the model of nitrogen oxide removal based on the assumption that the oxidation processes of free radicals are prevalent, the main final product of the NO_x conversion is nitric acid. In the
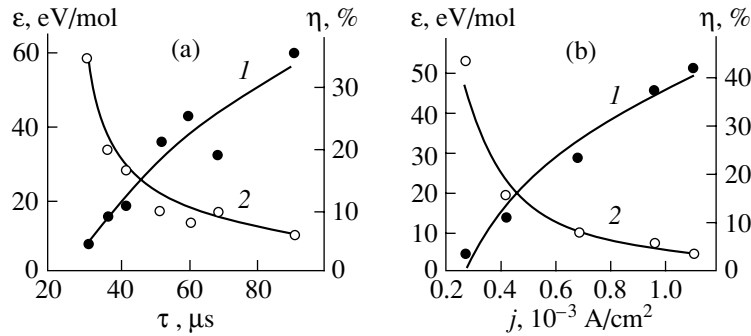
Fig. 3. (*1*) Decontamination factor η and (*2*) energy cost ε as functions of (a) the pulse duration τ and (b) current density *j* of the electron beam for an initial nitrogen oxide concentration of 1000 ppm.

experiments, when water vapor was added to the mixture, droplets of highly-diluted nitric acid arose on the chamber wall after irradiation. However, the technique and apparatus used did not allow us to carry out quantitative measurements. Therefore, we failed to correlate the number of nitric acid molecules with the concentration $\Delta[NO_x]$ of the impurity removed. No other products in either liquid or solid state were observed.

Figure 1 presents the characteristic dependences of the nitrogen oxide concentrations on the number of electron beam pulses. The initial impurity concentrations plotted on the ordinate axis correspond to $N = 0$. Curves *3* and *4* are the dependences of the NO_2 nitrogen dioxide and NO nitrogen monoxide concentrations, respectively. In the case of nitrogen monoxide, the dependence is linear, whereas, in the case of NO_2, it deviates from being linear. When the mixture is exposed to an electron beam, NO oxidizes to NO_2, whose concentration first slightly increases with the number of irradiation pulses and then decreases along with the NO concentration. Curve *1* corresponds to NO_x concentration, which is the sum of NO and NO_2 concentrations. Such behavior of the impurity concentrations allows us to suggest that the removal of impurities from the mixture is not related to reaction (1). Indeed, the decrease in the NO concentration in reaction (1) should be accompanied by an increase in the NO_2 concentration in accordance with the stoichiometric ratio of the oxides for this reaction (Fig. 1, curve *2*). In this case, the NO_x concentration, as well as the sum of all the oxide concentrations, would have been constant. However, this is not the case in our experiments. Thus, the irradiation of the mixture by an electron beam reduces the impurity concentrations and this reduction is not described by the simple reaction (1).

The nitrogen oxide removal is characterized by the energy ε spent on the oxidation of one impurity molecule. Figure 2 presents the dependences of ε (curve *1*) and the decontamination factor η (curve *2*) on the initial NO_x concentration in the mixture. In Fig. 2, the number of irradiation pulses is 300, the electron beam current density is 0.3×10^{-3} A/cm², and the pulse duration is

42 μs. The behavior of the energy cost ε complies with the behavior of the dependences in Fig. 1. At low NO_x concentrations, ε is fairly large and amounts to tens of electronvolts per molecule. Thus, for $[NO_x]_0 \sim 1000$ ppm, the energy cost is ~50 eV/mol. The increase in $[NO_x]_0$ leads to a decrease in the energy cost. Thus, at an initial nitrogen monoxide concentration of 5500 ppm, ε decreases to ~2 eV/mol. For comparison, the dissociation energies of an NO molecule and molecular oxygen are 6.5 and 5.12 eV, respectively [15]. Hence, it is unlikely that the direct electron-impact dissociation of the impurity is the main process of NO_x conversion.

These results confirm the above remark that the process of removing nitrogen oxides from a gas mixture is greatly affected by the initial impurity concentration. It was also shown that the electron beam parameters (the current density *j* and the pulse duration τ) affect the impurity conversion, namely, the decontamination factor η and the energy cost ε [16]. The dependences of the measured values of η and ε on the electron beam duration τ are shown in Fig. 3a for a current density of 0.7×10^{-3} A/cm². It is seen that both the decontamination factor and the energy cost per toxic molecule removed decrease with increasing pulse duration τ. The dependences of these quantities on the electron beam current density at a constant pulse duration (32 μs) demonstrate a similar behavior (Fig. 3b). The minimum energy cost is as low as 3–4 eV/mol, which corresponds to the minimum possible electron beam duration and current density allowed by the experimental facility used.

The above results cannot be explained (even qualitatively) within theoretical concepts based on the free-radical mechanism for NO_x oxidation. Indeed, according to this mechanism [1–3], the active particles, such as O, O_3, and OH and O_2H radicals, should play an important role in the conversion of nitrogen oxides. The concentrations of these agents increase with increasing energy deposition in a gas (i.e., with increasing beam current density at a constant pulse duration). In this case, both the number of the molecules removed and the decontamination factor should increase with the
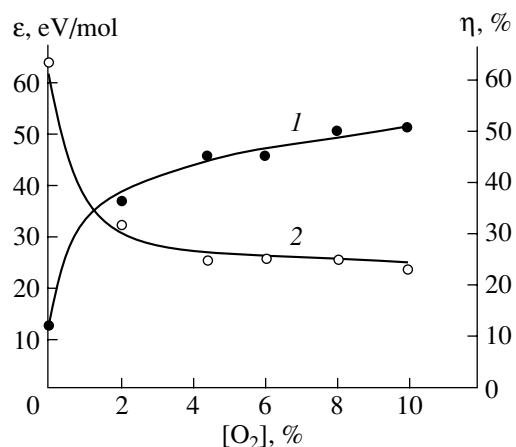
Fig. 4. (*1*) Energy cost ε and (*2*) decontamination factor η vs. oxygen concentration for an initial nitrogen oxide concentration of 600 ppm.
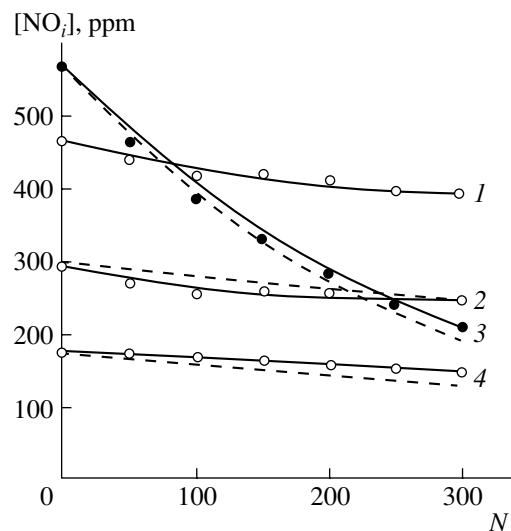
Fig. 5. Concentrations of (*1*) NO_x, (*2*) NO, (*3*) NO, and (*4*) NO_2 nitrogen oxides as functions of the number of irradiation pulses *N* for different oxygen contents in the gas mixture: $[O_2] = $ (*1, 2, 4*) 10 and (*3*) 0%. The solid lines show the experimental data, and the dashed lines show the computed results.

current density. However, the experimental results (Fig. 3b, curve *1*) do not show such a behavior.

Note that the number of the nitrogen oxide molecules removed and, consequently, the decontamination factor depend on the duration of ionizing exposure in the microsecond range (Fig. 3a, curve *1*). The typical characteristic times of chemical reactions with the participation of free radicals taken into account in the models of [1–3] are on the order of $10^{-3}–10^{-4}$ s [17, 18]. The experiments demonstrate that, in the case of pulsed ionization, the processes with characteristic times of $\sim 10^{-6}$ s are of importance. Such processes can be, e.g., the reactions of electron–ion and ion–ion recombina-

tion among electrons, negative oxygen ions, and impurity ions on the one hand and O_2^+, N_2^+, and other positive ions on the other hand. In ionized air at atmospheric pressure, the characteristic time of these reactions is about several microseconds [18, 19].

In addition to the electron beam parameters and the initial nitrogen oxide concentration, the characteristics of NO_x conversion also depend on the oxygen content in the mixture. The oxygen percentage in real flue gases varies in the range from several to ten percent. The dependences of the energy cost ε per impurity molecule removed and the decontamination factor η on the oxygen concentration are shown in Fig. 4. We used $N_2 : O_2 \approx$ (88–98) : (10–2) mixtures with a 100- to 1000-ppm nitrogen oxide admixture. When preparing the mixtures, the oxygen concentration was increased at the cost of increasing nitrogen concentration. It is seen in the figure that the increase in the oxygen concentration in the mixture reduces the number of the NO_x impurity molecules removed; i.e., it decreases the decontamination factor and increases the energy cost, thus worsening the removal characteristics.

Typical dependences of [NO], $[NO_2]$, and $[NO_x]$ concentrations (where $[NO_x] = [NO] + [NO_2]$) on the number of irradiation pulses are shown in Fig. 5. Curves *1, 2,* and *4* correspond to the 10% oxygen concentration, and curve *3* was obtained in the absence of oxygen. In all cases, the oxide concentrations decrease with the increasing number of electron beam pulses.

According to the existing free-radical models of NO_x conversion in an ionized gas, the increase in the oxygen concentration must accelerate the reactions of nitrogen oxide oxidation and, consequently, decrease the concentration of nitrogen oxides. This is not the case in our experiments. The results obtained indicate that the oxidation processes are not the only mechanism for nitrogen oxide conversion. Moreover, the highest reduction in the nitrogen oxide concentration (corresponding to the maximum decontamination factor) occurs in the oxygen-free mixtures (Fig. 5, curve *3*). An alternative to the free-radical mechanism for nitrogen oxide removal with nitric acid as a final product can be the dissociation of NO_x resulting in the creation of atoms and, subsequently, molecules of nitrogen and oxygen [20].

When fast electrons are injected into a nitrogen–oxygen mixture, their energy is spent not only on the ionization and excitation of the mixture components, but also on the dissociation of nitrogen. Knowing the formation energies of the mixture components, we can estimate their concentrations. Under our experimental conditions, the estimated concentration of atomic nitrogen produced due to the direct-impact dissociation of N_2 molecules by the fast beam electrons does not exceed 10^{12} cm^{-3}. The concentration of atomic nitrogen significantly increases (to $\sim 8 \times 10^{13}$ cm^{-3}) due to the

dissociative recombination of N_2^+ ions. Then, the reactions with the participation of atomic nitrogen and impurity molecules lead to a decrease in the nitrogen oxide concentration and the production of N_2 and O_2 molecules. The behavior of the nitrogen oxide concentration computed in accordance with a mechanism proposed in [20] is shown in Fig. 5 by the dashed lines. In calculations, we used the reaction rate constants from [8].

When studying the removal of sulfur oxides under the action of a microsecond electron beam, it was shown that applying a weak electric field ($E \sim 200$–500 V/cm) on an ionized model mixture intensifies the process of plasmochemical oxidation of SO_2 [7]. This is related to the appearance of an additional channel for the production of O_2^* vibrationally excited oxygen molecules, which participate in the chain mechanism [21]. It was interesting to find out whether the external electric field influences the conversion of nitrogen oxides. Such experiments were carried out at an electric field of $E \sim 0$–2000 V/cm in mixtures with the NO_x content ranging from 100 to 3300 ppm and with an addition of 0.1–1.0% water vapor. It turned out that, within the measurement accuracy, the experimental results obtained with and without an external field were approximately the same. This fact indicates the insignificance of O_2^* molecules in nitrogen oxide conversion.

CONCLUSION

The results obtained allow us to draw the following conclusions. The initial impurity concentration significantly affects nitrogen oxide conversion under the action of a pulsed electron beam. The minimum energy cost per one molecule removed ($\varepsilon_{min} \sim 2$ eV) is achieved at a high impurity concentration. The electron beam parameters and the oxygen content in the irradiated gas significantly affect the characteristics of the NO_x removal process. Oxidation reactions are not the only process of NO_x conversion in an ionized nitrogen–oxygen mixture. At low oxygen concentrations, the dissociation of nitrogen oxides is the most probable process. The mechanism governing NO_x conversion under the conditions of steady-state gas ionization and based on the reactions involving free radicals fails to account for the experimental results obtained with pulsed electron beams. To explain these results, it is necessary to develop an adequate non-steady-state model with allowance for plasmochemical reactions among neutral, charged, and excited particles, as well as reactions involving free radicals.

ACKNOWLEDGMENTS

REFERENCES

1. *Non-Thermal Plasma Techniques for Pollution Control*, Ed. by B. M. Penetrante and S. E. Schultheis (Springer-Verlag, Berlin, 1993), NATO ASI Ser., Ser. G **34**.

2. J. C. Person and D. O. Ham, Radiat. Phys. Chem. **31**, 1 (1988).

3. O. Tokunaga and N. Suzuki, Radiat. Phys. Chem. **24**, 145 (1984).

4. N. Frank, S. Hirano, and K. Kawamura, Radiat. Phys. Chem. **31**, 57 (1988).

5. A. G. Chmielewsky *et al.*, Radiat. Phys. Chem. **46**, 1063 (1995).

6. D. L. Kuznetsov, G. A. Mesyats, and Yu. N. Novoselov, in *Novel Application of Lasers and Pulsed Power*, Ed. by R. D. Curry, Proc. SPIE **2374**, 142 (1995).

7. D. L. Kuznetsov, G. A. Mesyats, and Yu. N. Novoselov, Teplofiz. Vys. Temp. **34**, 845 (1996).

8. E. I. Baranchikov, G. S. Belen'kiĭ, V. P. Denisenko, *et al.*, Dokl. Akad. Nauk SSSR **315**, 120 (1990).

9. A. V. Ignat'ev, D. L. Kuznetsov, G. A. Mesyats, and Yu. N. Novoselov, Pis'ma Zh. Tekh. Fiz. **18** (22), 53 (1992) [Sov. Tech. Phys. Lett. **18**, 745 (1992)].

10. G. V. Denisov, Yu. N. Novoselov, and R. M. Tkachenko, Pis'ma Zh. Tekh. Fiz. **24** (4), 52 (1998) [Tech. Phys. Lett. **24**, 146 (1998)].

11. Y. Nakagawa and H. Kawauchi, Jpn. J. Appl. Phys. **37**, 91 (1998).

12. A. M. Efremov, B. M. Koval'chuk, V. S. Tolkachev, *et al.*, Prib. Tekh. Éksp., No. 1, 99 (1987).

13. V. V. Generalova and M. N. Gurskiĭ, *Dosimetry in Radiation Technology* (Izd. Standartov, Moscow, 1981).

14. E. A. Khudyakova and A. P. Kreshkov, *Theory and Practice of Conductimetric Analysis* (Khimiya, Moscow, 1976).

15. A. A. Radtsig and B. M. Smirnov, *Reference Data on Atoms, Molecules, and Ions* (Atomizdat, Moscow, 1980; Springer-Verlag, Berlin, 1985).

16. G. V. Denisov, D. L. Kuznetsov, Yu. N. Novoselov, and R. M. Tkachenko, Pis'ma Zh. Tekh. Fiz. **24** (15), 47 (1998) [Tech. Phys. Lett. **24**, 601 (1998)].

17. V. N. Kondrat'ev, *Gas-Phase Reaction Rate Constants* (Nauka, Moscow, 1970).

18. L. I. Virin, R. V. Dzhagatspanyan, G. V. Karachevtsev, *et al.*, *Ion-Molecule Reactions in Gases* (Nauka, Moscow, 1979).

19. B. M. Smirnov, *Ions and Excited Atoms on Plasma* (Atomizdat, Moscow, 1974).

20. G. V. Denisov, Yu. N. Novoselov, and R. M. Tkachenko, Pis'ma Zh. Tekh. Fiz. **26** (16), 30 (2000) [Tech. Phys. Lett. **26**, 710 (2000)].

21. Yu. N. Novoselov, Pis'ma Zh. Tekh. Fiz. **19** (23), 58 (1993) [Tech. Phys. Lett. **19**, 760 (1993)].

The Simulation of Dendrite Growth and Partial Discharges in Epoxy Resin

M. D. Noskov*, A. S. Malinovski*, M. Sack**, and A. J. Schwab**

* Tomsk Polytechnical University, pr. Lenina 30, Tomsk, 634050 Russia
e-mail: noskov@hvd.tpu.ru
** University of Karlsruhe, D-76128 Karlsruhe, Germany
e-mail: sack@ieh.etec.uni-karlsruhe.de
Received October 26, 2000; in final form, March 13, 2001

Abstract—A self-consistent model of the development of dendrites and partial discharges in a dielectric solid under a variable voltage is suggested. Dendrites originate at sites with the enhanced local field strength and also where the dielectric is broken down under the action of partial discharges. The numerical simulation is used to quantitatively describe the space–time dynamics of the dendrites and partial discharges in epoxy resin for the tip–plane electrode configuration. The simulated data are compared with electrical measurements of partial discharges and with optical images of dendrite growth under the same conditions. © 2002 MAIK "Nauka/Interperiodica".

INTRODUCTION

The breakdown of polymeric insulation during long-term exposure to a variable voltage is associated with the evolution of gas-filled channels, so-called dendrites [1, 2]. Dendrites originate at sites with the enhanced local field strength (electrode surface irregularities, conducting inclusions, microcracks, pores, etc.). The breakdown takes place soon after a dendrite has covered the electrode spacing. Dendrites grow, because partial discharges in their channels break down the dielectric. Thus, the evolution of dendrites and partial discharges are interrelated processes, causing the breakdown of insulation. The dynamics of dendrite growth and the parameters of partial discharges depend on the physical and chemical properties of the dielectric, the frequency and amplitude of the applied voltage, electrode configuration, etc.

The study of dendrite growth has become of particular interest because of the recent advent of cables with the polymeric insulation that can withstand voltages as high as 400–500 kV. Using electronic and electron-optical tools, the researchers have determined the parameters of the partial discharges and the time-space characteristics of the dendrite growth [3–6]. Yet, in spite of the extensive studies, a quantitative theory of the dendrite growth due to the partial discharges has not been elaborated to this point. This is related to a variety of interrelated physical and chemical phenomena responsible for the dielectric breakdown and also to the fact that the growth of the dendrites is a complex stochastic process. In recent years, a number of mathematical models for the stochastic description of branching dendrite growth has been suggested [7–13]. The most interesting results have been obtained with dielectric

breakdown model [7–11] and the model of avalanche discharge [12, 13]. In the former, the evolution of the channels is assumed to be dependent on the electric field strength; namely, the growth probability is proportional to the local electric field. The field distribution is calculated with the Laplace equation. The dendrite potential is taken to equal the electrode potential or, otherwise, the voltage drop along the channel is specified. With various modification of the dielectric breakdown model, tree- and bushlike dendrites have been simulated [8, 10]. Also, the effect of various barriers and inclusions with different permittivities [10], conductivities [11], and dielectric strengths [9] have been analyzed. In the model of avalanche discharge [12, 13], microbreakdowns in the dielectric material due to electron avalanches is considered. The channel evolution is assumed to be related to random fluctuations of the local electric field. Within this model, various shapes of the dendrites have been predicted for different fluctuation ranges.

The disadvantage of both models is that they ignore partial discharges in the dendrite channels and, hence, do not describe the actual dynamics of the charge and field distributions in the dielectric. On the other hand, the available models of partial discharges (see, e.g., [14, 15]) do not touch upon the growth of dendrites and cannot be used to quantitatively characterize insulation breakdown. Therefore, the development of a unified model that covers both the mechanisms of partial discharges in the available channels and the dendrite growth due to the partial discharges seems to be quite topical. Such a model has been suggested in [16]. With it, we succeeded in discovering the basic mechanisms and laws behind the dendrite growth and in determining the parameters of the partial discharges [17]. In this

work, this model is extended for the quantitative characterization of the dendrite growth and the partial discharges in epoxy resin for the tip–plane electrode configuration subjected to a varying voltage. The results of simulation are compared with experimental data.

## A SELF-CONSISTENT MODEL OF DENDRITE GROWTH AND PARTIAL DISCHARGES

The Basic Postulates of the Model

Our self-consistent model considers the growth of dendrites as the spatial formation of branching channels. A new channel appears as a result of the breakdown of the dielectric surrounding the existing channels or at sites with the enhanced local field strength. The dielectric breakdown is associated with the partial discharges in the channels. It is assumed that the material damage is proportional to the discharge energy being released in the channel. The channel grows if the energy per unit channel length reaches the critical value W_c and the field strength near the channel damaged exceeds the critical strength E_c. The growth direction of a new channel stochastically depends on the local electric field strength: the probability that the channel will grow in the direction $\mathbf{n}$ varies as the square of the projection E_n of the local strength onto this direction if $E_n > E_c$ and equals zero otherwise. The random selection of the growth direction in the model reflects the inhomogeneity of the dielectric material and also the stochastic nature of the formation of a new channel. The introduction of the critical strength E_c means that there is some threshold strength for the formation of a new channel even if the material is damaged. The quadratic dependence of the growth probability on the field strength relates the channel growth to the electric field energy density.

A partial discharge in the dendrite channel is initiated when the local field along a channel segment exceeds the threshold initiating field E_{in}. The discharge is quenched when the local field drops below the quenching field E_q. The electric field is defined by both the electrode potential and the charge distribution along the channel. In the absence of the partial discharges, the channels are nonconducting. Charge transfer is accomplished only during the partial discharges. Since the partial discharge duration (10–100 ns) is much smaller than the characteristic time of dendrite growth ($>10^3$ s), it is assumed that the partial discharges proceed instantaneously [1].

Numerical Implementation of the Model

The three-dimensional numerical implementation of the model allows the computer-aided simulation of dendrite growth and partial discharges in a dielectric. The electrodes are a tip to which a variable voltage is applied and a grounded plane. The electric field is calculated with the charge method [18] based on the Cou-
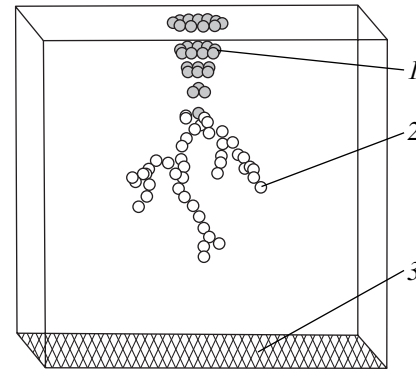
Fig. 1. Domain of simulation: *1*, tip; *2*, dendrite; and *3*, planar electrode.

lomb law and the principle of superposition as applied to electric fields produced by point charges. The field distribution for such an electrode configuration is simulated by placing a sufficiently large number of point charges on the surface of a cone representing the tip. The effect of the grounded planar electrode on the field distribution is taken in consideration by using the mirror image method. The value of each of the charges on the tip is calculated by iteration so as to obtain a given time variation of the applied voltage $\varphi(t)$. The dendrite channels are simulated by the branches of point charges separated by a fixed distance d from one another. Thus, a dendrite can be represented as a set of contacting spheres (Fig. 1). The sphere radius $r = d/2$ defines the minimal spatial scale of simulation and is an important parameter of the numerical model.

To describe the evolution of the dendrite growth and the partial discharges, the model uses discrete time steps Δt of equal width. The step Δt should be sufficiently small so that individual partial discharges can be distinguished. At each nth step corresponding to the time instant $t_n = n\Delta t$, the state of each of the spheres is characterized by the specific damage energy W_i^n and the electrical charge q_i^n (i is the sphere number). The potential φ_i^n of the ith sphere is calculated as the sum of the potential due to the charge q_i^n and the potentials due to all other charges (including the imaginary ones):

$$\varphi_i^n = \frac{q_i^n}{4\pi\varepsilon\varepsilon_0 r} + \sum_{j \neq 1} \frac{q_j^n}{4\pi\varepsilon\varepsilon_0 |\mathbf{r}_i - \mathbf{r}_j|}$$
$$- \sum_j \frac{q_j^n}{4\pi\varepsilon\varepsilon_0 |\mathbf{r}_i - \mathbf{r}_j^*|}. \tag{1}$$

Here, ε, ε_0 are the relative and absolute permittivities and $\mathbf{r}_j$, $\mathbf{r}_j^*$ are the coordinates of the real charge q_j and the associated specularly reflected charge, respectively.

At each of the steps, the following procedures are performed. (i) The instantaneous value of the voltage applied to the tip is changed, and the tip charge is recalculated; (ii) the partial discharges, as well as changes in the distributions of the specific damage energy and the charges on the spheres participating in the partial discharges, are simulated; and (iii) dendrite growth is simulated by adding new spheres to the existing structure or to the tip.

The charge on the tip is calculated as follows. The potential of each of the tip points is set equal to the instantaneous value of the applied voltage: $\varphi_i^n = \varphi(n\Delta t)$. The dendrite charges are taken to be fixed. Equation (1) is written for each of the tip points, and the resulting set of linear equations is solved by the Gauss method.

No more than one partial discharge is simulated at each of the steps. The dendrite channels are divided into segments formed by a pair of neighboring spheres. Each of the segments may be either conducting or nonconducting. The conducting (nonconducting) state corresponds to the presence (absence) of a discharge between the spheres. The simulation of a partial discharge includes the self-consistent calculation of the segment-state dynamics, charges, and sphere potentials. At the initial step of the iterative procedure (the iteration number $k = 0$), all the segments are nonconducting and the charges and specific damage energies of all the spheres equal those obtained at the preceding time step $n - 1$: $q_i^n(0) = q_i^{n-1}$, $W_i^n(0) = W_i^{n-1}$. The transition from the $(k-1)$th step to the kth step is as follows. First, the states of the segments are changed, if necessary. A segment passes to the conducting state or remains conducting if the absolute value of the potential difference between adjacent spheres i and j, $\Delta\varphi_{i,j}^n(k-1) = \varphi_i^n(k-1) - \varphi_j^n(k-1)$, exceeds the partial discharge initiation voltage $\varphi_{in} = E_{in}d$:

$$\left|\Delta\varphi_{i,j}^n(k-1)\right| > \varphi_{in}. \tag{2}$$

A segment passes to the nonconducting state or remains nonconducting if the absolute value of the potential difference is lower than the partial discharge quenching voltage $\varphi_q = E_q d$:

$$\left|\Delta\varphi_{i,j}^n(k-1)\right| < \varphi_q. \tag{3}$$

Then, the charge is transferred to the conducting segments. The value of the charge transferred $\Delta q_{i,j}^n(k)$ is proportional to the potential difference $\Delta\varphi_{i,j}^n(k-1)$ between the spheres:

$$\Delta q_{i,j}^n(k) = \lambda\Delta\varphi_{i,j}^n(k-1), \tag{4}$$

where λ is a numerical iteration parameter.

The parameter λ is selected so as to provide the maximal computation speed without affecting the sta-

bility of the iterative procedure. In this work, $\lambda = 0.3 \times 4\pi\varepsilon\varepsilon_0 r$. The charge transfer changes the charges on the spheres belonging to at least one conducting segment. New charges are calculated by the formula

$$q_i^n(k) = q_i^n(k-1) - \sum_j \Delta q_{i,j}^n(k), \tag{5}$$

where the summation is over all points j connected to a point i by the conducting segments. Once the charge transfer has been completed, the charges on the tip are recalculated so that the tip potential remains equal to the potential $\varphi(n\Delta t)$ corresponding to a given time step.

New potentials of the dendrite points, $\varphi_i^n(k)$, are calculated by formula (1) according to the new charge distributions over the tip and the dendrite. At each of the iteration steps, the damage energies of the spheres belonging to the conducting segments are increased. Only a part of the energy being released in the channels during the spatial discharges is spent on damage to the dielectric material. For simplicity, this fact is taken into account by appropriately raising the critical damage energy W_c; that is, W_c is now the energy per unit channel length that must be released during the partial discharges for a new channel to form. The increment of the specific damage energy equals the energy being released when the charge is transferred from the jth sphere to the ith sphere along the dendrite divided by the segment length d. Thus, after the kth iteration step, the damage energy $W_i^n(k)$ will be

$$W_i^n(k) = W_i^n(k-1) + \Delta q_{i,j}^n(k)\Delta\varphi_{i,j}^n(k-1)/d. \tag{6}$$

The iterative procedure is repeated until none of the segments is in the conducting state. After the nth iteration step, the charges q_i^n and the specific damage energies W_i^n of the dendrite points depend on the last iteration step k_l:

$$q_i^n = q_i^n(k_l), \quad W_i^n = W_i^n(k_l). \tag{7}$$

The partial discharge propagates through all the segments that were in the conductive state at least once during the iterative procedure. The amount Q^n of the partial discharge is calculated as the total change in all the charges on the tip and on the dendrite:

$$Q^n = \sum_i (q_i^n - q_i^{n-1}). \tag{8}$$

The dendrite growth is simulated by adding new spheres to the existing dendritic structure or to the tip. A new sphere can be attached to any of the dendrite spheres for which the specific damage energy W_i^n has reached the critical value W_c:

$$W_i^n \geq W_c. \tag{9}$$

With condition (9) met, an ensemble of possible positions of the new spheres attached to a given sphere is constructed. The number of the positions must be sufficiently large to realize the stochastic growth law adopted in the model. We constructed ensembles of ten positions. The possible positions are regularly arranged around a given sphere so that the new spheres do not cross the existing ones. The position of a new sphere is randomly selected with a probability P that is proportional to the $\Delta\varphi$ squared (where $\Delta\varphi$ is the potential difference between a given point and a position being considered) if the absolute value of $\Delta\varphi$ exceeds the critical value $E_c d$, $|\Delta\varphi| > E_c d$ (otherwise, the probability is zero); that is,

$$P = \begin{cases} \Delta\varphi^2/Z, & |\Delta\varphi| > E_c d \\ 0, & |\Delta\varphi| \le E_c d, \end{cases} \tag{10}$$

where Z is the normalizing factor defined as $Z = \Sigma(\Delta\varphi)^2$ (the summation is over all possible positions for which $|\Delta\varphi| > E_c d$).

If the growth has occurred, the value W_c is subtracted from the specific damage energy of that sphere to which a new one has been attached. The specific damage energy W_j^n and the charge q_j^n of a new sphere j are set equal to zero. Once the simulation has been completed, we pass to the next time step. The simulation lasts until the dendrite reaches the planar electrode. Initially, while the dendrite does not exist, a new sphere can be attached only to the tip. Since we do not consider the initiation of the dendrite, the tip is assigned the damage energy higher than W_c to make the growth possible.

Parameters of the Model

The basic model parameters are W_c, E_c, E_{in}, E_q, and r. They, along with the electrode configuration and the applied voltage parameters, are responsible for the features of the dendrite growth and the partial discharges. The parameters have the simple physical interpretation. The critical damage energy W_c is the energy needed for the formation of the channel of unit length. The critical field E_c is related to the channel formation due to the partial discharges. It may be much lower than the breakdown field of the undamaged dielectric. Both W_c and E_c depend on the physical and chemical properties of the dielectric and on the channel sizes. The threshold fields E_{in} and E_q specify the conditions for gas discharge initiation and quenching in the dendrite channels. It has been shown that the threshold fields in the thin channels increase as the channel diameter decreases and may considerably exceed the associated values for the discharge in free space. In addition, the threshold fields E_{in} and E_q vary with the gas composition and pressure in the channel. The radius r of the spheres used in the simulation not only specifies the

least spatial scale of simulation but also relates the charge and the potential of a sphere [the first term on the right of Eq. (1)]. Thus, the parameter r defines the size of the space charge region around the dendrite channels. The size of the region of the charge injected into the dielectric depends on the field strength and may exceed the channel diameter determined by optical means.

The model parameters can be determined by analyzing physical processes attendant on the dendrite growth or by comparing the results of simulation with experimental data. In our case, the parameters were determined by comparing test calculations with experimental data for the dendrite growth in epoxy resin. The parameters used in the simulations were $W_c = 0.07$ J/m, $E_c = 1.0 \times 10^8$ V/m, $E_{in} = 8 \times 10^7$ V/m, $E_q = 1.5 \times 10^7$ V/m, and $r = 10$ μm. The voltage parameters and the electrode configurations were the same as those used in the experiments.

MATERIAL AND EXPERIMENTAL TECHNIQUE

The experiments were performed with Araldit CY221 epoxy resin (HY2966 hardener) from Ciba Specialty Chemicals Inc. Prior to the preparation of the samples, a steel tip of radius 10 μm was inserted in a rectangular mold. For the experiments, the sample was placed on a planar grounded electrode inside a transparent vessel filled with transformer oil for insulation. The tip-plane spacing was 1.5 mm. A sinusoidal voltage with an effective value of 8 kV and a frequency of 50 Hz was applied to the tip.

The basic components of the computerized setup are a partial discharge measuring system, a video camera for shadow filming, and a photoelectric multiplier for detecting the optical radiation from the partial discharges (Fig. 2).

RESULTS AND DISCUSSION

At the initial stage of dendrite formation, one or several channels appear at the tip. The channels develop by stochastic branching and bending. Typical dendrite structures obtained by the simulation and in the experiments are shown in Fig. 3 for four successive instants of the growth. Structures of this type are called tree-like [1]. An increase in the voltage applied causes intense branching and, accordingly, the growth of bush-like structures. The dendrite shape can quantitatively be described in terms of fractal dimension [19]. In this work, the fractal dimension is estimated by the coverage method. The dendrite projection is covered by squares of side l. Then, the number $N(l)$ of squares that contain dendrite fragments is calculated. For a fractal structure of dimension D, the relationship

$$N(l) \propto l^{-D} \tag{11}$$

holds.
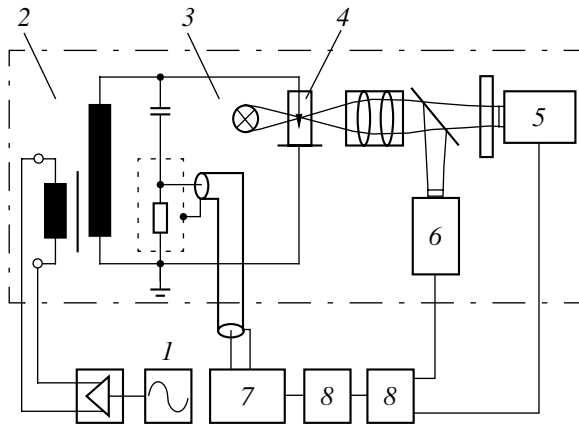
Fig. 2. Experimental setup: *1*, variable voltage source; *2*, step-up transformer; *3*, illuminating lamp; *4*, working chamber; *5*, photoelectric multiplier; *6*, video camera for shadow filming; *7*, partial discharge measuring system; and *8*, controlling computers.

Fig. 3. Dendrite growth: (a) simulation and (b) experiment. $t = (1)$ 2, (2) 4, (3) 6, and (4) 9 min.

Thus, the fractal dimension can be found from the slope of a straight line approximating the $\ln N(l)$ vs. $\ln l$ dependence. Dendrite fractal dimensions thus obtained vary between 1.45 and 1.55. Similar values for treelike structures were found in [20, 21]. The dendrite growth rate drops with time both upon the simulation and in the experiments. Figure 4 shows the time dependences of the dendrite length R (the distance from the tip to the farthest point of a dendrite). The slowing-down of dendrite growth at the initial stages has been noticed in many works [1, 3–5].

The growth of a dendrite is accompanied by partial discharges in its channels. The simulation indicates that the partial discharges are initiated at the tip and propagate towards the channel end. Each subsequent partial discharge propagates through a new channel. This can be explained by charge transfer and the reduction of the field along the channels having taken part in the preceding partial discharge. Typical discharge paths simulated are shown in Fig. 5a. Actual partial discharges radiate light. As follows from the radiation patterns obtained with an image-converter tube, different channels emit at different time instants (Fig. 5b). This observation can be viewed as a demonstration of the fact that different partial discharges occupy different channels.

The simulation shows that the amount Q of a partial discharge is proportional to the total length S of all channels covered by this discharge. For the same path length L (the distance between the tip and the farthest point of the path), the values of S may be different because of different numbers of the channels taking part in the partial discharge. Therefore, the path length defines only the maximal amount of the discharge. The relationship between the total path length S and the amount of partial discharges that was established during the simulation is corroborated by the experiments where the correlation between the discharge amount
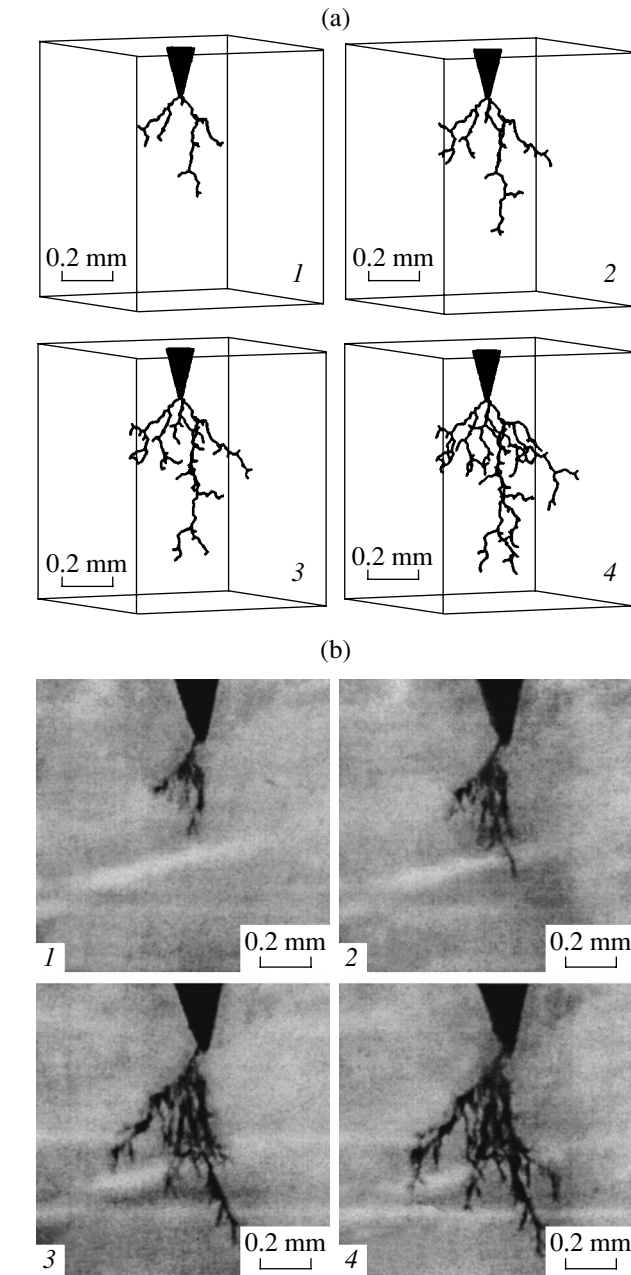
and the total length of radiating channels has been found [6]. During the dendrite growth, the maximal path length of the partial discharges increases and, accordingly, so does their maximal amount. Figure 6 demonstrates the time dependences of the maximal amount Q_{max} of the partial discharges that were observed upon the simulation and in the experiments.

For the same dendrite length, the amount of a partial discharge depends on the phase of the voltage applied. The maximal amount of the discharge rises with instantaneous value of the voltage. The polarity of the discharge coincides with the sign of the time derivative of
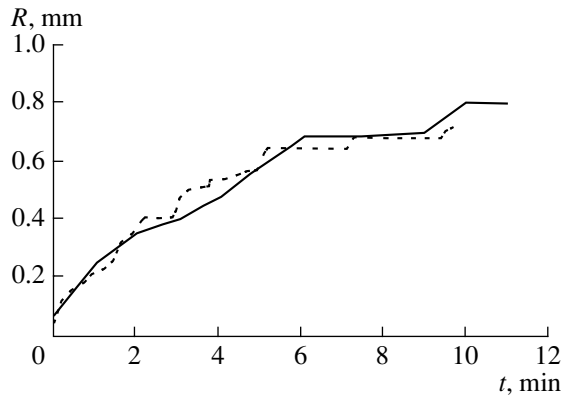
Fig. 4. Dendrite length vs. growth time. Continuous curve, experiment; dotted line, simulation.

the voltage. The correlation between the amount of the partial discharge and the phase (φ–q–n diagram) is shown in Fig. 7. φ–q–n diagrams of this type, which are often called wing-shaped, have been also observed in other works (see, e.g., [5]) when dendrites grew at a variable voltage. The discharge repetition rate increases with the absolute value of the time derivative of the voltage and reaches a maximum at phase angles of 0° and 180°. The experimental and simulated histograms showing the distribution of the partial discharges over the phase are depicted in Fig. 8.

The behavior of the partial discharges that was observed upon the simulation and in the experiments can be explained by the existence of the threshold fields responsible for the initiation and quenching of the discharges. The discharge starts if the field strength along the channels is higher than the critical value E_{in} and ceases when the field drops to E_q because of charge transfer. Thus, after the discharge, the voltage drop per unit channel length is roughly E_q. Therefore, the maximal discharge path length and, hence, the maximum possible amount of the partial discharge grow with voltage applied. This explanation has been corroborated experimentally. In [15], where the emission from the dendrite channels was recorded for various phases of the voltage, it was established that the emitting part of the dendrite expands with increasing absolute value of the voltage.

These results are in qualitative agreement with experimental data for the dendrite growth and partial discharges under similar conditions in epoxy resin and other polymers (see, e.g., [3–5]). This means that the model suggested adequately predicts the general laws of dendrite growth due to partial discharges and that the basic ideas of the model reflect real physical processes. The quantitative agreement of the simulated results and the experimental data for epoxy resin also counts in favor of the model. It should be noted that the model is to a great extent general, so that it can be used for describing the discharge-related dendrite growth in
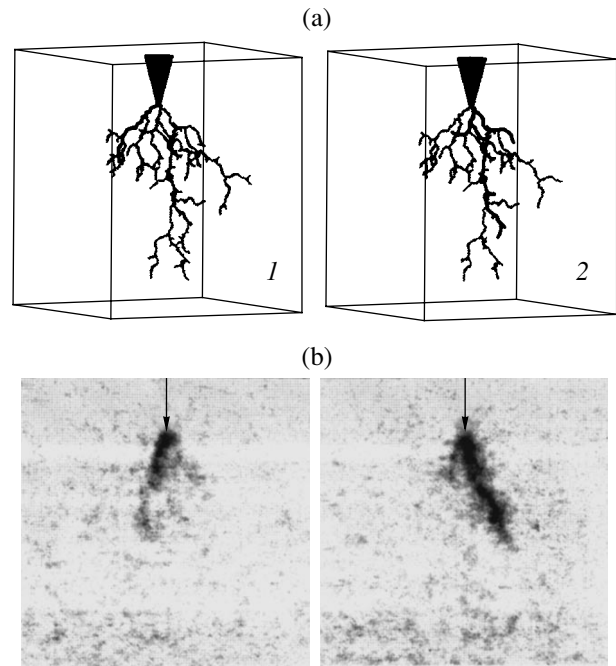

Fig. 5. (a) Partial discharge path and (b) dendrite channel emission pattern. The discharge paths are shown black and the dendrite channels, gray. The partial discharge amount is (*1*) 44 and (*2*) 181 pC. The arrow points to the tip position.
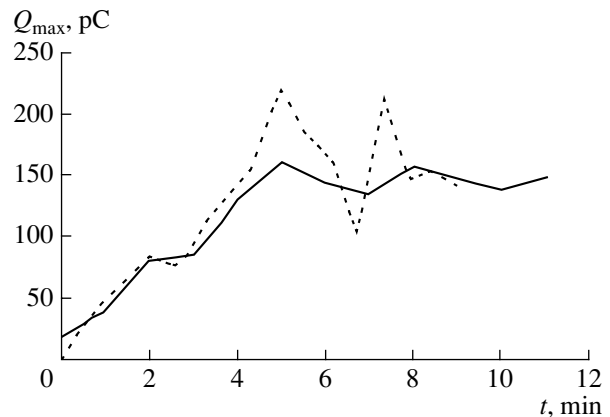

Fig. 6. Maximal amount of the partial discharge vs. time. Continuous curve, experiment; dotted line, simulation.

other polymers as well, such as polyethylene, polymethylmethacrylate, etc. Thus, the model can serve as the basis for a technique that allows us to measure the dielectric strength as a function of the electrode configuration, voltage parameters, and material properties. Another important application of our work is insulation diagnostics based on partial discharge analysis.

The model can be extended further by considering the wall conductivities of the channel and the dielectric, the dependences of the initiation and quenching fields on the pressure and channel radius, etc. This will make the model more accurate and enable us to study such effects as the transition from the treelike growth to the
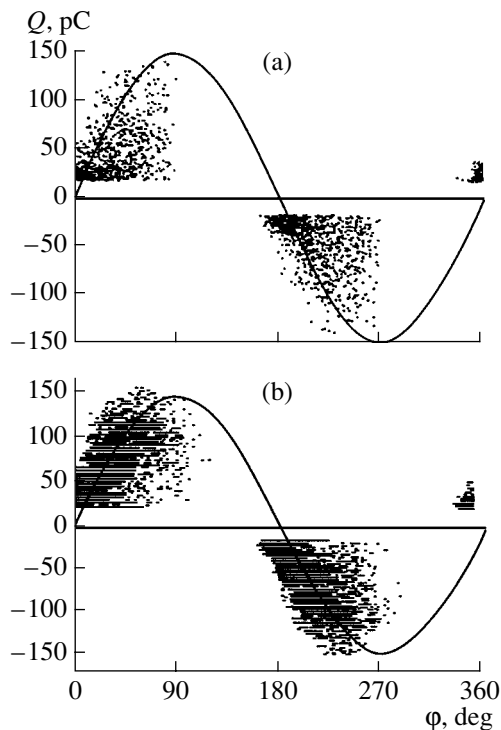
Fig. 7. Correlation between the partial discharge amount and the applied voltage phase (t = 9 min). (a) Simulation and (b) experiment.
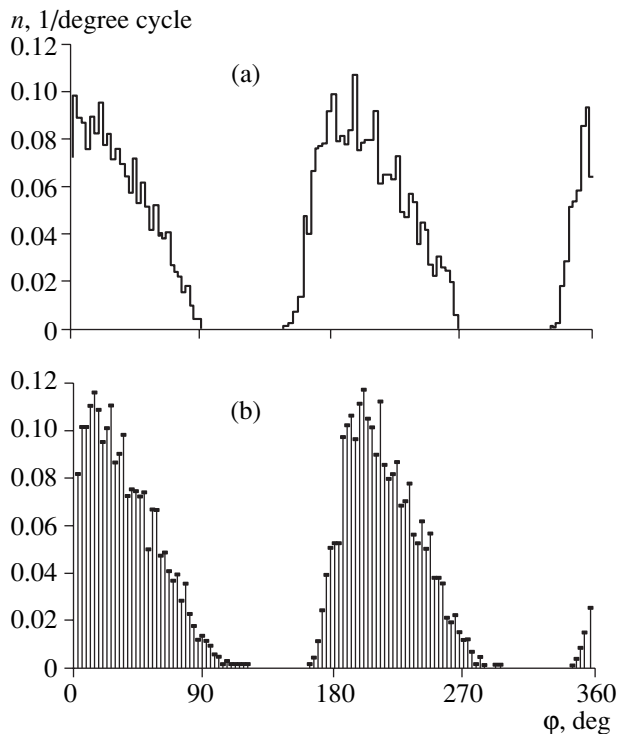
Fig. 8. Density function of the partial discharges n vs. applied voltage phase (t = 9 min). (a) Simulation and (b) experiment.

bushlike one as the voltage rises, the temporary ceasing and renewal of the discharges and dendrite growth, and others.

CONCLUSION

The self-consistent model of dendrite growth and partial discharges in polymeric insulation is developed. It is based on the energy approach, assumes the threshold character of channel expansion, and introduces the local initiation and quenching fields of the discharge. The nonlattice three-dimensional implementation of the model makes it possible to perform numerical experiments on the dendrite growth and partial discharge initiation for a tip–plane electrode configuration at a variable voltage.

We studied a possible correlation between the partial discharge parameters and dendrite growth in epoxy resin. The simulation shows that the amount of a partial discharge is proportional to the total length of the channels involved in this discharge. The maximal amount of the partial discharges grows with the dendrite length. As follows from the phase diagrams, within a period, the maximal amount of the discharge increases with the instantaneous value of the voltage, while the discharge repetition rate rises with the voltage frequency. The quantitative agreement between the simulated results and the experimental data confirm the adequacy of the model and its reliability.

REFERENCES

1. L. A. Dissado and J. C. Fothergill, *Electrical Degradation and Breakdown in Polymers* (Peregrinus, London, 1992).
2. G. S. Kuchinskiĭ, *Partial Discharges in High-Voltage Structures* (Énergiya, Leningrad, 1979).
3. J. V. Champion and S. J. Dodd, J. Phys. D **29**, 862 (1996).
4. J. V. Champion, S. J. Dodd, and J. M. Alison, J. Phys. D **29**, 2689 (1996).
5. Suwarno, Y. Suzuoki, F. Komori, *et al.*, J. Phys. D **29**, 2922 (1996).
6. Y. Ehara, M. Naoe, K. Urano, *et al.*, IEEE Trans. Dielectr. Electr. Insul. **5**, 728 (1998).
7. L. Niemeyer, L. Pietronero, and H. J. Wiesmann, Phys. Rev. Lett. **52**, 1033 (1984).
8. H. J. Wiesmann and H. R. Zeller, J. Appl. Phys. **60**, 1770 (1986).
9. P. J. Sweeney, L. A. Dissado, and J. M. Cooper, J. Phys. D **25**, 113 (1992).
10. V. R. Kukhta, V. V. Lopatin, and M. D. Noskov, Zh. Tekh. Fiz. **65** (2), 63 (1995) [Tech. Phys. **40**, 150 (1995)].

11. D. I. Karpov, V. V. Lopatin, and M. D. Noskov, Élektrichestvo, No. 7, 59 (1995).

12. L. A. Dissado and P. J. Sweeney, Phys. Rev. B **48**, 16261 (1993).

13. J. C. Fothergill, L. A. Dissado, and P. J. Sweeney, IEEE Trans. Dielectr. Electr. Insul. **1**, 474 (1994).

14. J. V. Champion and S. J. Dodd, J. Phys. D **31**, 2305 (1998).

15. K. Wu, Y. Suzuoki, T. Muzutani, *et al.*, J. Phys. D **33**, 1197 (2000).

16. A. S. Malinovski, M. D. Noskov, M. Sack, *et al.*, in *Proceedings of IEEE 6th International Conference on Conduction and Breakdown in Solid Dielectrics, 1998*, p. 305.

17. M. D. Noskov, A. S. Malinovski, M. Sack, *et al.*, IEEE Trans. Dielectr. Electr. Insul. **7**, 725 (2000).

18. H. Singer, H. Steinbigler, and P. Weiss, IEEE Trans. Power Appar. Syst. **93**, 1660 (1974).

19. K. Kudo, IEEE Trans. Dielectr. Electr. Insul. **5**, 713 (1998).

20. S. Kobayashi, S. Maruyama, H. Kawai, *et al.*, in *Proceedings of 4th International Conference on Properties and Applications of Dielectric Materials, 1994*, p. 359.

21. H. Uehara and K. Kudo, in *Proceedings of 6th International Conference on Conduction and Breakdown in Solid Dielectrics, 1998*, p. 309.

Translated by V. Isaakyan

BRIEF
COMMUNICATIONS

The Emission of Photons and the Dynamics of Submicrodefects on the Surface of Noble Metals

K. B. Abramova, V. I. Vettegren', I. P. Shcherbakov, and V. N. Svetlov

Ioffe Physicotechnical Institute, Russian Academy of Sciences,
ul. Politekhnicheskaya 26, St. Petersburg, 194021 Russia

e-mail: abramova.mhd@pop.ioffe.rssi.ru

Received February 15, 2001

Abstract—The shape and concentration of defects, as well as the intensity of electroluminescence from the back side of Cu, Ag, and Au samples with their front side irradiated by a laser shot, are studied experimentally. It is shown that there exists a correlation between the electroluminescence intensity and the concentration of radiation-induced defects, which is consistent with the dislocation model of luminescence. © *2002 MAIK "Nauka/Interperiodica"*.

INTRODUCTION

It has been discovered [1] that the mechanical failure of metals is accompanied by the emission of photons. Later, the same emission was found to attend metal stressing, including stressing by a laser shot [2, 3]. These findings were explained [4–6] by the release of the dislocation core energy when the dislocations emerge on the surface. The radiation spectra contain lines that are assigned to electron transitions in dislocation cores. The luminescence intensity has been found to be proportional to the dislocation density [2, 7, 8].

It has been demonstrated [9, 11] that upon stressing Cu, Mo, Pd, and Au samples, "nanodefects" (pyramidal indentations) with the sides parallel to the easy slip planes are generated. The generation of these defects was explained by the emergence of dislocations on the surface due to stressing [11]. The density of dislocations crossing the surface can be judged from the shape and concentration of the nanodefects.

Upon studying mechanoluminescence from the back side of the metal samples with their front side irradiated by laser pulses of equal power [7, 8, 12, 13], it was found that the luminescence intensity decreases with increasing number of irradiating pulses. It was speculated that the luminescence decays because the number of dislocations emerging on the surface decreases during the irradiation. The aim of this work is to check this supposition.

EXPERIMENTAL

We measured the luminescence from the surface of metal plates under stress cycling and examined their submicropattern after each of the cycles. The plates were made of 99.96% pure copper, 4*N* pure silver, and 4*N* pure gold. The thickness and the surface area of the samples were 0.1–0.5 mm and 30×30 mm^2, respectively.

The emission was detected by an FÉU-136 photoelectric multiplier sensitive to the 300–800 nm range. The multiplier operated in the analog mode and was integrated with a digital storage oscilloscope.

The front surface of the plates was stressed by applying 1.06-μm laser pulses of width 1.5 ms. Their energy in the free-running mode was 24 J. The diameter of the optical beam was 2 mm. The incident beam intensity was varied with neutral optical filters so as not to damage the material within the irradiated zone on the front surface and still provide reliable luminescence from the back side of the plate. In our experiments, the energy delivered to the sample per pulse was $P \approx 0.2 P_{thr}$ (≈ 0.7 J), where P_{thr} is the threshold energy, i.e., the energy at which a plasma torch appears on the back side.

The surface pattern was examined with an RTP-1 scanning tunnel profilometer designed in the Research Institute of Physics, St. Petersburg State University. To provide the desired resolution and the operating stability of the instrument, we used a diffraction grating onto which a gold layer was evaporated. The measuring tips were made of a tungsten wire by electrochemical etching. The adequacy of their shapes was estimated from the images of the finest defects on the topograms. The tip of the profilometer can travel normally to the surface by a distance of no more than 1 μm. To remove surface microirregularities higher than 1 μm, the samples were polished by diamond pastes and then rinsed by acetone and alcohol.

The condensation of water vapor from air on the surface produces an electrolyte. The ionic current flowing in the electrolyte excludes the possibility of imaging the surface. To suppress this effect, the samples were
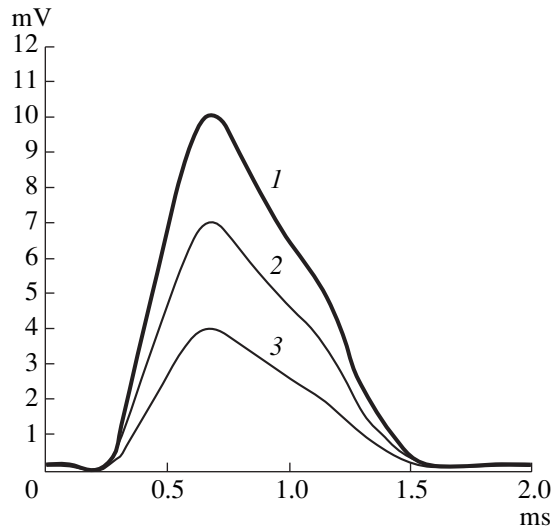
Fig. 1. Waveforms of photon emission signals from the photoelectric multiplier: *1*, first cycle; *2*, second cycle; and *3*, third cycle.

covered by polyurethane and the resulting space was purged by dry nitrogen.

RESULTS OF MEASUREMENTS

Figure 1 shows the waveforms of luminescence signals taken from the polished silver sample under cyclic stressing, i.e., when the sample was successively irradiated by three pulse of the same power. It is seen that the radiation intensity decreases from pulse to pulse. Within the dislocation model of luminescence, this is explained by a decrease in the number of dislocations emerging on the surface. Fragments of the topograms from the back side of the sample (whose luminescence signal is shown in Fig. 1) before and after irradiating the front side by the three laser pulses. In all of the topograms, similarly shaped defects marked by *1* in Fig. 2 are the basic surface features. The enlargement of such a defect is shown in Fig. 3. Their depth (height) vary from 15 to 30 nm, and their cross sizes lie between 50 and 200 nm. The 3D image of the defect looks like a
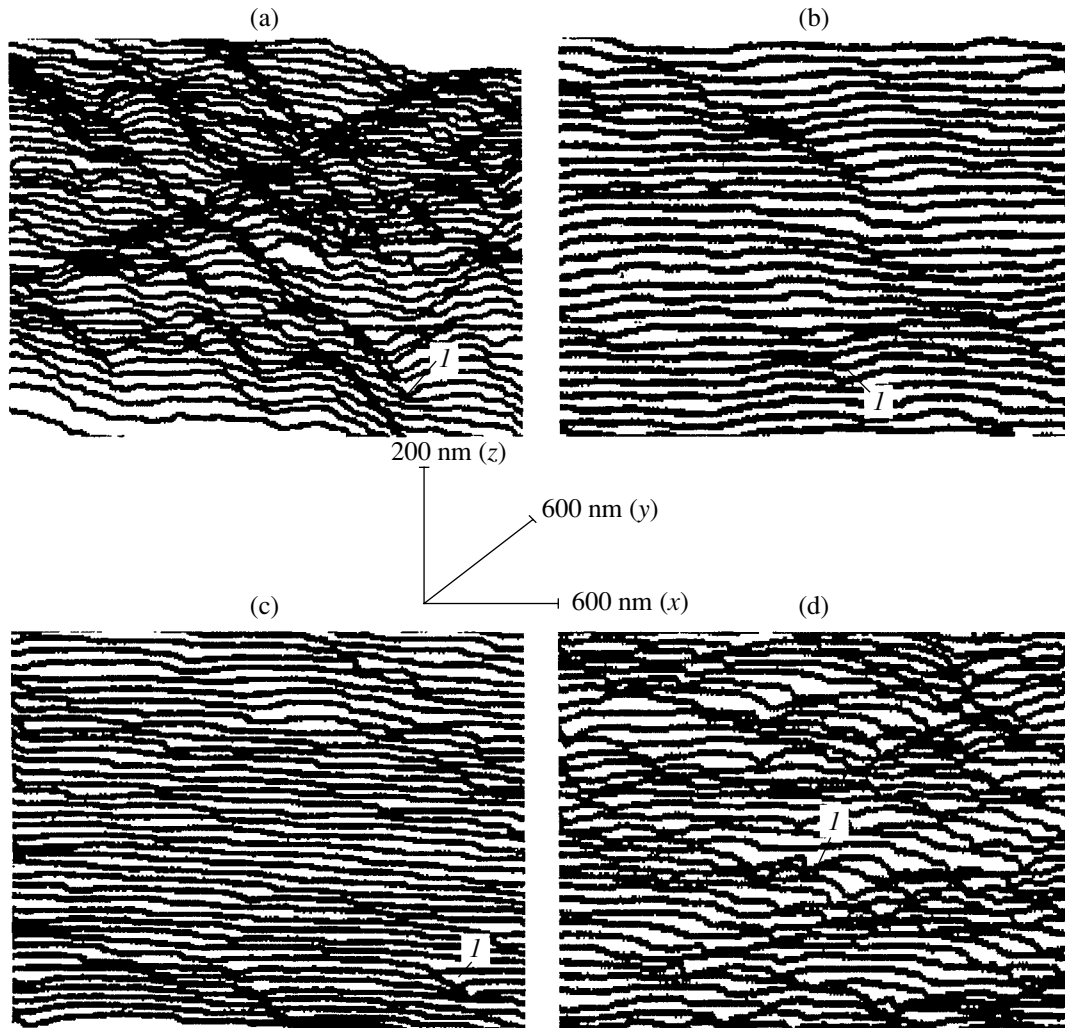
Fig. 2. Fragments of the topograms taken from the silver sample. (a) Nonirradiated polished surface; (b)–(d) surface after the first, second, and third cycle, respectively.
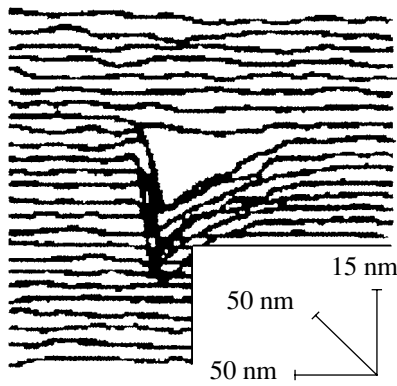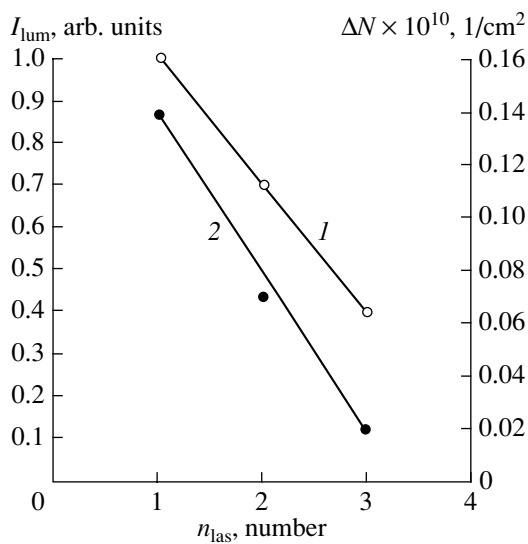
Fig. 3. Enlarged view of the resulting defect.

Fig. 4. (*1*) Concentration of the surface defects and (*2*) intensity of photon emission vs. cycle number.
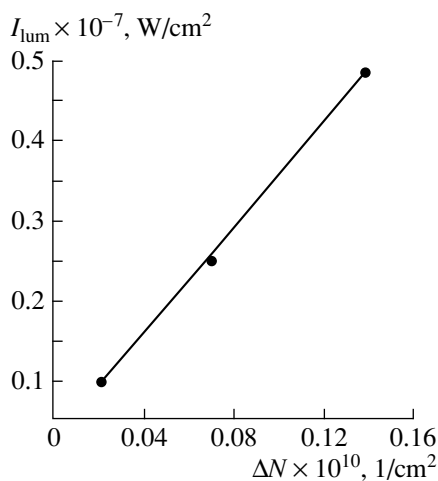

Fig. 5. Luminescence intensity vs. ΔN.

pyramidal indentation, three side walls of which are nearly perpendicular to the surface and the fourth one makes an angle of $\approx 30°$ with it. The vertex angle of the defect is $\approx 70°$. The orientation of the defect walls implies that the defects form (Figs. 2b–2d) when dislocations emerge on the surface under irradiation-induced stressing [9, 12, 13]. The same is also true for the defects appearing on the nonirradiated surface: it is known that during the mechanical polishing of the surface, dislocations move along the planes of easy slip, emerge on the surface, and produce the specific pattern shown in Fig. 2a [14].

It follows from the topograms that each stressing cycle changes the surface submicropattern, the changes being the same as upon long-term (static) stressing [10]. After the first cycle, the polished surface becomes even smoother. The coarse defects flatten, and the concentration of fine defects is low (Fig. 2b). One can assume that polishing generates loosely pinned dislocations at the surface. During stressing (during the first cycle in our work and within the initial half an hour of stressing in experiments described in [10]), they emerge on the surface. After the second laser pulse of the same power, the number of defects grows and they become closer spaced (Fig. 2c). Once the third pulse has been applied (Fig. 2d), the number of defects continues to grow and the defects coarsen [9–11]. To trace the defect evolution on the surface under stress cycling, we measured the total length of the nanodefect walls on the topograms [12] and calculated the difference ΔN between the values obtained after two successive cycles. The difference decreases with increasing number of the irradiating pulse

DISCUSSION

According to the dislocation theory of mechanoluminescence [4–6], its intensity is given by

$$I = C_1 N_d, \tag{1}$$

where

$$C_1 = \eta \langle h v \rangle \frac{P_h v_d}{a} \exp(-\alpha t);$$

η is the probability of radiative recombination, $\langle h v \rangle$ is the mean energy of photons emitted, α is a constant depending on the metal properties, t is time, P_h is the probability of hole generation, a is the lattice parameter, v_d is the dislocation velocity, and N_d is the mobile dislocation density.

Since it is assumed that luminescence occurs when the mobile dislocations emerge on the surface, we looked for a correlation between its intensity and the value of ΔN after each of the cycles. We measured the total length of the nanodefects and estimated the number of dislocations that had emerged on the surface from the value of ΔN [11].

Figure 4 plots ΔN (curve *1*) and the luminescence intensity (curve *2*) against the number of laser pulses. The correlation expected is obvious: both ΔN and the intensity linearly decrease with increasing number of pulses. Such behavior can be explained by a decrease in the number of dislocation sources.

Figure 5 demonstrates the dependence of the luminescence intensity (in absolute units obtained from Fig. 4 and the geometry of the experiments [11]) on ΔN. The intensity varies with ΔN as

$$I = C_2 \Delta N = C_1 \Delta N n. \tag{2}$$

From Fig. 5, it follows that $C_2 = 3.42 \times 10^{-17}$ W. Using the above sizes of the defect walls and the value of the Burgers vector for silver, $b \approx 0.3$ nm, we find the average number of dislocations per surface defect for $n \approx 600$. Then, $C_1 = C_2/n = 0.57 \times 10^{-19}$ W for one dislocation. This value agrees with the value $C_1 = 2.6 \times 10^{-19}$ W calculated from (1) by substituting the tabulated parameters for silver. The direct proportionality between the number of defects and the luminescence intensity was also observed for Cu and Au. For these metals, the calculated and experimental values of C_1 were also found to be close to each other.

CONCLUSION

Our experiments have shown that stress cycling of metal samples by laser pulses of equal power that leave the surface intact changes their surface microrelief and initiates luminescence from the back side. The luminescence intensity decreases from cycle to cycle. As the number of the irradiation cycles grows, the concentration of surface nanodefects changes but the rate of their accumulation drops.

The shape and the orientation of the nanodefects agree with the supposition that they form when mobile dislocations emerge on the surface. Starting from this supposition, we estimated the density of dislocations emerging under irradiation. It has been found that the rate of accumulation of dislocations decreases with time upon stress cycling.

It has been demonstrated that the mechanoluminescence intensity is proportional to the number of dislocations emerging on the metal surface. The proportionality coefficient is close to that calculated from the dislocation model of mechanoluminescence. Investigation into the mechanoluminescence process makes it possible to study the dynamics of mobile dislocations in the surface layer of metals.

REFERENCES

1. K. B. Abramova, V. P. Valitskiĭ, N. A. Zlatin, *et al.*, Zh. Éksp. Teor. Fiz. **71**, 1873 (1976) [Sov. Phys. JETP **44**, 983 (1976)].
2. K. B. Abramova and I. P. Shcherbakov, Zh. Tekh. Fiz. **64** (9), 76 (1994) [Tech. Phys. **39**, 901 (1994)].
3. A. M. Kondyrev, I. P. Shcherbakov, K. B. Abramova, and A. E. Chmel', Zh. Tekh. Fiz. **62** (1), 206 (1992) [Sov. Phys. Tech. Phys. **37**, 110 (1992)].
4. M. I. Molotskiĭ, Fiz. Tverd. Tela (Leningrad) **20**, 1651 (1978) [Sov. Phys. Solid State **20**, 956 (1978)].
5. M. I. Molotskii, Sov. Sci. Rev. **13** (3), 1 (1989).
6. B. R. Chandra, M. S. Ryan, Seema R. Simon, and M. Y. Ansari, Cryst. Res. Technol. **4**, 495 (1996).
7. K. B. Abramova, I. P. Shcherbakov, I. Ya. Pukhonto, and A. M. Kondyrev, Zh. Tekh. Fiz. **66** (5), 190 (1996) [Tech. Phys. **41**, 511 (1996)].
8. K. B. Abramova, I. P. Shcherbakov, and A. I. Rusakov, Zh. Tekh. Fiz. **69** (2), 137 (1999) [Tech. Phys. **44**, 259 (1999)].
9. V. I. Vettegren', V. L. Gilyarov, S. N. Rakhimov, and V. N. Svetlov, Fiz. Tverd. Tela (St. Petersburg) **40**, 668 (1998) [Phys. Solid State **40**, 614 (1998)].
10. V. I. Vettegren', S. N. Rakhimov, and V. N. Svetlov, Fiz. Tverd. Tela (St. Petersburg) **37**, 913 (1995) [Phys. Solid State **37**, 495 (1995)].
11. V. I. Vettegren', S. N. Rakhimov, and V. N. Svetlov, Fiz. Tverd. Tela (St. Petersburg) **40**, 2180 (1998) [Phys. Solid State **40**, 1977 (1998)].
12. K. B. Abramova, V. I. Vettegren', I. P. Shcherbakov, *et al.*, Zh. Tekh. Fiz. **69** (12), 102 (1999) [Tech. Phys. **44**, 1491 (1999)].
13. K. B. Abramova, A. M. Kondyrev, I. Ya. Puchonto, *et al.*, Proc. SPIE **3093**, 22 (1997).
14. I. I. Novikov, *Defects of Crystal Structure of Metals* (Metallurgiya, Moscow, 1975).

Translated by V. Isaakyan

The Structure of a Shock Electromagnetic Wave Synchronous with Several Waves Propagating in Coupled Transmission Lines with Different Types of Dispersion

A. B. Kozyrev

Institute of Physics of Microstructures, Russian Academy of Sciences, Nizhni Novgorod, 603600 Russia
e-mail: kozyrev@ipm.sci-nnov.ru
Received February 20, 2001

Abstract—A shock electromagnetic wave is studied numerically in the case when its structure is equally defined by two or more synchronous waves. The analysis relies on the previously suggested electrodynamic system of two coupled transmission lines one of which is normally dispersive (coaxial line) and the other is anomalously dispersive (ladder-type bar slow-wave structure). © 2002 MAIK "Nauka/Interperiodica".

The instability of the front of a shock electromagnetic wave (SEMW) propagating in nonlinear dispersive transmission lines has been extensively studied from the standpoint of its application for efficient direct video-to-radio pulse conversion [1–3]. However, these works restrict their analysis to the case in which the SEMW parameters (such as the duration of the leading edge) and the behavior of the field components as they approach constant values behind the SEMW front are governed by one synchronous wave. This work is the first to numerically detail the SEMW structure when it is equally defined by two or more synchronous waves. The analysis is exemplified by the electrodynamic system of two coupled transmission lines that has been proposed earlier (Fig. 1) [4]. One of these transmission lines (coaxial line) exhibits the normal dispersion; the other (ladder-type bar slow-wave structure), the anomalous dispersion.

As is known, when a step in the electromagnetic field propagates through a nonlinear transmission line filled with ferrite or with a medium exhibiting a nonlinear *C–V* characteristic, the slope of the wave front increases, producing an SEMW. In spectral terms, a nonlinear increase in the wave front slope means that higher frequency components are generated and the energy is transferred from lower to higher spectral components. Conversely, the stabilization of the SEMW front duration means the development of physical processes that absorb the energy and compensate for energy transfer to the high-frequency spectral components, thereby preventing a further increase in the SEMW front slope. As a rule, the SEMW structure is governed largely by the lowest-frequency mechanism. However, if the characteristic frequencies of different mechanisms are closely spaced or if the lowest-frequency mechanism is inefficient and cannot stabilize the SEMW front completely, one should expect that the SEMW structure will also depend on higher frequency

mechanisms of front stabilization. In particular, such a situation is observed in two coupled transmission lines with normal and anomalous dispersions [4]. The equivalent circuit is shown in Fig. 1. In such a system, within a certain range of its velocities, the SEMW is synchronous with two normal waves simultaneously. Numerical simulations have shown that, as a rule, the energy efficiency of the low-frequency wave excitation is low (because the SEMW weakly penetrates into the anomalously dispersive transmission line) and does not stabilize the SEMW front [4]. The authors of [4] considered a high-frequency system with $\omega_c \approx 24.4$ GHz (because their purpose was to increase the operating frequency of a pulsed RF oscillator built around a ferrite-filled transmission line). The final stabilization of the SEMW front occurred because of the incoherent magnetization reversal loss in the ferrite and the SEMW structure was defined by the low-frequency synchronous wave alone. The second (high-frequency) synchronous wave was not excited, because, in this case, the excitation would be a higher frequency process compared with the incoherent magnetization reversal of the ferrite.

This paper numerically simulates nonlinear SEMW propagation processes in a low-frequency system (in coupled transmission lines with the critical frequency $\omega_c \approx 6.24$ GHz), where the excitation of the second synchronous wave is a lower frequency process compared with the magnetization reversal of the ferrite and the SEMW front stabilizes by the excitation of two synchronous waves.

This situation is illustrated in Fig. 2, which plots the oscillograms of the voltage across the 600th cell of the normally and anomalously dispersive coupled transmission lines fed by a semiinfinite voltage pulse. In this case, the group velocities of the waves synchronous with the SEMW front are close to each other and the
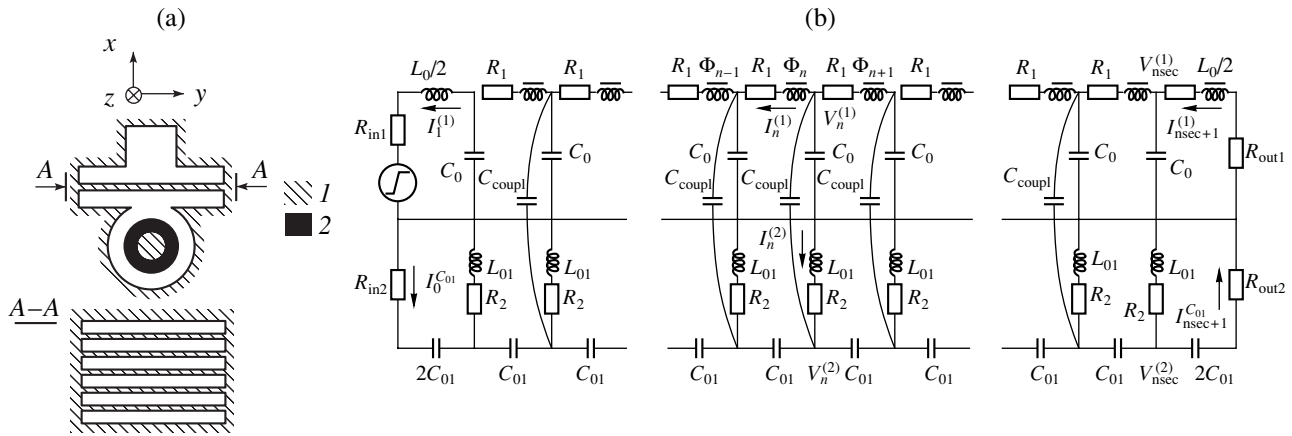
Fig. 1. (a) Design of the coupled quasi-coaxial transmission line and ladder-type bar slow-wave structure (T waveguide with an bar array: (*1*) metal and (*2*) ferrite) and (b) the equivalent circuit.
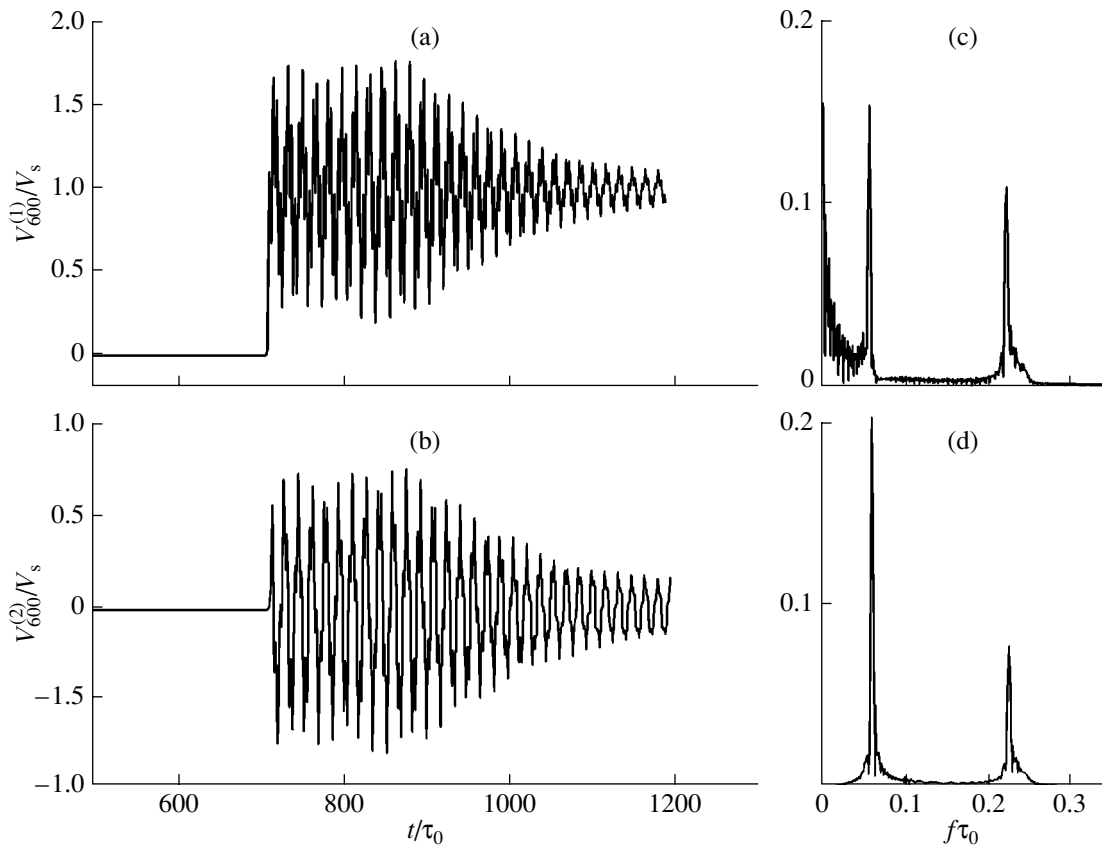
Fig. 2. Oscillograms of the voltage across the 600th cell of the line with the (a) normal ($V_{600}^{(1)}/V_s$) and (b) anomalous ($V_{600}^{(2)}/V_s$) dispersion normalized by the SEMW amplitude $V_s = 50$ kV and (c, d) their spectra for the relative SEMW velocity $v_s/v_0 = 0.85$ [4].

Parameters of the equivalent circuit are $\omega_c = 2/\tau_0 = 6.24$ GHz ($\tau_0 = (L_0C_0)^{1/2} = 0.32$ ns), $Z_0 = \sqrt{L_0/C_0} = 25$ Ω, $C_{01}/C_0 = 0.1$, $L_{01}/L_0 = 22.5$, and $C_{coupl}/C_0 = 0.16$.

energy of the oscillations generated is transferred with a constant rate equal to the difference between the SEMW velocity and the group velocities of the synchronous waves.

Thus, the oscillogram behind the SEMW front is the superposition of two radio pulses of equal length, as indirectly indicated by the same spectral widths of the peaks in the oscillogram's spectra. The shape of the

trailing edge of the radio pulse is defined by the dispersion of the group velocities of the synchronous waves [3]. The front has a sufficiently complex structure. Although its duration depends on the period of the lower frequency synchronous wave, the field (voltage) varies with a frequency on the order of that of the higher frequency synchronous wave in some regions of the front.

In the case we study, the energy efficiency of the synchronous wave excitation (the ratio of the power spent on the excitation of the synchronous wave to the total power applied to the SEMW front) is 0.46 for both waves. The total portion of the energy spent on the generation of the RF waves is high (92%). The remaining portion of the energy applied to the SEMW front goes into the incoherent magnetization reversal of the ferrite. Note that, in the electrodynamic system in the form of a transmission line with the next nearest links coupled [1, 2], the simultaneous synchronism with two waves (for example, with the lower frequency forward normal wave and the higher frequency backward spatial harmonic) can also be achieved at certain parameters of the transmission line and a certain SEMW velocity. However, in this case, the lower frequency synchronous wave is excited efficiently and completely stabilizes the SEMW front. The equally efficient excitation of two synchronous waves becomes impossible and only one frequency dominates in the spectrum of the oscillogram.

In Fig. 2, the group velocities of the synchronous waves have the same direction as the SEMW. As follows from the dispersion characteristics of the system of coupled transmission lines [4], the situation where the group velocities are directed oppositely may arise. The wavetrain produced in this case also has an irregular two-frequency structure depending on the group velocities of the synchronous waves and on their local dispersion. However, unlike the previous case, the energies of the waves are transferred in opposite directions, so that actually two radio pulses are generated simulta-neously. One of them travels to the input of the anomalously dispersive line and the other, to its output, so that they can be released on the associated matched loads.

To conclude, we note that, by analogy with coupled normally and anomalously dispersive lines, one can consider a multiconductor electrodynamic system consisting of several coupled transmission lines. Here, the SEMW structure will be defined by a specified number of synchronous waves, their excitation efficiency being controlled by the coupling between the lines. The excitation efficiency of the higher frequency waves will depend on the coefficients of line coupling and on the excitation efficiencies of the lower frequency waves. This would make it possible to generate multifrequency pulses with the energy uniformly distributed over the frequencies. The spectrum of such a radio pulse can be controlled electronically (by varying the bias current or the SEMW amplitude).

REFERENCES

1. A. M. Belyantsev, A. I. Dubnev, S. L. Klimin, *et al.*, Zh. Tekh. Fiz. **65** (8), 132 (1995) [Tech. Phys. **40**, 820 (1995)].

2. A. M. Belyantsev and A. B. Kozyrev, Zh. Tekh. Fiz. **70** (6), 78 (2000) [Tech. Phys. **45**, 747 (2000)].

3. A. M. Belyantsev and A. B. Kozyrev, Zh. Tekh. Fiz. **68** (1), 89 (1998) [Tech. Phys. **43**, 80 (1998)].

4. A. M. Belyantsev and A. B. Kozyrev, Zh. Tekh. Fiz. **71** (7), 79 (2001) [Tech. Phys. **46**, 864 (2001)].

Translated by A. Khzmalyan

Canonical Averaging of the Schrödinger Equation

A. G. Chirkov

St. Petersburg State Technical University,
ul. Politekhnicheskaya 29, St. Petersburg, 195251 Russia

Received June 13, 2001

Abstract—The representation of the Schrödinger equation in the form of a classical Hamiltonian system makes it possible to construct a unified perturbation theory that is based on the theory of canonical transformations and covers both classical and quantum mechanics. Also, the closeness of the exact and approximate solutions of the Schrödinger equation can be approximately estimated with such a representation. © *2002 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The main objective of investigation in the nonrelativistic quantum theory is the Schrödinger equation [1]

$$i\hbar\frac{\partial\Psi(q,t)}{\partial t} = \hat{H}\Psi(q,t), \tag{1}$$

where $i^2 = -1$; $\hbar = 1.054 \times 10^{-27}$ erg s is the Planck constant; $q = (q_1, q_2, \ldots, q_n)$ is, generally speaking, a point in the configurational space of the respective classical system; t is time; $\Psi(q, t)$ is a complex function with the integrable squared modulus; and $\hat{H}$ is the self-(adjoint (symmetric) operator acting in the Hilbertian) space.

Schrödinger equation (1) should be solved with the initial condition $\Psi(q, 0) = \Psi_0(q)$ and some boundary conditions.

The case treated in the perturbation theory arises if the operator $\hat{H}$ can be represented as the sum

$$\hat{H} = \hat{H}_0 + \varepsilon\hat{V}, \quad 0 < \varepsilon \ll 1 \tag{2}$$

of two self-adjoint operators. Here, problem (1) for the operator $\hat{H}_0$ is assumed to be exactly solvable (generating approximation), whereas the other operator is assumed to be small in a sense (perturbation) [2].

In such a representation, most physically interesting problems prove to be mathematically ill-posed, because perturbation operators are usually unbounded and even not self-adjoint. The latter fact is because physicists almost never distinguish the concepts of self-adjoint and symmetric operators. Therefore, the domain of definition for an operator in the Hilbertian space remains obscure, which prevents the use of methods of the spectral theory [2, 3]. Because of this, in practical calculations, it is hard to set the applicability condition of the perturbation theory and estimate the difference between the exact and approximate solutions.

However, it becomes possible to reduce the Schrödinger equation to the form of a classical Hamil-

tonian system (well-known in nonlinear mechanics) and accordingly to use rigorous and simpler methods of classical mechanics for its investigation. Consider Eq. (1) with Schrödinger operators (2), i.e., the problem

$$i\hbar\frac{\partial\Psi(q,t)}{\partial t} = (\hat{H}_0 + \varepsilon\hat{V})\Psi(q,t),$$
$$\Psi(q, 0) = \Psi_0(q), \tag{3}$$

where the operator $\hat{H}_0$ is time-independent and ε is a formal small parameter.

Along with problem (3), we consider the generating approximation, which is obtained from (3) at $\varepsilon = 0$:

$$i\hbar\frac{\partial\Psi^0(q,t)}{\partial t} = \hat{H}_0\Psi^0(q,t), \quad \Psi^0(q, 0) = \Psi_0(q). \tag{4}$$

The general solution of problem (4) found by the Fourier method can be represented in the form (for the sake of simplicity, the spectrum is assumed to be discrete)

$$\Psi^0(q, t) = \sum_{n=0}^{\infty} c_n^0\psi_n^0(q)\exp(-i\omega_n^0 t),$$
$$c_n^{(0)} = \int\Psi_0(q)\psi_n^{0*}(q)dq, \quad \omega_n^0 = E_n^0/\hbar, \tag{5}$$

where $\psi_n^0(q)$ and E_n^0 are the eigenfunctions and the eigenvalues of the problem

$$\hat{H}_0\psi_n^0(q) = E_n^0\psi_n^0(q) \tag{6}$$

(asterisk means complex conjugation).

It follows from the self-conjugacy of the operator $\hat{H}_0$ that the expansion

$$\Psi(q, t) = \sum_{n=0}^{\infty} c_n(t)\psi_n^0(q)\exp(-i\omega_n^0 t) \tag{7}$$

is valid for $\Psi(q, t) \in L^2$.

Substituting expansion (7) into (3), we find the equations for the expansion coefficients $c_n(t)$ and $c_n^*(t)$:

$$\dot{c}_n(t) = -i\varepsilon \sum_{m=0}^{\infty} v_{nm} c_m(t) \exp(-i\omega_{mn}^0 t),$$

$$\dot{c}_n^*(t) = i\varepsilon \sum_{m=0}^{\infty} v_{nm} c_m^*(t) \exp(i\omega_{mn}^0 t),$$
(8)

$$v_{mn}(t) = \frac{1}{\hbar} \int \psi_m^{0*}(q) \hat{V}(q,t) \psi_n^0(q) dq, \qquad (9)$$

where dot denotes the time derivative and $\omega_{mn}^0 = \omega_m^0 - \omega_n^0$.

Set of equations (8) proves to be Hamiltonian (in the classical meaning) with the Hamiltonian

$$\varepsilon H_1(c, c^*, t) = -i\varepsilon \sum_{n,m=0}^{\infty} v_{nm} c_n^* c_m \exp(i\omega_{nm}^0 t), \quad (10)$$

which describes the classical distributed system with the infinite number of internal resonances. The Hamiltonian property is provided by the fact that the matrix v_{nm} is Hermitian (the perturbation operator is self-adjoint). We separate the fundamental resonance $\omega_n^0 = \omega_m^0$ to represent Hamiltonian function (10) in the form

$$\varepsilon H_1(c, c^*, t) = -i\varepsilon v_{nn} c_n c_n^*$$
$$- i\varepsilon \sum_{n,m=0}^{\infty} {}' v_{nm} c_n^* c_m \exp(i\omega_{nm}^0 t).$$
(11)

Now we pass from the variables c_n and c_n^* to the real variables action and angle, I_n and ψ_n, by the formulas

$$c_n = \sqrt{I_n} \exp(-i\psi_n), \quad I_n = c_n c_n^*,$$

$$c_n^* = \sqrt{I_n} \exp(i\psi_n), \quad \psi_n = -\arctan \frac{(c_n - c_n^*)}{i(c_n + c_n^*)}.$$
(12)

Then, set (8) is represented in the form

$$\dot{I}_n = \varepsilon 2 \sum_{m=0}^{\infty} {}' \sqrt{I_n I_m} \operatorname{Im}\{ v_{nm} \exp[-i(\psi_m - \psi_n + \omega_{mn}^0 t)]\},$$

$$\dot{\psi}_n = \varepsilon v_{mn} + \varepsilon \sum_{m=0}^{\infty} {}' \sqrt{\frac{I_n}{I_m}}$$
(13)

$$\times \operatorname{Re}\{ v_{nm} \exp[-i(\psi_m - \psi_n + \omega_{mn}^0 t)]\};$$

Hamiltonian function (12), in the form

$$\varepsilon H_1(I, \psi, t) = \varepsilon \sum_{m,n=0}^{\infty} v_{mn} \sqrt{I_n I_m}$$
$$\times \exp[-i(\psi_m - \psi_n + \omega_{mn}^0 t)].$$
(14)

In (13), prime means that the fundamental resonance (i.e., the terms with $n = m$) is separated.

The Planck constant does not appear in sets (8) and (13) explicitly; hence, these are classical Hamiltonian systems with the infinite number of internal resonances. The methods of their investigation were developed by Bogolyubov, Mitropol'skiĭ, Grebenikov, Arnold, and their disciples [4–8]. In particular, the estimation of the difference between the exact and approximate solutions by norm and finding the applicability condition for the averaging method for these systems are given by the Los' theorem [4, 9], which is a generalization of the Bogolyubov theorem for the case of an infinitely dimensional coordinate Hilbertian space.

Using the method of averaging, one can construct a solution of sets (8) and (13), which approximates the exact one with any given accuracy.

The canonical form of sets (8) and (13) makes it possible to pass to the evolutionary equations by using Hamiltonian functions (10), (11), or (14) alone, i.e., by calculating the average Hamiltonian function. For example, for (10) and (11), the second-order approximation $\overline{H}^{(2)}$ for the averaged Hamiltonian is constructed using the formulas

$$\overline{H}^{(2)}(\bar{c}, \bar{c}^*) = \varepsilon \overline{H}_1(\bar{c}, \bar{c}^*) + \varepsilon^2 \overline{H}_2(\bar{c}, \bar{c}^*),$$

$$\overline{H}_1 = \langle H_1 \rangle, \quad \overline{H}_2 = -\left\langle \frac{\partial \tilde{H}_1}{\partial \bar{c}^*} \frac{\partial \{H_1\}}{\partial \bar{c}} \right\rangle,$$
(15)

where $\bar{c}_n$ and $\bar{c}_n^*$ are the evolutionary components of the variables c_n and c_n^*, respectively, and

$$\langle f \rangle = \lim_{T \to \infty} \frac{1}{T} \int_{t_0}^{t_0+T} f(\bar{c}, \bar{c}^*, t) dt,$$

$$\tilde{f}(\bar{c}, \bar{c}^*, t) = f(\bar{c}, \bar{c}^*, t) - \langle f \rangle,$$
(16)

$$\{f\} = \int \tilde{f}(\bar{c}, \bar{c}^*, t) dt.$$

Here, an arbitrary function of the slow variables $\bar{c}$ and $\bar{c}^*$ is put equal to zero.

Note that $\overline{H}$ is the integral of motion for the averaged equations, i.e., the adiabatic invariant [7, 8].

The first approximation $\bar{c}_n^{(1)}$ for the expansion coefficients c_n is found using the formula $c_n^1 = \bar{c}_n$, where

$\bar{c}_n$ satisfies the equations

$$\dot{\bar{c}}_n = \varepsilon \frac{\partial \overline{H}_1}{\partial \bar{c}_n^*}. \qquad (17)$$

The second-order approximation $c_n^{(2)}$ for the expansion coefficients c_n is calculated using the formulas

$$c_n^{(2)} = \bar{c}_n + \varepsilon \frac{\partial \{H_1\}}{\partial \bar{c}_n^*}, \qquad (18)$$

where the second-order approximation for the evolutionary components $\bar{c}_n^{(2)}$ is derived from the equations

$$\dot{\bar{c}}_n = \varepsilon \frac{\partial \overline{H}_1}{\partial \bar{c}_n^*} + \varepsilon^2 \frac{\partial \overline{H}_2}{\partial \bar{c}_n^*}. \qquad (19)$$

Thus, one can identify the variables c_n and their evolutionary components only in the first approximation. The relations obtained make it possible to perform calculations in classical and quantum mechanics by unified formulas.

REFERENCES

1. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory* (Fizmatgiz, Moscow, 1963; Pergamon, New York, 1977, 3rd ed.).
2. T. Kato, *Perturbation Theory for Linear Operators* (Springer-Verlag, Berlin, 1966; Mir, Moscow, 1972).
3. M. C. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1978; Mir, Moscow, 1982), Vol. 4.
4. Yu. A. Mitropol'skiĭ, *Averaging Method in Nonlinear Mechanics* (Naukova Dumka, Kiev, 1971).
5. Yu. A. Mitropol'skiĭ and B. I. Moseenkov, *Asymptotic Solutions of Partial Differential Equations* (Vishcha Shkola, Kiev, 1976).
6. E. A. Grebenikov, *Averaging Method in Applied Problems* (Nauka, Moscow, 1986).
7. V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Nauka, Moscow, 1989; Springer-Verlag, New York, 1989).
8. V. I. Arnold, *Supplementary Chapters to the Theory of Ordinary Differential Equations* (Nauka, Moscow, 1978).
9. F. S. Los', Ukr. Mat. Zh. **2** (3), 87 (1950).

Translated by D. Zhukhovitskiĭ

BRIEF
COMMUNICATIONS

A Dislocation Mechanism of Friction
between a Nanoprobe and Solid Surface

S. Sh. Rekhviashvili

Institute of Applied Mathematics and Automation, Design Office of the Research Center,
Russian Academy of Sciences, Nal'chik, 360000 Russia
Received June 18, 2001

Abstract—A dislocation mechanism of friction between an atomic-force microscope (AFM) probe and an atomically smooth solid surface is put forward. In this mechanism, the contact region is represented by an edge dislocation. The triboacoustic emission measured with an AFM shows the dislocation nature of friction. The friction force is calculated for a parabolic tip. © *2002 MAIK "Nauka/Interperiodica".*

INTRODUCTION

According to modern concepts, direct proportionality between the friction force and the load for macroscopic rubbing bodies (the Amonton law) stems from the superposition of many micro- and nanocontacts between friction surfaces. As a result, the actual total contact area is several orders of magnitude smaller than the apparent area [1]. At present, friction in nanocontacts is investigated by an AFM, which is among the most powerful instruments used for studying solid surfaces at the atomic or molecular level and measuring ultralow forces (of the order of 1 pN), as well as in nanotechnology. An AFM mechanically scans the surface (in the contact or contactless mode) with a special sensor, a cantilever, involving a support and an elastic tipped microbeam.

The atomism of friction is still in its infancy. To date, most theoretical works either proceed from poorly substantiated model representations, which do not allow us to explicitly express the friction force [2] or use numerical simulation by the molecular dynamics method [3, 4] and the Monte Carlo technique [5, 6]. The results of both dramatically depend on the number of atoms in a system and on the approximation of the interaction potential. However, in the latter case, one can simulate the images and observe the processes occurring immediately in the contact area.

In this paper, a dislocation mechanism of friction between the nanoprobe and an atomically smooth surface is put forward for the first time. Here, the contact area is represented by an edge dislocation enclosing an area Ω. This assumption is based on the following.

(i) As follows from X-ray diffraction data, the dislocation displacement is a multiple of the Burgers vector or, virtually, of the lattice spacing. A dislocation moves jumpwise in the field of the Peierls–Nabarro forces, which causes the sticking–glide effect, often observed in experiments with AFMs.

(ii) From the dislocation theory, it is known that the tangential stress is minimal upon gliding close-packed planes or planes with incommensurate structures. This seems to be the reason for low values of friction factors and wear rates under solid friction at the nanostructure level.

(iii) The simulation of contact interaction between an AFM probe and a solid surface by molecular dynamics methods and the Monte Carlo procedure confirms the formation of the "bad" crystal region [3–6].

(iv) Edge dislocation glide generates, as a rule, low-frequency acoustic waves. Within the wave zone, the velocity vector of elements of the medium is proportional to the integral [7]

$$\mathbf{V} \sim \int \frac{\partial^2}{\partial t^2} j_{ks} d\mathbf{r},$$

where j_{ks} is the dislocation flux density tensor (single-component in our case) and $\mathbf{r}$ is the radius vector of a dislocation.

As follows from this expression, acoustic emission must be observed under the nonstationary motion of the probe, i.e., when the second time derivative of the dislocation flux density is nonzero. As will be shown later, this fact is confirmed in experiments.

EXPERIMENTAL

To verify the dislocation mechanism of friction, we used a specially designed AFM whose block diagram is shown in Fig. 1a. A rectangular pulse from driving generator *1* is applied to piezoelectric element *2*. As a result, thin tungsten *3*, fixed on the piezoelectric element, executes mechanical vibration, slipping over the surface of specimen *4*. The vibration amplitude and the slip velocity of the probe depend on the voltage applied to the piezoelectric element and on the triggering signal frequency. During scanning, pickup *5* measures the

intensity of sound waves generated by the friction of the tip with the surface. The signal from the pickup is applied to low-noise amplifier *6*, tunable band-pass filter *7*, and then to oscilloscope *8*.

The waveform of the triboacoustic signal from the AFM is shown in Fig. 1b, where *1* is the signal applied to the piezoelectric element and *2* is the output signal. The leading and trailing edges of triggering pulse *1* correspond to two end positions of the probe. In either position, the piezoelectric element abruptly stops, while the probe executes damped vibrations near the surface. As follows from experiments, the acoustic emission intensity nearly linearly grows with the load applied to the probe. The peaks in curve *2* correspond to the turning points of the probe. In other words, the triboacoustic emission peaks when the probe acceleration is maximum, which is typical of the dislocation mechanism of acoustic wave generation [7].

ANALYSIS

The friction force is calculated starting from the classical concept of edge dislocation. The work needed to displace a dislocation is $A = \tau b \Omega$, where τ is the shear stress, b is the Burgers vector, and Ω is the area enclosed by the dislocation [8]. Let an elementary slip (microslip) of the AFM probe be equal to b; then, the tangential force is given by

$$F = \tau \Omega. \qquad (1)$$

The force needed to break interatomic bonds under isobaric isothermal conditions is

$$\Delta F = \frac{\Delta G}{d N_A},$$

where ΔG is the Gibbs energy, d is the interatomic spacing, and N_A is the Avogadro number.

In the contact area, the number of atoms is $N = n_s \Omega$, where n_s is the sheet atom concentration; therefore, in order to break the whole contact, one should apply the force

$$F = \Delta F N = \frac{\Delta G n_s \Omega}{d N_A}. \qquad (2)$$

Comparing (1) and (2) gives

$$\tau = \frac{n_b}{N_A} \Delta G, \qquad (3)$$

where $n_b = n_s/d$ is the volume concentration of the atoms.

The shear modulus vs. heat of sublimation for various solids is shown in Fig. 2, which approves the validity of formula (3) (i.e., the direct proportionality between τ and ΔG). The shear stress was assumed to be proportional to the shear modulus; the Gibbs energy, to the heat of sublimation. The data points are taken from [9, 10], and the regression line is given by $y(x) = $
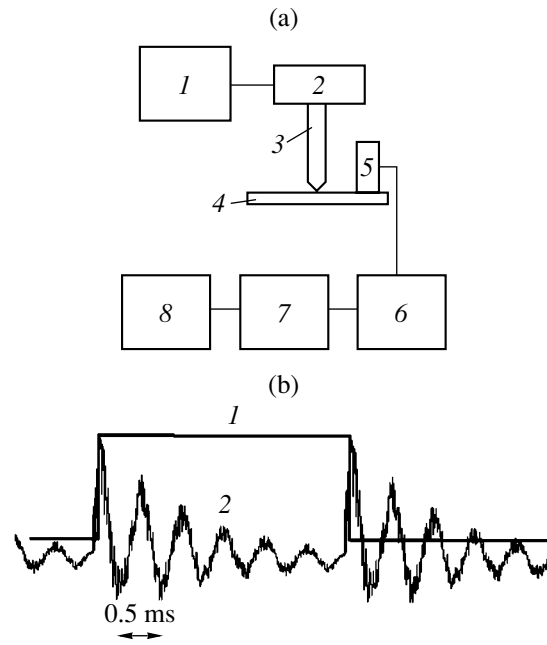
Fig. 1. (a) AFM block diagram and (b) waveform of the triboacoustic signal upon interaction between the tungsten probe and the single-crystal Si (111) surface. The radius of curvature of the probe is about 100 nm; load, no more than 0.1 μN.
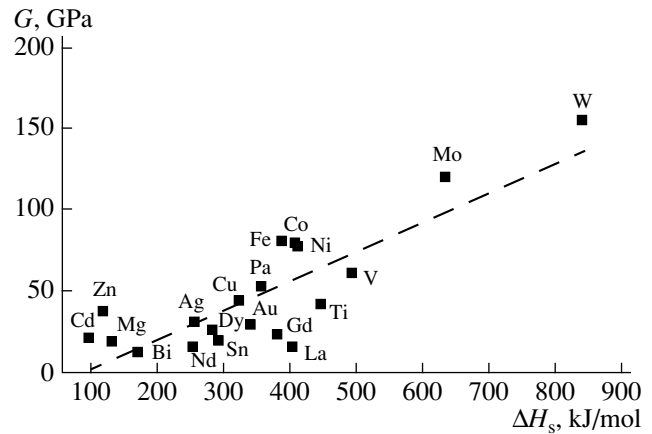

Fig. 2. Correlation between the shear modulus and heat of sublimation.

0.186x − 19.132. The correlation factor was found to be 0.842.

The experiments with AFMs indicate that the friction is nonzero even in the absence of an external load [11, 12]. This is due to the fact that after unloading, the adhesive force and the capillary force, which appears during the AFM operation in a humid atmosphere, still act. In calculating the contact area Ω, these forces can be taken into account in the Deryagin–Muller–Toporov approximation [13]. As a result, we obtain the friction

force in the form of

$$F = \frac{\pi n_b \Delta G}{N_A} \left[\frac{R}{E} (F_\perp + 2\pi R\gamma + F_c) \right]^{2/3},$$

$$\frac{1}{E} = \frac{3}{4} \left[\frac{1-\mu_1^2}{E_1} + \frac{1-\mu_2^2}{E_2} \right],$$

(4)

where R is the radius of curvature of the probe tip; $F_\perp$ is the load; γ is the specific energy of adhesion for flat surfaces; F_c is the capillary force; and $E_{1,2}$ and $\mu_{1,2}$ are the moduli of elasticity and the Poisson ratios for the tip and the specimen, respectively.

Note that the friction force depends on the load as $F \sim F_\perp^{2/3}$, which is typical of experiments with AFMs [11, 12].

CONCLUSION

In this paper, various physical processes in the AFM probe–specimen system were analyzed. The experimental and theoretical investigations performed completely confirmed the dislocation nature of the interaction between an atomically smooth surface and the nanoprobe.

One more point is noteworthy. When the glide of a little cluster of atoms over the crystal surface was numerically simulated within the framework of the dislocation model, it was found that the AFM images should possess the typical fractal properties: (i) self-similarity when the scale (in our case, the size of the contact area) changes and (ii) the fractional Hausdorff dimension in the range between 2 and 3. The images may be heavily distorted, up to the total disappearance or inversion of the contrast, when the maxima and minima on the scan change places. These phenomena call for special consideration.

REFERENCES

1. *Physics Encyclopedia*, Ed. by A. M. Prokhorov (Sov. Éntsiklopediya, Moscow, 1983), pp. 765–766.
2. G. V. Dedkov, Pis'ma Zh. Tekh. Fiz. **24** (19), 44 (1998) [Tech. Phys. Lett. **24**, 766 (1998)]; G. V. Dedkov, Zh. Tekh. Fiz. **70** (7), 96 (2000) [Tech. Phys. **45**, 909 (2000)].
3. U. Landman *et al.*, Science **248**, 454 (1990).
4. A. V. Pokropivnyĭ *et al.*, Pis'ma Zh. Tekh. Fiz. **22** (2), 1 (1996) [Tech. Phys. Lett. **22**, 46 (1996)]; Zh. Tekh. Fiz. **67** (12), 70 (1997) [Tech. Phys. **42**, 1435 (1997)].
5. B. S. Good and A. Banerjea, J. Phys.: Condens. Matter **8**, 1325 (1996).
6. S. Sh. Rekhviashvili, in *Proceedings of IV All-Russia Symposium "Mathematical Simulation and Computer Technologies," Kislovodsk, 2000*, Ed. by A. A. Samarskiĭ, Vol. 2, Part 1, p. 25.
7. V. D. Natsik and K. A. Chishko, Fiz. Tverd. Tela (Leningrad) **14**, 3126 (1972) [Sov. Phys. Solid State **14**, 2678 (1972)].
8. J. Friedel, *Dislocations* (Pergamon, Oxford, 1964; Mir, Moscow, 1967).
9. *Handbook of Physical Quantities*, Ed. by I. S. Grigoriev and E. Z. Meilikhov (Énergoatomizdat, Moscow, 1991; CRC Press, Boca Raton, 1997).
10. A. R. Regel' and V. M. Glazov, *The Periodic Law and Physical Properties of Electronic Melts* (Nauka, Moscow, 1978).
11. H. Tanaka *et al.*, Thin Solid Films **342**, 4 (1999).
12. R. E. Carpick and M. Salmeron, Chem. Rev. **97**, 1163 (1997).
13. N. A. Burnham and A. J. Kulik, in *Surface Forces and Adhesion: Handbook of Micro and Nanotribology* (CRC Press, Boca Raton, 1997), p. 21.

Translated by B. Malyukov