Technical Physics, Vol. 50, No. 6, 2005, pp. 673–679. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 6, 2005, pp. 1–7.

Electric Field and the Charge Distribution on the Surface of an Insulator in a Vacuum

V. K. Belyaev

National Technical University Kiev Polytechnical Institute,
ul. Peremogi 37, Kiev, 03056 Ukraine
e-mail: bel_vk@ua.fm

Received August 3, 2004

Abstract—The free charge steady-state distribution over the insulator surface that arises in a strong electric field in a vacuum can be found by solving the boundary-value problem for the electrostatic field strength if the angle between the field vector and vacuum–insulator interface is given. A general solution to this boundary-value problem is derived for the case of an in-plane field and rectilinear interfaces. Laws of charge and field formation that follow from the solution obtained are considered. Formulas for the electric field strength and charge density in terms of elementary functions are obtained for a number of particular cases. Power-type expressions for the electric field and a critical angle between the electrode and insulator surface that describe the field behavior and charge distribution near the vacuum–insulator–electrode contact are derived. *© 2005 Pleiades Publishing, Inc.*

INTRODUCTION

A free electric charge arises on the surface of an insulator placed in a strong longitudinal field in a vacuum. This phenomenon is one of those governing the dielectric strength over the surface of vacuum insulation [1, 2]. Taking into account field distortions due to this charge is necessary both for designing vacuum equipment and for due interpretation of experimental data [3–5]. The emergence of the free charge is related to secondary-emission-induced processes on the insulator surface that occur as secondary electron avalanches travel over it. The avalanches are generated near the contact of the insulator surface with the negative electrode when a critical field strength is attained in this region. Moving along the insulator in a longitudinal electric field, the electrons interact with the insulator surface, forming a positive free charge on it. The electric field of this surface charge influences the electron paths and the energy of electron–surface interaction, thereby modifying the charge distribution. The charge distribution is stabilized if the average energy of electron interaction with the surface becomes equal to the lower of two possible energy values (W_1) at which the coefficient of secondary emission from the insulator surface equals unity. The condition for charge stability was obtained from analysis of the path of an averaged electron with regard to the cosine distribution of the angle at which secondary electrons escape from the insulator surface [6]. This condition can be recast for the electric field components on the surface in the form

$$E_n/E_\tau = M = \text{const},$$
$$|M| = \sqrt{2W_0/(W_1 - W_0)} = \tan(\pi\beta_\sigma), \quad (1)$$

where E_n and E_τ are the normal and tangential components of the electric field vector on the insulator surface, M is a constant the absolute value of which depends on the secondary emission properties of the insulator surface, W_0 is the energy with which the electrons leave the surface, and β_σ is the smallest angle between the field vector and the surface (expressed as fractions of π) (Fig. 1).

According to this condition, the free charge is so distributed over the insulator surface that the ratio between the field normal and tangential components

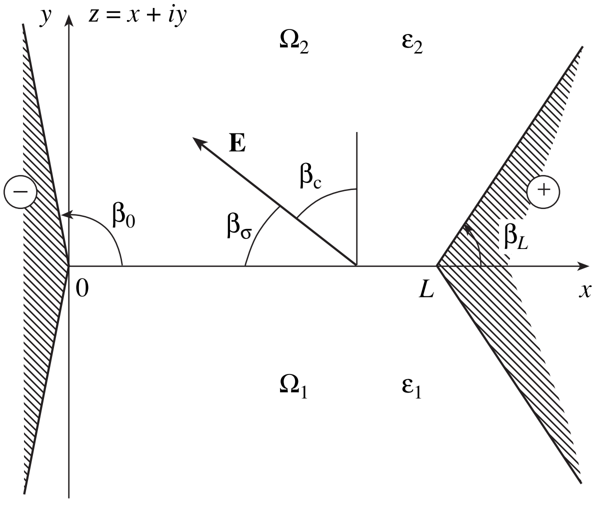


Fig. 1. Model scheme used in the calculation of the in-plane field for the case when condition (1) of charge stability is fulfilled on the insulator–insulator interface.

remains constant throughout the surface; i.e., the tilting angle of the field vector is specified.

Stability condition (1) can be applied for calculating the electric field in the system and the steady-state distribution of the charge over the insulator surface. The arising problem of calculating the electrostatic field at a preset tilting angle of the field vector differs from standard boundary-value electrophysical problems of first, second, and third kind, and its solution requires another approach.

In this work, the author derives an analytical solution to the problem stated for in-plane electric field $\mathbf{E} = \mathbf{e}_xE_x + \mathbf{e}_yE_y$ and rectilinear interfaces between the insulators and electrodes (Fig. 1). Two insulating media (one of them is a vacuum) occupy the interelectrode space so that the insulator—insulator interface of length L divides the space into regions Ω_1 and Ω_2. Constant voltage U applied to the electrodes produces an electric field parallel to the insulator–insulator interface with average strength $E_0 = U/L$ in the electrode gap. We assume that condition (1) is fulfilled on the side of region Ω_2 (vacuum). First, we consider a solution to the problem for the case of a known charge distribution over the insulator—insulator interface (hereafter interface), which is then used in analysis of a solution to the basic problem (where the distribution is unknown).

1. CHARGE DISTRIBUTION OVER THE INSULATOR SURFACE IS KNOWN

When the charge distribution over the insulator surface (hereafter surface) is known, calculation of the electric field reduces to the solution of the electrostatic equations

$$\mathrm{curl}\mathbf{E} = 0, \quad \mathrm{div}(\varepsilon\mathbf{E}) = 0$$

subject to the standard matching conditions at the interface between like media, which take into account the presence of a free charge with known density σ,

$$\varepsilon_2 E_{2n} - \varepsilon_1 E_{1n} = \sigma, \quad E_{1\tau} = E_{2\tau}.$$

Here, E_{kn} and $E_{k\tau}$ ($k = 1, 2$) are the normal and tangential components of the electrostatic field vector in region Ω_k and ε_k is the permittivity of the medium in region Ω_k.

The electrostatic equations enable introduction of a piecewise analytical function of complex field strength $E_k(z) = E_{kx} - iE_{ky}$ ($k = 1, 2$) in complex plane $z = x + iy$. Functions E_2 and E_1 are analytical in the upper and lower regions, respectively. After the function introduced has been mapped into the canonical domain, the boundary-value problem is stated a solution to which can be obtained using the methods of the theory of analytical functions [7]. For electrodes of an arbitrary shape (not necessarily linear) but symmetric about the interface, a solution to the field equation subject to the boundary conditions will have the form

$$E_2(z) = \frac{-f'(z)}{\pi(\varepsilon_1+\varepsilon_2)\sqrt{1-f^2(z)}} \times \int_0^L \frac{\sigma(x)\sqrt{1-f^2(x)}}{f(x)-f(z)}dx - E_0\frac{Lf'(z)}{\pi\sqrt{1-f^2(z)}}, \quad (2)$$

$$E_1(z) = \overline{E_2(\bar{z})}.$$

Here, function $f(z)$ conformally maps upper region Ω_2, where the field is sought, onto the upper half of complex plane $\zeta = \xi + i\eta$ in such a way that the point $z = 0$ turn into the point $\zeta = -1$ and the point $z = L$ turns into the point $\zeta = +1$ and $f'(z)$ is the derivative of mapping function $f(z)$.

The first term in (2) is due to the free charge at the interface, and the second one is the solution to the homogeneous problem ($\sigma(x) = 0$) and is completely defined by the geometry of the domain being considered.

For rectilinear interfaces between the regions (Fig. 1), we have

$$E_k(z) = \frac{C(-1)^{k+1}(1+f(z))^{\alpha_1}}{(\varepsilon_1+\varepsilon_2)L(1-f(z))^{\alpha_3}} \times \int_0^L \frac{\sqrt{1-f^2(x)}\sigma(x)d(x)}{f(x)-f(z)} - E_0C\frac{(1+f(z))^{\alpha_1}}{(1-f(z))^{\alpha_3}},$$

$$g(\zeta) = \frac{L}{\pi C}\int_{-1}^{\zeta}(1+\zeta)^{\beta_0-1}(1-\zeta)^{-\beta_L}d\zeta, \quad (3)$$

$$C = \frac{2^{\beta_0-\beta_L}}{\pi}B(\beta_0, 1-\beta_L),$$

$$\alpha_1 = \frac{1}{2}-\beta_0, \quad \alpha_3 = \frac{1}{2}-\beta_L,$$

where $g(\zeta)$ is the function mapping the upper half-plane onto the domain considered (i.e., the function inverse to $f(z)$); β_0 and β_L are the respective angles that the negative electrode (passing through point 0) and the positive electrode (passing through point L) make with a selected direction on the surface; and $B(\ldots)$ is the beta function.

The field at the interface is given the Sokhotsky–Plemel formulas [8, 9]. At the point (x_0, $y = 0$) at the interface (on the side of domain Ω_2),

$$E_2(x_0, 0) = \frac{-C}{(\varepsilon_1+\varepsilon_2)L}\frac{(1+f(x_0))^{\alpha_1}}{(1-f(x_0))^{\alpha_3}}$$

$$\times \int_0^L \frac{\sqrt{1-f^2(x)}\,\sigma(x)d(x)}{f(x)-f(x_0)} \tag{4}$$

$$-E_0 C \frac{(1+f(x_0))^{\alpha_1}}{(1-f(x_0))^{\alpha_3}} - \frac{i\sigma(x_0)}{(\varepsilon_1+\varepsilon_2)}.$$

The first two terms and the last one define, respectively, the x and y components of the field at the surface. Expression (4) shows that a free charge with density σ at the flat interface between two insulators (at the boundary between symmetric regions) generates a field with normal component $E_n = E_y = \sigma/(\varepsilon_1 + \varepsilon_2)$. The validity of this relationship was rigorously proven in [7] and is easy to explain qualitatively: according to the symmetry of the problem being considered, the interface coincides with the field line in the interelectrode space when free charge σ is absent. The normal field component at the interface can be generated only by an additionally introduced free charge distributed over the interface, which relationship (4) indicates.

By exactly calculating the singular integral appearing in the solution (the Cauchy integral along the boundary (–1, 1) in the canonical domain), one can obtain explicit expressions for the field strength. In the case of the electrodes normal to the interface ($\beta(z) = \beta_0 = \beta_L = \pi/2$), these expressions for a number of typical charge distributions $\sigma(x)$ (although they can simulate virtually any distribution) are presented in [7].

2. CONDITION (1) OF CHARGE STABILITY ON THE SURFACE IS FULFILLED, BUT THE CHARGE CONFIGURATION IS UNKNOWN

The problem of finding the strength of an in-plane electric field with the tilting angle of the field vector known was solved by reducing it to the Riemann boundary-value problem for a piecewise analytic function.

Considering the electric field strength in a domain covering two insulators bounded by tilted electrodes (Fig. 1) and assuming that condition (1) is fulfilled at the interface on the side of Ω_2, we get the following boundary conditions:

$$\left.\begin{aligned} \frac{E_{2y}(z)}{E_{2x}(z)} &= M = -\tan(\pi\beta_\sigma), \quad z \in \text{interface} \\ E_{2y}(z) &= -\cot(\pi\beta_0)E_{2x}(z), \ z \in \text{negative electrode} \\ E_{2y}(z) &= -\cot(\pi\beta_L)E_{2x}(z), \ z \in \text{positive electrode.} \end{aligned}\right\} \tag{5}$$

Conditions (5) correspond to the Hilbert homogeneous boundary-value problem [8, 9] for analytic function $E_2(z) = E_{2x} - iE_{2y}$ in Ω_2. In the theory of analytical functions, such a problem is also referred to as the boundary-value problem of directional derivative.

Having mapped the domain considered onto the upper half-plane $\zeta = \xi + i\eta$ (by means of mapping function $f(z)$ inverse to $g(\zeta)$), we arrive at the following Hilbert problem for function $F_2(\zeta) = F_{2\xi} + iF_{2\eta} = E_2(g(\zeta))$:

$$\left.\begin{aligned} F_{2\eta} &= -MF_{2\xi}, \ |\xi| < 1, \quad \eta = 0 \\ F_{2\eta} &= \cot(\pi\beta_0)F_{2\xi}, \ \xi < -1, \quad \eta = 0 \\ F_{2\eta} &= \cot(\pi\beta_L)F_{2\xi}, \ \xi > 1, \quad \eta = 0. \end{aligned}\right\} \tag{6}$$

Now we pass to the Riemann problem [8, 9] for function $F(\zeta)$ that corresponds to Hilbert problem (6) by analytically extending the function over the entire plane by virtue of the principle of symmetry (the bar above the function denotes the complex conjugate),

$$F(\zeta) = \begin{cases} F_2(\zeta) & \text{if} \quad \mathrm{Im}(\zeta) > 0 \\ \overline{F_2(\bar{\zeta})} & \text{if} \quad \mathrm{Im}(\zeta) < 0. \end{cases}$$

The Riemann problem is stated as

$$F^+(\xi) = G(\xi)F^-(\xi),$$

$$G(\xi) = \left\{\begin{array}{l} G_1 = \dfrac{\tan(\pi\beta_0) + i}{\tan(\pi\beta_0) - i}, \ \xi < -1, \ \eta = 0 \\ G_2 = \dfrac{1 - iM}{1 + iM} = \dfrac{1 + \tan(\pi\beta_\sigma)}{1 - \tan(\pi\beta_\sigma)}, \\ \qquad \xi < 1, \ \eta = 0 \\ G_3 = \dfrac{\tan(\pi\beta_L) + i}{\tan(\pi\beta_L) - i}, \quad \xi > 1, \quad \eta = 0. \end{array}\right\} \tag{7}$$

Here, $F^+(\xi)$ and $F^-(\xi)$ are the limiting values of $F(\zeta)$ as it approaches the interface from left and right, respectively (relative to the clockwise direction) and $G(\xi)$ is the coefficient in the boundary-value Riemann problem. Solution $F(\zeta)$ should be limited at infinity so that $F(\zeta \longrightarrow \infty) \neq 0$ if $\beta_0 = \beta_L$ and $F(\zeta \longrightarrow \infty) = 0$ if $\beta_0 \neq \beta_L$ (we assume that $\beta_0 \geq \beta_L$).

The problem thus stated is the homogeneous boundary-value Riemann problem with a discontinuous coefficient and rectilinear infinite boundary coincident with real axis 0ξ. We seek for a general solution by determining the canonical function (the index of the problem is zero) [9]. The coefficients in a general solution are found from the condition of boundedness of the solution at infinity and an integral relationship which a solution to the problem must meet,

$$\int_0^L E_x(x)dx = -E_0L$$

or, in the canonical domain,

$$\int_{-1}^1 F_\xi(\zeta)g'(\xi)d\xi = -E_0L,$$

Behavior of the field at characteristic points

Angular relationships	Behavior of field		
	$z = 0$, negative electrode	$z = L$, positive electrode	$z \longrightarrow \infty$
1) $\beta_0 > \beta_L \geq \beta_c$	$\longrightarrow \infty$	0	$\longrightarrow 0$
2) $\beta_0 > \beta_c > \beta_L$	$\longrightarrow \infty$	$\longrightarrow \infty$	$\longrightarrow 0$
3) $\beta_c \geq \beta_0 > \beta_L$	0	$\longrightarrow \infty$	$\longrightarrow 0$
4) $\beta = \beta_0 = \beta_L \geq \beta_c$	$\longrightarrow \infty$	0	Limited and $\neq 0$
5) $\beta = \beta_0 = \beta_L < \beta_c$	0	$\longrightarrow \infty$	Limited and $\neq 0$

where $g'(\zeta)$ is the derivative of mapping function $g(\zeta)$.

Eventually, the solution to Riemann problem (7) is given by

$$F(\zeta) = E_0 C(-1)^{\beta_\sigma}\frac{(1+\zeta)^{\alpha_{12}}}{(1-\zeta)^{\alpha_{32}}},$$

where

$$\alpha_{12} = \alpha_1 - \beta_\sigma, \quad \alpha_{32} = \alpha_3 - \beta_\sigma.$$

The expression for the field strength in the plane of variable z in domain Ω_2, on which side stability condition (1) is fulfilled, has the form

$$E_2(z) = E_0 C(-1)^{\beta_\sigma}\frac{(1+f(z))^{\left(\frac{1}{2}-\beta_\sigma\right)-\beta_0}}{(1-f(z))^{\left(\frac{1}{2}-\beta_\sigma\right)-\beta_L}},$$

where

$$(-1)^{\beta_\sigma} = \cos(\pi\beta_\sigma) + i\sin(\pi\beta_\sigma) = \frac{1-iM}{\sqrt{1+M^2}}.$$

This solution is also valid in the case of asymmetric regions Ω_1 and Ω_2.

If the problem is symmetric about the interface, the field strength in the adjacent region (Ω_1) can also be determined: $E_1(z) = \overline{E_2(\bar{z})}$. Note that, if the domains are symmetric about the interface, the field component normal to it (the y component) is generated by the surface charge alone, $E_y = \sigma/(\varepsilon_1 + \varepsilon_2)$. Then, comparing the expression obtained for the field strength with (4), we find the free charge density distribution over the surface that meets stability condition (1),

$$\sigma(x) = E_0 C(\varepsilon_1 + \varepsilon_2)\sin(\pi\beta_\sigma) \times \frac{(1+f(x))^{\left(\frac{1}{2}-\beta_\sigma\right)-\beta_0}}{(1-f(x))^{\left(\frac{1}{2}-\beta_\sigma\right)-\beta_L}}. \quad (8)$$

3. ANALYSIS OF THE EXPRESSIONS OBTAINED

From boundary conditions (5) to the boundary-value problem and the solutions obtained, it follows that the field strength near the surface, where stability condition (1) is provided, does not depend on the permittivity of both the vacuum (Ω_2) and the insulator (Ω_1). The field strength depends on applied voltage U (i.e., on average field strength $E_0 = U/L$), the emissivity of the surface, and the geometry of the system.

The electric field near the vacuum–insulator–electrode contact may either have integrable singularities at the points $z = 0$ and $z = L$ or be limited at these points, depending on the relationship between angle β_σ and tilting angles β_0 and β_L of the electrodes. The critical value of the tilting angle of the electrodes at which the behavior of the field at the points of contact changes is $\beta_c = 1/2 - \beta_\sigma$ (in fractions of π). This value corresponds to the angle between the field vector at the interface and the positive direction of the normal to the interface.

The table shows the field behavior according to the relationship between tilting angles β_0 and β_L of the electrodes and critical angle β_c (the solution should be limited at infinity).

The solutions obtained above make it possible to grasp an idea of the field variation (asymptotes) near the vacuum–insulator–electrode contact when the surface charge is stable. The associated expressions in the form of power-type dependences are the following.

Near the negative electrode ($z \longrightarrow 0$),

$$E_2(d) = \mathrm{const}(d)^{\frac{\beta_c}{\beta_0}-1} = \mathrm{const}(d)^{\left(-\frac{1-\beta_c}{2-\beta_e}-\frac{2\beta_\sigma}{2-\beta_e}\right)},$$

where d is a small neighborhood of the contact ($d = z/L$) and β_e is the expansion angle of the electrode ($\beta_e = 2\beta_0$). The field grows indefinitely, if $\beta_0 > \beta_c = 1/2 - \beta_\sigma$ (cases (1), (2), and (4) in the table), or approaches zero, if $\beta_0 \leq \beta_c$ (cases (3) and (5)).

Near the positive electrode ($z \longrightarrow L$),

$$E_2(d) = \mathrm{const}(d)^{\frac{1-\beta_c}{1-\beta_L}-1} = \mathrm{const}(d)^{\left(-\frac{1-\beta_c}{2-\beta_e}+\frac{2\beta_\sigma}{2-\beta_e}\right)},$$

where d is a small neighborhood of the contact ($d = 1 - z/L$) and β_e is the expansion angle of the electrode ($\beta_e = 2(\pi - \beta_L)$). The field grows indefinitely, if $\beta_L < \beta_c = 1/2 - \beta_\sigma$ (cases (1) and (4)), or approaches zero, if $\beta_L \geq \beta_c$ (cases (2), (3), and (5)).

The constants in the asymptotic expressions are defined by average field strength E_0, as well as by angles β_σ, β_0, and β_L. The tilting angle of the electrode at which the field strength is considered affects the value of the constants to a greater extent than the angle of the opposite electrode. If the opposite electrode represents a sharp edge (the expansion angle is zero), const = $2E_0(-1)^{\beta_\sigma}/\pi\beta_e$. In this relationship, the exponent is changed by $2\beta_\sigma/(2-\beta_e)$ compared with a similar power-type dependence of the field strength near the contact in a system without a charge at the interface.

The above relationships suggest that, to prevent the field near the triple contact between the vacuum, insulator, and negative electrode from being increased when the charge becomes steady, angle β_0 between the electrode and insulator surfaces must be smaller than its critical value, $\beta_0 < \beta_c = 1/2 - \beta_\sigma$ (Fig. 2).

The surface charge density, unlike the field strength, significantly depends on permittivity ε of the insulator. An increase in ε or M (the quantity characterizing the emissivity of the surface) raises the charge density. The shape of the charge distribution is independent of the voltage (average field strength E_0) (Fig. 3).

It should be noted that the relationship between the field strength and charge density on the surface, $E_n = \sigma/(\varepsilon_1 + \varepsilon_2)$, is valid only if the field is symmetric about the interface. Unjustified attempts to extend this relationship for other types of symmetry (e.g., for axisymmetric systems, intricate surface of the insulator, etc.) to calculate the field or find the charge distribution will yield incorrect results.

As follows from Fig. 3, the distributions of the potential and charge density over the surface differ considerably, which should be taken into account when the charge distribution is found by probe measurements.

The behavior of the charge density near the interface–electrode contact is similar to that of the field strength.

Total charge Q on the surface that provides the fulfillment of condition (1) is obtained by integration of (8) and varies in proportion to voltage U applied to the electrode gap, and its density is proportional to average field E_0 along the interface.

Irrespective of the tilting angles of the electrodes, we have

$$Q = U(\varepsilon_1+\varepsilon_2)\tan(\pi\beta_\sigma) = -UM(\varepsilon_1+\varepsilon_2),$$

$$Q_s = \frac{1}{L}\int_0^L \sigma(x)dx = -E_0 M(\varepsilon_1+\varepsilon_2).$$

4. PARTICULAR CASES

From the formulas obtained above, one can derive expressions for the field strength and charge density for typical tilting angles in terms of elementary functions.

(i) $\beta_0 = \beta_L = 1/2$. The field strength has an integrable singularity at the point $z = 0$ and is limited far from the interface, $E(z \longrightarrow \infty) = -E_0$. Thus, we have

$$E_2(z) = -E_0(-1)^{\beta_\sigma}\left(\frac{1+\cos(\pi z/L)}{1-\cos(\pi z/L)}\right)^{\beta_\sigma},$$

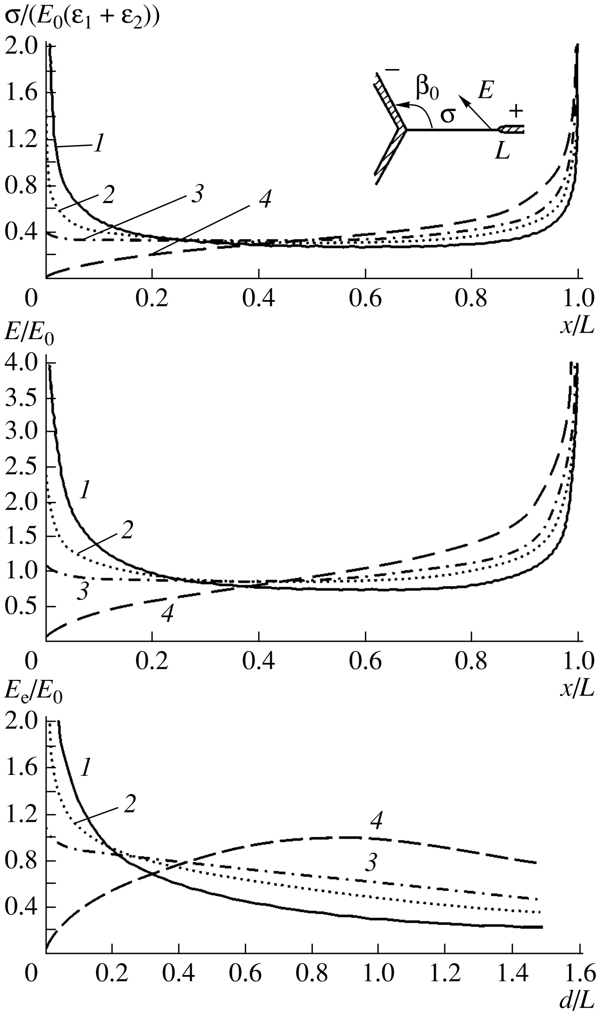


Fig. 2. Distributions of field strength E and charge density σ over the insulator–insulator interface, as well as the distribution of field strength E_e over the surface of the negative electrode at tilting angles β_0 of this electrode equal to (*1*) 3/4 π (135°), (*2*) 1/2 π (90°), (*3*) 0.4 π (72°), and (*4*) 1/4 π (45°). The positive electrode is an edge ($\beta_L = 0$). $M = -0.4$, $\beta_\sigma = 0.12\pi$ (22°), and $\beta_c = 0.38\pi$ (68°). d is the distance along the electrode from the point of contact with the interface.

$$\frac{\sigma(x)}{E_0(\varepsilon_1+\varepsilon_2)} = \sin(\pi\beta_\sigma)\left(\frac{1+\cos(\pi z/L)}{1-\cos(\pi z/L)}\right)^{\beta_\sigma}.$$

(ii) $\beta_L = 0$. At the point of contact $z = L$, the field strength has an integrable singularity of order $(1/2 - \beta_\sigma)$ and approaches zero far away from the interface,

$$E_2(z) = -E_0\frac{(-1)^{\beta_\sigma}}{\pi\beta_0}\frac{\left(\frac{z}{L}\right)^{\frac{\frac{1}{2}-\beta_\sigma}{\beta_0}-1}}{\left(1-\left(\frac{z}{L}\right)^{\frac{1}{\beta_0}}\right)^{\frac{1}{2}-\beta_\sigma}},$$

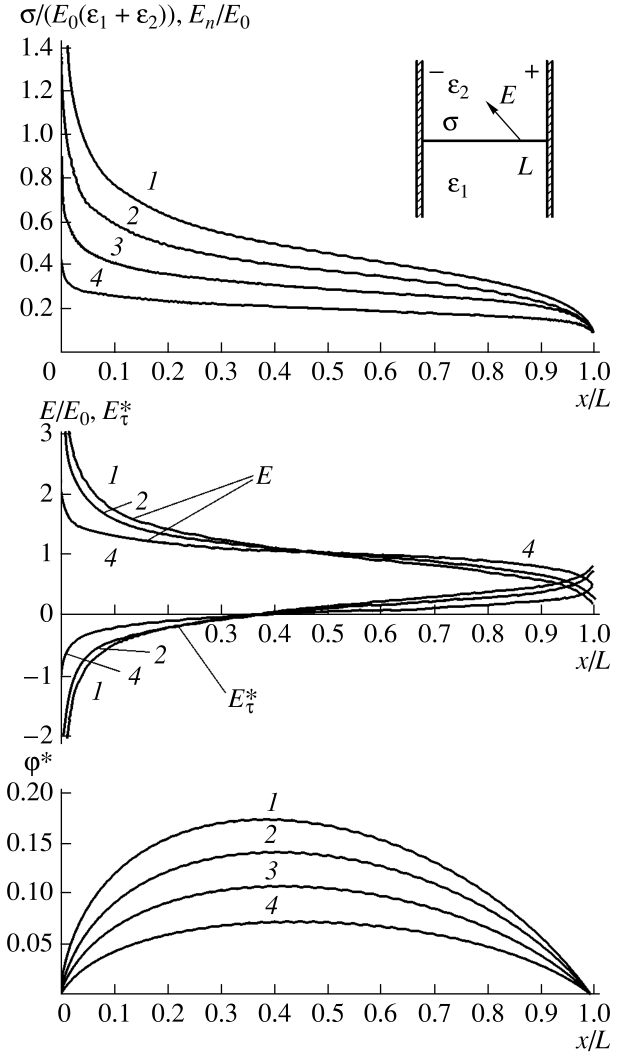


Fig. 3. Distributions of charge density σ, field strength E, field components E_n and E_τ, and field potential φ at the insulator–insulator interface for $\beta_L = 1/2\ \pi$; M = (*1*) –0.5, (*2*) –0.4, (*3*) –0.3, and (*4*) –0.2 (these values are typical of insulators); and β_σ = (*1*) 0.15 π (27°), (*2*) 0.12 π (22°), (*3*) 0.093 π (17°), and (*4*) 0.063 π (11°). $E^*_\tau = E_\tau/E_0 + 1$ and φ* are associated with the charge alone; i.e., external field E_0 is disregarded).

$$\frac{\sigma(x)}{E_0(\varepsilon_1+\varepsilon_2)} = \frac{\sin(\pi\beta_\sigma)}{\pi\beta_0}\frac{\left(\frac{x}{L}\right)^{\frac{\frac{1}{2}-\beta_\sigma}{\beta_0}-1}}{\left(1-\left(\frac{x}{L}\right)^{\frac{1}{\beta_0}}\right)^{\frac{1}{2}-\beta_\sigma}}.$$

(iii) $\beta_0 = 1$. The field strength has an integrable singularity of order $(1/2 + \beta_\sigma)$ at the point $z = 0$. Away from the interface, the field strength approaches zero,

$$E_2(z) = -E_0\frac{(-1)^{\beta_\sigma}}{\pi(1-\beta_L)}\frac{\left(1-\frac{z}{L}\right)^{\frac{\frac{1}{2}+\beta_\sigma}{1-\beta_L}-1}}{\left(1-\left(1-\frac{z}{L}\right)^{\frac{1}{1-\beta_L}}\right)^{\frac{1}{2}+\beta_\sigma}},$$

$$\frac{\sigma(x)}{E_0(\varepsilon_1+\varepsilon_2)} = \frac{\sin(\pi\beta_\sigma)}{\pi(1-\beta_L)}\frac{\left(1-\frac{x}{L}\right)^{\frac{\frac{1}{2}+\beta_\sigma}{1-\beta_L}-1}}{\left(1-\left(1-\frac{x}{L}\right)^{\frac{1}{1-\beta_L}}\right)^{\frac{1}{2}+\beta_\sigma}}.$$

(iv) $\beta_0 = 1$ and $\beta_L = 0$. The field strength has integrable singularities at both points of contact, $z = 0$ and $z = L$, and tends to zero away from the interface,

$$E_2(z) = -E_0\frac{(-1)^{\beta_\sigma}}{\pi\left(\frac{z}{L}\right)^{\frac{1}{2}+\beta_\sigma}\left(1-\frac{z}{L}\right)^{\frac{1}{2}-\beta_\sigma}},$$

$$\frac{\sigma(x)}{E_0(\varepsilon_1+\varepsilon_2)} = \frac{\sin(\pi\beta_\sigma)}{\pi\left(\frac{x}{L}\right)^{\frac{1}{2}+\beta_\sigma}\left(1-\frac{x}{L}\right)^{\frac{1}{2}-\beta_\sigma}}.$$

The associated distributions of the charge density and field strength are presented in Figs. 2 and 3.

CONCLUSIONS

The results obtained in this work can be summarized as follows. The relationships for the electric field are valid both for symmetric and asymmetric bordering domains (the form of region Ω_1 adjacent to the vacuum is insignificant, since the boundary-value Hilbert problem, rather than the conjugate problem, is solved). At the same time, the charge distributions are valid only for domains symmetric about the insulator–vacuum interface, because they were derived using the expression $E_n = \sigma/(\varepsilon_1 + \varepsilon_2)$, which is otherwise inapplicable.

With this in mind and also taking into consideration the small size of the contact area, we may assume that the expressions for the field near the vacuum–insulator–electrode contact and critical angle β_c apply to a variety of real situations where this contact is shielded by the electrodes (including the case of out-of-plane fields).

The approach used in this study (calculation of the field with an angle between the field vector and the interface known) has proved itself to be efficient in the case of rectilinear interfaces. It can also be used for more intricate boundaries, as well as for a variable tilting angle of the field vector. The mapping function and the canonical function in the general solution, however, take a more complicated form in this case.

REFERENCES

1. S. P. Bugaev and G. A. Mesyats, *Pulsed Discharge in Insulators*, Ed. by G. A. Mesyats (Nauka, Novosibirsk, 1985), pp. 4–25 [in Russian].
2. J. M. Wetzer, IEEE Trans. Dielectr. Electr. Insul. **4**, 349 (1997).
3. I. D. Chalmers, J. H. Lei, B. Yang, *et al.*, IEEE Trans. Dielectr. Electr. Insul. **2**, 225 (1995).
4. O. Yamamoto, T. Hara, I. Nakanishi, *et al.*, IEEE Trans. Electr. Insul. **28**, 707 (1993).
5. O. Yamamoto and M. Fukuda, in *Proceedings of the 13th International Symposium on High Voltage Engineering, Netherlands, 2003.*
6. H. Boersch, H. Hamisch, and W. Ehrlich, Z. Angew. Phys. **15**, 518 (1963).
7. V. K. Belyaev, Tekh. Élektrodinamika, No. 6, 40 (1987).
8. F. D. Gakhov, *Boundary-Value Problems* (Nauka, Moscow, 1977; Dover, New York, 1990).
9. N. I. Muskhelishvili, *Singular Integral Equations* (Nauka, Moscow, 1968).

Translated by M. Lebedev

Technical Physics, Vol. 50, No. 6, 2005, pp. 680–684. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 6, 2005, pp. 8–12.
Original Russian Text Copyright © 2005 by Ignatenko, Danilyuk, Borisenko.

THEORETICAL AND MATHEMATICAL PHYSICS

Tunnel Magnetoresistance Oscillations in a Ferromagnet–Insulator–Ferromagnet System

S. A. Ignatenko, A. L. Danilyuk, and V. E. Borisenko

Belarussian State University of Informatics and Radio Electronics,
ul. P. Brovki 6, Minsk, 220027 Belarus
e-mail: s2ign@tut.by

Received October 4, 2004

Abstract—A model of spin-dependent transport of electrons through a ferromagnet–insulator–ferromagnet structure is developed. It takes into account the image forces, tunnel barrier parameters, and effective masses of an electron tunneling in the barrier and in the ferromagnetic electrode in the free electron approximation. Calculations for an iron–aluminum oxide–iron structure show that, with an increase in the bias voltage, the tunnel magnetoresistance decreases monotonically and then breaks into damped oscillations caused by the interference of the electrons' wave functions in the conduction region of the potential barrier. The image forces increase the tunnel magnetoresistance by two or three times. *© 2005 Pleiades Publishing, Inc.*

INTRODUCTION

Structures consisting of alternating ferromagnetic and insulating layers are a subject of extensive research [1, 2], since they show considerable promise for integrated memory devices. The conductivity of such structures can be controlled by applying a magnetic field. The basic characteristic of such structures is the tunnel magnetoresistance (TMR).

To calculate the TMR, a number of models were developed [3–10], which differ in the way the potential barrier parameters and bias voltage are taken into account. However, these models fail in explaining neither the experimentally observed sharp decrease in and negative values of the TMR with increasing bias nor TMR oscillations. Some of the restrictions inherent in these models have been removed in [10], where simulation is based on the solution of the three-dimensional Schrödinger equation. Yet, an adequate physical interpretation of the effects observed, in particular, magnetoresistance oscillations, is today lacking.

In a real metal–insulator–metal structure, the shape of the potential barriers is other than rectangular due to image forces [11]. For tunneling electrons, both the height and width of the barrier vary. Since the TMR value considerably depends on the barrier parameters [1, 2], image forces should be taken into account for an adequate description of electron transport in a metal–insulator–metal structure. These forces were disregarded in the models suggested earlier [3–7, 10].

The purpose of this work is to theoretically investigate the effect of image forces on the spin-dependent transport of electrons in a ferromagnet–insulator–ferromagnet structure with allowance for bias potential and also to give a physical interpretation for the nonlinear variation of the TMR with increasing external bias. Numerical simulation is carried out for a structure where iron and aluminum oxide serve as a ferromagnet and insulator, respectively.

MODEL

The structure being simulated consists of two identical ferromagnets separated by a tunnel-thin nonmagnetic insulating layer (Fig. 1). The application of a potential generates a current consisting of two components formed by spin-up electrons and spin-down electrons, respectively. In the absence of scattering, these components do not mix up. Electrons with a certain spin that are emitted from one ferromagnet tunnel toward the other ferromagnet, where they can occupy

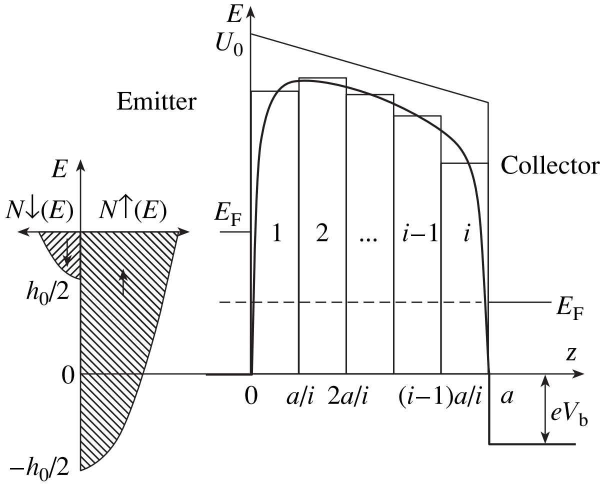


Fig. 1. Potential profile of a ferromagnet–insulator–ferromagnet structure subjected to external bias V_b with regard to image forces.

1063-7842/05/5006-0680$26.00 © 2005 Pleiades Publishing, Inc.

vacancies with the same spin polarization. If the ferromagnets are magnetized in parallel, resistance R_p of the structure is minimal; for antiparallel magnetization, resistance R_ap is maximal. This is because the degree of spin polarization of electrons depends on the magnetization vector direction [3]. The TMR value depends on the difference between resistances R_p and R_ap,

$$\mathrm{TMR} = \frac{R_\mathrm{ap} - R_\mathrm{p}}{R_\mathrm{ap}}. \tag{1}$$

It is schematically shown in Fig. 1 that image forces [11] smooth out the potential barrier in the insulator at the insulator–ferromagnet interfaces. Such a shape of the potential barrier is involved in the model suggested. Moreover, unlike the earlier models [3–7, 10], we take due account of the transverse motion of electrons and the difference between the effective masses of electrons with different spin polarization in the barrier and in the ferromagnet.

In the free electron approximation, the band spectrum of spin-polarized charge carriers in the ferromagnet is described by a parabola for each spin component. The difference between the bottom of the band of spin-up polarized electrons and the bottom of the band of spin-down polarized electrons reflects the degree of spin polarization of electrons in the ferromagnet and is quantitatively described by the value of molecular field h_0 [4–7, 10].

If the layers of the structure are orthogonal to the z axis and its dimensions in the xy plane are sufficiently large for the quantization of the electron motion along the interfaces to be neglected, the wave function of electrons has the form

$$\Phi_\sigma = \psi_\sigma(z) e^{ik_\parallel \rho}, \tag{2}$$

where $\psi_\sigma(z)$ is the longitudinal wave function (in the z-axis direction), $k_\parallel$ is the transverse wavevector in the xy plane, $\rho = (x, y)$ is the coordinate in the barrier plane, and σ is the electron spin (up ↑ or down ↓).

As an electron tunnels through the potential barrier, the transverse component of the energy remains unchanged and so $k_\parallel$ may be taken to be fixed.

The one-electron stationary Schrödinger equation for longitudinal wave functions $\psi_\sigma(z)$ in an ith domain (Fig. 1) has the form

$$\left[-\frac{\hbar^2}{2m^*_{i\sigma}}\frac{\partial^2}{\partial z^2} + \frac{\hbar^2 k_\parallel^2}{2m^*_{i\sigma}} + U_0 - h_{0i}\sigma_i - eV_i - e\varphi_i\right]\psi_\sigma(z) = E\psi_\sigma(z), \tag{3}$$

where $\hbar$ is the Planck constant e is the electron charge, $m^*_{i\sigma}$ is the effective mass of an electron in the ith domain, U_0 is the potential barrier height, V_i is the applied field in the ith domain, $h_{0i}\sigma_i$ is the internal exchange energy [5, 7], h_{0i} is the molecular field ($h_0 = 0$ inside the insulator), σ_i is the Pauli matrix, φ_i is the image force potential in the ith domain, and E is the electron total energy.

The electron effective mass in each of the layers is isotropic and different in the ferromagnet and insulator. Quantities h_{0i} and σ_i are taken along the magnetization axis of the ith domain.

We assume that wave function $\psi_\sigma(z)$ in the ith domain is a linear combination of two waves corresponding to the free motion of the electron in opposite directions; namely, in the emitter region, there are the incident and reflected waves, and, in the collector, only the transmitted wave. The barrier region is divided into n layers, each with its own wave function written in accordance with Eq. (3). The wavevectors in the emitter, collector, and barrier are given as follows:

$$\begin{gathered}
k_{\sigma\mathrm{em}} = \frac{\sqrt{2m^*_{\mathrm{f}\sigma}\left(E_z \pm \frac{h_0}{2}\right)}}{\hbar}, \\
k_{\sigma\mathrm{col}} = \frac{\sqrt{2m^*_{\mathrm{f}\sigma}\left(E_z \pm \frac{h_0}{2} + eV_\mathrm{b}\right)}}{\hbar}, \\
k_i = \frac{\sqrt{2m^*_\mathrm{in}\left(E_z - U_0 + \frac{i}{n}eV_\mathrm{b} + \frac{1.15e^2\ln 2}{8\pi\varepsilon\varepsilon_0 a}\frac{n^2}{i(n-i)} - \left(1 - \frac{m^*_\mathrm{in}}{m^*_{\mathrm{f}\sigma}}\right)E_\parallel\right)}}{\hbar},
\end{gathered} \tag{4}$$

where $i = 1, 2, \ldots, n$; $m^*_{\mathrm{f}\sigma}$ and m^*_in are the effective masses of the electron in the ferromagnet and insulator, respectively; E_z and $E_\parallel$ are the longitudinal and transverse components of the energy, respectively; V_b is the bias voltage applied to the structure; ε_0 is the permittivity of a vacuum; ε is the relative permittivity of the barrier material; and a is the barrier thickness. The fourth term in the expression for the wavevector in the barrier

Parameters of the Fe/Al_2O_3/Fe structure

Parameter	Value
Fermi level E_{F}	1.3 eV [13]
Molecular field h_0	1.9 eV [10, 13]
Effective mass of a spin-up electron in Fe, $m^*_{\mathrm{f}\uparrow}/m_{\mathrm{e}}^{-}$	1.27 [13]
Effective mass of a spin-down electron in Fe, $m^*_{\mathrm{f}\downarrow}/m_{\mathrm{e}}^{-}$	1.36 [13]
Effective mass of an electron in the barrier (Al_2O_3), $m^*_{\mathrm{d}}/m_{\mathrm{e}}^{-}$	0.4 [12]
Potential barrier width a	1–5 nm
Relative permittivity ε of nanodimensional aluminum oxid	5–10
Potential barrier height in the absence of bias U_0	3–5 eV

describes the effect of image forces [11], and the term $(1 - m^*_{\mathrm{in}}/m^*_{\mathrm{f}\sigma})E_{\|}$ shows the contribution of the electron energy transverse component at the ferromagnet–insulator interface.

At the boundaries of all the domains, wave functions $\psi_\sigma(z)$ and their first derivatives $\psi'_\sigma(z)/m^*$ must meet the continuity conditions.

Thus, we have a set of $2(n+1)$ equations, which will be solved by the transfer matrix method [12]. The transfer matrix relates the wave amplitudes in the emitter and collector. In contrast to [10], where the 4×4 matrices are applied, we will use 2×2 matrices separately for each spin orientation. Such a choice, which is due to the incoherent propagation of spins, allows us to significantly simplify calculation.

The wave transmission coefficient for the structure studied is equal to the ratio between the flux densities of the transmitted and incident waves.

To calculate the tunnel current, we sum up the tunneling probabilities over the electron distribution in the emitter by the Tsu–Isaki formula [12],

$$I = \frac{2\pi e}{\hbar^3} m^*_{\mathrm{f}\sigma} \iint T(E_z, E_{\|})(f_1 - f_3)\,dE_z\,dE_{\|}, \tag{5}$$

where I is the tunnel current and f_1 and f_3 are the Fermi–Dirac functions in the emitter and collector regions, respectively.

Summation is over two independent spin components. For the zero temperature, expression (5) reduces to the form

$$I = \frac{2\pi e}{\hbar^3} m^*_{\mathrm{f}\sigma} \left(\int\limits_{E_{\mathrm{F}} - eV_{\mathrm{b}}}^{E_{\mathrm{F}}} \int\limits_{0}^{E_{\mathrm{F}} - E_z} T(E_z, E_{\|})\,dE_{\|}\,dE_z \right. \\ \left. + \int\limits_{0}^{E_{\mathrm{F}} - eV_{\mathrm{b}}} \int\limits_{E_{\mathrm{F}} - E_z - eV_{\mathrm{b}}}^{E_{\mathrm{F}} - E_z} T(E_z, E_{\|})\,dE_{\|}\,dE_z \right), \tag{6}$$

where E_{F} is the Fermi energy of the emitter.

In terms of tunnel currents I_{p} and I_{ap}, which correspond to the parallel and antiparallel magnetizations of the ferromagnets, expression (1) transforms to

$$\mathrm{TMR} = \frac{I_{\mathrm{p}} - I_{\mathrm{ap}}}{I_{\mathrm{p}}}. \tag{7}$$

Numerical calculations by the model proposed are made for a Fe/Al_2O_3/Fe structure, since the parabolic band model works well as applied to free electrons in iron and aluminum oxide is commonly used as a tunnel-thin insulator.

COMPUTATIONAL RESULTS AND DISCUSSION

The parameters of the Fe/Al_2O_3/Fe structure simulated are presented in the table. It is known that the tunnel current for iron electrodes is determined largely by free *d* electrons [14]. This is consistent with the earlier hypothesis [15] that the wavevectors of spin-up and spin-down electrons in iron at the Fermi level are equal to 0.109 and 0.042 nm^{-1}, respectively. In other bands, the Bloch states do not have appropriate symmetry for effective coupling through the ferromagnet–insulator interface and quickly die down in the potential barrier [13, 14].

Fermi energy E_{F} and molecular field h_0 were determined from the positions of the band bottoms for spin-up and spin-down electrons by first-principles calculations [13]. The width of the potential barrier (Al_2O_3) was varied from 1 to 5 nm, because, if the width is less than 1 nm, it is difficult to grow a continuous film and, if the width exceeds 5 nm, the tunnel current detected is negligible [2, 12].

Figure 2 shows the dependence of the tunnel current on the external bias for a barrier width of 1 nm and barrier height $U_0 = 3.6$ eV. At low voltages, the parallel-magnetization current is higher (curve *1*), because the barrier penetrability for spin-up electrons is greater than for spin-down electrons and the spin polarization of the tunnel current is positive (Fig. 1). For biases of 0.6–1.7 V, the parallel-magnetization current becomes higher the TMR (see (7)) changes. The change of the collector magnetization means polarization reversal of the collector electrons. For tunneling spin-up electrons, this implies a decrease in the potential barrier penetrability; for spin-down electrons, the situation is reverse. Since the positive polarization dominates in the emitter electron current ($k_{\uparrow \mathrm{em}} > k_{\downarrow \mathrm{em}}$), the magnetization reversal results in a decrease of the total current.

Figure 3 shows the TMR versus bias dependence for various widths of the potential barrier without regard to image forces. For low biases, the TMR is positive and nearly maximal. As the bias increases, the TMR drops, approaching monotonically a negative maximum, and then exhibits damped oscillations. The wider the tunnel barrier, the lower the small-bias TMR and the oscillation frequency and amplitude.

The dependence of the TMR on the external bias can be divided into two, monotone and oscillating, parts. It seems that different mechanisms are responsible for the tunnel current in them. For the monotone part, the decrease in the TMR and the transition to the negative region are associated with a sharper growth of the penetration coefficient for antiparallel magnetization compared with parallel magnetization. Here, the spin polarization of electrons in the emitter has a less significant effect on the TMR.

Consider two components of the coefficient of electron penetration through the potential barrier. One depends on the relationship between the wavevectors in the ferromagnets; the other, on the relationship between the intensities of the transmitted and incident waves (the element of the transfer matrix) [12]. With increasing the bias, the first component grows for both parallel and antiparallel magnetization. At a certain bias value, these components come closer to each other. This is because the electron tunneling depends on the applied electric field more strongly than on the spin polarization of the electrons in the emitter and collector. Thus, as the bias grows, the TMR decreases monotonically, because the wavevectors of the electrons come closer to each other.

The second component, which defines the ratio of the transmitted and incident wave intensities, is always greater for antiparallel magnetization. For either magnetization, these intensities grow in proportion to the bias. With regard to the first component, this causes a monotone growth of the penetration coefficients, the growth being more significant for antiparallel magnetization. Hence, the monotone decline in the TMR with the transition to the negative region (Fig. 3).

When the voltage corresponds to the maximal negative value of the TMR, the wave functions of the electrons interfere consistently (in other words, become coherent). Then, as the bias grows, the increase in the transmitted intensity for the antiparallel configuration slows down, as a result of which the penetration coefficients for the two spin configurations come closer and the TMR rises. In this case, the interference of the wave functions in the range $eV_b > U_0 - E_F$ begins to exert a dominating influence on the electron tunneling.

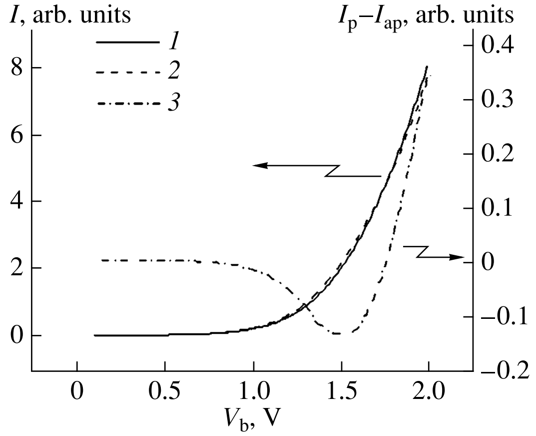


Fig. 2. *I–V* characteristic of the Fe/Al_2O_3/Fe structure for (*1*) parallel and (*2*) antiparallel magnetization. (*3*) Current-difference curve.

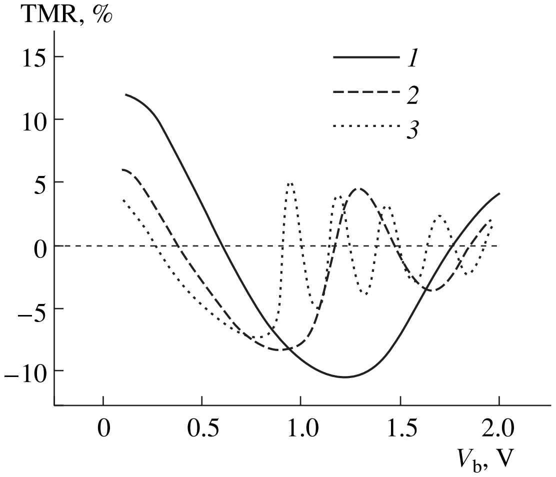


Fig. 3. TMR of the Fe/Al_2O_3/Fe structure vs. the bias for Al_2O_3 thicknesses of (*1*) 1, (*2*) 2, and (*3*) 5 nm.

Under these conditions, the upper level electrons pass through both the band gap and conduction band of the insulator. In the conduction band of the insulator, they have a real wavevector. Some of the electrons reflect from the insulator–collector interface and interfere with incident electron waves. The narrower the barrier, the higher the bias needed to produce a sufficiently long conduction region in the insulator through which the electrons pass. Therefore, the negative value of TMR is maximal for the 1-nm-wide barrier.

In the region of TMR oscillations, the contribution of the wave function interference to the conductivity prevails for both parallel and antiparallel magnetizations of the ferromagnets. Interference mismatch causes the transmission coefficients to oscillate. However, the peaks and dips of the transmission coefficients arise under different conditions for the different magnetizations, which results in TMR oscillations. Thus, the basic reason for TMR oscillations is a dominating con-

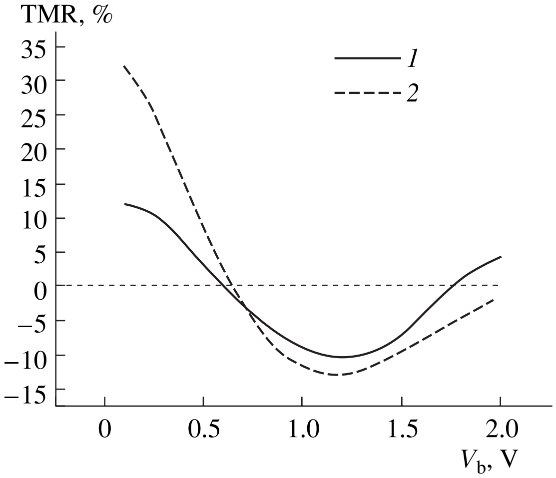


Fig. 4. TMR of the $Fe/Al_2O_3/Fe$ structure vs. the bias for a = 1 nm (*1*) without and (*2*) with taking into account image forces.

tribution of the interference between the incident and reflected waves to the conductivity of the structure and a mismatch in the interference conditions at the opposite magnetizations of the emitter and collector.

As the bias grows, the TMR oscillation amplitude decreases and the oscillation period increases. The decrease in the amplitude can be explained by the fact that, as the bias grows, the difference in the transmission coefficients for the two different magnetizations of the ferromagnets becomes smaller. This is observed at biases in the range 1–2 V, where the transmission coefficient of the electrons occupying the upper energy levels is close to unity. The increase in the TMR oscillation period is associated with an extension of the barrier at energies below the Fermi level in the emitter. Note that the TMR oscillates only if the voltage applied to the structure exceeds the difference between the potential barrier height and the Fermi level; i.e., $V_b > U_0 - E_F$.

Figure 4 shows the effect of the image force potential on the bias dependence of the TMR for $\varepsilon = 10$. In this case, the TMR value depends on two factors. The first one is the TMR growth caused by the decrease in the effective barrier thickness. The other is a decrease in the TMR due to the change in the barrier shape [16]. The first factor prevails. Therefore, taking into account image forces is generally expected to raise the TMR. For instance, for $Fe/Al_2O_3/Fe$ structure with a barrier thickness of 1 nm, the TMR value increases by a factor of 2.5 when $V_b \longrightarrow 0$ (Fig. 4). Note that, for the electrons tunneling from the Fermi level at a = 1 nm, the effective barrier width decreases by approximately 20%. A change in the potential barrier shape modifies the conditions under which the interference arises. This shows up in an insignificant shift of the TMR's negative maximum.

CONCLUSIONS

Our model of spin-dependent transport of electrons through a ferromagnet–insulator–ferromagnet structure and, in particular, the numerical simulation of the tunnel currents and TMR in a $Fe/Al_2O_3/Fe$ structure demonstrate that the tunnel barrier parameters, effective mass of electrons tunneling in the barrier and in the ferromagnet, and image forces all have a considerable influence on the behavior of the structure.

It is shown that, as the bias grows, the magnetoresistance decreases monotonically, since the wavevectors of electrons with different spin polarization in the collector come close to each other. It reaches a maximal negative value when the wave functions start interfering in a consistent manner. Magnetoresistance oscillations are caused by the periodically varying (as the bias grows) conditions for electron wave interference in the collector for the oppositely magnetized ferromagnets. Taking into account image forces leads to a considerable rise in the magnetoresistance at low biases, since the potential barrier shrinks effectively and changes shape in this case.

ACKNOWLEDGMENTS

This work was financially supported by the Ministry of Education of Belarus (project no. 04-3065) and the Belarusian Foundation for Basic Research (project no. F04M-039).

REFERENCES

1. J. S. Moodera, T. H. Kim, C. Tanaka, *et al.*, Philos. Mag. B **80**, 195 (2000).
2. G. A. Prinz, Science **282**, 1660 (1998).
3. M. Julliere, Phys. Lett. A **54A**, 225 (1975).
4. J. M. MacLaren, X. G. Zhang, and W. H. Butler, Phys. Rev. B **56**, 11827 (1997).
5. J. C. Slonchewsi, Phys. Rev. B **39**, 6995 (1989).
6. J. Zhang and R. M. White, J. Appl. Phys. **83**, 6512 (1998).
7. X. Zhang, B.-Z. Li, G. Sun, *et al.*, Phys. Rev. B **56**, 5484 (1997).
8. T. Miyazaki, N. Tezuka, S. Kumagai, *et al.*, J. Phys. D **31**, 630 (1998); M. Sharma, S. X. Wang, and J. H. Nickel, Phys. Rev. Lett. **82**, 616 (1999).
9. P. LeClair, H. J. M. Swagten, J. T. Kohlhepp, *et al.*, Appl. Phys. Lett. **76**, 3783 (2000).
10. S. S. Liu and G. Y. Guo, J. Magn. Magn. Mater. **209**, 135 (2000).
11. J. G. Simmons, J. Appl. Phys. **34**, 1793 (1963).
12. D. Ferry and S. Goodnick, *Transport in Nanostructures* (Cambridge Univ. Press., Cambridge, 1997).
13. A. H. Davis and J. M. MacLaren, J. Appl. Phys. **87**, 5224 (2000).
14. W. H. Butler, X.-G. Zhang, T. C. Schulthess, *et al.*, J. Appl. Phys. **85**, 5834 (1999).
15. M. B. Stearns, J. Magn. Magn. Mater. **5**, 167 (1977).
16. F. Montaigne, M. Hehn, and A. Schuhl, Phys. Rev. B **64**, 144402 (2001).

Translated by A. Pankrat'ev

Technical Physics, Vol. 50, No. 6, 2005, pp. 685–691. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 6, 2005, pp. 13–18.
Original Russian Text Copyright © 2005 by Makaryan, Molevich.

GASES AND LIQUIDS

Weak Shock Waves in Negative-Dispersion Nonequilibrium Media

V. G. Makaryan and N. E. Molevich
Korolev State Aerospace University, Samara, 443086 Russia
e-mail: molevich@mb.ssau.ru

Received July 30, 2004

Abstract—The "frozen" and equilibrium shock adiabats for a gas with sustained steady-state nonequilibrium are constructed accurate to the second order of smallness. With these adiabats, the pattern of and stability conditions for weak shock waves in negative-dispersion nonequilibrium media, where the speed of low-frequency (equilibrium) sound exceeds that of high-frequency (frozen) sound, are considered. On the basis of a model nonlinear equation describing the evolution of gasdynamic perturbations in low-dispersion media, the nonstationary evolution of shock waves at a negative low-frequency nonlinearity coefficient is analyzed. This situation corresponds to a low-frequency adiabat convex upwards. It is shown that a step autowave may arise in this case whose amplitude is entirely specified by the nonequilibrium parameters of the medium and correlates with the point where the low-frequency and high-frequency adiabats intersect. In addition, it is found that the initial unsteady shock wave may split into two steady ones: a step autowave followed by a steady smooth-front expansion shock. *© 2005 Pleiades Publishing, Inc.*

INTRODUCTION

The acoustics of thermodynamically nonequilibrium media (such as a vibrationally excited gas, nonisothermal plasma, chemically active mixtures, media with nonequilibrium phase transitions, gaseous suspensions with different temperatures of the components, etc.) is appreciably different from the acoustics of equilibrium media [1]. In these media, the inversion of the second (bulk) viscosity coefficient ($\xi < 0$) and dispersion (when the low-frequency sound propagates faster than the high-frequency one, $c_0 > c_\infty$) is possible [1–13]. Negative-viscosity media are acoustically active. In addition, the low-frequency coefficient of hydrodynamic nonlinearity is no longer defined as $\Psi_0 \sim (\gamma_0 + 1)/2$ in this case; instead, it becomes a complicated function of the stationary degree of nonequilibrium of the medium and may even be negative [5, 14]. These new acoustic properties should be taken into account in analysis of various gasdynamic phenomena taking place in these media. It is obvious that they will influence primarily the pattern of gasdynamic perturbations with a small but finite amplitude. In [15, 16], the shock pattern at $\Psi_0 > 0$ was investigated by solving a model nonlinear equation derived under the condition that a nonequilibrium medium has a weak dispersion. In Sect. 1 of this paper, the steady-state pattern of weak shock waves in nonequilibrium media at $\Psi_0 > 0$ and $\Psi_0 < 0$ is studied by analyzing the shock adiabats in simple terms.

In media with anomalous thermodynamic properties where $\Psi \sim (d^2\tilde{P}/d\tilde{V}^2)_S < 0$, compression shock waves, unlike expansion shock waves, cannot exist [17, 18]. If nonlinearity coefficients Ψ have opposite signs in a pressure range considered, complicated regimes may arise where diffuse waves adjacent to discontinuities or a combination of several shock waves are observed [17]. In Sect. 2, it is shown that such regimes may also occur in thermodynamically nonequilibrium media where the low-frequency and high-frequency nonlinearity coefficients have opposite signs, $\Psi_0 < 0$ and $\Psi_\infty > 0$.

1. STEADY-STATE STRUCTURE OF WEAK SHOCK WAVES IN NEGATIVE-DISPERSION NONEQUILIBRIUM MEDIA

Slow relaxation processes, such as chemical reactions, slow energy transfer between various degrees of freedom, and so on, taking place in a gas cause a considerable expansion of a shock wave and greatly complicates its pattern [17–21]. In relaxation gasdynamics, two shock adiabats drawn through a given initial point are considered. One corresponds to total equilibrium of the final states of the gas and, therefore, is called the equilibrium adiabat. The other, referred to as "frozen," assumes that the relaxation processes do not proceed at all. For small-amplitude waves, these adiabats can be represented accurate to the second order of smallness in

1063-7842/05/5006-0685$26.00 © 2005 Pleiades Publishing, Inc.

change of the specific volume, $V_2 - V_1$, in the form

$$P_2 - P_1 = \rho_1 c_0^2\left[-\frac{V_2 - V_1}{V_1} + \Psi_0\frac{(V_2 - V_1)^2}{V_1^2}\right],$$
$$P_2 - P_1 = \rho_1 c_\infty^2\left[-\frac{V_2 - V_1}{V_1} + \Psi_\infty\frac{(V_2 - V_1)^2}{V_1^2}\right]. \qquad (1)$$

Here, c_0, Ψ_0 and c_∞, Ψ_∞ are the low-frequency (equilibrium) and high-frequency (frozen) speeds of sound and nonlinearity coefficients, respectively; P_1 and P_2 are the pressures in the initial and final states; and $\rho_1 = 1/V_1$ is the gas density in state 1. In states close to equilibrium, the relaxation equation has the form

$$\frac{dE}{dt} = -\frac{E - E_e}{\tau}, \qquad (2)$$

where E is a physical quantity characterizing the state of the body and E_e is its value in equilibrium.

In this case, for polytropic gas media, we have $c_0 = \sqrt{\gamma_0 T/m}$, $c_\infty = \sqrt{\gamma_\infty T/m}$, $\Psi_0 = (\gamma_0 + 1)/2$, and $\Psi_\infty = (\gamma_\infty + 1)/2$ with $\gamma_0 < \gamma_\infty$. Thus, relaxation process (2) always leads to a positive dispersion of sound. Besides, up to the second-order of smallness, shock adiabats (1) coincide with the corresponding Poisson adiabats. Figure 1 shows the qualitative behavior of the equilibrium (curve *e*) and "frozen" (curve *f*) shock adiabats drawn through given initial point *1*. Here, $\tilde{P} = (P_2 - P_1)/\rho_1 c_\infty^2$ and $\tilde{V} = (V_2 - V_1)/V_1$ are the dimensionless increments of the pressure and specific volume in the shock wave. The shock wave velocity is defined by the slope of the chord connecting the initial (*1*) and final (*2*) states: $\tilde{D}^2 = D^2/c_\infty^2 = -\tilde{P}/\tilde{V}$.

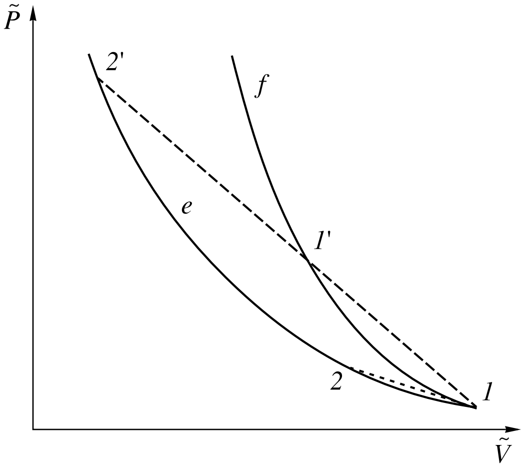


Fig. 1. Frozen and equilibrium adiabats in a relaxing gas.

At a low shock wave velocity such that $c_0/c_\infty < \tilde{D} < 1$, chord *1–2* is located between the tangents to curves *e* and *f* at point *1* (Fig. 1). As shown in [17–20], in this case, the shock wave merely expands (the higher the second viscosity coefficient $\xi \sim \rho(c_\infty^2 - c_0^2)\tau$, the more the expansion of the shock wave).

At $\tilde{D} > 1$, the chord connects points *1*, *1*', and *2*'. Point *1*' corresponds to the state where the relaxation to state *1* is "frozen." Compression of the gas from state *1* to state *1*' is stepwise and from state *1*' to final state *2*' gradual [17–21].

Now let us consider the propagation of a shock wave through a gaseous medium in the steady-state nonequilibrium state, $E > E_e$. The simplest model of such a medium assumes an exponential relaxation law similar to (2),

$$\frac{dE}{dt} = -\frac{E - E_e}{\tau} + Q. \qquad (3)$$

For a vibrationally excited gas, E is the energy of the vibrational degrees of freedom of the molecules, E_e is its equilibrium value, τ is the vibrational relaxation time, and Q is the energy source sustaining thermal nonequilibrium in the system (in particular, optical or electric pumping). Then, the low-frequency adiabat exponent and the nonlinearity coefficient depend on the steady-state degree of nonequilibrium $S = Q\tau/T$,

$$\gamma_0 = \frac{C_{P0}}{C_{V0}},$$

$$\Psi_0 = \left[\frac{S\tau'(1+S)}{C_{P0}C_{V0}} + \frac{1 + 2C_{V0}}{2C_{V0}} - \frac{S(1+S)^2}{2C_{P0}C_{V0}^2}\tau''\right],$$

where $C_{V0} = C_{V\infty} + C_{\rm vib} + S\tau'$, $C_{P0} = C_{P\infty} + C_{\rm vib} + S(\tau' + 1)$, $C_{V\infty}$ and $C_{P\infty}$ are the low-frequency and high-frequency heat capacities at constant volume and pressure in the vibrationally excited gas [5, 14], $C_{\rm vib} = dE_e/dT$ is the equilibrium vibrational heat capacity, $\tau' = \partial\ln\tau/\partial\ln T$, and $\tau'' = T^2\tau^{-2}\partial^2\tau/\partial T^2$.

The nonlinearity coefficient is defined as $\Psi_0 = (1 + \gamma_0)/2$ only when the steady-state nonequilibrium of the medium is absent; i.e., when $S = 0$. Therefore, at $S \neq 0$, the equilibrium shock adiabat differs from the Poisson adiabat even in the second-order of smallness.

In media with steady-state thermodynamic nonequilibrium, the dispersion of speed of sound may be negative. If $\tau' < 0$ in the frame of the model considered, low-frequency speed of sound c_0 exceeds high-frequency one c_∞ when the degree of nonequilibrium satisfies the inequality $C_{\rm vib}/(C_{V\infty} - \tau') < S < -(C_{V\infty} + C_{\rm vib})/\tau'$. In [7–13], the conditions for negative dispersion of speed of sound were found for other models of relaxation of internal degrees of freedom of molecules, as well as for chemically active mixtures, multiphase media, and a weakly ionized gas. It was found that nonlinearity coefficient Ψ_0 as a function of the degree of nonequilibrium S and $\tau(T)$ may be both positive and negative [14].

The shock wave pattern in a negative-dispersion medium can be studied by the approach described above. The equilibrium and frozen shock adiabats (see (1)) at $c_0 > c_\infty$ and $0 < \Psi_0\gamma_0/\gamma_\infty < \Psi_\infty$ are presented in Fig. 2 (curves *e* and *f*, respectively). At such Ψ_0, the adiabats intersect. The slope of the chord drawn through the point where the adiabats meet (Fig. 2, straight line *1*) defines the critical shock wave velocity,

$$\tilde{D}_{\mathrm{cr}} = \sqrt{1 + \Psi_\infty \tilde{\rho}_{\mathrm{cr}}}. \tag{4}$$

In a weak shock wave propagating with such a velocity, the density dimensionless increment is given by

$$\tilde{\rho}_{\mathrm{cr}} \approx -\tilde{V}_{\mathrm{cr}} = \frac{m}{\Psi_\infty - \gamma_0\Psi_0/\gamma_\infty}, \tag{5}$$

where $m = (c_0^2 - c_\infty^2)/c_\infty^2$.

At shock wave velocities $\tilde{D} \geq \tilde{D}_{\mathrm{cr}}$ (the slope of the corresponding chord is larger than that of straight line *1*), the shock wave structure is typical of relaxing media with positive dispersion, because the point of intersection between the chord and the frozen adiabat is to the right of the point of intersection with the equilibrium adiabat. Therefore, the medium is first rapidly compressed to the value specified by the point of intersection between the corresponding chord and the frozen adiabat and then is gradually compressed to the final state specified by the intersection of the chord with the equilibrium adiabat. The related shock wave pattern is shown in Fig. 3 (curve *1*). Note that this pattern cannot arise if $\Psi_0\gamma_0/\gamma_\infty > \Psi_\infty$, since the shock adiabats do not intersect.

When $\tilde{D}_{\mathrm{cr1}} < \tilde{D} < \tilde{D}_{\mathrm{cr}}$ (the slope of the chord is larger than that of straight line *2* but smaller than the slope of straight line *1* in Fig. 2), the shock wave structure is radically different (Fig. 3, curve *2*). In this range, the frozen adiabat is to the left of the equilibrium one. Here, fast compression to the value determined by the intersection of the corresponding chord with the frozen adiabat is followed by gradual expansion to the final state specified by the intersection between the chord and equilibrium adiabat.

The value of $\tilde{D}_{\mathrm{cr1}} = \sqrt{1 + m + \gamma_0\Psi_0\tilde{\rho}_{\mathrm{cr1}}/\gamma_\infty}$ is found from the condition that the slope of tangent *3* to adiabat *f* at point $\tilde{V}_{\mathrm{cr1}}$ equals that of chord *2* drawn through the same point to adiabat *e* (Fig. 2); i.e., $\tilde{D}_{\mathrm{cr1}}^2 = -(d\tilde{P}/d\tilde{V})_{f,\,V = V_{\mathrm{cr1}}} = 1 - 2\Psi_\infty\tilde{V}_{\mathrm{cr1}}$ and, hence, $\tilde{\rho}_{\mathrm{cr1}} \approx -\tilde{V}_{\mathrm{cr1}} = m/(2\Psi_\infty - \gamma_0\Psi_0/\gamma_\infty)$. Note that the values of ρ_{cr} and ρ_{cr1} obtained in this work differ from those obtained by other method [15, 16] only by factor γ_0/γ_∞ multiplying Ψ_0. The reason for this discrepancy is explained in the Sect. 2.

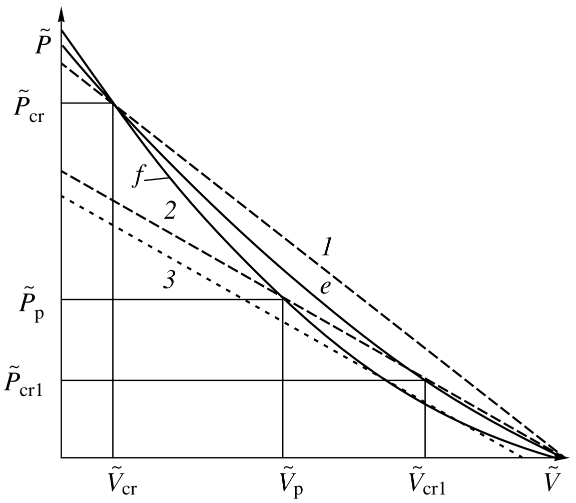


Fig. 2. Shock adiabats at $c_0 > c_\infty$ and $0 < \Psi_0\gamma_0/\gamma_\infty < \Psi_\infty$.

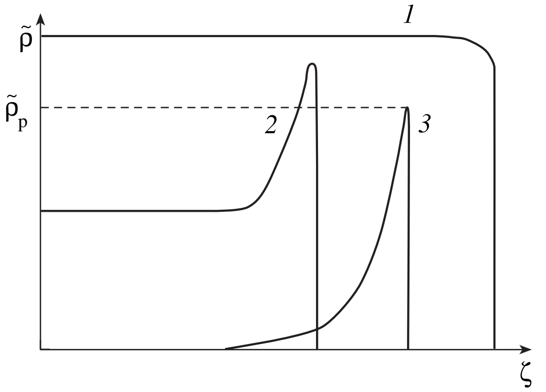


Fig. 3. Shock wave pattern in a negative-dispersion medium at $\Psi_0 > 0$. ζ is the spatial coordinate.

The slope of tangent *3* determines the velocity of an infinitesimal perturbation (the velocity of high-frequency sound) propagating behind the shock wave front. This velocity must exceed that of the shock wave in order that the propagation be stable. Therefore, shock waves with velocities $\tilde{D} \leq \tilde{D}_{\mathrm{cr1}}$ will be evolutionarily unstable. Earlier, it was shown [15] that such an unstable wave in a negative-dispersion medium disintegrates into a sequence of steady-state shock-wave pulses. These pulses are autowaves, because their shape, velocity, and amplitude are independent of the velocity of the initial unstable shock wave. The pulse amplitude, which is specified by the point of intersection between chord *2* and adiabat *f* (Fig. 2), is found from the relationship $\rho_{\mathrm{p}} = 2\tilde{\rho}_{\mathrm{cr1}}$, and the pulse velocity

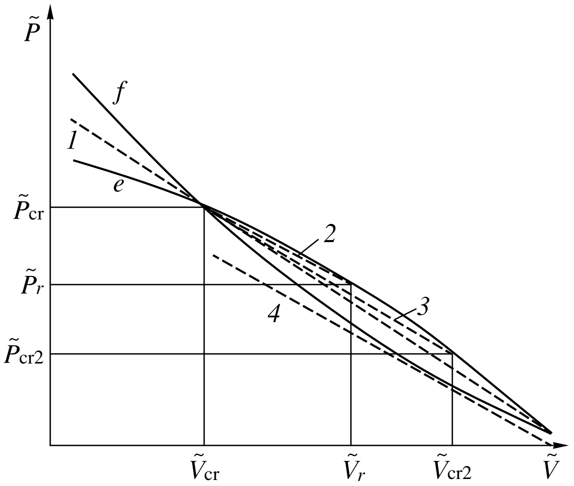


Fig. 4. Shock adiabats at $\Psi_0 < 0$.

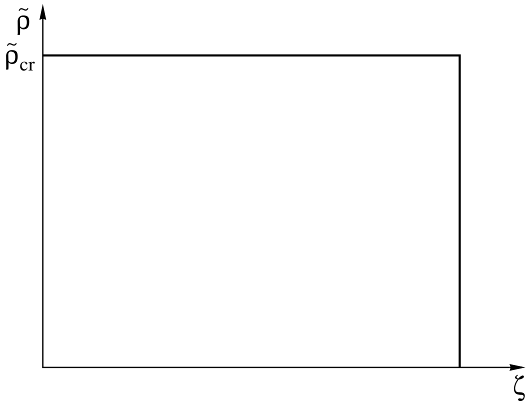


Fig. 5. Autowave at $\Psi_0 < 0$.

equals $\tilde{D}_{\rm cr1}$. The pulse has a discontinuous leading edge and an exponentially falling (with a rate of $\sim\Psi_0/2\Psi_\infty c_\infty\tau$ [15]) trailing edge.

Now consider a possible steady-state pattern of compression shock waves at a negative equilibrium nonlinearity coefficient Ψ_0. The qualitative behavior of shock adiabats (1) at $\Psi_0 < 0$ is shown in Fig. 4. Here, the speed of low-frequency sound behind (before) the shock front is always lower (higher) than the shock wave velocity (found, as usual, from the slope of the chord connecting initial state *1* with the final state). Because of this, the shock waves are evolutionarily unstable.

In the following section, it will be shown that, at $\Psi_0 < 0$, evolutionarily stable gasdynamic structures still exist. First of all, these are waves with amplitude $\tilde{\rho}_{\rm cr}$ (5) (corresponding to the point where the adiabats meet) that propagate with velocity $\tilde{D}_{\rm cr}$ (4) (determined by the slope of chord *1*, Fig. 4). The form of this wave is shown in Fig. 5. Formally, the equilibrium adiabat is unnecessary to describe its structure.

2. EVOLUTION OF GASDYNAMIC PERTURBATIONS AT $\Psi_0 < 0$

Disintegration of an unstable wave into a series of autowave pulses was investigated in [15, 16] by numerically solving a model equation with positive nonlinearity coefficients. Using the same approach, we will study the unsteady evolution of small-amplitude gasdynamic perturbations at $\Psi_0 < 0$. The initial model equation has the form

$$C_{V\infty}\tau\left(\tilde{\rho}_{tt} - c_\infty^2\tilde{\rho}_{xx} - c_\infty^2\Psi_\infty\tilde{\rho}^2_{xx} - \frac{\mu_\infty}{\rho}\tilde{\rho}_{xxt}\right)_t + C_{V0}\left(\tilde{\rho}_{tt} - c_0^2\tilde{\rho}_{xx} - c_0^2\Psi_0\tilde{\rho}^2_{xx} - \frac{\mu_0}{\rho}\tilde{\rho}_{xxt}\right) = 0,$$

$$\mu_\infty = \frac{4\eta}{3} + \chi M\left(\frac{1}{C_{V\infty}} - \frac{1}{C_{P\infty}}\right),\qquad \mu_0 = \frac{4\eta}{3} + \chi M\left(\frac{1}{C_{V0}} - \frac{1}{C_{P0}}\right). \tag{6}$$

Here, μ_∞ and μ_0 are the high- and low-frequency viscosity–heat conduction coefficients; η and χ are the shear viscosity coefficient and thermal conductivity, respectively; and M is the molecular weight. Equation (6) was derived in [22] up to the second order of smallness, i.e., up to θ^2. It describes the evolution of small-amplitude ($\tilde{\rho} \sim \theta$) perturbations in a homogeneous medium in terms of exponential relaxation model (3) in the weak-dispersion ($m \sim \theta$) and low-dissipation ($\tilde{\mu}_{0,\infty} = \mu_{0,\infty}/\tau c_\infty^2\rho \sim \theta$) approximation. For copropagating waves ($\zeta = (x - c_\infty t)/c_\infty\tau$, $y = \theta t/\tau$), Eq. (6) transforms into

$$\left(\tilde{\rho}_y + \frac{\Psi_\infty}{2}\tilde{\rho}^2_\zeta - \tilde{\mu}_\infty\tilde{\rho}_{\zeta\zeta}\right)_\zeta - \frac{C_{V0}}{C_{V\infty}}\left(\tilde{\rho}_y - \frac{C_{V0}}{C_{V\infty}}\tilde{\xi}\tilde{\rho}_\zeta + \frac{\gamma_0\Psi_0}{2\gamma_\infty}\tilde{\rho}^2_\zeta - \tilde{\mu}_0\tilde{\rho}_{\zeta\zeta}\right) = 0, \tag{7}$$

where $\tilde{\xi} = -C_{V\infty}m/2C_{V0}$ is the dimensionless second viscosity coefficient.

In [15, 16], the discrepancy between $\gamma_0\Psi_0/\gamma_\infty$ in (7) and Ψ_0 was neglected at $m \ll 1$, which led to the above-mentioned disagreement between the results obtained by different methods. This discrepancy can be avoided if factor γ_0/γ_∞ appears in model equation (2) in explicit form.

The initial perturbation had the form of a step. The results of numerical simulation of Eq. 7 at $\Psi_0 < 0$ are shown in Figs. 6–8.

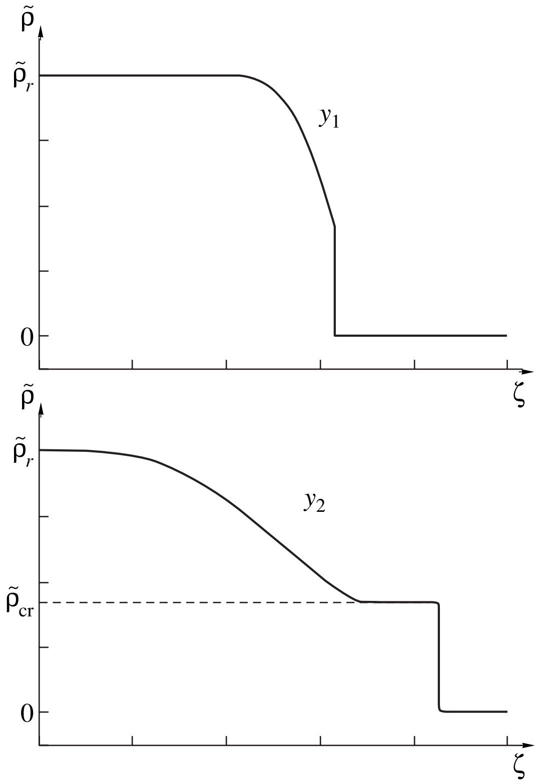


Fig. 6. Evolution of the step at $\Psi_0 < 0$ and $\tilde{\rho}_r > \tilde{\rho}_{cr}$. y_1 and y_2 ($y_2 > y_1$) are dimensionless time instants.

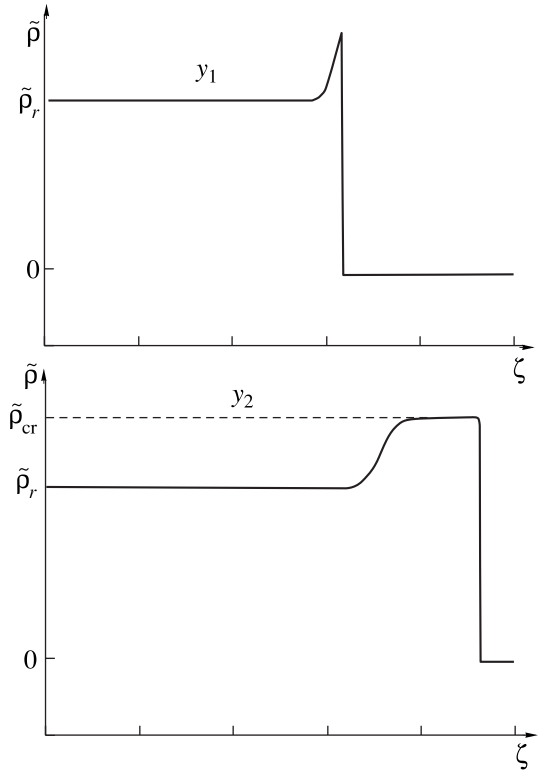


Fig. 7. Evolution of the step at $\Psi_0 < 0$ and $\tilde{\rho}_{cr2} < \tilde{\rho}_r < \tilde{\rho}_{cr}$.

When the step had initial amplitude $\tilde{\rho}_r = \tilde{\rho}_{cr}$, it merely transformed in the course of evolution into an autowave with amplitude $\tilde{\rho}_{cr}$ (Fig. 5). A small but finite value of high-frequency viscosity–heat conduction coefficient $\tilde{\mu}_\infty$ of causes the front to broaden, while a nonzero value of $\tilde{\mu}_0$ changes the amplitude of the autowave, $\tilde{\rho}_{cr} = \tilde{\mu}_\Sigma C_{V0}/C_{V\infty}(\Psi_\infty - \gamma_0\Psi_0/\gamma_\infty)$. Here, $\tilde{\mu}_\Sigma = \tilde{\xi} + \tilde{\mu}_0$ is the low-frequency coefficient of total viscosity. Note that the autowave exists only if $\tilde{\mu}_\Sigma < 0$. In [15], the same conclusion was drawn regarding the existence of structures shown in Fig. 3 by curves *2* and *3*.

When the step had initial amplitude $\tilde{\rho}_r > \tilde{\rho}_{cr}$ (Fig. 6), it disintegrates into a gradually expanding unsteady compression wave and the already mentioned autowave.

When the step had initial amplitude $\tilde{\rho}_{cr2} < \tilde{\rho}_r < \tilde{\rho}_{cr}$ (Fig. 7) where $\tilde{\rho}_{cr2} = \tilde{\rho}_{cr}\Psi_\infty/(2\Psi_\infty - \gamma_0\Psi_0/\gamma_\infty)$, it evolved into two steady-state waves: an autowave, as before, and an expansion shock wave. Both waves propagate in the same direction. The expansion shock wave velocity is determined by the slope of the chord drawn through points $\tilde{V}_r$ and $\tilde{V}_{cr}$ on equilibrium adiabat *e* (straight line *2* in Fig. 4),

$$\tilde{D}_r = \sqrt{\frac{\tilde{P}_{cr} - \tilde{P}_r}{\tilde{V}_r - \tilde{V}_{cr}}} = \sqrt{1 + \Psi_\infty\tilde{\rho}_{cr} + \Psi_0\frac{\gamma_0}{\gamma_\infty}\tilde{\rho}_r}.$$

At $\tilde{\rho}_r < \tilde{\rho}_{cr}$, the inequality $c_{01} < \tilde{D}_r < c_{02}$ is fulfilled and the condition $\tilde{\rho}_r > \tilde{\rho}_{cr2}$ is necessary for another inequality, $\tilde{D}_r < \tilde{D}_{cr2} = c_{\infty 2}$, to be met. Here, $c_{01} = \sqrt{1 + m + 2\Psi_0\gamma_0\tilde{\rho}_{cr}/\gamma_\infty}$ and $c_{02} =$

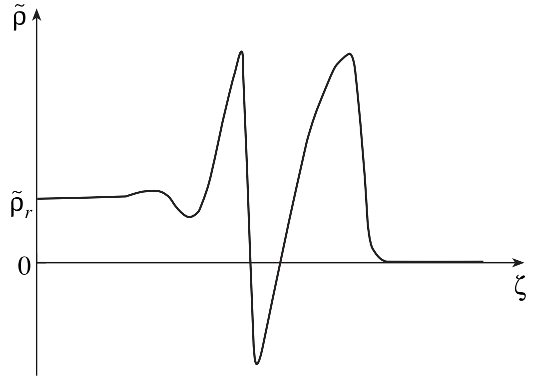


Fig. 8. Initial stage of splitting a step with small amplitude $\tilde{\rho}_r < \tilde{\rho}_{cr2}$.

$\sqrt{1 + m + 2\Psi_0\gamma_0\tilde{\rho}_r/\gamma_\infty}$ are the speeds of low-frequency sound, respectively, before and behind the expansion shock wave front and $c_{\infty 2} = \sqrt{1 + 2\Psi_\infty\tilde{\rho}_{\text{cr}2}}$ is the speed of high-frequency sound behind the front of the expansion shock wave at $\tilde{\rho}_r = \tilde{\rho}_{\text{cr}2}$. In Fig. 4, velocities $\tilde{D}_{\text{cr}2}$ and $c_{\infty 2}$ are defined by the condition that chord *3* connecting points $\tilde{V}_{\text{cr}}$ and $\tilde{V}_{\text{cr}2}$ on the equilibrium adiabat is parallel to tangent *4* to the frozen adiabat that is drawn through point $\tilde{V}_{\text{cr}2}$. These inequalities provide the evolutionary stability of the expansion shock wave traveling in the gas precompressed by the autowave. At $\Delta\tilde{\rho} = \tilde{\rho}_{\text{cr}} - \tilde{\rho}_r \ll \tilde{\rho}_{\text{cr}}$, the expansion shock wave has the form of a symmetric (about the mean level) jump: $\tilde{\rho} \sim \tanh(\zeta/2\Delta)$, where $\Delta \sim 2\tilde{\mu}_\Sigma/\Psi_0\Delta\tilde{\rho}$. As $\Delta\tilde{\rho}$ grows, the shape of the jump deforms and becomes asymmetric about the mean level. Such a front pattern is also characteristic of weak compression shock waves, $c_0/c_\infty < \tilde{D} < 1$ (chord *1–2* in Fig. 1), traveling in a medium with relaxation equation (2) [19, 20].

Expansion shock wave velocity $\tilde{D}_r$ at $\Psi_0 < 0$ is lower than that of the autowave, $\tilde{D}_{\text{cr}}$. Therefore, a combination of the two waves is stable; however, the distance between their fronts increases with time.

The simulation of unsteady evolution of a small step with amplitude $\tilde{\rho}_r < \tilde{\rho}_{\text{cr}2}$ by means of Eq. (7) led to the fast development of numerical instability. The initial stage of the step evolution is shown in Fig. 8.

DISCUSSION

In this paper, we (i) analyzed the form of shock adiabats and their associated weak shock wave patterns in a nonequilibrium medium where the low-frequency (equilibrium) speed of sound exceeds the high-frequency (frozen) speed of sound and (ii) pioneered investigation of complicated regimes arising when the low- and high-frequency nonlinearity coefficients have opposite signs. It is shown that, in a thermodynamically nonequilibrium medium, steady-state structures appreciably different from shock waves propagating through equilibrium media may exist. This fact has been corroborated by numerous experiments (see, for example, [23–25]. In particular, splitting of the shock wave front and emergence of precursors was noted; however, it still remains unclear whether nonequilibrium or inhomogeneity of the medium is the key factor responsible for their formation [26–28]. The above reasoning disregards the inhomogeneity of a real nonequilibrium medium and, therefore, is idealized. The joint effect of nonequilibrium and inhomogeneity of a medium on the shock wave pattern should be considered in terms of more sophisticated models, which may be tested, in particular, by passing to the limit of the idealized structures discussed above. In general, our results suggest that steady-state nonequilibrium strongly influences the evolution of a gasdynamic perturbation by imparting new viscous, dispersion, and nonlinear properties to the medium.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project no. 05-02-96500), the Federal Program "Integration of Science and Higher Education of Russia in 2002–2006," the Russian–American Program "Basic Research and Higher Education (BRHE)," CRDF (project SA-014-02), the Ministry of Education of the Russian Federation, and the Administration of Samara Oblast.

REFERENCES

1. N. E. Molevich, AIAA Pap., No. 1020, 1 (2004).
2. H. J. Bauer and H. E. Bass, Phys. Fluids **16**, 988 (1973).
3. A. A. Borisov, *Research on Hydrodynamics and Heat Exchange* (ITF SO AN SSSR, Novosibirsk, 1976), pp. 94–95.
4. E. Ya. Kogan and N. E. Molevich, Izv. Vyssh. Uchebn. Zaved. Fiz. **29** (7), 53 (1986).
5. E. Ya. Kogan and N. E. Molevich, Zh. Tekh. Fiz. **56**, 941 (1986) [Sov. Phys. Tech. Phys. **31**, 573 (1986)].
6. A. I. Osipov and A. V. Uvarov, Vestn. Mosk. Univ., Ser. 3: Fiz., Astron. **28** (6), 52 (1987).
7. N. E. Molevich and A. N. Oraevskiĭ, Zh. Éksp. Teor. Fiz. **94**, 128 (1988) [Sov. Phys. JETP **67**, 504 (1988)].
8. N. A. Dunaevskiĭ, S. A. Zhdanok, A. P. Napartovich, *et al.*, Prikl. Mekh. Tekh. Fiz., No. 4, 33 (1988).
9. I. P. Zavershinskiĭ, E. Ya. Kogan, and N. E. Molevich, Akust. Zh. **38**, 702 (1992) [Sov. Phys. Acoust. **38**, 387 (1992)].
10. N. E. Molevich and A. N. Oraevskiĭ, Tr. Fiz. Inst. Akad. Nauk SSSR **222**, 45 (1992).
11. V. N. Malnev and A. V. Nedospasov, in *Proceedings of the 1st International Workshop on Perspectives of MHD and Plasma Technologies in Aerospace Application, Moscow, 1999,* pp. 128–130.
12. N. E. Molevich and V. E. Nenashev, Akust. Zh. **46**, 520 (2000) [Acoust. Phys. **46**, 450 (2000)].
13. N. E. Molevich, Akust. Zh. **49**, 229 (2003) [Acoust. Phys. **49**, 189 (2003)].
14. E. Ya. Kogan and N. E. Molevich, Akust. Zh. **39**, 951 (1993) [Acoust. Phys. **39**, 505 (1993)].
15. V. G. Makaryan and N. E. Molevich, Pis'ma Zh. Tekh. Fiz. **29** (18), 11 (2003) [Tech. Phys. Lett. **29**, 752 (2003)].
16. V. G. Makaryan and N. E. Molevich, Izv. Ross. Akad. Nauk, Mekh. Zhidk. Gaza, No. 5, 181 (2004).
17. Ya. B. Zel'dovich and Yu. P. Raizer, *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena* (Nauka, Moscow, 1966; Academic, New York, 1967).

18. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 6: *Fluid Mechanics* (Nauka, Moscow, 1986; Pergamon, New York, 1987).
19. O. V. Rudenko and S. I. Soluyan, *Theoretical Foundations of Nonlinear Acoustics* (Nauka, Moscow, 1975; Plenum Consultants Bureau, New York, 1977).
20. G. A. Ostroumov, *Fundamentals of Nonlinear Acoustics* (Leningr. Gos. Univ., Leningrad, 1967) [in Russian].
21. J. F. Clarke and M. McChesney, *The Dynamics of Real Gases* (Butterworths, London, 1964; Mir, Moscow, 1967).
22. N. E. Molevich, Sib. Fiz.-Tekh. Zh., No. 1, 133 (1991).
23. A. I. Klimov, G. I. Mishin, A. B. Fedotov, and V. A. Shakhovatov, Pis’ma Zh. Tekh. Fiz. **15** (20), 31 (1989) [Sov. Tech. Phys. Lett. **15**, 800 (1989)].
24. S. A. Bystrov, V. I. Ivanov, and F. V. Shugaev, Fiz. Plazmy **15**, 558 (1989) [Sov. J. Plasma Phys. **15**, 324 (1989)].
25. A. Yu. Gridin and A. I. Klimov, Khim. Fiz. **12**, 363 (1993).
26. A. S. Baryshnikov, I. V. Basargin, and M. V. Chistyakova, Zh. Tekh. Fiz. **71** (3), 17 (2001) [Tech. Phys. **46**, 287 (2001)].
27. S. O. Macheret, Yu. Z. Ionich, and N. V. Chernysheva, Phys. Fluids **13**, 2693 (2001).
28. A. I. Klimov, Doctoral Dissertation (Inst. High Temp. RAS, Moscow, 2002).

Translated by N. Mende

Technical Physics, Vol. 50, No. 6, 2005, pp. 692–697. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 6, 2005, pp. 19–24.
Original Russian Text Copyright © 2005 by Mat'ev, Pikulev.

GASES AND LIQUIDS

Quasi-Potential Gas Flow in Nuclear-Pumped Flowing-Gas Laser Cells

V. Yu. Mat'ev and A. A. Pikulev

All-Russia Research Institute of Experimental Physics, Russian Federal Nuclear Center,
Sarov, Nizhnii Novgorod Oblast, 607190 Russia
e-mail: pikulev@expd.vniief.ru

Received September 2, 2004

Abstract—The influence of the boundary conditions on the two-dimensional gas velocity distribution in nuclear-pumped flowing-gas laser cells is considered for low energy deposits. It is demonstrated that, for any boundary conditions, the two-dimensional flow separates into two quasi-one-dimensional components and the size of the area of influence of the boundary conditions is comparable to the laser cell width. *© 2005 Pleiades Publishing, Inc.*

INTRODUCTION

In nuclear-pumped laser cells, the interior is covered by thin uranium layers the fission fragments of which irradiate the gas contained in the cell. In a flowing-gas laser (see figure), the gas flows through the cell, as the name suggests [1, 2]. Here, the x axis is directed along the gas flow, the cell length along the flow is $L \approx 0.1$ m, and the cell width is $2h \approx 0.02$ m. The optical axis of the laser is orthogonal to the plane of the figure.

The general pattern of the two-dimensional gas flow in nuclear-pumped laser cells is considered in [3, 4]. The method of separation of variables is applied to analytically calculate the two-dimensional gas velocity for low (compared to the internal energy of the gas) energy deposits from the fission fragments under the assumption that the gas velocity field is irrotational and so can be described by a scalar potential satisfying the Poisson equation [3]. The emphasis is on hermetic cells, where the boundary conditions for the gas velocity are straightforward (the normal component of the velocity vanishes at the cell walls). In the works cited, as well as in [2], flowing-gas cells are considered only in the speculative case of an unperturbed gas flow incident on the energy deposition area (bounded by the uranium layers) "from infinity." However, flowing-gas lasers may be equipped with gas-cooling radiators (placed immediately ahead of or behind the uranium layers) [1] or with other components of the gas system, which form different boundary conditions at the inlet and/or outlet of the cell. This brings up the question of how to attack such problems more correctly and how much the gas flow pattern in the laser cell will change.

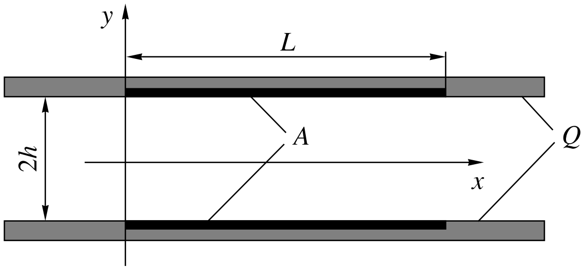
In this paper, the main results obtained in [3, 4] are generalized to the case of arbitrary boundary conditions for the gas flow velocity (at both the inlet and outlet of the cell) and the effect of the boundary conditions on the flow pattern in the laser channel is investigated. It is demonstrated that the main conclusion drawn in [3, 4] that the two-dimensional gas flow separates into two quasi-one-dimensional components remains valid for any boundary conditions.

GASDYNAMIC MODEL

The gas is assumed to be ideal, nonviscous, and thermally nonconductive (viscosity and thermal conductivity are of considerable importance only in a narrow near-wall layer), and the gas pressure is set uniform throughout the cell (the Mach number is $M \leq 10^{-2} \ll 1$) [3, 4]. Under these conditions, the time variation of the gas density for a fixed Lagrangean coordinate is described by the simple thermodynamic relation [3, 4]

$$\frac{\gamma P}{\rho}\frac{d\rho}{dt} = -\frac{dP}{dt} - (\gamma - 1)\frac{\delta Q}{\delta V \delta t}, \tag{1}$$

where P is the gas pressure, ρ is the gas density, γ is the adiabatic exponent, and δQ is the energy absorbed by infinitesimal gas volume δV during infinitesimal time period δt (the energy deposit is assumed to be known [5]).



Nuclear-pumped laser: *Q*, cell; *A*, uranium layer.

With regard to the continuity equation

$$\frac{1}{\rho}\frac{d\rho}{dt} + \operatorname{div}\mathbf{v} = 0, \quad \mathbf{v} = (u, w),$$

where **v** is the gas velocity with components $v_x = u$ and $v_y = w$, Eq. (1) can be recast as [3, 4, 6]

$$\operatorname{div}\mathbf{v} = G(x, y, t),$$

$$G(x, y, t) = \frac{1}{\gamma P}\left\{(\gamma - 1)\left(\frac{\delta Q}{\delta V \delta t}\right) - \frac{dP}{dt}\right\}. \quad (2)$$

If the gas flow in the cell is steady, dP/dt may be set equal to 0.

The key gasdynamic factor of nuclear-pumped laser cells is the energy deposited into the gas by irradiating fission fragments of the uranium layers. If the energy deposit is not too high, the associated gas flow may be viewed as irrotational [3, 4] (the incoming gas flow is supposed to be vortex-free). In this case, the velocity is completely specified by scalar potential $\Phi(x, y, t)$,

$$\mathbf{v}(x, y, t) = \nabla\Phi,$$

which, according to (2), satisfies the Poisson equation [3, 4]

$$\Delta\Phi = G(x, y, t). \quad (3)$$

This equation does not involve time in explicit form. The velocity distribution is quasi-stationary: it changes in line with change in the gas pressure or energy deposit distribution. Hereafter, argument t is omitted and function $G(x, y)$ is assumed to be given.

GENERAL SOLUTION FOR THE VELOCITY POTENTIAL

At the walls, the normal component of the gas velocity is equal to zero. In the hermetic cells, whose inner surface is closed, this condition suffices to solve elliptic equation (3). In the case of flowing-gas cells, where the surface is basically unclosed, the situation is more complicated. However, here, at horizontal walls (see figure), the vertical component of the velocity vanishes, $w(x, y) = 0$, for $y = \pm h$ (as well as in hermetic cells). A solution to Eq. (3) that satisfies this condition can be obtained by applying the method of separation of variables and expanding the velocity potential into the Fourier series,

$$\Phi(x, y) = \frac{\varphi_0(x)}{2} + \sum_{n=1}^{\infty}\varphi_n(x)\cos(a_n y),$$

$$\varphi_n(x) = \frac{2}{h}\int_0^h \Phi(x, y)\cos(a_n y)dy, \quad a_n = \frac{\pi n}{h}.$$

This solution has the form [3, 4]

$$u(x, y) = \frac{\partial\Phi}{\partial x} = U(x) + \sum_{n=1}^{\infty} u_n(x)\cos(a_n y), \quad u_n(x) = \frac{d\varphi_n}{dx}, \quad (4)$$

$$w(x, y) = \frac{\partial\Phi}{\partial y} = \sum_{n=1}^{\infty} w_n(x)\sin(a_n y), \quad w_n(x) = -a_n\varphi_n(x), \quad (5)$$

where

$$u_n(x) = \frac{\exp(a_n x)}{2}\left\{A_n + \int_{-\infty}^{x} g_n(\xi)\exp(-a_n\xi)d\xi\right\} + \frac{\exp(-a_n x)}{2}\left\{B_n + \int_{-\infty}^{x} g_n(\xi)\exp(a_n\xi)d\xi\right\}, \quad (6)$$

$$w_n(x) = -\frac{\exp(a_n x)}{2}\left\{A_n + \int_{-\infty}^{x} g_n(\xi)\exp(-a_n\xi)d\xi\right\} + \frac{\exp(-a_n x)}{2}\left\{B_n + \int_{-\infty}^{x} g_n(\xi)\exp(a_n\xi)d\xi\right\}, \quad (7)$$

$$U(x) = U_0 + \frac{2}{h}\int_0^x\int_0^h G(\xi, y)dyd\xi,$$

$$g_n(x) = \frac{2}{h}\int_0^h G(x, y)\cos(a_n y)dy.$$

Here, U_0 is the flowing gas velocity (the gas velocity at the inlet of the cell); $U(x)$ is the transversely averaged longitudinal velocity of the gas, which depends on the transversely averaged energy deposit; and $g_n(x)$ are the coefficients of expansion of function $G(x, y)$ into the Fourier series. The values of A_n and B_n are determined by the boundary conditions for the gas velocity at the entrance to and exit from the cell.

By renormalizing constants A_n, this solution can be put in a more convenient form,

$$A_n = H_n\exp(-a_n L) - \int_{-\infty}^{\infty} g_n(\xi)\exp(-a_n\xi)d\xi.$$

For simplicity, the energy deposition area is assumed to be bounded by the cell dimensions,

$$g_n(x) = 0 \quad \text{for} \quad x < 0 \quad \text{and} \quad x > L.$$

Then, the solution given by (6) and (7) takes the form

$$\begin{aligned} u_n(x) &= \frac{1}{2}\{B_n\exp(-a_nx) + H_n\exp[-a_n(L-x)]\} \\ &+ \frac{1}{2}\left\{\int\limits_0^x g_n(\xi)\exp[-a_n(x-\xi)]d\xi \right. \\ &\left. - \int\limits_x^L g_n(\xi)\exp[-a_n(\xi-x)]d\xi\right\}, \end{aligned} \tag{8}$$

$$\begin{aligned} w_n(x) &= \frac{1}{2}\{B_n\exp(-a_nx) - H_n\exp[-a_n(L-x)]\} \\ &+ \frac{1}{2}\left\{\int\limits_0^x g_n(\xi)\exp[-a_n(x-\xi)]d\xi \right. \\ &\left. + \int\limits_x^L g_n(\xi)\exp[-a_n(\xi-x)]d\xi\right\}. \end{aligned} \tag{9}$$

Here,

$$\begin{aligned} u_n(0) &= \frac{1}{2}[H_n\exp(-a_nL) + B_n] - \frac{1}{2}\Psi_n, \\ u_n(L) &= \frac{1}{2}[B_n\exp(-a_nL) + H_n] + \frac{1}{2}\Xi_n; \end{aligned} \tag{10}$$

$$\begin{aligned} w_n(0) &= \frac{1}{2}[B_n - H_n\exp(-a_nL)] + \frac{1}{2}\Psi_n, \\ w_n(L) &= \frac{1}{2}[B_n\exp(-a_nL) - H_n] + \frac{1}{2}\Xi_n; \end{aligned} \tag{11}$$

$$\psi_n = \int\limits_0^L g_n(\xi)\exp(-a_n\xi)d\xi,$$

$$\begin{aligned} \Xi_n &= \int\limits_0^L g_n(\xi)\exp[-a_n(L-\xi)]d\xi \\ &= \int\limits_0^L g_n(L-\zeta)\exp(-a_n\zeta)d\zeta. \end{aligned}$$

Presented in this form, the solution is more convenient for both qualitative analysis and numerical calculations, because it contains only damped exponentials, whereas the terms in the first braces of presentations (6) and (7) for this solution should be handled with care, since they are multiplied by an increasing exponent.

QUALITATIVE ANALYSIS OF THE SOLUTION

Prior to analysis, let us make some useful simplifications. Constants H_n and B_n are not too large *a fortiori* (they do not exceed the gas velocity in order of magnitude) and the value of $\exp(-a_nL)$ is vanishingly small ($a_nL \geq \pi L/h \sim 10\pi$). Therefore, boundary conditions (10) and (11) may be simplified to

$$u_n(0) \approx \frac{1}{2}(B_n - \Psi_n), \quad u_n(L) \approx \frac{1}{2}(\Xi_n + H_n), \tag{12}$$

$$w_n(0) \approx \frac{1}{2}(B_n + \Psi_n), \quad w_n(L) \approx \frac{1}{2}(\Xi_n - H_n). \tag{13}$$

Constants B_n are thus seen to depend actually on only the gas velocity at the inlet and constants H_n, on the gas velocity at the outlet. Therefore, constants B_n and H_n are determined independently of each other for any boundary conditions at the inlet and outlet, which allows unification of a set of possible solutions.

Next, the energy deposit (and, hence, expansion coefficients $g_n(x)$ for function $G(x, y)$) changes appreciably only near the ends of the uranium layer over a distance of h (where h is the half-width of the cell) from the inlet and outlet. In the remaining part of the cell, coefficients $g_n(x)$ vary along the flow smoothly, the variation being insignificant over a distance of order h. Over such distances, the exponentials involved in integrals (8) and (9) turn the integrand virtually into zero: for $|x - \xi| = 2h$, we have

$$a_n|x-\xi| \geq 2\pi, \quad \exp(-2\pi) \approx 0.002.$$

Therefore,

$$\begin{aligned} &\int\limits_0^x g_n(\xi)\exp[-a_n(x-\xi)]d\xi \\ &\approx g_n(x)\int\limits_0^x \exp[-a_n(x-\xi)]d\xi = \frac{g_n(x)}{a_n}[1-\exp(-a_nx)], \end{aligned}$$

$$\begin{aligned} &\int\limits_x^L g_n(\xi)\exp[-a_n(\xi-x)]d\xi \\ &\approx g_n(x)\int\limits_x^L \exp[-a_n(\xi-x)]d\xi \\ &= \frac{g_n(x)}{a_n}\{1-\exp[-a_n(L-x)]\}. \end{aligned}$$

Thus, in a major part of the cell, the solution given by (8) and (9) may be represented in the form

$$u_n(x) \approx \frac{1}{2}\{B_n \exp(-a_n x) + H_n \exp[-a_n(L-x)]\} + \frac{g_n(x)}{2a_n}\{\exp[-a_n(L-x)] - \exp(-a_n x)\}, \tag{14}$$

$$w_n(x) = \frac{1}{2}\{B_n \exp(-a_n x) - H_n \exp[-a_n(L-x)]\} + \frac{g_n(x)}{2a_n}\{2 - \exp[-a_n(L-x)] - \exp(-a_n x)\}. \tag{15}$$

A similar estimate can be derived for integrals Ψ_n and Ξ_n,

$$\Psi_n \cong \frac{g_n(0)}{a_n}[1 - \exp(-a_n L)] \approx \frac{g_n(0)}{a_n},$$

$$\Xi_n \cong \frac{g_n(L)}{a_n}[1 - \exp(-a_n L)] \approx \frac{g_n(L)}{a_n}.$$

This estimate is cruder, since coefficients $g_n(x)$ near the uranium layer vary still noticeably. However, the solution given by (14) and (15) can be (at least, approximately) extended up to the ends of the cell in this way.

Even at a short distance from the inlet and outlet (more specifically, over a distance of the cell width), all the exponentials in solution (14)–(15) become negligible. Hence, we have

$$u_n(x) \longrightarrow 0, \quad w_n(x) \longrightarrow \frac{g_n(x)}{a_n} \text{ for } 2h \le x \le L - 2h. \tag{16}$$

We may thus conclude that, in a major part of the cell, the longitudinal velocity of the gas is, in essence, equal to its transversely averaged value $U(x)$, which depends on the transversely averaged energy deposit. For the transverse velocity, we have, according to (5) and in view of (16) and (2),

$$\frac{\partial w}{\partial y} = \frac{\partial}{\partial y}\sum_{n=1}^{\infty}\frac{g_n(x)}{a_n}\sin(a_n y)$$

$$= \sum_{n=1}^{\infty} g_n(x)\cos(a_n y) = G(x,y) - \langle G(x,y)\rangle \tag{17}$$

$$= \frac{\gamma-1}{\gamma P}\left\{\left(\frac{\delta Q}{\delta V \delta t}\right) - \left\langle\frac{\delta Q}{\delta V \delta t}\right\rangle\right\},$$

where $\langle\ldots\rangle$ means averaging across the cell.

Thus, in the cell (except for the narrow end regions), the longitudinal gas velocity is almost uniform in the transverse direction and the transverse velocity depends on the difference between the local and transversely averaged energy deposits. Such a separation of the two-dimensional flow into two quasi-one-dimensional components (as demonstrated in [3, 4] with a gas flow incident on the energy deposition area "from infinity") is seen to take place in flowing-gas cells under any boundary conditions (at both the inlet and the outlet).

DISCUSSION OF THE BOUNDARY CONDITIONS

General solution (8)–(9) is applicable to both hermetic and flowing-gas cells. The formal difference is that, in the former case, $U_0 = 0$, whereas, in the latter case, the two-dimensional flow described by solution (8)–(9) is added to the general flow pattern. Such an additivity property of the velocity field is due to the linearity of the equation for the velocity potential in the gasdynamic model considered.

For hermetic cells, the boundary conditions along the x axis are straightforward,

$$u_n(0) = u_n(L) = 0, \tag{18}$$

hence, in view of (10),

$$B_n = \frac{\Psi_n + \Xi_n \exp(-a_n L)}{1 - \exp(-2a_n L)} \approx \Psi_n,$$
$$H_n = -\frac{\Xi_n + \Psi_n \exp(-a_n L)}{1 - \exp(-2a_n L)} \approx -\Xi_n \tag{19}$$

in accordance with (12). Then, by virtue of (13), we have

$$w_n(0) \approx \frac{1}{2}(B_n + \Psi_n) = \Psi_n \cong \frac{g_n(0)}{a_n},$$
$$w_n(L) \approx \frac{1}{2}(\Xi_n - H_n) = \Xi_n \cong \frac{g_n(L)}{a_n}. \tag{20}$$

Theoretically, the same conditions may also be set for flowing-gas cells. Here, the flow immediately takes the quasi-one-dimensional form specified by (16) and (17). Thus, if the longitudinal velocity is uniform across the cell at the inlet, it remains such downstream throughout the cell (except for a narrow region at the outlet, where the flow pattern may be changed according to the boundary conditions at the outlet). However, such "quasi-hermetic" conditions are purely speculative.

The "zero" boundary conditions at the inlet,

$$u_n(0) = w_n(0) = 0, \tag{21}$$

which correspond to the unperturbed gas flow, would seemingly be natural. According to (12) and (13), this is possible only when $\Psi_n = 0$ (and, hence, $B_n = 0$). In this case, $g_n(x) = 0$ (the energy deposit is transversely uniform), the transverse velocity equals zero, in accordance with (17), and the flow is one-dimensional.

Thus, formulation of the boundary conditions in flowing-gas cells needs no special discussion. In [3, 4], where a gas flow incident on the energy deposition area "from infinity" is considered, zero boundary conditions (21) are set at $x \longrightarrow -\infty$ and constants A_n and B_n are determined by joining the solutions for the domains $x < 0$ and $x > 0$. In the domain $x < 0$ (no energy deposit), the solution may be obtained in the same form as (6) and (7),

$$\begin{aligned} u_n(x) &= \frac{1}{2}[C_n \exp(a_n x) + D_n \exp(-a_n x)], \\ w_n(x) &= \frac{1}{2}[-C_n \exp(a_n x) + D_n \exp(-a_n x)]. \end{aligned} \tag{22}$$

Here, one should put $D_n = 0$ in order that the velocity profile be uniform as $x \longrightarrow -\infty$: $u(x \longrightarrow -\infty) = U_0$ and $w(x \longrightarrow -\infty) = 0$. Combining (22), (12), and (13) at $x = 0$, we obtain

$$B_n = 0, \quad C_n = -\Psi_n. \tag{23}$$

Constants H_n can be determined by assuming that the cell extends "to infinity" beyond the energy deposition area ($x > L$). Beyond this area (at $x \longrightarrow +\infty$), the solution can be written in a form similar to (22), namely,

$$u_n(x) = \frac{1}{2}[E_n \exp[a_n(x-L)] + F_n \exp[-a_n(x-L)]],$$

$$w_n(x) = \frac{1}{2}[-E_n \exp[a_n(x-L)] + F_n \exp[-a_n(x-L)]].$$

From the boundedness of the solution at infinity, it follows that $E_n = 0$. Combining the above two expressions with (12) and (13) at $x = L$, we get

$$H_n = 0, \quad F_n = \Xi_n.$$

This solution is of interest, since it implies that the unperturbed velocity profile starts deforming before it approaches the energy deposition area: in a subsonic flow, any variation of the flow conditions in the channel is perceived in advance. On the other hand, away from the cell (at $x \longrightarrow +\infty$), all the changes in the gas velocity that were acquired in the energy deposition area disappear except for the increase in the transversely uniform longitudinal velocity. Outside the energy deposition area, the two-dimensional effects fade out over a distance approximately equal to the channel width. Certainly, these conclusions are valid only in the framework of the initial approximations (the potential flow). Actually, the longitudinal pressure gradient would cause transverse splitting of the longitudinal gas velocity.

It can be said that the above-mentioned joining of the energy deposition area with "empty" areas sets "soft" boundary conditions, since the gas velocity at the inlet (and outlet) is determined by joining with solutions in other areas, instead of being strictly specified.

However, such an approach to the problem of setting boundary conditions in flowing-gas cells is not unique. Just ahead of the energy deposition area (and immediately after it), there may be devices that strictly specify other boundary conditions (such as, parallel-plate gas-cooling radiators [1]). The form of the solution to Eq. (3) obtained here (see (8)–(13)) allows for any influence on the gas velocity at both the inlet and outlet of the cell.

For instance, in the case of radiators with closely spaced plates, one can assume that the gas velocity at the outlet of a radiator (and at the inlet of the next radiator) has only the longitudinal component, while the transverse component is equal to zero, $w_n(0) = w_n(L) = 0$. Then, (12) and (13) yield

$$B_n = -\Psi_n, \quad H_n = \Xi_n, \tag{24}$$

with

$$\begin{aligned} u_n(0) &\approx \frac{1}{2}(B_n - \Psi_n) = -\Psi_n \cong -\frac{g_n(0)}{a_n}, \\ u_n(L) &\approx \frac{1}{2}(\Xi_n + H_n) = \Xi_n \cong \frac{g_n(L)}{a_n}. \end{aligned} \tag{25}$$

The boundary conditions thus stated are opposite to "quasi-hermetic" conditions (18)–(20).

THE CASE OF LONGITUDINALLY UNIFORM ENERGY DEPOSITION

Let us demonstrate the influence of the three types of boundary conditions mentioned above with the simple and illustrative case when the energy deposit is longitudinally uniform within the cell, $G = G(y)$, and vanishes outside it. Here,

$$g_n(x) = \begin{cases} g_n = \text{const}, & 0 < x < L \\ 0, \quad x < 0, \quad x > L, & \end{cases} \tag{26}$$

because the pressure is uniform. The velocity distribution may be calculated by formulas (14) and (15), which are exact, rather than approximate, in this case. We restrict our consideration to the vicinity of the inlet, since, outside this region, the "inlet" boundary conditions have no influence on the flow and the flow pattern at the outlet is symmetric. Then,

$$\begin{cases} u_n(x) \approx \frac{1}{2}B_n \exp(-a_n x) - \frac{g_n}{2a_n}\exp(-a_n x) \\ 0 < x \sim h \ll L \\ w_n(x) = \frac{1}{2}B_n \exp(-a_n x) + \frac{g_n}{2a_n}[2 - \exp(-a_n x)], \end{cases}$$

with

$$\Psi_n = \Xi_m \approx \frac{g_n}{a_n}.$$

Then, for quasi-hermetic boundary conditions (18)–(20), we obtain

$$u_n \approx 0, \quad w_n = \frac{g_n}{a_n}, \quad 0 < x \sim h \ll L.$$

In this case, the transverse velocity is independent of the longitudinal coordinate,

$$\frac{\partial w}{\partial y} = \frac{\partial}{\partial y}\sum_{n=1}^{\infty}\frac{g_n}{a_n}\sin(a_n y)$$

$$= \sum_{n=1}^{\infty} g_n \cos(a_n y) = G(y) - \langle G(y) \rangle.$$

For soft boundary conditions (23), we get

$$u_n(x) \approx -\frac{g_n}{2a_n}\exp(-a_n x),$$

$$w_n(x) = \frac{g_n}{2a_n}[2 - \exp(-a_n x)], \quad 0 < x \sim h \ll L.$$

Herein,

$$\left.\frac{\partial w}{\partial y}\right|_{x=0} = \frac{\partial}{\partial y}\sum_{n=1}^{\infty}\frac{g_n}{2a_n}\sin(a_n y)$$

$$= \frac{1}{2}\sum_{n=1}^{\infty} g_n \cos(a_n y) = \frac{1}{2}[G(y) - \langle G(y) \rangle],$$

and the transverse velocity at the inlet is exactly half its limiting value (17) far away from the inlet.

Finally, for longitudinal boundary conditions (24) and (25), we have

$$u_n(x) \approx -\frac{g_n}{a_n}\exp(-a_n x),$$

$$w_n(x) = \frac{g_n}{a_n}[1 - \exp(-a_n x)], \quad 0 < x \sim h \ll L,$$

and the transverse velocity at the inlet is equal to zero.

CONCLUSIONS

The two-dimensional gas velocity distribution in nuclear-pumped flowing-gas laser cells at low energy deposits that was obtained earlier [3, 4] for a particular form of the boundary conditions at the inlet is generalized to the case of arbitrary boundary conditions at both the inlet and outlet. Three types of boundary conditions are considered, namely, quasi-hermetic, soft (derived by joining), and longitudinal conditions. The last-named conditions are the most adequate if cooling radiators are placed just ahead of the inlet (and immediately after the outlet) of the cell. The conditions at the inlet and outlet may be specified independently by combining different situations in an arbitrary way.

It is demonstrated that the effect of the boundary conditions is tangible only in the neighborhood of the inlet and outlet. The length of these regions is close to the channel width. Outside these neighborhoods, the two-dimensional flow separates into two quasi-one-dimensional components, as has been shown in papers [3, 4] for specific boundary conditions.

REFERENCES

1. V. V. Borovkov, B. V. Lazhintsev, V. A. Nor-Arevyan, *et al.*, Kvantovaya Élektron. (Moscow) **22**, 1187 (1995).
2. L. M. Montierth, W. A. Neuman, D. W. Nigg, and B. J. Merrill, J. Appl. Phys. **69**, 6776 (1991).
3. V. Yu. Mat'ev, Zh. Tekh. Fiz. **73** (3), 28 (2003) [Tech. Phys. **48**, 303 (2003)].
4. V. Yu. Mat'ev, in *Proceedings of the 2nd Conference on Nuclear-Excited Plasma and Problems of Nuclear-Pumped Lasers, Arzamas-16, 1995*, Vol. 1, pp. 430–442.
5. V. Yu. Mat'ev, V. V. Borovkov, and S. P. Mel'nikov, Zh. Tekh. Fiz. **71** (1), 79 (2001) [Tech. Phys. **46**, 76 (2001)].
6. J. R. Torczynski, J. Fluid. Mech. **201**, 167 (1989).

Translated by A. Pankrat'ev

Technical Physics, Vol. 50, No. 6, 2005, pp. 698–704. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 6, 2005, pp. 25–31.
Original Russian Text Copyright © 2005 by Potanin.

GAS DISCHARGES, PLASMA

Effect of the Ion Distribution over Longitudinal Velocities on the Efficiency and Separating Parameters of the ICR Method of Isotope Separation

E. P. Potanin
Russian Research Centre Kurchatov Institute, pl. Kurchatova 1, Moscow, 123182 Russia
e-mail: potanin@imp.kiae.ru

Received April 15, 2004; in final form, September 28, 2004

Abstract—In the context of the ICR method of isotope separation, resonance RF heating of the ions in an electric field propagating along a constant magnetic field while simultaneously rotating in the direction perpendicular to it is calculated in a linear approximation. The analysis is carried out for two types of the initial ion distribution function over longitudinal velocities: a function proportional to the first power of the velocity in the range of low velocities and a shifted semi-Maxwellian distribution function. The distribution function of the ions over transverse velocities is calculated under the assumption that their initial distribution over transverse velocities is Maxwellian. The ion fluxes onto the collector plates are estimated by integrating the corresponding ion distribution functions over the allowed range of the longitudinal and transverse velocities and the transverse coordinates of the guiding centers of the ion trajectories in front of the extractor. In the first part of the paper, calculations are carried out for a model binary mixture of isotopes with mass numbers of 6 and 7. The effect of the shape of the ion distribution function over longitudinal velocities on the heating efficiency and on the concentration of the target isotope ions at the collector, as well as the effect of the longitudinal ion temperature on the width of resonance curves for the ion heating efficiency, is investigated. In the second part, a study is made of the selectivity of heating of isotope ions in a gadolinium plasma, in particular, the effect of the longitudinal magnetic field on the selectivity of heating of neighboring isotopes with atomic masses of 156, 157, and 158 is examined. *© 2005 Pleiades Publishing, Inc.*

INTRODUCTION

The isotopes of chemical elements that form no gaseous compounds under normal conditions are usually separated by the electromagnetic method [1], which is, however, of low productivity. Plasma methods based on systems with a traveling magnetic field or on plasma centrifuges [2, 3] might have a high level of productivity, but it seems inexpedient to use them for isotope production because of the comparatively low values of separation coefficients that are achievable in such devices and because of the fairly high energy consumption [4]. A plasma method based on selective ICR heating of the ions of the target isotope [5–11] makes it possible to achieve comparatively high separation coefficients and thus may hold promise for industrial applications.

CALCULATIONS OF THE ION HEATING EFFICIENCY. SEPARATING PARAMETERS OF THE COLLECTOR FOR SEPARATION OF BINARY ISOTOPIC MIXTURES

The ICR method is based on an isotopically selective ICR heating of the ions of the target isotope, followed by the extraction of this isotope from the plasma flow under conditions such that the collisions between particles are negligible. The schematic diagram of a version of the separator that is most suitable for industrial applications is illustrated in Fig. 1. A steady plasma flow (P) from the plasma source (the design of different types of plasma sources was described in detail in [8]) passes through the region where a solenoid (*1*) creates a uniform magnetic field and where an RF antenna (*2*) producing the heating field is located. A system for collecting the target material is arranged behind the heating region. The ion-collecting system consists of equidistant collector plates (*3*) and a waste plate (*5*). The collector plates, which are parallel to the

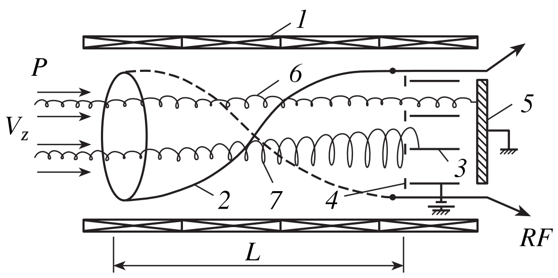


Fig. 1. Schematic diagram of the device.

1063-7842/05/5006-0698$26.00 © 2005 Pleiades Publishing, Inc.

plasma flow and are protected by front screens (*4*), are intended for extracting the heated ions (with a large gyroradius) of the target isotope (*7*). The waste plate, which is perpendicular to the plasma flow, is aimed at depositing the “cold” (nontarget) ions (*6*). In order to increase the degree of separation, the collecting plates can be held at a positive repulsing potential U, which substantially reduces the nontarget ion flux. An ICR separator is capable of yielding a considerably larger amount of the target material (in comparison to industrial electromagnetic separators based on ion beams), because the plasma (and, accordingly, the flow of the isotopes to be processed) is free of fundamental restrictions associated with the positive space charge of an ion beam.

When the plasma passes through the ICR heating region, the ion distribution over transverse velocities changes appreciably. Knowledge of the distribution function of the ions in a plasma flow that has passed through the heating region is important because, with this information, it is possible not only to estimate the separating parameters of the collector system but also to determine the selective properties of the RF system. In some ICR devices (see, e.g., [10]), use is made of inductive antennas in the form of twisted multiphase cylindrical helices—the so-called helical (or spiral) antennas. For a four-phase antenna in which the ac current in each subsequent phase is shifted by $\pi/2$, the alternating electric field within the cylindrical heating region can be represented as a circularly polarized wave that travels in the positive direction along the z axis and whose polarization vector rotates with an angular velocity ω,

$$E_x = E\cos(\omega t - Kz - \varphi), \tag{1}$$

$$E_y = -E\sin(\omega t - Kz - \varphi), \tag{2}$$

where $K = 2\pi/\lambda$ is the wavenumber, λ is the spatial period of the electric field, E is the field oscillation amplitude, and φ is the initial phase.

In this approximation (which can be regarded as linear), the alternating components of the magnetic field of the wave are ignored. As the ions are heated, their transverse temperature increases and gives rise to an alternating electrostatic field, which reduces the heating field produced in the plasma by the antenna. In [12], it was shown that the maximum density $n_{i\,\max}$ of the target ions during their heating can be estimated from the relationship

$$n_{i\,\max} \approx \frac{m_i c^2 K^3 V_T}{4\pi e^2 \omega},$$

where m_i is the mass of an ion, c is the speed of light in vacuum, and V_T is the ion thermal velocity.

Setting $m_i = 10^{-23}$ g, $K = 10^{-1}$ cm^{-1}, $\omega = 4 \times 10^6$ s^{-1} (which corresponds to the ion gyrofrequency at a magnetic induction of $B = 0.25$ T), and $V_T = 1.6 \times 10^6$ cm/s, we obtain $n_{i\,\max} \cong 10^{12}$ cm^{-3}. If the ion density is lower than the critical density, then the inverse effect of the motion of plasma particles on the vacuum antenna field can be ignored (the case of relatively low plasma densities) and representations (1) and (2) correctly describe the heating electric field in the plasma at radii of $r \le R/2$, where R is the antenna radius [12]. For the plasma density $n = 10^{12}$ cm^{-3}, a mean ion energy on the order of $\theta = 10$ eV, and the relative ion atomic mass $A = 6$, the mean time between Coulomb collisions of ions, estimated from the relationship [8]

$$\tau_{ii} \approx 2 \times 10^6 \frac{\theta^{3/2}\sqrt{A}}{n},$$

is equal to $\tau_{ii} \cong 1.7 \times 10^{-4}$ s, which is much longer than the time of flight of ions through the heating region in the Sirena device [10] ($\tau_f \cong 4 \times 10^{-5}$ s). During their flight time, the mean energy of the ions increases to about 40 eV. For a binary isotopic mixture, the concentration of the target ions of 8%, and the cross-sectional area of the plasma column of 0.03 m^2, the RF power that goes into the heating of target ions with the density $n_t = 8 \times 10^{10}$ cm^{-3} in a longitudinal ion flow with the mean velocity $V_z = 2 \times 10^4$ m/s is on the order of 10^2 W.

Let us consider ICR heating of a longitudinal ion flow with an initial Maxwellian distribution over transverse velocities and with the same longitudinal velocities V_z (in what follows, such a flow will be referred to as an ion beam) in an electric field given by relationships (1) and (2), i.e., in a field rotating in the azimuthal direction and traveling along the z axis. In this case, the ion distribution function over transverse velocities $V_\perp$ at the exit from the ICR heating region is described by the expression [13]

$$f_\perp(V_\perp) = \left(\frac{m}{2\pi k T_{\perp 0}}\right) \times \exp\left(-\frac{m}{2kT_{\perp 0}}(V_\perp^2 + V_0^2)\right) I_0\left(\frac{mV_0 V}{kT_{\perp 0}}\right), \tag{3}$$

where m is the mass of an ion, k is Boltzmann’s constant, $T_{\perp 0}$ is the initial transverse ion temperature, and I_0 is the zero-order modified Bessel function.

For a heating regime in which the transverse energy of the ions is much higher than their initial thermal energy, the quantity V_0 is given by the formula

$$V_0 \cong \frac{2E\omega_0}{B(\omega - KV_z - \omega_0)} \sin\left((\omega - KV_z - \omega_0)\frac{L}{2V_z}\right), \tag{4}$$

where B is the induction of the main magnetic field, ω is the heating field frequency, ω_0 is the ion gyrofrequency, and L is the length of the heating region. Distribution function (3) is normalized to unity and depends on both the longitudinal velocity of the ion beam and the heating field frequency. When the initial longitudinal ion velocities are not the same and obey a distribution $f_z(V_z)$, the total ion distribution function can be represented as the product

$$f = nf_{\perp}(V_{\perp})f_z(V_z), \qquad (5)$$

where n is the ion number density. It should be noted that distribution function (5) is very non-Maxwellian.

In order to determine the ion fluxes onto the collector plate, we utilize a method that was mainly developed in [13–15] for calculating the separation of binary isotopic mixtures. This method consists essentially in finding the fluxes of all isotopic species by integrating the ion flux densities over the permissible ranges of values of the transverse and longitudinal ion velocities ($V_{\perp}$, V_z) and of the coordinates y_0 of the guiding centers of the ion trajectories in front of the collector plate with allowance for the known ion velocity distribution function $f(V_{\perp}, V_z)$. Unlike in [13], we will obtain the fluxes of the ion species onto the plates by integrating the corresponding distribution functions. In this way, we will not use the results of numerical integration over the coordinates of the guiding centers of the ion trajectories but will determine the ranges of the permissible values of these coordinates in advance [14]. According to [14], the total density of the transverse flux of any ion species onto the collector plate is equal to

$$j = \int f\omega_0 V_{\perp} dV_{\perp} dV_z dy_0. \qquad (6)$$

For a binary isotopic mixture, the density of the target (first) ion species,

$$C = \frac{n_1}{n_1 + n_2}$$

(where n_i are the number densities of the ion species), can be found from the relationship

$$\frac{C}{1-C} = \frac{C_0}{1-C_0}\frac{j_1 n_{20}}{j_2 n_{10}}\sqrt{\frac{m_2}{m_1}}, \qquad (7)$$

where C_0 is the initial concentration of the target ion species and n_{10} and n_{20} are the initial number densities of the first (target) and the second ion species.

In [14, 15], the separating parameters of the collector were calculated in terms of an approximate model under the following two assumptions: first, the ions of a flow that has passed through the heating region obey Maxwellian distribution functions over transverse velocities and, second, the transverse temperatures of both ion species, $T_{\perp 1}$ and $T_{\perp 2}$, are known. In the present paper, the fraction of the ions heated above a given energy and the concentration of the target isotope at the front edge of the collector plate are calculated from formulas (3)–(7) as functions of the heating field frequency under the assumption that the ions in the source obey several most likely types of distribution (which play the role of the initial ion distributions). This formulation of the problem makes it possible to reveal how the separation efficiency depends on the shape of the distribution function over longitudinal velocities, $f_z(V_z)$, for different relationships between the length L of the heating region and the wavelength λ of the current in the heating solenoid. For this purpose, it is expedient to introduce the parameter p, which is equal to the ratio of the wavelength λ to the length L of the heating region. Since the plasma flowing out of a plasma source is bounded on one side, it is nonequilibrium. Consequently, the ion distribution over longitudinal velocities should differ strongly from a conventional symmetric Maxwellian distribution. The distribution functions that most closely correspond to the actual distributions are asymmetric functions or shifted semi-Maxwellian functions.

We begin by investigating the heating efficiency in the case of a model non-Maxwellian distribution function that depends linearly on the longitudinal ion velocity in the low-velocity range,

$$f_z(V_z) = \frac{mV_z}{kT_z}\exp\left(-\frac{mV_z^2}{2kT_z}\right), \qquad (8)$$

where T_z is the effective longitudinal ion temperature.

If we introduce the dimensionless parameters

$$y = \frac{V_{\perp}}{(2kT_{\perp 0}/m)^{1/2}}, \quad x = \frac{V_z}{(2kT_z/m)^{1/2}},$$

$$n_L = \left(\frac{2kT_z}{m}\right)^{1/2}\frac{2}{L\omega_0}, \quad n_E = \frac{2E}{B\left(\frac{2kT_{\perp 0}}{m}\right)^{1/2}},$$

$$y_{\Omega} = \frac{n_E \sin\left((\Omega - \pi n_L x/p - 1)\frac{1}{nx}\right)}{(\Omega - \pi n_L x/p - 1)}, \quad p = \frac{\lambda}{L}.$$

From formulas (5), (6), and (8) we can then also obtain an expression for the relative fraction η of the ions of an isotopic species that are heated to transverse energies above a given energy $W_{\min}\left(y_1 = \sqrt{\frac{W_{\min}}{kT_{0\perp}}}\right)$:

$$\eta = 4\int_0^{\infty} dx \int_{y_1}^{\infty} xy\exp(-(y^2 + y_{\Omega}^2 + x^2))I_0(2yy_{\Omega})dy. \qquad (9)$$

Let us first consider a model isotopic mixture of ions with mass numbers of 6 and 7. Figure 2 illustrates the results of calculating the quantity η (which can be called the heating efficiency) for an isotope with a relative atomic mass of 6, the remaining parameters being the following: $W_{\min} = 40$ eV; the amplitude of the heating electric field produced by the antenna is E = 50 V/m; the longitudinal ion temperatures are T_z = 5, 10, and 20 eV; the length of the heating region is L = 0.8 m; the induction of the main magnetic field is B =

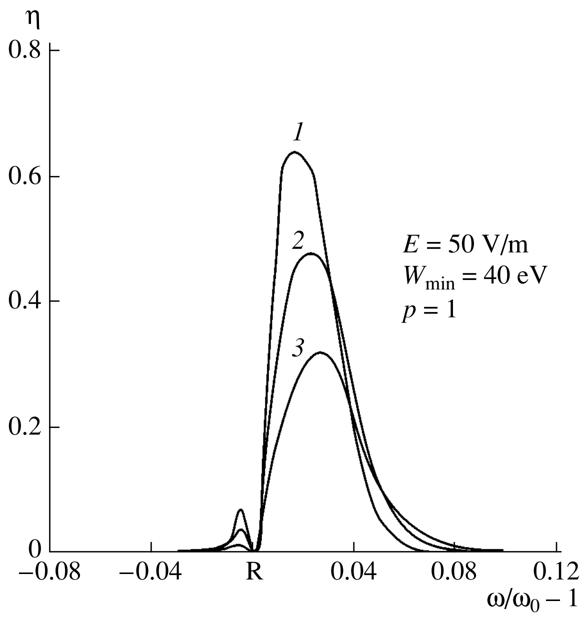


Fig. 2. Ion heating efficiency η vs. heating field frequency for distribution function (8), for $p = 1$, and for different longitudinal ion temperatures: T_z = (*1*) 5, (*2*) 10, and (*3*) 20 eV.

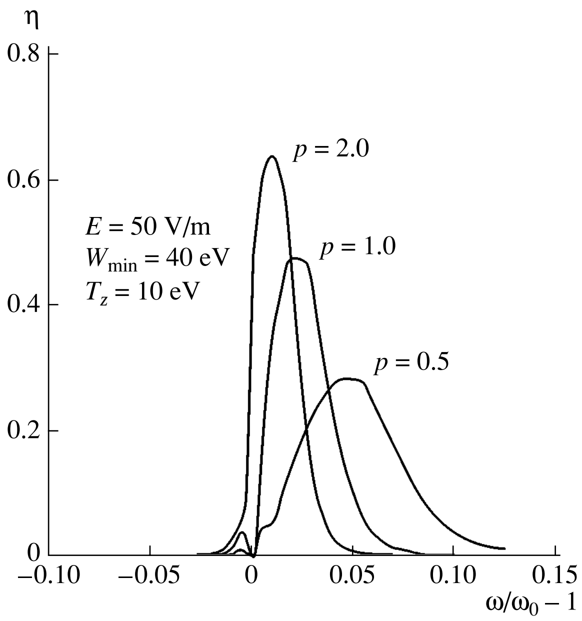


Fig. 3. Ion heating efficiency η vs. heating field frequency for distribution function (8), for T_z = 10 eV, and for different values of the parameter *p*.

0.25 T; and $p = 1.0$. Note that the heating efficiency η is seen to depend on the longitudinal ion temperature. Analogous dependences are shown in Fig. 3, which illustrates the results calculated for T_z = 10 eV, for a fixed length *L* of the heating region, and for different values of the parameter *p*. It should be noted that, in the range of *p* values under analysis, the maximum heating efficiency decreases markedly with decreasing wavelength λ. The numerical calculations were carried out under the assumption that the electric field is unchanged. It should be taken into account, however, that the electric field changes with increasing wavelength when the antenna current is maintained constant. The fact that the width of the resonance curves increases with decreasing wavelength is attributed to the Doppler broadening.

Relationship (7) implies that, in order to find the concentration of the target isotope ions with a number density *n* at the front edge of the collector plate, it is necessary to determine the density of the corresponding ion flux:

$$j = \frac{4n}{\pi}\left(\frac{2kT_{\perp 0}}{m}\right)^{1/2} \times \int_0^\infty dx \int_0^\infty xy^2 \exp(-(y^2 + y_\Omega^2 + x^2)) I_0(2yy_\Omega)\,dy. \tag{10}$$

The dependence of the concentration of an isotope of mass number 6 at the front edge of the collector plate on the dimensionless frequency of the electric field is shown in Fig. 4, which illustrates the results of calculations for the initial density $C_0 = 0.08$, for a one-wavelength antenna ($p = 1$), and for different longitudinal ion temperatures. The width of the resonance curves is seen to increase slightly with increasing longitudinal temperature T_z. It is of interest to compare these results

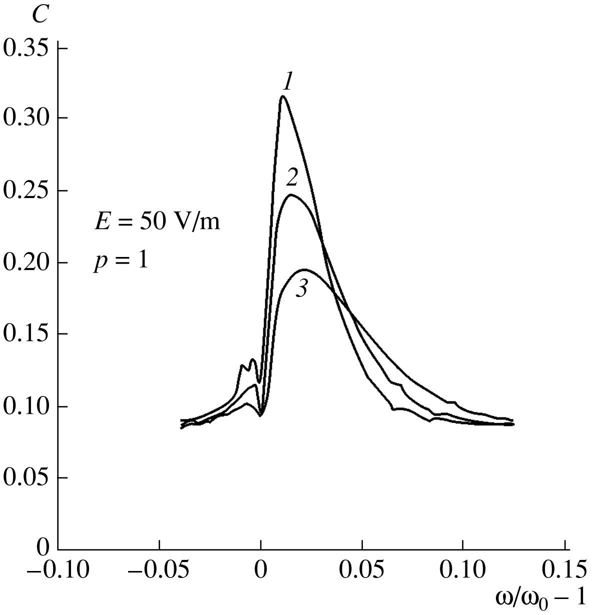


Fig. 4. Concentration of an isotope of mass number 6 at the front edge of the collector plate vs. heating field frequency for distribution function (8), for $C_0 = 0.08$, and for different longitudinal ion temperatures: T_z = (*1*) 5, (*2*) 10, and (*3*) 20 eV.

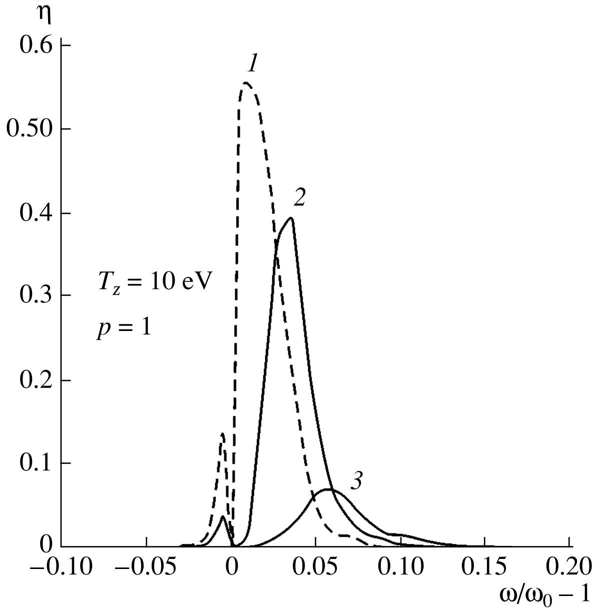


Fig. 5. Heating efficiency η vs. heating field frequency for $W_{\min}$ = 40 eV, E = 50 V/m, and for different values of the shift parameter in distribution (11): V_{0z} = (*1*) 0 (a zero velocity shift), (*2*) 10^4, and (*3*) 2×10^4 m/s.

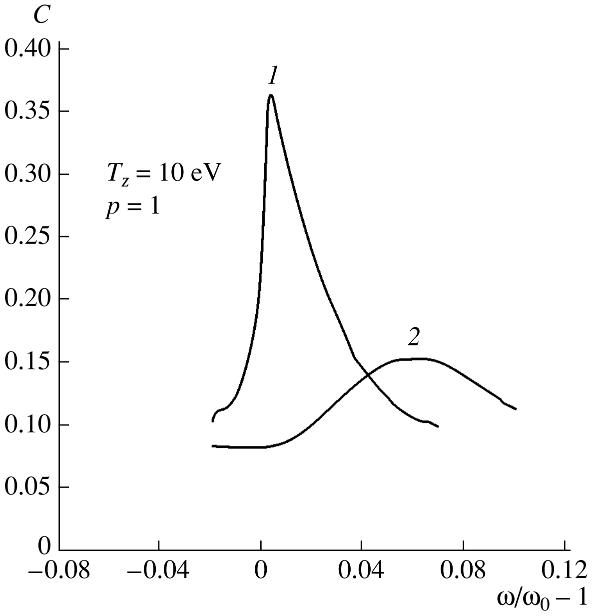


Fig. 6. Concentration of an isotope of mass number 6 vs. heating field frequency for E = 50 V/m and for two different values of the shift parameter: V_{0z} = (*1*) 0 and (*2*) 2×10^4 m/s.

with those in the case of a shifted semi-Maxwellian distribution function over longitudinal velocities,

$$f_z(V_z) = 2\left(\frac{m}{2\pi kT_z}\right)^{1/2}\exp\left(-\frac{m(V_z - V_{0z})^2}{2kT_z}\right), \quad V_z \ge V_{0z}$$
$$\text{and } f_z(V_z) = 0 \quad \text{at} \quad V_z < V_{0z}, \tag{11}$$

where V_{0z} is the velocity shift parameter.

In this case, the expression for the quantity η takes the form

$$\eta = \frac{4}{\sqrt{\pi}} \times \int_{x_0}^{\infty} dx \int_{y_1}^{\infty} y \exp(-(y^2 + y_\Omega^2 + (x - x_0)^2)) I_0(2yy_\Omega)dy, \tag{12}$$

where

$$x_0 = \frac{V_{0z}}{(2kT_z/m)^{1/2}}.$$

In Fig. 5, the solid curves represent the quantity η calculated from expression (12) for shift parameters $V_{0z} = 10^4$ and 2×10^4 m/s and the dashed curve was calculated for a zero velocity shift. As can be seen, an increase in the shift parameter by a factor of only 2 (from 10^4 to 2×10^4 m/s) leads to a substantial decrease in the efficiency of RF heating of an isotope of mass number 6. This is well confirmed by the results of calculating the separating parameters of the collector of heated particles. For a shifted semi-Maxwellian distribution function over longitudinal velocities, the ion flux density has the form

$$j = \frac{4n}{\pi^{3/2}}\left(\frac{2kT_{\perp 0}}{m}\right)^{1/2}\int_{x_0}^{\infty} dx \times \int_0^{\infty} y^2 \exp(-(y^2 + y_\Omega^2 + (x - x_0)^2)) I_0(2yy_\Omega)dy. \tag{13}$$

Figure 6 shows the density of the target isotope at the front edge of the collector plate for an unshifted (V_{0z} = 0) and a shifted ($V_{0z} = 2 \times 10^4$ m/s) distribution function. We see that, at high velocities of the plasma flow, not only does the resonance shifts toward higher frequencies (because of the Doppler effect), but also the ion separation effect becomes less pronounced. The latter result has a clear physical meaning: as the shift parameter increases at a fixed temperature T_z, the ion heating time decreases. The heating efficiencies calculated from two different ion distributions over longitudinal velocities, namely, distributions (8) and (11), for the same mean longitudinal velocity of the ion flow differ insignificantly (see Fig. 7).

It should be noted that, under the same heating conditions, the half-widths of the resonance curves (i.e., the full widths at half-maximum of the peaks) for the heating efficiency η and the density C of the target isotope are nearly the same.

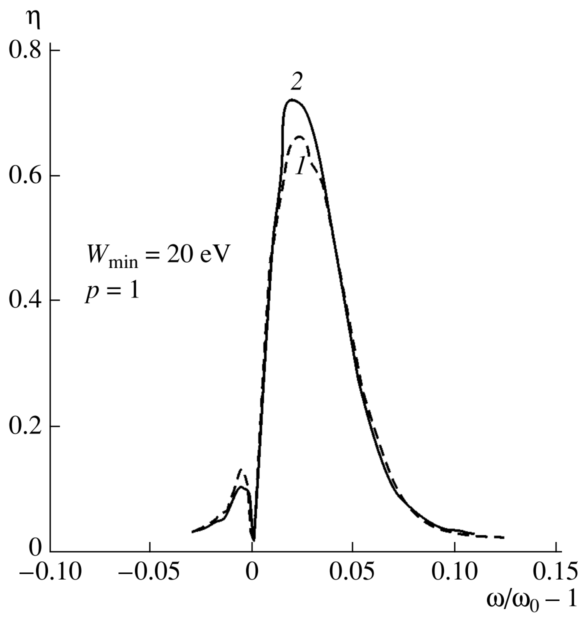


Fig. 7. Heating efficiency η vs. heating field frequency for two different ion distributions over longitudinal velocities (curve *1* for distribution (8), and curve *2* for distribution (11)), for the same mean longitudinal ion velocity $U = 6 \times 10^3$ m/s, and for $E = 50$ V/m and $T_z = 10$ eV.

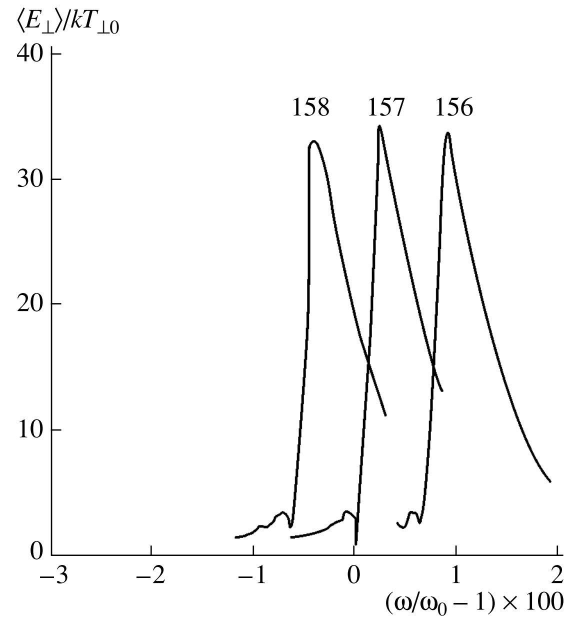


Fig. 8. Mean transverse energy of Gd isotope ions vs. heating field frequency for $B = 1$ T, $E = 15$ V/m, $L = 4$ m, $\lambda = 4$ m, $T_{\perp 0} = 5$ eV, and $T_z = 10$ eV.

MASS RESOLUTION IN THE CASE OF HEAVY ELEMENTS

The calculations described above were carried out for a binary isotopic mixture such that the mass difference between the isotopes to be separated is relatively large. In this case, the effective mass resolution can be provided by comparatively weak magnetic fields. Let us estimate the selectivity of the heating of a multispecies mixture of gadolinium isotopes. Note parenthetically that the first experimental attempt to separate gadolinium isotopes by the ICR method was made in France [16]. In the United States, preliminary results from experiments on the separation of the ^{157}Gd isotope in gram quantities have recently been reported in [17]. Figure 8 illustrates the results of calculation of how the mean energy of ions with atomic masses close to that of the ^{157}Gd isotope depends on the heating field frequency. The calculations were carried out for the following parameter values: $p = 1$, the amplitude of the transverse electric field in the plasma was $E = 15$ V/m, the induction of the main magnetic field was $B = 1$ T, the length of the heating region was $L = 4$ m, and the effective longitudinal ion temperature was $T_z = 10$ eV. The curves in Fig. 8 clearly shows that, for the given, comparatively weak, longitudinal magnetic field of $B = 1$ T, the resolution of the peaks of the resonance energy absorption is low (in terms of the Rayleigh criterion, which is known in optics, this situation is at the threshold of mass resolution: the peak in one curve occurs essentially at the same position as the minimum in the neighboring curve). The resonance is also observed to be Doppler-shifted toward higher frequencies (in a magnetic field with the induction $B = 3$ T, the gyrofrequency ω_0 of a ^{157}Gd isotope ion is about 1.8×10^6 s^{-1}); this effect is attributed to the chosen propagation direction of the heating wave and the chosen direction of the mean velocity of the plasma flow [10].

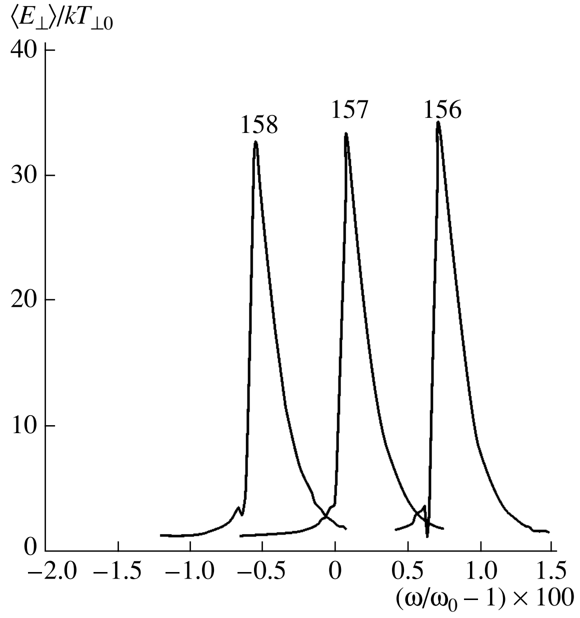


Fig. 9. The same as in Fig. 8, but for a magnetic induction of 3 T.

Figure 9 illustrates the results of similar calculations carried out for the magnetic induction $B = 3$ T. As is seen, an increase in the induction of the main magnetic field raises the ion heating selectivity: the differences between the gyrofrequencies of the ions of different isotopes increase while the half-widths of the resonance curves remain the same as in the previous case. The latter circumstance can play a decisive role when other (additional) line broadening mechanisms (such as plasma instabilities or collisions between ions) that cannot be taken into account in the model used here operate in the plasma.

CONCLUSIONS

(i) It is shown that, in the case of an ion distribution function over longitudinal velocities that is linear in the low-velocity range, the efficiency of heating of the target isotope ions depends substantially on their longitudinal temperature; for a constant heating electric field, it is found that, in the range of values of the parameter p from 2 to 0.5, the width of the resonance heating curve increases with decreasing wavelength; it is also found that the width of resonance curves for the heating efficiency and for the density of the target isotope ions depends on the longitudinal ion temperature only slightly.

(ii) It is shown that, in the case of a shifted semi-Maxwellian distribution function over longitudinal velocities, a twofold increase in the velocity shift parameter leads to a substantial decrease in the heating efficiency.

(iii) It is demonstrated that, in the case of a multispecies gadolinium plasma, an increase in the longitudinal magnetic field results in a corresponding increase in the heating selectivity.

ACKNOWLEDGMENTS

I am grateful to V.G. Pashkovskiĭ for his help in carrying out numerical simulations.

REFERENCES

1. V. G. Bondarenko and R. N. Kuz'min, in *Isotopes: Properties, Production, and Application*, Ed. by V. Yu. Baranov (IzdAT, Moscow, 2000), pp. 219–236 [in Russian].
2. A. I. Karchevskiĭ and E. P. Potanin, in *Isotopes: Properties, Production, and Application*, Ed. by V. Yu. Baranov (IzdAT, Moscow, 2000), pp. 254–267 [in Russian].
3. A. I. Karchevskiĭ, Yu. A. Muromkin, and E. P. Potanin, in *Isotopes: Properties, Production, and Application*, Ed. by V. Yu. Baranov (IzdAT, Moscow, 2000), pp. 268–285 [in Russian].
4. E. P. Potanin, Doctoral Dissertation (Moscow, 1997).
5. G. A. Avkar'yan, V. A. Namiot, and A. A. Rukhadze, Pis'ma Zh. Tekh. Fiz. **1**, 820 (1975) [Sov. Tech. Phys. Lett. **1**, 356 (1975)].
6. J. M. Dawson, H. C. Kim, D. Arnush, *et al.*, Phys. Rev. Lett. **37**, 1547 (1976).
7. M. Mussetto, T. E. Romesser, D. Dixon, *et al.*, in *Proceedings of the 10th IEEE International Conference on Plasma Science, San Diego, 1983*, p. 70.
8. Yu. A. Muromkin, Itogi Nauki Tekh., Ser.: Fiz. Plazmy **12**, 83 (1991).
9. A. C. La Fontaine and P. Louvet, in *Compte rendu des Journeess sur les Isotopes Stables, Saclay, France, 1993*, p. 332.
10. A. I. Karchevskiĭ, V. S. Laz'ko, Yu. A. Muromkin, *et al.*, Fiz. Plazmy **19**, 411 (1993) [Plasma Phys. Rep. **19**, 214 (1993)]; Preprint No. 5239/7, IAE (Kurchatov Institute of Atomic Energy, Moscow, 1990).
11. A. I. Karchevskiĭ and Yu. A. Muromkin, in *Isotopes: Properties, Production, and Application*, Ed. by V. Yu. Baranov (IzdAT, Moscow, 2000), pp. 237–253 [in Russian].
12. D. M. Berger, W. A. Newcomb, J. M. Dawson, *et al.*, in *Proceedings of the 2nd International Conference on Peaceful Application of Atomic Energy: Invited Papers on Hot-Plasma Physics and Thermonuclear Reactions, Geneva, 1958* (Atomizdat, Moscow, 1959), pp. 181–192.
13. A. L. Ustinov, Preprint No. 5354/6, IAE (Kurchatov Institute of Atomic Energy, Moscow, 1991).
14. A. I. Karchevskiĭ and E. P. Potanin, Fiz. Plazmy **21**, 416 (1995) [Plasma Phys. Rep. **21**, 394 (1995)].
15. A. I. Karchevskiĭ and E. P. Potanin, Fiz. Plazmy **22**, 1146 (1996) [Plasma Phys. Rep. **22**, 1044 (1996)].
16. A. Pailloux, A. Compant La Fontaine, and P. Louvet, in *Proceedings of the 6th Workshop on Separation Phenomena in Liquids and Gases, SPLG-98, Nagoya, 1998*, p. 332.
17. M. L. Grossbeck, J.-P. Renier, and T. Bigelow, in *Final Report of NERI Project*, No. 99-0074 (2003).

Translated by O. Khadin

Technical Physics, Vol. 50, No. 6, 2005, pp. 705–710. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 6, 2005, pp. 32–37.

GAS DISCHARGES, PLASMA

Experimental Study and Computer Simulations of the Implantation of Laser Plasma Ions in Pulsed Electric Fields

V. N. Nevolin, V. Yu. Fominskiĭ, A. G. Gnedovets, and V. E. Koshmanov

Moscow Engineering Physics Institute, Kashirskoe sh. 31, Moscow, 115409 Russia

e-mail: kash@retech.ru

Received September 20, 2004

Abstract—Results are presented from experimental studies of pulsed plasma flows generated by nanosecond laser pulses with an intensity of 7×10^8 W/cm^2 from a solid-state target in a strong electric field. The current pulses through the laser target and the depth distributions of the iron ions implanted in a silicon substrate to which a negative high-voltage pulse was applied are measured. The physical processes occurring in laser plasma with an initial iron ion density of 6×10^{10} cm^{-3} are simulated numerically by the particle-in-cell method for different delay times and different shapes of the accelerating high-voltage pulse. The model developed allows one to calculate the ion flows onto the processed substrate, the electron flows onto the target, and the energy spectra of the implanted ions. The results from computer simulations are found to be in good agreement the experimental data. © *2005 Pleiades Publishing, Inc.*

INTRODUCTION

Implantation of high-energy (up to several tens of kiloelectronvolts) ions has been utilized over the past few decades in the microelectronics technology [1] and for modifying the surface properties of structural materials [2]. The important role played by ion implantation in modern high technology has stimulated the search for new simple and inexpensive methods for the implantation processing of materials by high-energy ion beams. The employment of a pulsed laser plasma as an ion source opens new possibilities for the ion-implantation processing of materials and the ion-assisted deposition of coatings with the use of a relatively simple scheme of ion acceleration [3–6].

The use of intense laser pulses ensures the congruent evaporation of the target material and the high degree of its ionization. These properties of laser radiation were utilized in [6] in designing an ion-beam implanter with a laser-plasma ion source. To obtain a large-diameter ion beam, the plasma boundary in that source was fixed with the help of a fine-mesh metal grid. It was noted in [6] that, along with obvious advantages, such an ion source has some drawbacks, which, however, can be overcome by using another scheme of ion beam formation. As the plasma flows from the target to the substrate, a negative high-voltage pulse is applied to the latter; this leads to the acceleration of ions and their subsequent implantation in the substrate. The choice of the target (or the set of several different targets) unambiguously determines the composition of the ion beam. The main controllable parameters of the laser ion-implantation process are the laser intensity, the parameters of the high-voltage pulse (including the time at which it is applied to the substrate), and the geometry and mutual arrangement of the target and substrate. Variations in these parameters can substantially affect the density and temperature of the laser plasma, the configuration and strength of the electric field between the target and substrate, and the energy and dose of the ion beam.

The scheme of ion beam formation with an unstabilized plasma boundary [3, 5] is fairly simple from the technical standpoint. It allows one to deposit ion-assisted coatings and is not subject to the so-called "linear-beam" effect. As is case of pulsed ion implantation from a gas-discharge plasma [7], the application of a negative high-voltage pulse to a body embedded in plasma makes it possible to implant ions in rather complicated surfaces.

In many respects, the innovative potential of the new method is determined by the possibility of performing relevant computer simulations, which substantially simplify the optimization of the technological process. To develop an adequate numerical model, it is necessary to thoroughly investigate the expansion of a pulsed laser plasma in a strong electric field. Today, the physics of this process still remains poorly studied.

The goal of the present paper was to experimentally and theoretically examine the dynamics of a pulsed laser plasma and the implantation of ions under the action of a strong external electric field, as well as to clarify the physics of the processes determining the formation of an ion beam from the plasma cloud for different delay times and different shapes of the accelerating high-voltage pulse.

1. PROCEDURE OF ION IMPLANTATION FROM A LASER PLUME

Figure 1 shows a schematic of the experimental setup for pulsed ion implantation from a laser plume. A focused laser pulse with a wavelength of 1.06 μm and duration of 10 ns was incident on an iron (99.99%) target at an angle of 45°. The target was mounted in a steel vacuum chamber, which was evacuated by a turbomolecular pump down to a pressure of 10^{-4} Pa. A distance of 8 cm from the target, a silicon substrate was installed. The substrate was connected to the generator of negative high-voltage pulses. The repetition rate of the laser and high-voltage pulses was 10 Hz. The rise time of the triangular high-voltage pulses was 1 μs and the decay time was longer than 10 μs. The amplitude of the voltage pulses was up to 40 kV.

The target was installed on a vacuum input connected to a computer recording the current pulses generated in the laser plasma when applying the high-voltage pulses. The high-voltage pulses were recorded with the help of an antenna connected to the computer. To investigate the free expansion of the laser plasma into a vacuum, the parameters of the ion component were measured by a grid detector [4]. The detector was set a distance of 8 cm from the laser target, at the place where the processed substrate was then installed.

After ion implantation, the silicon plates were chemically treated to remove the deposited Fe film and then were analyzed by the method of backward Rutherford scattering (BRS) of helium ions. The energy of ions in the probing beam was 1.5 MeV, and the scattering angle was 105°. Note that the objects under examination are of interest from the standpoint of the formation of iron-silicide electron structures [8]. These objects also allow one to apply the highly informative BRS method to measuring the depth distribution of the implanted ions. To measure the depth distribution of iron ions in silicon, the BRS spectra were processed with the help of the RUMP code. It should be noted that the dose of the implanted ions did not exceed 10^{15} cm^{-2}; therefore, the effect of ion sputtering on the formation of the ion depth distribution was insignificant.

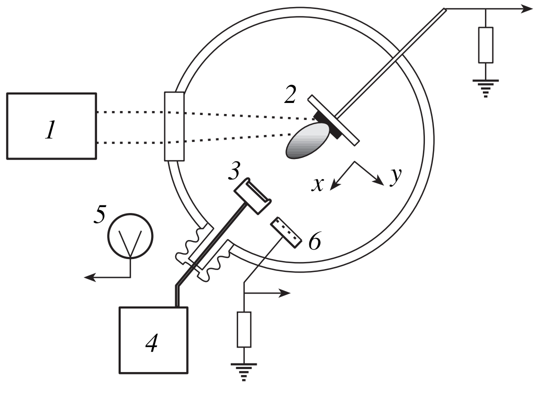


Fig. 1. Schematic of the experimental setup for ion implantation from a pulsed laser plume: (*1*) laser, (*2*) laser target, (*3*) substrate for ion implantation, (*4*) generator of negative high-voltage pulses, (*5*) antenna, and (*6*) ion detector.

2. EXPERIMENTAL STUDY OF THE LASER PLASMA AND ION-IMPLANTED LAYERS

Experiments with a laser plasma in an external electric field showed that stable regimes of ion implantation could be achieved for moderate-intensity laser pulses with an energy of ~20 mJ. In this case, high-voltage electrical breakdowns could be avoided and the plasma load insignificantly affected the output parameters of the high-voltage source. All the experiments were carried out at a fixed intensity of laser radiation on the target (about 7×10^8 W/cm^2).

Grid-detector measurements of the free expansion of the laser plasma into a vacuum showed that the maximum ion energy was no higher than 400 eV. The peak of the ion energy distribution function was at 150–200 eV. According to these measurements, the front of the ion component traveled a distance of about 2 cm from the target over 1 μs. The maximum ion density in a 2-cm-wide plasma layer was about 6×10^{10} cm^{-3}. The measurements of the angular ion velocity spectrum showed that, under our experimental conditions, the plasma was ejected in the form of a relatively narrow plume. The amplitude of the ion signal decreased appreciably as the detector was deflected by an angle of 30° with respect to the plume axis, which was normal to the target surface. Therefore, in developing the numerical model and performing computer simulations, it was assumed that, 1 μs after the laser pulse, all the plasma parameters changed only slightly in the direction perpendicular to the axis of the laser plume.

The parameters of the electric pulses measured on the laser-irradiated target depended substantially on the time at which the high-voltage pulses were applied. When the accelerating voltage was applied 1 μs after the laser pulse (Fig. 2), the target signal increased during 1 μs and then gradually decreased during 6–7 μs. Within this time interval, the accelerating voltage decreased from 40 to 20 kV. The increase in the delay time resulted in a decrease in the duration of the target signal. Thus, at a delay time of 3 μs, the pulse duration decreased to 2.5 μs and the signal amplitude increased by almost twice, whereas the accelerating voltage decreased by only 10 kV. Evidently, in this case, one can expect the formation of the narrowest energy distribution of the implanted ions. A further increase in the delay time resulted in a gradual decrease in the ampli-

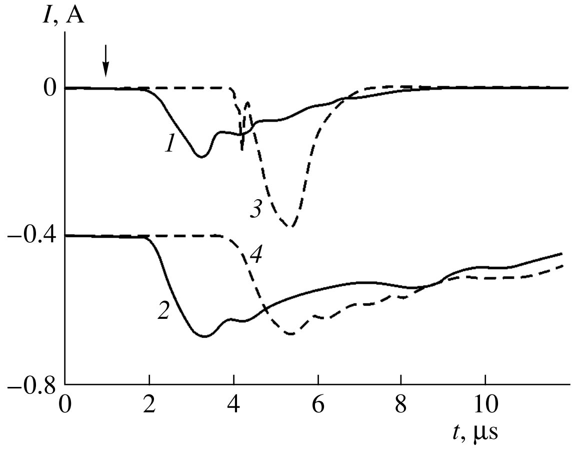


Fig. 2. Measured waveforms of the current pulses through the laser target (*1, 3*) and the detected signals from the antenna recording the high-voltage pulses (*2, 4*). The high-voltage pulse is applied to the substrate (*1, 2*) 1 and (*3, 4*) 3 μs after the laser pulse (marked by the arrow).

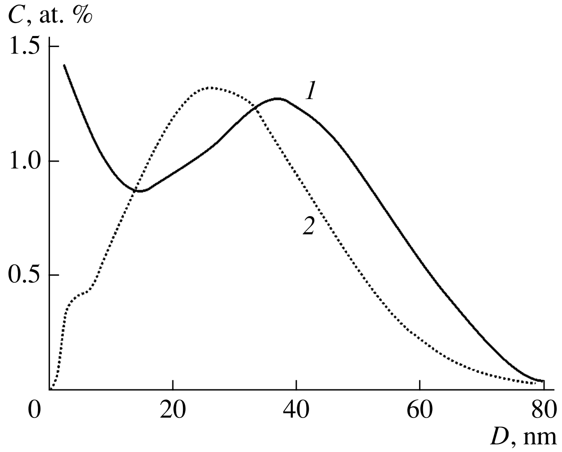


Fig. 3. Depth distribution of the iron ions implanted in silicon from a pulsed laser plasma under the action of a high-voltage pulse applied to the substrate with a delay time of 3 μs: (*1*) the measured depth distribution and (*2*) the depth distribution calculated using the ion energy distribution determined from the mathematical model.

tude of the target signal. The accompanying variations in the signal duration were less pronounced.

BRS measurements of the depth distribution of the implanted iron ions in silicon showed that the narrowest distribution (which corresponded to the minimum spread in the ion ranges) was achieved at a delay time of 3 μs. The iron ions had a characteristic bell-shaped depth distribution with a maximum at a depth of about 40 nm (Fig. 3, curve *1*). At other delay times, the distribution was broader and the total dose of the implanted ions was lower. The implantation depth coincided well with the calculated ranges of iron ions with energies of ~40 keV in amorphous silicon. A relatively large amount of implanted ions was observed in the surface layer of the substrate. This can be explained by the broadening of the energy distribution of the iron ions toward the low-energy range. This may also be attributed to the ion-induced intermixing of the iron atoms deposited from the laser plume with silicon. The reason for such a broadening and its influence on the shape of the ion depth distribution will be considered below.

3. MATHEMATICAL MODEL FOR SIMULATING THE DYNAMICS OF THE LASER PLASMA AND ION FLOWS

The expansion of the laser plasma into a vacuum in an external electric field was simulated numerically by the particle-in-cell (PIC) method [9]. The PIC method, which is known to be an efficient tool for solving plasma physics problems, has been successfully used to simulate various plasma technological processes, including ion implantation from gas-discharge plasmas [10].

Simulations of the physical processes in plasma by the PIC method consist in simultaneously solving the equations of motion for electrons and ions from a certain sample representing the total ensemble of the plasma particles and Poisson's equation for the potential of the self-consistent electric field calculated at the mesh points of the spatial grid.

The narrow angular width of the plasma flow after the end of the laser pulse (and especially after the high voltage is switched off) allows one to ignore the radial expansion of the plasma and to solve a one-dimensional problem.

The formation time of the laser plume (which is close to the duration of the laser pulse $\tau \sim 10$ ns) is much less than the time of plasma expansion. Therefore, starting from a certain instant $t_0 \gg \tau$, the propagation of the plume to the substrate can be considered independently of its formation. The initial distribution of the plasma density and velocity were specified using the results from the above measurements: over a time of 1 μs after the end of the laser pulse, the plasma expanded over a distance of ~2 cm from the target and its density acquired a bell-shaped profile.

The velocity distributions of the plasma particles at $t = t_0 = 1$ μs were assumed to be Maxwellian with the parameters N_j, T_j, and u corresponding to the densities and temperatures of the plasma components and the plasma flow velocity, respectively. The initial ion and electron densities $N_j(x, t = t_0)$ within the interval from 0 to 2 cm were approximated by parabolas with zero values at the ends of the interval and a peak value of $N_{jm} = 6 \times 10^{10}$ cm^{-3} at $x = 1$ cm. The mean plasma flow velocity was assumed to be $u = 10^6$ cm/s. For the ion and electron temperatures, we used values typical of laser experiments: $kT_j = 1$ eV.

As the boundary conditions for Poisson's equation, we used the known values of the potential at the

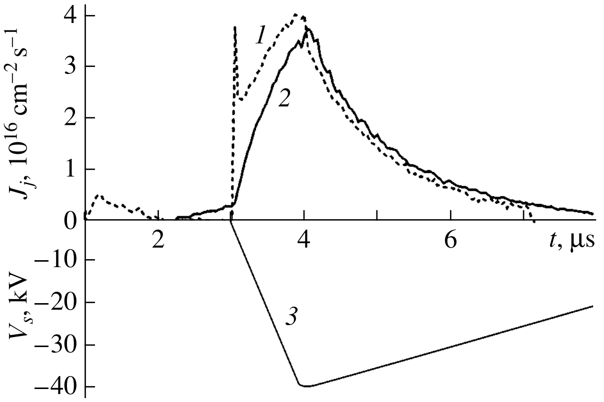


Fig. 4. Time evolution of the particle flux densities J_j from the laser plasma expanding from the target to the substrate: (*1*) electron flux onto the target and (*2*) ion flux onto the substrate. The high-voltage pulse V_S (*3*) is applied 3 μs after the laser pulse.

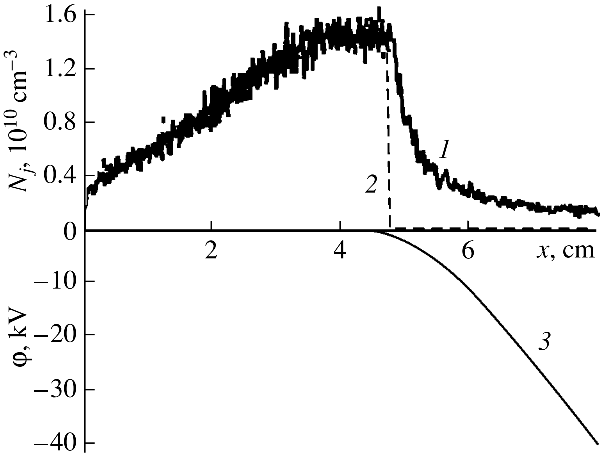


Fig. 5. Spatial distributions of the (*1*) ion and (*2*) electron densities and (*3*) the electric potential φ in the laser plasma expanding from the target ($x = 0$) to the substrate ($x = 8$ cm) at the instant $t = 4$ μs. The high-voltage pulse is applied 3 μs after the laser pulse.

grounded target $\varphi(x = 0, t) = 0$ and at the substrate $\varphi(x = x_s, t) = V_s(t)$, to which high-voltage pulses of different shapes were applied with a certain time delay τ_V with respect to the laser pulse.

When solving the kinetic problem numerically, the time step Δt and the cell size of the spatial grid Δx are determined by the Debye screening radius $r_D = (\varepsilon_0 kT_e/N_e e^2)^{1/2}$ and the plasma frequency $\omega_e = (N_e e^2/\varepsilon_0 m_e)^{1/2}$. For the solutions obtained by the PIC method to be stable, a number of conditions must be satisfied, in particular, $\Delta x < r_D$ and $\Delta t < 2\pi/\omega_e$. In our case, the plasma was modeled by 10^5 particles of both species (electrons and ions) and the time and spatial steps were 4×10^{-11} s and 2×10^{-3} cm, respectively, the total number of cells being 4×10^3.

Below, we present the results from numerical simulations of the expansion of a laser plume consisting of Fe$^+$ ions and electrons into a vacuum. The calculations were carried out for a triangular high-voltage pulse with a shape close to that of an actual high-voltage pulse: the voltage increased in absolute value to 40 kV over 1 μs and then decreased over 8 μs. The distance between the target and the substrate was set to 8 cm.

The computer simulations of the free expansion of the laser plasma showed that the ion component reached the substrate about 3 μs after plasma formation. Figure 4 shows the calculated electric signals from the target and substrate in the regime of ion implantation for a time delay of $\tau_V = 3$ μs between the laser pulse and the application of the high-voltage pulse. It should be noted that, after applying the accelerating voltage, the electron flux onto the target almost coincides with the ion flux onto the substrate. This is explained by the fundamental property of plasma to maintain quasineutrality in the volume occupied by it. After applying the high-voltage pulse, the ions at the leading edge of the plasma are drawn by the accelerating field onto the substrate. In contrast, the electrons are retarded by the electric field and cannot reach the substrate. To balance the escape of ions from the plasma (i.e., to maintain plasma quasineutrality), the excess electrons leave the plasma from its rear side and fall onto the target. As a result, the electric signals from the substrate and the targets are close to one another during the entire phase of plasma expansion in the external field (of course, these signals have different signs). Thus, the simulation results confirm the validity of monitoring the laser plasma in an external field by the target signals.

The characteristic structure of the laser plasma expanding in an external accelerating electric field is illustrated in Fig. 5. The simulation results show that, in the bulk of the plasma, the external field is almost completely screened and the plasma is quasineutral over almost the entire phase of its expansion from the target. It can be seen from Fig. 5 that, in a layer of width 4.5 cm, the ion density (curve *1*) coincides with the electron density (curve *2*). At the front of the expanding plasma, the charges are separated, and an ion layer forms between the plasma and the substrate. The ions arrive at the substrate from the front of the laser plume, where the electric field is very strong.

Figure 6 shows the calculated time behavior of the ion density in the laser plasma expanding in an external electric field. The ion current increases as the front of the laser plasma approaches the substrate (Fig. 4). Over this time period, the accelerating voltage passes its peak value of 40 kV and starts to decrease. After this, the plasma front begins to propagate backward (from the substrate to the target) and the region occupied by the space charge expands. At the same time, the plasma as a whole continues moving from the target toward the substrate.

Our simulations show that, at the same accelerating voltage, the energy spectrum and the time during which the ion beam acts on the substrate depend substantially on the time at which the high-voltage pulse is applied. The calculated time evolution of the ion flux onto the substrate for different delay times of the high-voltage pulse are shown in Fig. 7. It should be noted that, irrespective of the instant at which the high-voltage pulse is applied, the process of implantation terminates in a time of 7–8 μs. The increase in the delay time of the high-voltage pulse leads to a substantial decrease in the duration of the high-energy ion beam. At a delay time of about 4 μs, the calculated duration of the ion signal decreases to ≈3 μs and its amplitude increases nearly by twice as compared to the 1-μs delay time. This agrees well with the above experimental results. At large delay times, the signal of the accelerated ions appears after a relatively weak signal from the ions that reach the substrate before applying the high-voltage pulse. This signal is not observed experimentally because, within this time interval, the electrons arrive at the substrate together with the ions.

The model developed allows one to predict the energy spectrum of the implanted ions under various experimental conditions. The energy distributions of the implanted ions at different delay times of the high-voltage pulse are shown in Fig. 8. It can be seen that the increase in the delay time from 1 to 3 μs leads to a decrease in the maximum ion energy from 42 to 38 keV. The fraction of high-energy ions in this case increases. It should be noted that, under certain conditions, the energy of the beam ions can exceed the “nominal” energy determined by the amplitude of the applied high-voltage pulse. Such high-energy ions can arise for the following reasons: (i) due to the interaction with the electrons escaping from the plasma; (ii) due to a short-term jump in the plasma potential at the instant when the external field, efficiently acting on the highly mobile electron component, is applied to the substrate; and (iii) due to the additional acceleration of a fraction of ions in the positively charged ion layer under the action of repulsive Coulomb forces.

The above spectra were used to calculate the model depth distributions of iron ions in silicon (Fig. 3). The model depth distributions were found by calculating the depth distributions of fixed-energy ions with a step of 1 keV. For this purpose, the SRIM (Stopping and Range of Ions in Matter) computer code was utilized. The calculated fixed-energy depth distributions were then summed with allowance for the ion energy distribution (Fig. 8). To compare the measured and calculated depth distributions, they were normalized to their peak values. It can be seen from Fig. 3 that, at large depths, the calculated distribution differs somewhat from the measured one. The increase in the measured ion concentration at large depths may be attributed to two factors: First, doubly charged ions accelerated to energies exceeding 40 keV could be present in the laser plume. Second, the ions implanted normally to the substrate

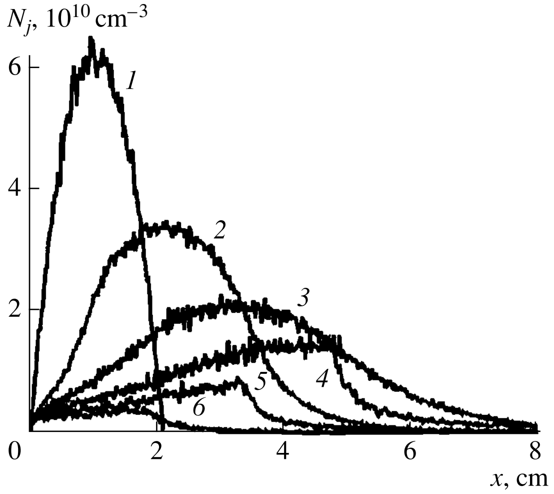


Fig. 6. Time evolution of the ion density in the laser plasma in an external electric field applied 3 μs after the laser pulse. Numerals by the curves show the time *t* in μs.

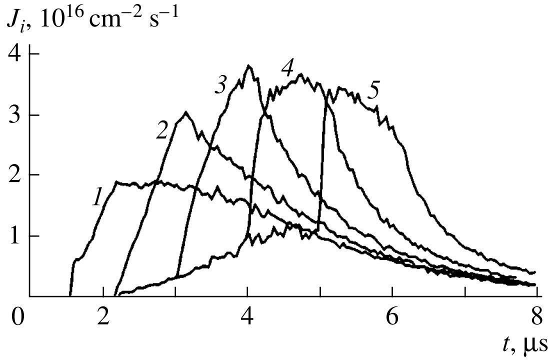


Fig. 7. Ion current signals for different delay times between the laser and high-voltage pulses. Numerals by the curves show the delay time τ_V in μs.

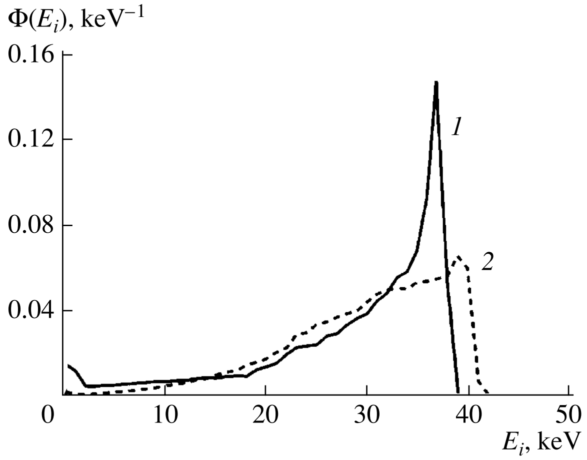


Fig. 8. Calculated energy distributions of the ions implanted in the substrate from the laser plasma for different delay times between the laser and high-voltage pulses: (*1*) 3 and (*2*) 1 μs.

surface could fall into the channels of the silicon crystal lattice. It is known [1] that, in this case, the implantation depth increases.

CONCLUSIONS

It has been found experimentally that the current characteristics of the laser plasma expanding from the target to the substrate depend substantially on the time at which the high-voltage accelerating pulse is applied. The amplitude of the target current increases appreciably as the delay time is increased from 1 to 3 μs. A further increase in the delay time leads to a gradual decrease in the target current. The duration of the current pulse, in this case, gradually decreased from 8 to 2.5 μs. When using triangular high-voltage pulses with an amplitude of 40 kV and a duration of 10 μs, variations in the current characteristics were accompanied by a change in the depth distribution and dose of the implanted ions.

Computer simulations have shown that the ion component of the plasma freely expanding into a vacuum approaches the substrate ~2 μs after the formation of the laser plume. When a negative high-voltage pulse is applied to the substrate, a positively charged ion layer forms at the leading edge of the plasma (between the plasma and the substrate) and an ion beam is generated in this region. The layer width, the spatial distributions of the ions and the electric potential in it, and the current characteristics of the ion beam depend on the dynamics of the expanding laser plasma and the time at which the external field is applied. The calculated parameters of the ion beam agree well with the measured ones.

Computer simulations have also shown that the shape of the electron current pulse to the target is similar to that of the ion current pulse to the substrate. This confirms the validity of monitoring the ion flux onto the substrate by the current signals from the laser target.

Calculations of the energy distributions of the implanted ions with the help of the model developed have shown that the maximum ion energy decreases from 42 to 38 keV as the delay time increases from 1 to 3 μs. This is accompanied by an increase in the fraction of high-energy ions. A comparison of the measured and calculated depth distributions of the implanted ions has shown that the model can be successfully used to predict the energy characteristics of the generated ion beams. This is of great importance for implementing the laser-plasma ion implantation technique in various technological processes.

ACKNOWLEDGMENTS

This work was supported in part by INTAS (grant no. 03-51-4206) and the RF Presidential Program “Support of Leading Scientific Schools” (grant no. NSh-97.2003.2).

REFERENCES

1. *Ion Implantation in Semiconductors and Other Materials: Collection of Articles* (Plenum, New York, 1973; Mir, Moscow, 1980).
2. Yu. A. Bykovskiĭ, V. N. Nevolin, and V. Yu. Fominskiĭ, *Ion and Laser-Assisted Implantation into Metallic Materials* (Énergoatomizdat, Moscow, 1991) [in Russian].
3. V. Yu. Fominski, V. N. Nevolin, R. I. Romanov, *et al.*, J. Appl. Phys. **89**, 1449 (2001).
4. V. Yu. Fominski, V. N. Nevolin, R. I. Romanov, *et al.*, Thin Solid Films **422**, 39 (2002).
5. B. Qi, R. M. Gilgenbach, M. C. Jones, *et al.*, J. Appl. Phys. **93**, 8876 (2003).
6. Yu. K. Al’tudov, Yu. A. Bykovskiĭ, and V. N. Nevolin, Zh. Tekh. Fiz. **50**, 178 (1980) [Sov. Phys. Tech. Phys. **25**, 105 (1980)].
7. J. R. Conrad, J. L. Radtke, R. A. Dodd, *et al.*, J. Appl. Phys. **62**, 4591 (1987).
8. C. Fetzer, I. Dézsi, A. Ventomme, *et al.*, J. Appl. Phys. **92**, 3688 (2002).
9. C. K. Birdsall and A. B. Langdon, *Plasma Physics via Computer Simulation* (IOP, Bristol, 1991).
10. P. K. Chu, S. Qin, C. Chan, *et al.*, Mater. Sci. Eng., R **17**, 207 (1996).

Translated by A. Sakharov

Technical Physics, Vol. 50, No. 6, 2005, pp. 711–714. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 6, 2005, pp. 38–41.
Original Russian Text Copyright © 2005 by Vasilyev, Ivanov-Omskii, Bogush.

Interstitial Carbon Showing up in the Absorption Spectra of Natural Diamonds

E. A. Vasilyev*, V. I. Ivanov-Omskii, and I. N. Bogush***
** CNIGRI Yakutian Research Geological Enterprise, ALROSA Ltd., Mirnyĭ, 678170 Russia*
e-mail: vasilyev@cnigri.alrosa-mir.ru
*** Ioffe Physicotechnical Institute, Russian Academy of Sciences,*
Politekhnicheskaya ul. 26, St. Petersburg, 194021 Russia

Received July 8, 2004

Abstract—Natural diamonds are studied by Fourier-transform IR (FTIR) spectroscopy, and it is shown that *B*2 centers in them form through intermediate stages, which are accompanied by the appearance of absorption bands with maxima near 1550 and 1526 cm^{-1}. The concentration of interstitial carbon atoms in the centers responsible for these bands may be several times higher than the concentration of the interstitials in *B*2 defects. *© 2005 Pleiades Publishing, Inc.*

INTRODUCTION

Nitrogen in diamonds is present largely as a substitutional impurity, causing the formation of defects *A*, *B*1, or *C*, which can be identified from the characteristic absorption spectra in the IR range. Here, *C* is a substitutional nitrogen atom, and *A* refers to two neighboring substitutional nitrogen atoms. Defect *B*1 is likely to consist of four substitutional atoms and a vacancy. Nitrogen also enters into defects *B*2 (complex objects that lie in {100} planes and range from several nanometers to several tens of micrometers in size), which are formed by carbon and, partly, by nitrogen interstitials, as was first supposed in [1]. However, the earlier model of defects *B*2 [2] assumes that they consist of only nitrogen atoms. In the IR range, these defects produce a *B*2 absorption band with a maximum in the range 1359–1380 cm^{-1} [2].

In synthetic diamonds, nitrogen is present in the form of defects *C*. High-temperature annealing causes the $C \longrightarrow A \longrightarrow B1$ transformations, during which nitrogen forms larger aggregates [3, 4]. The *A*-to-*B*1 transformation is accompanied by the appearance of carbon interstitials, which then generate defects *B*2 [1, 5]. Late in aggregation, *B*2 decomposes into dislocation loops and octahedral voids [6] containing molecular nitrogen [7]. In 1986, Woods [1] set off a group of “ordered” crystals, where the *B*2 concentration is proportional to the *B*1 concentration. The existence of “disordered” crystals was explained by the decomposition of defects *B*2. However, this fact may be given another explanation: interstitials do not necessarily enter into *B*2 centers, and there may exist other interstitial defects. In this case, the concentration of defects *B*1 should be proportional to the sum of interstitial carbon in the form of *B*2 and in the form of other as yet unidentified defects.

In the IR region, there are several absorption bands that are assigned to defects containing carbon interstitials. The band at 1570 cm^{-1} is attributed to a defect with two interstitials [8]. The bands near 1411, 1557, and 1903 cm^{-1}, as well as several repetitions (5*RL*), are identified as an interstitial combined with another intrinsic defect [8]. The band at 1530 cm^{-1}, arising upon irradiation and annealing, is due to vibration involving five carbon atoms [8]. An isolated carbon interstitial in diamond is mobile at temperatures below room temperature, which favors complexation [8]. In natural diamonds of the *IaAB* type, weak bands with maxima near 1524 and 1550 cm^{-1} (first detected by Sutherlend [9]) are often observed; however, their origin and correlation with the known centers have not been studied in detail. Supposedly, this is because their intensities are too low against the background of the vibrational–rotational absorption spectrum of water molecules in the range 1300–1900 cm^{-1}, which is much more intense than the noise. Detailed data for the bands with maxima near 1524 and 1550 cm^{-1}, which appear upon irradiation or subsequent annealing [8, 10], are not available in the literature. It was only speculated [5] that the frequencies of local vibrations that are attributed to complexes of interstitials lie in this range.

The purpose of this work is to reveal a correlation, if any, between the bands at 1526 and 1550 cm^{-1} (*E* and *F* bands for brevity), nitrogen defects *A* and *B*1, and interstitial defects *B*2.

RESULTS

We studied 38 natural octahedral diamonds 2–4 mm in size. The concentrations of impurity nitrogen in the form of defects *A* and *B*1 were calculated from the IR absorption spectra using the calibration values avail-

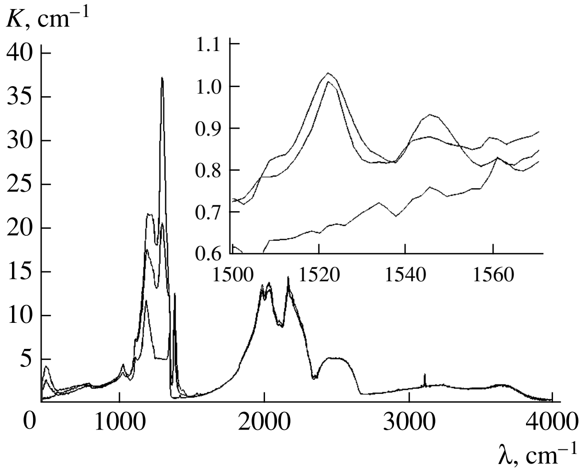


Fig. 1. Absorption spectra of the *IaAB*- and *IaB*-type diamonds. The inset shows fragments of the spectra in the range 1500–1560 cm^{-1}.

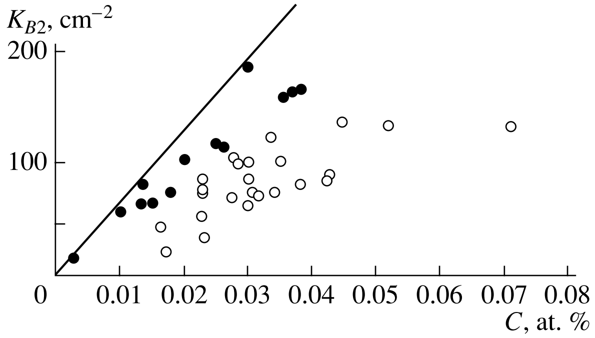


Fig. 2. Integral absorption coefficient K_{B2} vs. the concentration of nitrogen in the form of *B*1. The solid line is taken from [1]. The filled circles refer to ordered samples.

able from [1, 11, 12]. Our absorption spectra, some of which are shown in Fig. 1, were recorded with an FSM-1201 FTIP spectrometer in the range 400–4000 cm^{-1} with a resolution of 2 cm^{-1}. They were averaged over 32 scans. Centers *B*2 were characterized by integral absorption coefficient K_{B2} and the position of the band maximum. The *E* and *F* bands were approximated by the Lorentzian distribution so as to more exactly determine the position and height of the peak, as well as the integral absorption coefficient.

The maximum absorption coefficient was calculated to be 1.3 cm^{-1} for the *E* band and 1.6 cm^{-1} for the *F* band. In the Lorentzian approximation of the shapes of these bands, the wavenumbers of their maxima are 1524 ± 3 and 1548 ± 3 cm^{-1}; their half-widths, 11 ± 2 and 12.6 ± 4 cm^{-1}, respectively; and the maximal integral absorption coefficients are K_E = 29 cm^{-2} and K_F = 17 cm^{-2}. The absorption spectra of ten samples contain a weak line with a maximum near 1561 cm^{-1}, which sometimes has an inflection point in the shoulder of the *F* line. The absorption spectra for most of the crystals also contain a regular set of weak bands in the range 1420–1470 cm^{-1}, where the line at 1435 cm^{-1} also attributed to the *B*2 centers [1] stands out. Note that, in the absorption spectra of five *IaAB*-type samples, the *E*, *F*, and 1435 cm^{-1} bands are absent. In these samples, the maximum of the *B*2 band is shifted toward longer wavelengths (1362–1365 cm^{-1}) and the profiles of the bands are symmetric and describable by the Lorentzian curve. In the rest of the samples, the position of the maximum varies from 1366 to 1376 cm^{-1} and the band profiles are asymmetric with an extension toward shorter wavelengths.

We found that the *E* and *F* bands are absent in the *IIa*-type crystals, *IaA*-type crystals, and *IaB*-type crystals without *B*2. These bands occur only in the intermediate crystals of type *IaAB*, which agrees with published data [1] and indicates that they are related to defects *A* and *B*1. There are also crystals of type *IaAB*, whose spectra do not have these bands. If the *B*2 band in the spectra of the *IaB* crystals is absent, the bands under study are also lacking.

Figure 2 plots the integral absorption coefficient of the *B*2 band versus the concentration of nitrogen in the form of *B*1 defects. It is seen that four crystals may well be called ordered and nine more samples are close to this group. It also follows from Fig. 2 that the integral absorption coefficient of the *B*2 band may be several times smaller than the limiting value and that the concentration of *B*2 defects does not correspond to the concentration of the carbon interstitials emerging at the formation of *B*1 defects.

Figure 3 plots K_E and K_F versus K_{B2}, demonstrating an increase in absorption coefficients K_E and K_F with the concentration of *B*2 defects. The absence of direct proportionality indicates that the bands and defects are not related. However, the dependences of positions ν_0 of the maxima of the *E* and *F* bands on the ratios K_E/K_{B2} and K_F/K_{B2}, where K_{B2} is the absorption coefficient of the *B*2 band (Fig. 4), suggest a certain interplay between the three centers.

DISCUSSION

The results obtained indicate that, although the *E* and *F* bands are somewhat related to the *B*2 centers, they *E* and *F* centers have another nature. Two formation mechanisms for these two absorption bands may be suggested.

(1) The bands are caused by aggregates of interstitials, which are precursors to the *B*2 centers. In this case, the concentration of interstitials in the *E* and *F* centers, $C_{i(E)} + C_{i(F)}$, should be added to concentration $C_{i(B2)}$ of interstitials in the *B*2 defects so as to reach one-fourth of nitrogen concentration $C_{N(B1)}$ in the *B*1 centers.

(2) The absorption is caused by dislocations or voids formed as a result of the decomposition of the *B*2 cen-

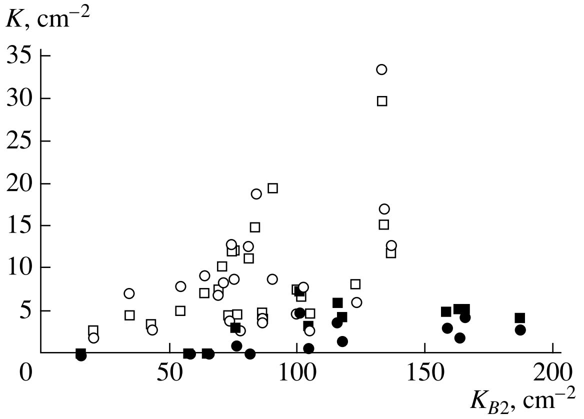


Fig. 3. Integral absorption coefficients for the (□) E and (○) F bands vs. the integral absorption coefficient for the $B2$ band. For the filled symbols, see Fig. 2.

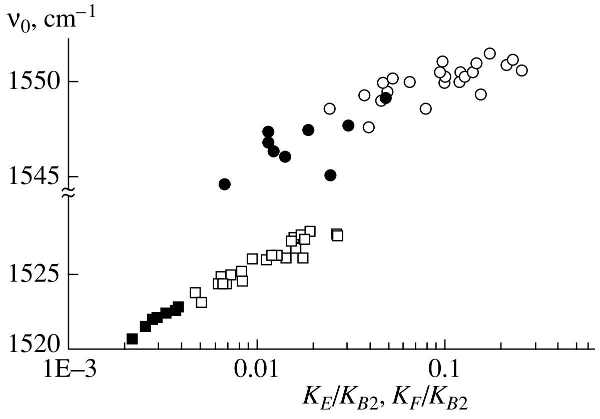


Fig. 4. Positions of the maxima of the (□) E and (○) F bands vs. the ratios of their integral absorption coefficients to the integral absorption coefficient of the $B2$ band. For the filled symbols, see Fig. 2.

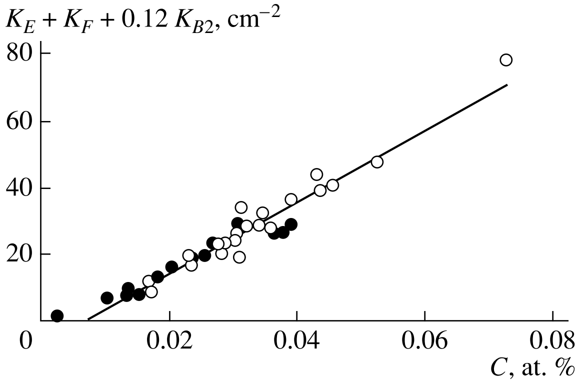


Fig. 5. Sum of integral absorption coefficients K_E, K_F, and K_{B2} vs. the concentration of nitrogen in $B1$ centers. The solid line is a linear approximation.

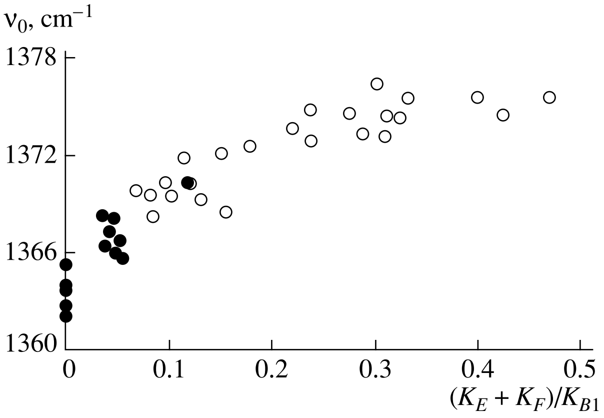


Fig. 6. Position of the $B2$ band maximum vs. the ratio of the sum of the absorption coefficients for the E and F bands to that for the $B1$ band.

ters. If the decomposition of the $B2$ centers is complete, the absorption coefficient in the bands must be maximum. However, as noted above, the E and F bands in the *IaB* crystals, where absorption coefficient K_{B2} is minimal, are absent.

Based on the assumption that all $B1$ defects arise in the grown crystal and that most carbon interstitials C_i enter into the defects $B2$, E, and F, one can estimate the C_i concentration in these defects. Using regression analysis, we derived an empiric relationship between the sum of the integral absorption coefficients of the K_E and K_F bands, integral absorption coefficient of the $B2$ band, and concentration of nitrogen in the $B1$ defects,

$$C_{\mathrm{N}(B1)} = 10.5(K_E + K_F + 0.12K_{B2}), \qquad (1)$$

which is shown in Fig. 5. The estimates are the following: C_{iB2}(at.%) = $3.6 \times 10^{-5}K_{B2}$ and $C_{i(E+F)} = 2.6 \times 10^{-4}(K_E + K_F)$.

The absorption cross section determined for the $B2$ centers agrees with the value found in [2]. Then, as applied to the diamonds studied, the maximal concentration of interstitials in the defects giving rise to the E and F absorption bands is estimated as $2.6 \times 10^{-4} \times 65 = 0.017$ at. % (Figs. 3, 4). In the $B2$ defects, the concentration of interstitials is $3.6 \times 10^{-5} \times 130 = 0.0046$ at. % at a concentration of nitrogen in the form of $B1$ defects of 0.073 at. %.

It is known that the greater the shift of the $B2$ band toward shorter wavelengths, the larger the size of the centers [2]. The size of the centers depends on the annealing temperature and time and, thus, is a measure of aggregation of interstitials rather than of nitrogen atoms. In other words, the larger the $B2$ centers are, the weaker the E and F bands are bound to be. The relationship between the ratio $(K_E + K_F)/K_{B1}$ and the position of the $B2$ band maximum (Fig. 6) validates this statement.

CONCLUSIONS

(1) The absence of direct proportionality between the concentrations of *B*1 and *B*2 defects may be due to the incomplete formation of *B*2 defects rather than their decomposition.

(2) The *B*2 defects form through intermediate stages, which are accompanied by the emergence of absorption bands with maxima near 1550 (*E*) and 1526 cm^{-1} (*F*).

(3) The *E* and *F* bands are attributed to defects other than *B*2 but also consisting of carbon interstitials. Changes in the positions of the maxima of the *E* and *F* bands indicate a complex composition of these defects and a certain interplay with *B*2 defects.

(4) The concentration of interstitials in the centers responsible for the *E* and *F* bands may be several times higher than the concentration of carbon interstitials in *B*2.

REFERENCES

1. G. S. Woods, Proc. R. Soc. London, Ser. A **407**, 219 (1986).
2. E. V. Sobolev, V. I. Lisoivan, and S. V. Lenskaya, Zh. Strukt. Khim. **9**, 1029 (1968).
3. Yu. A. Klyuev, A. M. Naletov, V. I. Nepsha, *et al.*, Zh. Fiz. Khim. **56**, 524 (1982).
4. T. Evans and Z. Qi, Proc. R. Soc. London, Ser. A **381**, 159 (1982).
5. J. P. Goss, B. J. Coomer, R. Jones, *et al.*, Phys. Rev. B **67**, 165208 (2003).
6. T. Evans, I. Kiflawi, W. Luyten, *et al.*, Proc. R. Soc. London, Ser. A **449**, 295 (1995).
7. I. Kiflawi and J. Bruley, Diamond Relat. Mater. **9**, 87 (2000).
8. A. T. Collins, Diamond Relat. Mater. **8**, 1455 (1999).
9. G. B. B. M. Sutherlend, D. E. Blackwell, and W. G. Simeral, Nature **174**, 901 (1954).
10. G. S. Woods, Philos. Mag. B **50**, 673 (1984).
11. S. R. Boyd, I. Kiflawi, and G. S. Woods, Philos. Mag. B **69**, 1149 (1994).
12. S. R. Boyd, I. Kiflawi, and G. S. Woods, Philos. Mag. B **72**, 351 (1995).

Translated by K. Shakhlevich

Technical Physics, Vol. 50, No. 6, 2005, pp. 715–718. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 6, 2005, pp. 42–45.
Original Russian Text Copyright © 2005 by Novodvorsky, Khramova, Wenzel, Bartha.

SOLID-STATE ELECTRONICS

Size Effects of Static Conduction in Thin Tantalum Films

O. A. Novodvorsky*, O. D. Khramova*, C. Wenzel, and J. W. Bartha****

** Institute of Problems of Laser and Information Technologies, Russian Academy of Sciences,*
Svyatoozerskaya ul. 1, Shatura, Moscow Oblast, 140700 Russia

*** Institute of Semiconductor and Microsystem Technology, Dresden University of Technology,*
Dresden, D-01062 Germany

e-mail: onov@laser.nictl.msk.su

Received August 23, 2004

Abstract—In thin tantalum films grown by pulsed laser deposition in a vacuum, the resistivity versus thickness dependence is found to oscillate. The oscillation periods equal 5.0 and 5.6 nm for α- and β-Ta films, respectively. This observation is interpreted as the quantum-size effect in thin tantalum films. *© 2005 Pleiades Publishing, Inc.*

INTRODUCTION

In contemporary microelectronics, the strong tendency toward decreasing the feature size and increasing the interconnection density generates the need to replace aluminum interconnections with copper interconnections and oxide and/or nitride insulators with those with a lower permittivity [1]. This has given impetus to investigating the parameters of thin-film copper conductors with thicknesses close to the mean free path of an electron in copper (39 nm). The aim of these investigations is to elucidate the feasibility of applying copper conductors the size of which is smaller than the mean free path of electrons in integrated circuits [2].

Because of a high mobility of copper atoms in metals and semiconductors, one faces the problem of preventing copper diffusion into silicon and silicon oxide. In other words, application of copper interconnections necessitates the formation of stable diffusion barriers entirely covering a copper conductor. To date, a number of methods of applying tantalum, a promising barrier material, have been proposed. It is known that tantalum has two crystalline phases: body-centered cubic (α-Ta) and tetragonal (β-Ta). The physical properties of α-Ta, such as a low resistivity ($\rho = 13\ \mu\Omega$ cm), high chemical stability, and low diffusion coefficient, make it a suitable material for coatings. The diffusion rate of copper atoms in the β-Ta phase ($\rho = 160–180\ \mu\Omega$ cm) is considerably higher [3].

Thick layers used as interconnections or for filling trenches are routinely grown with standard physical and chemical deposition methods. However, in growing thin layers with tailored properties (specifically those serving as diffusion barriers in ultrasmall devices), the method of laser plasma deposition acquires a special importance. Pulsed laser emission provides controllable mass transfer during evaporation (laser ablation) and, thereby, precision control of the layer thickness during deposition. A high energy of the particles being deposited makes it possible to grow both good crystalline films and amorphous films several nanometers thick. Moreover, the presence of high-energy particles in the laser erosion torch enables one to obtain high-quality films even at room temperature.

Tantalum films grown have either the α-phase, β-phase, or their combination [4]. In [5], techniques of tantalum thin film physical deposition were analyzed. In particular, it was established that the energy spectrum of ions plays an essential role during laser plasma deposition. On a thoroughly cleaned silicon surface, this method forms the α-phase of tantalum. However, under the conditions of a low torch ionization and low average energy of ions, the β-phase of tantalum grows. This demonstrates the promise of lasers for producing thin Ta films and also the significance of controlling the ion energy spectrum in a laser torch during deposition [6].

The size effects in thin metal films have been the subject of extensive research. Thin copper, silver, gold, and lead films grown on the silicon surface, as well as silver, indium, and gallium films grown on gold or silver sublayers, were examined [2]. The layer-by-layer growth of some of the films exhibited the oscillating resistance versus thickness dependence, which was interpreted as the quantum size effect [7] or periodic variation in film roughness [8].

Thus, to produce tantalum diffusion barriers for ultrathin copper lines, it is necessary to consider not only classical size effects but also quantum-size effects, since the latter are responsible for the oscillating thickness dependence of the conductivity in thin Ta films. In this study, we measured *in situ* the resistance of thin Ta films obtained by pulsed laser deposition as a function of their thickness (ranging between several nanometers and 50 nm).

1063-7842/05/5006-0715$26.00 © 2005 Pleiades Publishing, Inc.

EXPERIMENTAL

Thin Ta films were prepared by the method of pulsed laser deposition. This method makes it possible to precisely control the film thickness during growth by varying the number of laser pulses [9]. Owing to a rapid instantaneous growth rate (≈0.1 mm/s), a high vacuum is effectively maintained during the deposition of each layer, so that residual gas atoms have no time to be adsorbed on the film surface [10]. Surface contamination arising between the pulses is removed by high-energy (100 eV or higher) particles from the frontal part of the laser torch [11]. Because of a high nucleation rate and high nucleus density, this method makes it possible to grow ultrathin (several nanometers thick) continuous films. In addition, since the amount of the evaporating material per pulse is small, this method allows precision control of the film thickness by varying the number of laser pulses.

In the deposition experiments, we used a KrF excimer laser ($\lambda = 248$ nm) with a pulse energy of 100 mJ at pulse repetition rate $f = 10$ Hz. A quartz lens ($F = 20$ cm) focused the laser radiation into a 0.5-mm^2 spot on the target surface. The substrates used were (100) silicon plates 20 × 25 mm in size. The substrate was mounted parallel to the target at a distance of 25 mm in such a way that the center of the deposition area was exactly above the irradiation zone on the target. To determine the growth rate (and, accordingly, an increment of the film thickness) per pulse, we constructed (in each particular case) step calibration curves of the film thickness versus the number of pulses. The height of the steps was measured using a Tencor Alpha Step 500 profilometer. As follows from [12], measurements taken by the two- and four-point probe techniques coincide when the resistance of the sample exceeds 1 kΩ. We measured the conductance of the films *in situ* by the two-point probe technique. Prior to deposition, we applied tantalum or platinum contact pads on the substrate that were 5 mm distant from each other. A voltmeter was connected to the contacts by silver clamps. The test film was deposited through a 5 × 10-mm hole in a mask, the 10-mm side lying between the electrodes. Thus, the resistance of a 5 × 5-mm square on the tantalum film was measured *in situ* (during the deposition process). The measurements were taken under two deposition conditions. In the first case, the film was deposited directly on the substrate; in the second, on a 7.5-nm-thick niobium sublayer. It was shown [13] that the tantalum film deposited on a niobium sublayer (interface) thicker than 3 nm has a bcc lattice (α-Ta).

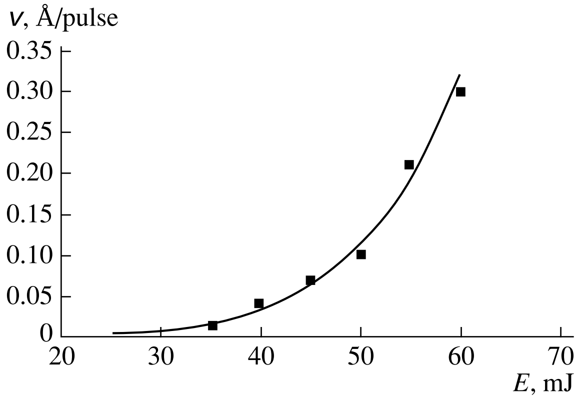


Fig. 1. Deposition rate υ of tantalum vs. laser pulse energy *E*.

RESULTS AND DISCUSSION

At laser ablation, the deposition rate depends on various parameters, in particular, on the peak pulse energy and the substrate–target distance. The deposition rates of niobium and tantalum were determined at peak pulse energies of up to 60 mJ for film thicknesses to 60 nm. A number of steps were deposited with the number of pulses per step varying between 500 and 4000. Under the measurement conditions used (50 mJ/pulse), the deposition rates of Nb, α-Ta, and β-Ta were found to be 0.015, 0.011, and 0.024 nm/pulse, respectively. Figure 1 plots the deposition rate of α-Ta against the pulse energy with the target irradiation geometry kept fixed. As was mentioned above, the surface area of the focusing spot was 0.5 nm^2. It can be seen that the dependence

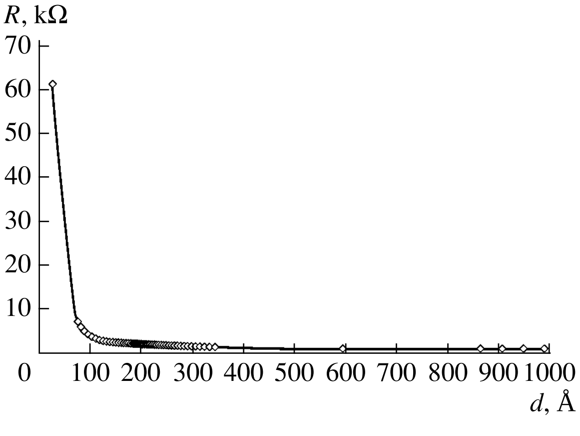


Fig. 2. Resistance *R* of the β-Ta film vs. its thickness *d*.

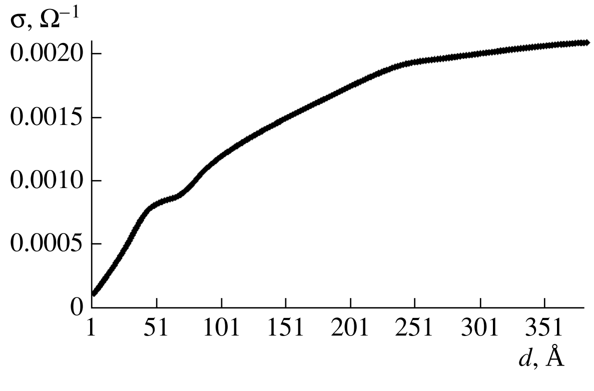


Fig. 3. Conductivity σ of the α-Ta film vs. its thickness *d*.

is nonlinear. When the pulse energy was less than 30 mJ, the above range of the number of pulses was insufficient for the film to grow.

The typical thickness dependence of the resistance of the tantalum film deposited on the silicon substrate at room temperature is shown in Fig. 2. The point where the resistance sharply drops manifests the formation of a continuous film.

Figure 3 shows the thickness dependence of the conductance of the α-Ta film. The film was deposited on the niobium underlayer; therefore, the zero-thickness conductance is other than zero. The conductance is seen to vary slightly when the film thickness exceeds 35 nm. Figure 4 shows the thickness dependence of the resistivity of the β-Ta film. Subtracting the contribution due to the classical size effect (which can be approximated by a quadratic polynomial within a finite interval of thicknesses) from the experimental resistivity curve, we find the contribution of the quantum size effect to the thickness dependence of the resistivity. As follows from Fig. 5, where the quantum-size contribution for the α-Ta is presented, the associated curve is of a well-defined oscillating character.

The oscillation periods for the α-Ta and β-Ta films were 5.8 and 5.0 nm, respectively. It was observed [9] that, with increasing the thickness of a bismuth film, the electric conductivity oscillates with a period of about 25 nm. The de Broglie wavelength for tantalum is shorter than that for bismuth: according to our data, its room-temperature value equals 11.6 and 10.0 nm, respectively, for the α-Ta and β-Ta. Theoretical estimates [14] of the de Broglie wavelength, $\lambda = h/P_F$, where $P_F = h(3n/8\pi)^{1/3}$ is the electron momentum on the Fermi surface, h is the Planck's constant, and n is the electron concentration (cm^{-3}), give $\lambda = 5$ Å. Such discrepancies between the measured and calculated values of the de Broglie wavelength were also observed for bismuth, where the oscillation period in a number of experiments reached 200 nm [15]. At the same time, the de Broglie wavelength estimated on the Fermi surface is 40 nm [9]. Under the assumption that the concentration of carriers varies in proportion to the resistivity, the de Broglie wavelength for the β-Ta must increase to 2.25 nm. This value is several times lower than our measurement. Supposedly, the Fermi energy changes [16] in small-size objects like quantum-size films.

In our opinion, similar oscillating dependences were also observed in the works where *in situ* measurements of the resistance of copper films [2] and Nichrome alloy [17] were carried out. It seems likely that this effect merely escaped the attention of the authors of the works cited and the scale of their graphs did not allow them to reliably measure the oscillation period. Making allowance for the quantum effects [16], one can expect that the de Broglie wavelength in thin metal films is longer than in bulk metals.

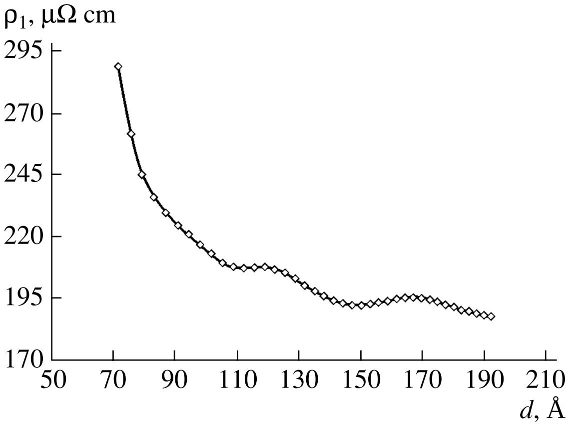


Fig. 4. Resistivity ρ_1 of the β-Ta film vs. its thickness d.

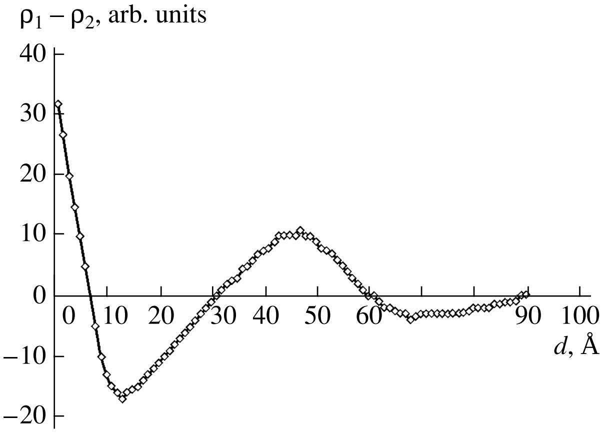


Fig. 5. Dependence of the value $\rho_1 - \rho_2$ on the thickness of the α-Ta film (ρ_2 is the average resistivity of the film).

CONCLUSIONS

Thin tantalum films were grown by the method of pulsed laser deposition. On the background of the monotonic decrease in the film resistivity with increasing film thickness, which is due to the classical size effect, the oscillating dependence is found with oscillation periods of 5.0 and 5.6 nm, respectively, for the α-Ta and β-Ta films. Such behavior is explained by the quantum-size effect in tantalum films thinner than 20 nm.

The observation of the quantum-size effect in *in situ* measurements of the resistance of the tantalum films indicates that the method of laser plasma deposition makes it possible to produce high-quality films homogeneous in structure and uniform in thickness.

ACKNOWLEDGMENTS

This work was supported by the Ministry of Education and Research of Germany, project RUS 00/217.

REFERENCES

1. *Metal Based Thin Films for Electronics*, Ed. by K. Wetzig and C. M. Schneider (Wiley-VCH, Berlin, 2003), Chap. 2.
2. H.-D. Liu, Y.-P. Zhao, G. Ramanath, *et al.*, Thin Solid Films **384**, 151 (2001).
3. D. Gupta, in *Encyclopedia of Applied Physics* (VCH Publ., 1993), Vol. 5, pp. 75–86.
4. D. W. Matson, M. D. Merz, and E. D. McClanhan, J. Vac. Sci. Technol. A **10**, 1791 (1992).
5. R. A. Roy, P. Catania, K. L. Saenger, *et al.*, J. Vac. Sci. Technol. B **11**, 1921 (1993).
6. O. A. Novodvorsky, O. D. Khramova, C. Wenzel, *et al.*, J. Appl. Phys. **94**, 3612 (2003).
7. V. B. Sandomirskiĭ, Zh. Éksp. Teor. Fiz. **52**, 158 (1967) [Sov. Phys. JETP **25**, 101 (1967)].
8. G. Palasantzas, Y. P. Zhao, G. C. Wang, *et al.*, Phys. Rev. B **61**, 11109 (2000).
9. A. N. Zherikhin, G. Yu. Shubnyi, L. N. Kherizina, *et al.*, Poverkhnost, No. 6, 79 (2000).
10. S. V. Gaponov and N. N. Salashchenko, Elektron. Promyshlennost, No. 1, 11 (1976).
11. O. A. Novodvorsky, O. D. Kharamova, C. Wenzel, *et al.*, Opt. Lasers Eng. **32**, 449 (2000).
12. S. U. Jen, C. C. Yu, C. H. Liu, *et al.*, Thin Solid Films **434**, 316 (2003).
13. D. W. Face and D. E. Prober, J. Vac. Sci. Technol. A **5**, 3408 (1987).
14. N. W. Ashkroft and N. D. Mermin, *Solid State Physics* (Holt, Rinehart and Winston, New York, 1976; Mir, Moscow, 1979).
15. Yu. F. Komnik, *The Physics of Metal Films: Size and Structure Effects* (Atomizdat, Moscow, 1979) [in Russian].
16. N. S. Rytova, Fiz. Tverd. Tela (Leningrad) **8**, 725 (1966) [Sov. Phys. Solid State **8**, 580 (1966)].
17. I. H. Kazi, P. M. Wild, T. N. Moore, *et al.*, Thin Solid Films **433**, 337 (2003).

Translated by Yu. Vishnyakov

Technical Physics, Vol. 50, No. 6, 2005, pp. 719–726. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 6, 2005, pp. 46–53.
Original Russian Text Copyright © 2005 by Kholodkov, Golant.

OPTICS, QUANTUM ELECTRONICS

Er^{3+} Ion Photoluminescence in Silicate Glasses Obtained by Plasma-Chemical Deposition in a Low-Pressure Microwave Discharge

A. V. Kholodkov and K. M. Golant

Research Center of Fiber Optics, Prokhorov General Physics Institute,
Russian Academy of Sciences, Moscow, 119991 Russia
e-mail: artem@fo.gpi.ru

Received April 21, 2004

Abstract—The luminescent properties of Er^{3+} ions embedded in silicate glass matrices with F, N, K, Al, P, Ge, P + Al, P + K, and Ge + Al admixtures are studied. Glass samples with an erbium concentration of up to 9×10^{20} cm^{-3} are synthesized by plasma-assisted CVD. The spectra, kinetics, and relative quantum efficiency of Er^{3+} photoluminescence are estimated by exciting the samples with 514.5-nm Ar^+ laser radiation. It is shown that the luminescent properties of the activator in such high-erbium unfused glasses are superior to those of the same activator in fused glasses of the same composition. This effect is attributed to suppression of clustering in the glasses prepared by low-temperature CVD, which arises because of a limited mutual solubility of the oxides in the melt. The efficient composites with an erbium concentration of up to 4×10^{20} cm^{-3} obtained in this work can be used as an active medium of waveguide lasers and amplifiers. © *2005 Pleiades Publishing, Inc.*

INTRODUCTION

The Er^{3+} ion is widely used as an activator in fiber-optic amplifiers and lasers, since the luminescence band of the $^4I_{13/2}$–$^4I_{15/2}$ metastable laser transition falls into the basic telecommunication range near 1.55 μm. However, the intrinsically high efficiency of these lasers and amplifiers declines with an increase in the activator concentration because of the well-known clustering effect. As a result, the production of high-quantum-efficiency Er^{3+}-doped active optical waveguides that meet the requirements of integrated optics becomes a challenging task. The advantages of integrated optical circuits over fiber-optic ones are a low amount of manual labor (all circuit features are made by lithography), small size, and low cost.

To make active elements based on short-optic-path waveguides, it is necessary that the activator ion concentration in the glass be one order of magnitude higher than in the fibers. However, as the activator ion concentration grows, the mean ion spacing shrinks and additional channels of excitation relaxation become significant. One of these channels is associated with the up-conversion effect [1, 2]: interaction of two nearby excited erbium ions with subsequent nonradiative relaxation of one of them [3]. In fused glasses, including silica fibers, the reason for a sharp increase in the concentration of pairs of nearby Er^{3+} ions with increasing erbium concentration is the poor solubility of Er_2O_3 in SiO_2.

The up-conversion effect observed at elevated concentrations of the activator can be combated in two ways, either eliminating various reasons for the effect (for details, see [4]). The former way is to introduce Al, P, or alkali metals into the SiO_2 matrix. These admixtures loosen the structural network of the glass, thereby reducing the rate of excitation migration between Er^{3+} ions and improving the erbium solubility. The latter way is to apply low-temperature (without fusion) CVD synthesis of the glass, which makes activator ions statistically uniformly build in the glass network, thus preventing the formation of erbium clusters. Clearly, the combination of these approaches will provide the highest efficiency in suppressing up-conversion.

Today, magnetron sputtering, ion implantation, and plasma-chemical deposition are the most popular techniques for obtaining unfused glasses in the form of planar Er-doped layers [5, 6]. The glass formation temperature in these techniques varied from 300 to 500°C. However, the "refinement" of the film structure takes place, as a rule, upon additional annealing at ≈1000°C.

Of special interest among these techniques is surface plasma-chemical vapor deposition (SPCVD) [7], which forms a transparent glass layer at substrate temperatures in the range 1000–1200°C. In this case, increased process temperatures make subsequent annealing unnecessary. A significant advantage of SPCVD is a higher rate of glass deposition compared with the other techniques. It was shown [8] that the SPCVD may serve as a technological platform for creating active integrated optical waveguides.

In this work, we investigated the luminescence due to Er ions embedded in SPCVD-grown bulk unfused glasses of different composition to a concentration of 9×10^{20} cm^{-3}. The luminescence spectra and kinetics were studied at wavelengths of 0.98 and 1.53 μm. Identical conditions of luminescence excitation and recording made it possible to compare the quantum efficiency of excitation of the $^4I_{13/2}$–$^4I_{15/2}$ radiative transition in the glasses.

EXPERIMENTAL

The doped glasses used were deposited on the inner surface of quartz silica tubes using a setup intended for preparation of preforms for fiber drawing. In this version of the SPCVD process, a halogenide–oxygen mixture kept at a pressure of 0.5 Torr is applied through a tube of diameter 20 mm (a wall thickness of 2 mm) toward a stationary plasma column. The column is sustained inside the tube by surface microwave plasma wave propagation. When the reagent mixture enters the discharge region, the halogenides dissociate to form oxides, which are deposited on the inner surface of the tube. Thus, a 3- to 5-cm-long deposition zone arises at the head of the plasma column. To provide a 25- to 30-cm-long glass layer uniform in thickness, the deposition zone is periodically displaced back and forth along the tube by controllably varying the length of the plasma column. The plasma column length variation (plasma scanning) frequency was 20 Hz; the deposition zone displacement amplitude, 25 cm. (For a detailed description of the SPCVD setup, see [9]).

The composition of the glass deposited was specified by the ratio between the flow rates of halogenides ($SiCl_4$ and others). The reagent ratios were set by flow-rate controllers and controllers of the solid reagent temperature. The total flow rate of the gaseous reagents provided a rate of glass deposition of 2–3 μm/min. The thickness of the glass deposited was varied from 150 to 300 μm.

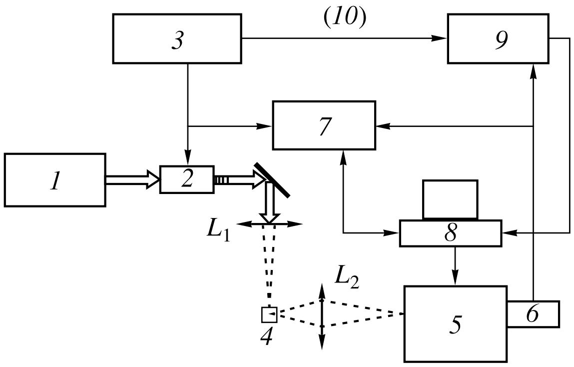


Fig. 1. Setup for studying the luminescent properties of the glasses.

An additional advantage of the SPCVD process over the other techniques of low-temperature synthesis of activated glasses is that it allows for obtaining glasses with a low concentration of Cl and an extremely low (≈1 ppm) concentration of OH hydroxyl groups, which adversely affect erbium luminescence [10].

The temperature of the substrate tube is a basic parameter governing the composition and properties of the glass. In the SPCVD process, the tube is heated by the plasma from within and by additional heating elements from without. During deposition, the mean time the inner surface of the tube is exposed to the plasma at the ends of the scan zone differs greatly, so that the temperature distribution along the substrate surface is highly nonuniform.

In our experiments, the desired temperature at the outer surface of the tube was kept accurate to ±1% with a feedback loop in the heater circuit. It was measured with an IRCON pyrometer at the site of the tube corresponding to the center of the scan zone.

The glass samples cut from the cross section of the tube had the form of 2-mm-thick semirings with polished end faces. They were selected from several experiments aimed at obtaining Er-doped glasses of different composition (see table). In a number of them, the temperature nonuniformity along the tube resulted in the corresponding composition nonuniformity with the ratio of the reagents entering the reactor kept unchanged. Because of this, the samples taken for spectroscopic studies were cut from different sections of the tube. The composition of the glass in each of the sections was examined with the X-ray microanalyzer of a JEOL JSM-5910LV electron microscope.

The luminescence from the samples was studied with the experimental setup shown in Fig. 1, which allows one to record luminescence spectra and measure the luminescence time. The radiation from cw Ar^+ laser *1* passes through a modulator consisting of electrooptic shutter *2* and driving generator *3* of rectangular pulses, is compressed by quartz lens L_1 and is applied to sample *4* vertically (parallel to the entrance slit of monochromator *5*). The modulator interrupts the pump radiation by rectangular pulses with leading-edge and trailing-edge times of 1 μs or shorter. Luminescence is excited over the deposited layer normally to the plane of the semiring. The diameter of the laser beam incident on the sample was no more than 150 μm, i.e., smaller than the thickness of the deposit. With regard to the focal length of lens L_1, ≈50 cm, the variation of the beam diameter over the sample can be neglected. Quartz lens L_2 transfers the image of the zone irradiated to the entrance slit of the monochromator. At the exit from the monochromator, the radiation is recorded by GaAs photodiode *6* (at wavelengths of 1000–1700 nm) or an FÉU-62 photomultiplier (in the wavelength range

Composition of the samples under study and the kinetic characteristics of photoluminescence due to the $^4I_{13/2}$–$^4I_{15/2}$ transition

Host	Er, ppm	F, wt %	N, at. %	K, mol %	Al, mol %	Ge, mol %	P, mol %	τ, ms
SiO_2, T_{depos} = 1200°C	1020							11.6
	990							11.5
	1280							11.5
F : SiO_2, T_{depos} = 1150°C	800	3.5						***
	640	3.2						***
	800	2						***
K : SiO_2, T_{depos} = 1050°C	2600			0.6				12
	1200			0.4				12.9
	1000			0.2				12.8
N : SiO_2, T_{depos} = 1230°C	240		3 +1.5 wt %Cl					
Al : SiO_2, T_{depos} = 1110–1230°C	1980				0.8			9.9
	1950				0.8			9.9
	2550				1.2			9.6
	3200				1.0			10.0
	3500				1.0			10.0*
	5400				1.2			9.5*
	7500				1.6			9.1*
P : SiO_2, T_{depos} = 1100°C	13000						6.5	5.8***
	13400						5.5	5.4***
	3600						6.5	8.9**
	3200						5.2	7.9**
	1800						4.6	8.7*
Ge : SiO_2, T_{depos} = 1220°C	1800					11		11.6*
	2300					15		11.6*
	3600					4		10.5**
	6450					4		11.3**
P, Al : SiO_2, T_{depos} = 1050°C	710				2.2		6	9.2
	2600				2.9		9.4	9.2*
	4100				2.6		10.5	8.9*
	5660				2.7		11.9	8.1*
	7100				2.7		12.8	7.74*
P, K : SiO_2, T_{depos} = 1020°C	2400			0.74			3.5	2.9
	3600			0.69			3.5	3.4*
	5360			0.52			5.3	6.3**
	7100			0.2			7.65	7.5**
	10760			0.1			10.1	7.11*
Ge, Al : SiO_2, T_{depos} = 1150°C	5600				2.6	16.1		8.9*
	5000				4	17.9		9
	2900				5.2	17.5		8.9

Note: */**/*** Slight/moderate/major contribution of the fast exponential to the kinetics of 1.53-μm luminescence.

800–1100 nm). The luminescence spectra are taken with EG&G5209 synchronous detector *7* controlled by computer *8*. The luminescence time is measured with SR250 gated integrator *9* synchronized (*10*) with the modulator. Near 1.53 μm, the measurements were made at a modulation frequency of 9 Hz; near 0.98 μm, at 110 Hz. The mean power of the pump beam at the output of the modulator was 250 mW.

SPCVD SYNTHESIS OF MULTICOMPONENT GLASSES

As has already been mentioned in the introduction, the SPCVD technology of glasses has a number of features, which show up most vividly in obtaining doped glasses [11]. It is known, for example, that the introduction of such elements as F, Ge, K, Al, and P into the silicate matrix to modify the glass reduces its viscosity. As the temperature of the substrate tube grows, compounds of these elements are desorbed from the surface at a higher rate, which decreases the doping level in the glass being deposited. A temperature gradient along the tube typical of the SPCVD process causes a corresponding longitudinal nonuniformity in the glass composition.

Since $ErCl_3$ evaporated at 1000°C in our experiments, we had to keep the temperature of the tube at a relatively high level throughout the deposition zone in order to prevent the condensation of the reagent. This may be a reason for an extremely low potassium concentration in the glasses deposited at this temperature (see table). Apparently, here we are facing the ability of alkali metals to significantly fluidize silicate glasses [12].

The temperature dependence of the incorporation efficiency of phosphorus in the temperature range 1000–1200°C is also very strong. Therefore, its concentration was controlled largely by the temperature of the tube during deposition.

A specific feature of nitrosilicate glass deposition is that the oxidation of silicon tetrachloride proceeds under the oxygen deficiency conditions [13]. Under such conditions, to reach a high level of codoping by other elements, such as Er, Ge, Al, etc., is a challenge because of the process chemistry. Furthermore, when oxygen is deficient in the glass, the concentration of harmful Cl rises drastically. As a result, the maximal erbium concentration in the nitrosilicate glass was as low as 240 ppm (1.6×10^{19} cm^{-3}).

The efficiency of incorporation of the elements into the silicate matrix depends not only on the temperature of the substrate tube and relative flow rate of the reagents but also on the composition of the glass being deposited. We relate this effect to the type of defect generated in the structural network of the glass by an element embedded in the glass. Note also that the incorporation efficiency correlates with the solubility of the corresponding oxide in SiO_2. For example, the efficiency of incorporation of erbium into the silicate and alumina–silicate matrices during SPCVD was ≈20% in our experiments. Due to a considerable structural disturbance, chemisorbed erbium and aluminum atoms turn out to be weakly bonded to oxygen on the silica glass surface. Weak bonding favors their desorption or substitution of silicon atoms for them. An increase in the aluminum halogenide flow rate versus that of $SiCl_4$ enhances light scattering in the glass, indicating the presence of the concentration homogeneity limit in SPCVD alumina–silicate glass. The addition of Ge generates network defects favorable to the incorporation of erbium and aluminum, as is confirmed by the rise in the incorporation efficiency for these elements to ≈50%. With phosphorus added to the silicate matrix, the incorporation efficiency of erbium rises still further (up to ≈100%). As follows from the table, codoping by phosphorus and aluminum doubles the maximal concentrations of both elements, presumably because a P–Al cluster is more readily incorporated into the silicate network. It should be noted that the obtaining of Al- and P-codoped SiO_2 glasses by means of the standard (nonplasma) MCVD technique is a big technological bugaboo.

Summing up, it may be said that the SPCVD technique as applied to multicomponent glasses is in a sense self-regulating; that is, the atomic packing is not too highly disturbed by thermal fluctuations and the structure of the glass is dense and perfect.

EXPERIMENTAL RESULTS

Figure 2 shows the luminescence spectra of erbium ions embedded in the fluorosilicate matrix. That taken at a low temperature is seen to consist of a number of closely spaced sharp lines. The long luminescence time at 0.98 μm indicates that, in this glass, unlike the rest of the glasses studied in this work, the multiphonon relaxation channel associated with the $^4I_{11/2}$ level is combated. The kinetics of luminescence near 1.53 μm features rapid relaxation mechanisms. The luminescence spectrum of Er^{3+} ions in the nitrosilicate matrix (Fig. 3) also comprises a series of sharp lines, which particularly stand out as the pump modulation frequency increases. In our opinion, these lines can be assigned to erbium ions with the shortest lifetime of the $^4I_{13/2}$ level. The spectrum at 0.98 μm has an extra peak at 1.001 μm. Unfortunately, we did not manage to determine the luminescence time in this case because of a low intensity of the signal.

Figures summarize the luminescence spectra taken from all the samples at 1.53 μm. They are normalized by their peak values for comparison convenience. Erbium ions in the F : SiO_2, Al : SiO_2, and Al, Ge : SiO_2 matrices have the broadest spectra (50, 47, and 47 nm wide, respectively). The spectrum taken from the K : SiO_2 matrix, conversely, is the narrowest (24 nm wide), and the luminescence time here is the longest. Such radical modifications of the luminescence pattern take place despite a low potassium concentration in the glass.

The luminescence spectra taken from the phosphorus-containing samples are similar to each other, but the luminescence kinetics in them is different. In the P, K : SiO_2 and P : SiO_2 glasses, the $^4I_{13/2}$ level lifetime decreases markedly with decreasing phosphorus concentration and increasing erbium concentration. At the same time, the lifetime of this level in P, Al : SiO_2 depends only on the erbium concentration.

DISCUSSION

It was shown [14] that Er^{3+} ion clustering and its associated up-conversion are responsible for the slow component in the kinetics of luminescence at 0.98 μm. However, we did not observe this effect in the samples studied. The reasons may be the following: (i) the effect is too weak to be detected, (ii) the effect is distinct only in highly clustered glasses, and (iii) the fast processes observed in the kinetics of 1.53-μm luminescence are related to multiphonon transitions rather than to up-conversion.

Determining the absolute quantum efficiency of luminescence in thin layers is a bottleneck, since the absorbed pump power is difficult to measure. Since the luminescence excitation and recording conditions in our experiments were reproduced with an accuracy of 5%, we succeeded in making comparative estimates of the luminescence quantum efficiency. Figure 5 plots the relative quantum efficiencies of luminescence ($^4I_{13/2} \longrightarrow {}^4I_{15/2}$, $\lambda = 1.53$ μm) against the erbium concentration. The relative quantum efficiency was calculated by the formula

$$q_i = \frac{\mathrm{Lum}_i N_\mathrm{m}}{\mathrm{Lum}_\mathrm{m} N_i} = \frac{\sigma_i Q_i(H_{11/2}, {}^4I_{13/2})}{\sigma_\mathrm{m} Q_\mathrm{m}(H_{11/2}, {}^4I_{13/2})}, \qquad (1)$$

where $\sigma_{i,\,\mathrm{m}}$ are the composition-dependent cross sections of electron excitation by 514.5-nm radiation, $Q_{i,\,\mathrm{m}}$ are the probabilities that electrons excited to the level $H_{11/2}$ will transit to the level $^4I_{13/2}$ and then execute a radiative transition to the ground state, and N_i and N_m are the Er^{3+} ion concentrations in the ground state. Quantities Lum_m and N_m correspond to the sample exhibiting a maximal efficiency.

In the calculation, the total erbium concentration in the samples was used instead of the erbium ion concentration in the ground state (N_i, N_m). This restricts the validity of formula (1) to the case of weak optical excitation. As is known, the adverse effect of up-conversion on the luminescence quantum efficiency grows with excitation level. Of practical interest are the parameters of the glasses at an excitation level close to inversion. Therefore, the excitation level in our experiments was 50–60%, according to the pump power used. In this case, the error in the quantum efficiency calculated by formula (1) may reach 20% because of different rates of excitation relaxation to the ground level.

Let us see how the deposition conditions influence the luminescent properties of erbium in the glasses. The structure and composition of SPCVD silica glasses are governed by chloride heterogeneous oxidation, the mobility of chemisorbed atoms being dependent on

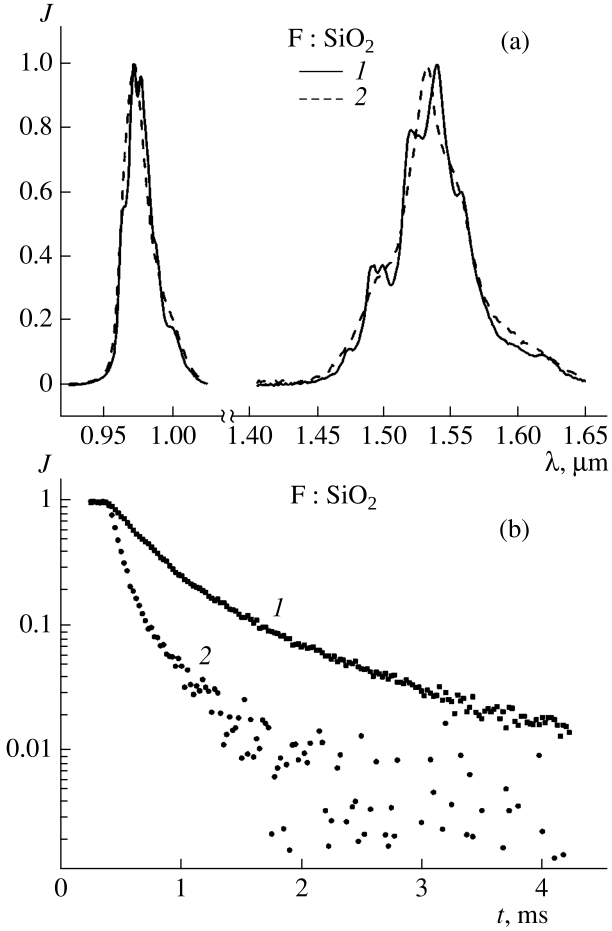


Fig. 2. (a) Spectra and (b) kinetics of Er^{3+} ion luminescence due to the $^4I_{11/2}$–$^4I_{15/2}$ transition (0.98 μm) for the fluorosilicate glass deposited at (*1*) low and (*2*) high temperature. The luminescence kinetics is approximated by a sum of two exponentials with indices (*1*) $t_1 = 0.3$ ms and $t_2 = 1.23$ ms and (*2*) $t_1 = 0.09$ ms and $t_2 = 0.48$ ms.

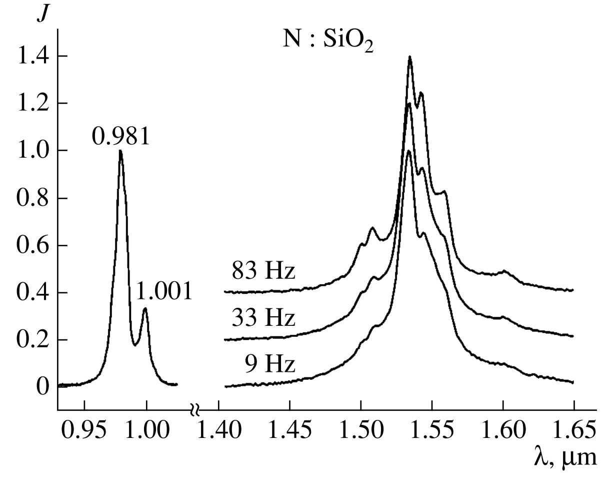


Fig. 3. Erbium luminescence spectra in the nitrosilicate glass at 0.98 and 1.53 μm for different pump modulation frequencies.

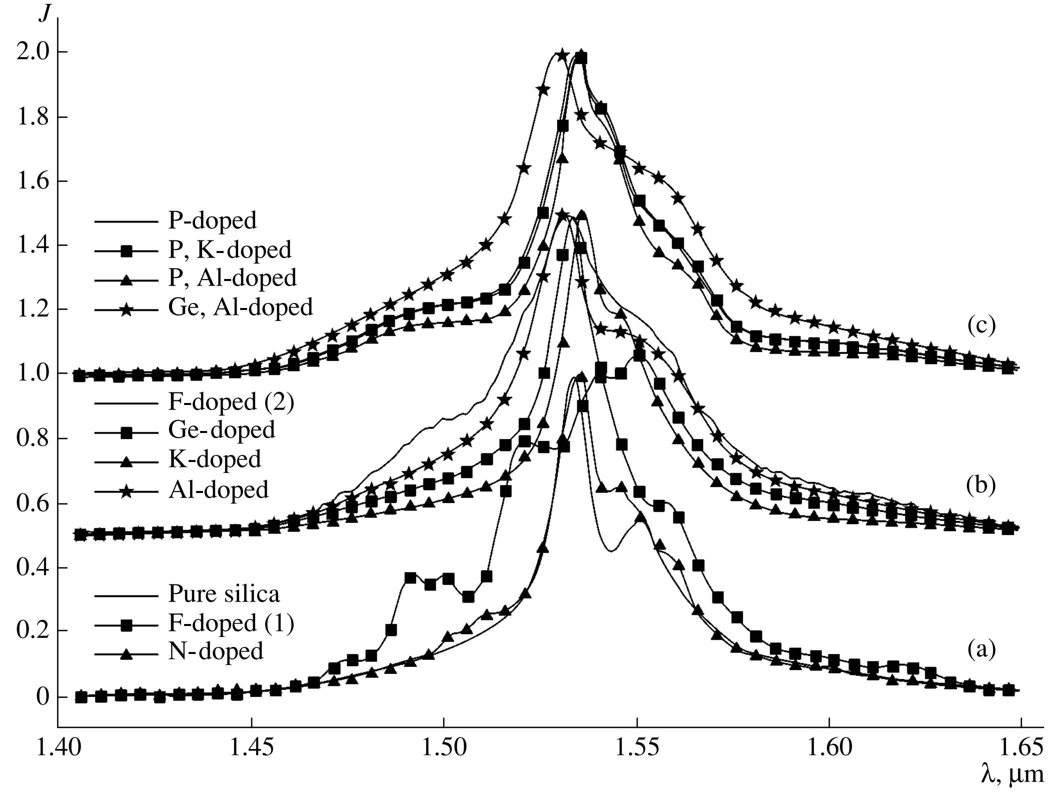


Fig. 4. Erbium ion luminescence spectra near 1.53 μm taken from the glasses of different composition.

substrate surface temperature T_{surf}. Three situations are worth considering here.

(1) $T_{\text{surf}} < T_{\text{Er,Si}}$ and T_X (X = Al, P, K, or Ge), where T_X is the characteristic temperature activating motion of the corresponding element in the glass. In this case, the atom mobility can be neglected and the structural network of the glass is governed by the particles from the gaseous phase that are “frozen” in the substrate surface. Under conditions (1), the glass is a thermodynamically nonequilibrium system. The complex luminescence spectra (Figs. 2, 3, and 5) reflect the contributions from different Er^{3+} ions in the structurally unstable anionic environment. For example, the complex shapes of the spectra shown in Fig. 2 indicate that the erbium ions are in the same anion environment, and the long lifetime of the state $^4I_{11/2}$, which is responsible for 0.98-μm luminescence, points to the presence of fluorine atoms in the environment [1]. The changes in the luminescence spectra taken from the nitrosilicate matrix seem to be due to chlorine and nitrogen atoms. Interestingly, when plotted as a function of the erbium concentration, the luminescence efficiency of all the glasses deposited under the conditions $T_{\text{surf}} < T_{\text{Er,Si}}$, T_X is fitted by a single straight line (Fig. 5). Thus, such conditions wipe out the difference between the glasses of different composition in this respect.

(2) $T_X < T_{\text{surf}} < T_{\text{Er,Si}}$. In this case, dopant atoms have a chance to occupy thermodynamically equilibrium and energetically favorable positions in the network of the glass. It is likely that setting the equilibrium structure is

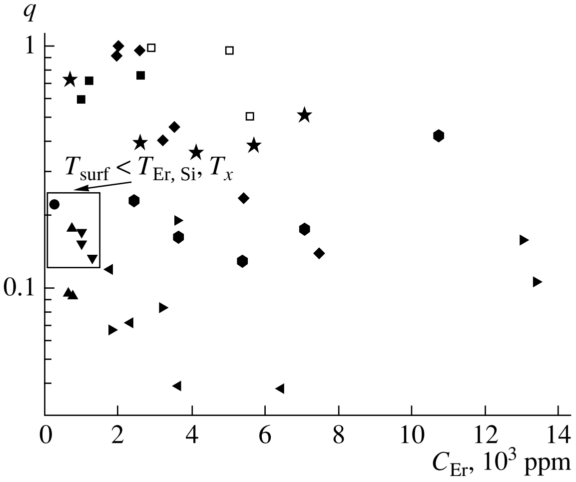


Fig. 5. Relative quantum efficiency of 1.53-μm luminescence (1000 ppm = 6.6×10^{19} cm^{-3}). (■) K-doped sample, (●) N-doped sample, (▲) F-doped sample, (▼) pure silica, (◆) Al-doped sample, (◀) Ge-doped sample, (▶) P-doped sample, (⬢) (P + K)-doped sample, (★) (P + Al)-doped sample, and (□) (Ge + Al)-doped sample.

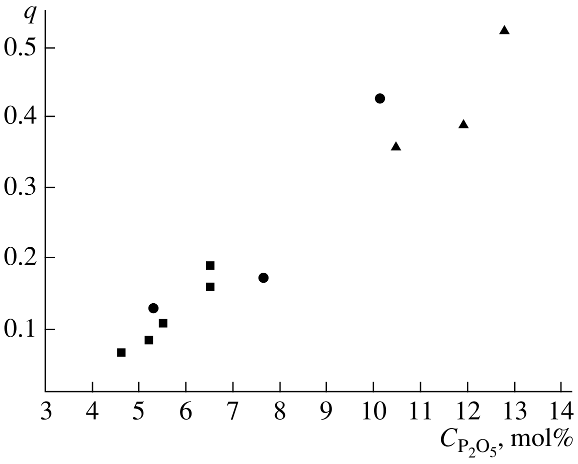


Fig. 6. Relative quantum efficiency of 1.53-μm luminescence vs. the phosphorus concentration. (■) P-doped, (●) (P + K)-doped, and (▲) (P + Al)-doped samples.

also accompanied by the motion of erbium atoms but only in the neighborhood of their initial positions. In this state, the structure of the specific glass volume (i.e., the volume containing a single erbium ion) is equilibrium and the glass as a whole may be viewed as thermodynamically quasi-equilibrium. Accordingly, the luminescence spectrum takes on the attributes of the typical luminescence spectrum of Er^{3+} ion in fused glasses. Almost all the glasses studied in this work were deposited under these temperature conditions. The spectrum (Fig. 3) and kinetics of 0.98-μm luminescence for the fluorosilicate matrix suggest that fluorine atoms leave the neighborhood of erbium at these temperatures.

(3) $T_{\text{surf}} > T_{\text{Er}}$ or T_{Si}. Under these conditions, erbium ions, when moving in the forming network of the glass, may generate clusters or even cause precipitation of Er_2O_3 in SiO_2 if their concentration is high.

The above concept of forming a homogeneous (cluster-free) structure of the glass may be invalid in some cases. Specifically, if the dopant concentration is high and $T_X \ll T_{\text{surf}}$ or the time of temperature action is long, the bulk of the glass may get mixed up well and thermodynamic equilibrium with clusters may set in the system even in the absence of activating motions of Er ions or Si atoms.

The luminescence pattern in the phosphosilicate glass is intriguing. Figure 6 plots the erbium luminescence quantum efficiency against the phosphorus concentration in the glass. From 5 mol % of phosphorus on, the quantum efficiency rises monotonically. Interestingly, while the high-phosphorus samples exhibit the best luminescence properties, those with a low phosphorus content are inferior to the other glasses in terms of the luminescent activity.

Figure 7 demonstrates changes in the spectrum and kinetics of erbium ion luminescence that arose when the samples were heated in the flame of an oxygen–propane burner. Softening of the glass is seen to cause erbium clustering. The luminescence spectrum broadens, and up-conversion signs appear in the kinetics pattern.

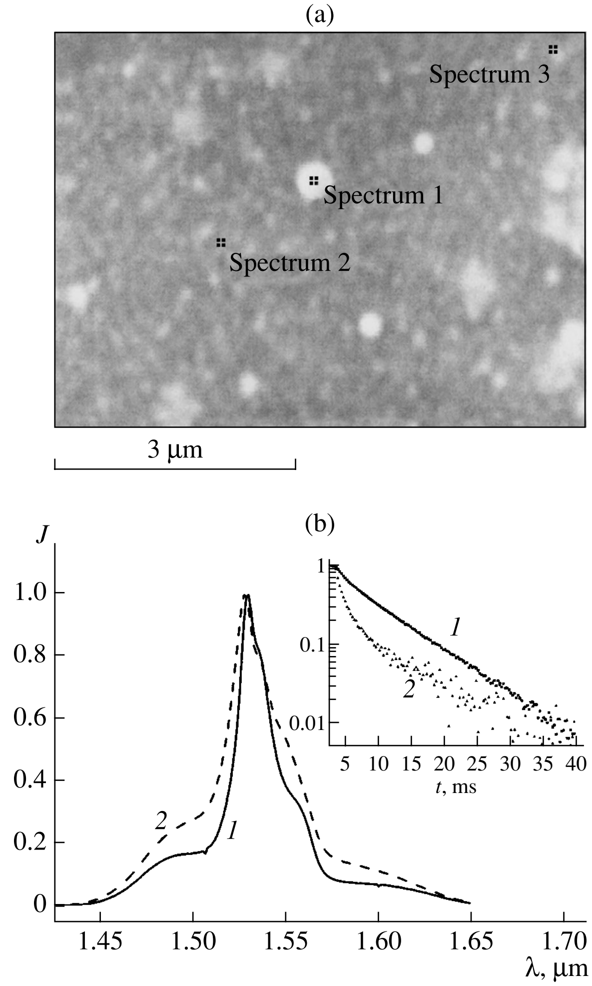


Fig. 7. Effect of fusion of the Al_2O_3(2.7 mol %)–P_2O_5(12.8 mol %)–Er(7100 ppm) glass on (a) clustering and (b) Er^{3+} luminescence parameters. The bright spots in panel (a) correspond to high-erbium areas. Panel (b) shows the luminescence spectra and kinetics for the sample obtained by plasma-chemical deposition (*1*) before and (*2*) after high-temperature treatment.

CONCLUSIONS

It was shown experimentally that moderate-temperature plasma-chemical deposition may serve as a basis for a new high-erbium glass synthesis technology. A high efficiency of luminescence due to the $^4I_{13/2}$–$^4I_{15/2}$ transition in these glasses offers considerable scope for designing active integrated optical systems. Of most interest are Al- and K-doped glasses obtained by the plasma-chemical technique. In these materials, the quantum efficiency of luminescence remains high as the erbium concentration grows to at least 5000 ppm (3.3×10^{20} cm^{-3}). Halogens (F and Cl) introduced into the matrix adversely affect the quantum efficiency and reduce the concentration threshold of ion erbium clustering. It is demonstrated with the fluoro- and phosphosilicate composites that the luminescent properties of Er ions in unfused SPCVD glasses strongly depend on the substrate temperature during deposition. The optimal temperature is that at which the atomic structure in the neighborhood of an Er^{3+} ion becomes equilibrium but the activation motion of Er^{3+} ions in the matrix is still suppressed.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, grant no. 04-02-16441.

REFERENCES

1. M. P. Hehlen, N. J. Cockroft, T. R. Gosnell, *et al.*, Opt. Lett. **22**, 772 (1997).
2. N. V. Nikonorov, A. K. Przhevuskii, M. Prassas, *et al.*, Appl. Opt. **38**, 6284 (1999).
3. M. P. Hehlen, N. J. Cockroft, T. R. Gosnell, *et al.*, Phys. Rev. B **56**, 9302 (1997).
4. A. K. Przhevuskii and N. V. Nikonorov, Opt. Mater. **21**, 729 (2003).
5. Y. B. Choi, S. H. Cho, and D. C. Moon, Opt. Lett. **25**, 263 (2000).
6. A. Polman, M. A. Marcus, and D. C. Jacobson, Mater. Res. Soc. Symp. Proc. **244**, 381 (1992).
7. D. Pavy, M. Moisan, F. Ould-Saada, *et al.*, in *Proceedings of the 12th European Conference on Optical Communication, Barselona, 1986*, pp. 19–22.
8. A. V. Kholodkov and K. M. Golant, in *Proceedings of the Optical Fiber Communications (OFC) Conference and Exhibition, Los Angeles, 2004*, Vol. 2, pp. 545–547.
9. E. M. Dianov, K. M. Golant, V. I. Karpov, *et al.*, Opt. Mater. **3**, 181 (1994).
10. Y. Yingchao, A. J. Faber, and H. de Waal, J. Non-Cryst. Solids **181**, 283 (1995).
11. A. V. Kholodkov, K. M. Golant, and I. V. Nikolin, Microelectron. Eng. **69**, 365 (2003).
12. M. E. Lines, J. Non-Cryst. Solids **171**, 209 (1994).
13. E. M. Dianov, K. M. Golant, R. R. Khrapko, *et al.*, J. Lightwave Technol. **13**, 1471 (1995).
14. A. V. Belov, in *Proceedings of the 11st International Conference on Integrated Optics and Optical Fiber Communication Combined with the 23rd European Conference on Optical Communication (IOOC-ECOC), 1997*, pp. 51–53.

Translated by V. Isaakyan

Technical Physics, Vol. 50, No. 6, 2005, pp. 727–731. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 6, 2005, pp. 54–58.
Original Russian Text Copyright © 2005 by Kurochkin, Ryabtsev, Neizvestny.

OPTICS, QUANTUM ELECTRONICS

Experimental Setup for Quantum Cryptography by Means of Single Polarized Photons

V. L. Kurochkin, I. I. Ryabtsev, and I. G. Neizvestny
Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences,
pr. Akademika Lavrent'eva 13, Novosibirsk, 630090 Russia
e-mail: kurochkin@isp.nsc.ru

Received June 16, 2004

Abstract—Pioneering experiments on single-photon quantum cryptography that are performed with a tailored setup are reported. The key is transferred by pulsed semiconductor lasers, which encode the polarization state of the photons in two mutually nonorthogonal bases. Photon detectors are based on C30902S silicon avalanche photodiodes. For a laser pulse repetition rate of 100 kHz and a mean number of photons per pulse of about 0.2, the key generation rate reaches ≈4 kbit/s. The error rate in the key does not exceed 1%. *© 2005 Pleiades Publishing, Inc.*

From the practical standpoint, quantum cryptography is today the most mature area of quantum informatics. Quantum cryptography allows for absolutely secure data transfer between legitimate users of communication lines. The security of data transfer, combined with the complete impossibility of unauthorized access, relies on fundamental laws of nature, unlike those cryptography approaches using mathematical methods, where the information basically may be decoded. In accordance with the mathematically proven Shannon statement [1], a message being transferred cannot be decoded if it is encoded by a random one-shot key whose length equals that of the message, provided that the key is known by legitimate users alone. However, the problem in this situation is how to transport the key to remote users. Generally speaking, classical communication methods cannot provide secure transfer of the key over accessible data channels, since there are techniques of inconspicuous listening with subsequent decoding.

The ideas of quantum physics and quantum informatics as applied to long-range data communication may be used to attack the problem of transfer of an absolutely random key through accessible data channels with security assurance. The absolute security provided by quantum cryptography follows from the forbiddings of quantum physics that are imposed on a metering device: (i) it is impossible to extract information about nonorthogonal states without disturbance [2] and (ii) it is impossible to clone an unknown quantum state (the no-cloning theorem) [3]. From these forbiddings, it follows that, if single quantum objects are employed as data carriers, any intervention into the data transfer process undertaken by an unauthorized person will inevitably cause irreversible changes in the quantum states of the objects, from which the fact of intervention can be established.

Bennet and Brassard in 1984 [4] were the first to justify the principles of quantum cryptography and to devise a communications protocol. The first experimental demonstration of their concept [5] attracted much interest worldwide and gave an impetus to extensive research in this field. In [5], a key dissemination protocol was suggested, which was represented as a secured sequence of zeros and unities enciphered by single photons polarized in two mutually nonorthogonal bases. Later, this protocol was named BB84. Subsequently, a protocol based on phase coding using a Mach–Zender interferometer found wide application (for the basic ideas behind this protocol, see [2]). A radically new way of disseminating a quantum key through entangled states that is based on the Einstein–Podolsky–Rozen effect was suggested by Ekert [6]. The ideas of quantum cryptography turned out to be so promising that many research groups worldwide immediately launched R&D work on creating commercial devices. A comprehensive review of theoretical and experimental works in this field is presented in [7]. Since the publication of that review, a variety of new approaches to organizing data channels have been suggested. For example, the rudiments of relativistic quantum cryptography were described in [8, 9], a time–frequency scheme of encoding was suggested in [10], and the concept of using a phase shift between two sequential single photons to generate a key was put forward in [11, 12]. Using single-photon quantum cryptography, a quantum key has been disseminated to advantage over several tens of kilometers both along fiber-optic links and in open space. It should be noted that the early protocols still remain the most efficient in terms of organization of real quantum data channels. The concept of phase encoding has been implemented in 67- [13], 100- [14], and 150-km-long [15] fiber-optic communication lines. The polarization encoding protocol BB84, which was also used in this work, was applied to organize communication through open space over a distance of 10 [16] and 23 km [17]. Today, the feasibility of communicating with satellites is being discussed [18].

In this paper, we report experiments on generation of a quantum key with the help of an original setup for quantum cryptography. The key was disseminated by encoding the polarization states of single photons emitted by pulsed semiconductor lasers in two mutually nonorthogonal bases (the BB84 protocol [4]). The experiments were aimed at testing techniques of generating single photons in a given quantum state with subsequent detection of these photons via discriminating them in initial states at a low level of erroneous measurements. We also developed a model of capturing all the photons by an unauthorized person and demonstrated the absolute impossibility of inconspicuous penetration of an unauthorized person into the quantum communication line. As detectors, we used arrays of high-speed single-photon counters based on C30902S silicon photodiodes with active avalanche quenching.

Let us outline the generation of a quantum key using the BB84 protocol [4, 7]. A transmitter (usually called Alice) produces single photons linearly polarized in two mutually nonorthogonal bases; in one (vertical–horizontal basis), the photons are of 0° and 90° polarizations; in the other (diagonal basis), the photons are of 45° and –45° polarizations. Alice and a receiver (Bob) come to an agreement about a code assigned to each of the polarization in binary notation. For example, photons with 0° and 45° polarizations mean “0,” while those with 90° and –45° polarizations mean “1.” During a session, Alice sends a sequence of randomly polarized photons (0°, 45°, 90°, or –45°) and Bob records the photons arrived and randomly selects a measurement basis for each of them. Then, using an extra (unsecured) channel, Bob informs Alice of the basis in which he has made the measurement but does not tell the results of this measurement. Since the photon can take values either “0” or “1,” the message about recording the photons, while sent over the unsecured channel, carries no information for an outside listener (usually called Eva). In answer to the message, Alice tells whether the basis was correctly chosen for each of the photons. Keeping only correct-basis measurements during the session, Alice and Bob generate a unique random sequence of zeros and unities, of which a secured key is then formed.

An important step in quantum cryptographic data transfer is a test for possibility of capturing information transmitted through a quantum channel. To this end, Alice and Bob, using the unsecured transmission line, make a checking comparison of a randomly selected part of the key generated. If listening was absent, the codes coincide. The error in the code is due to the noise of the detector and imperfections in the data channel. If an outside listener penetrates into the channel and reads out the information, he will have to generate the same photon (it is remembered that the information is transferred by a single photon!) and send it to Bob again. Then, according to the no-cloning theorem as applied to the state of an arbitrary quantum object [3], the listener will irreversibly break the polarizations of the photons and will not be able to exactly reproduce them, causing a discrepancy between Alice’s and Bob’s codes. As a result, the error that will be revealed by comparing the codes in the unsecured channel will far exceed the error inherent in undisturbed data transfer. In this way, the fact of unauthorized penetration into the quantum channel will be disclosed and the legitimate users will be able to take necessary precautions against intervention. In essence, two mutually nonorthogonal bases and a relatively low bit rate are needed to assure security. Note also that data transfer through a quantum communication line implies using one-photon laser pulses, since many-photon pulses will allow Eva to inconspicuously direct some of the photons to her photodetectors and, hence, the fact of penetration will pass unnoticed.

Our experimental setup is depicted in Figs. 1 and 2. A transmitting unit (Fig. 1) comprises four ILPN-210 semiconductor lasers *2*, each generating a pulsed radiation with one of the four polarizations (0°, 45°, 90°, and –45°). The pulses 8–10 ns wide are combined to form a single pulse by means of mirrors *5*, attenuated at the exit by absorbing filters *6*, and directed toward the receiver through a 70-cm-long air gap protected against light. The lasers are fed by current-modulated power supply *1* and are thermally stabilized with semiconductor Peltier microcoolers *3*. Each of the lasers generates a pulse of coherent 830-nm radiation when a control signal from a computer arrives at its power source. The attenuated laser pulses arrive at the input of the receiver (Fig. 2) and are then split into two beams by 50% beam splitter *1*. The direct (undeflected) beam (Fig. 2) is directed to Glan polarizing splitting prism *3*, which directs each of the photons with the vertical–horizontal basis to its dedicated photodetector *4* (the splitting factor is 10^4 or higher). The photons with the diagonal basis equiprobably (50%) fall on either of the two photodetectors. The other part of the input beam, which is deflected by the beam-splitting plate at the entrance, first passes through the $\lambda/2$-plate, which rotates the polarization of the photons by 45°. As a result, the photons that initially had the diagonal basis are distributed by the beam splitter (with a high splitting factor) among the respective photodetectors. The photons with the other basis equiprobably fall on either of these two photodetectors.

The receiver allows the transmitter to be tuned so that each laser pulse, having passed through the exit attenuator, has no more than one photon. Under such conditions, the photon momentum distribution obeys the Poisson statistic. In quantum cryptography, a signal is regarded as single-photon if average number $\bar{n}$ of photons per pulse varies between 0.1 and 0.2 [7]. If the

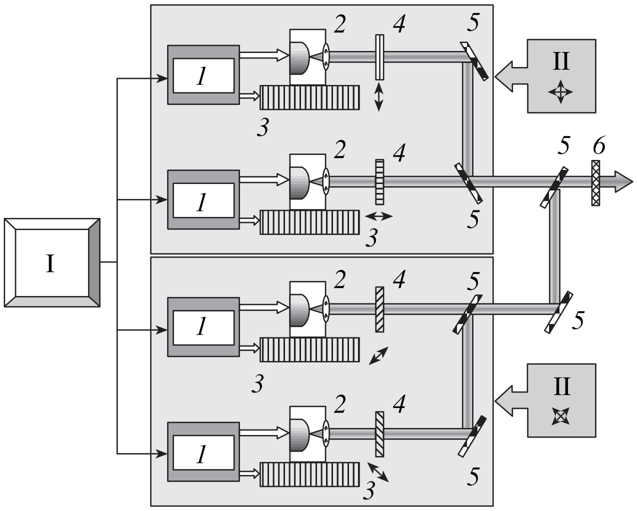


Fig. 1. Transmitter of the experimental setup for single-photon quantum cryptography. (*1*) Power supply for semiconductor laser, (*2*) semiconductor laser, (*3*) Peltier microcooler, (*4*) polarizer (Glan prism), (*5*) mirror, and (*6*) absorbing filter. I, computer; II, basis.

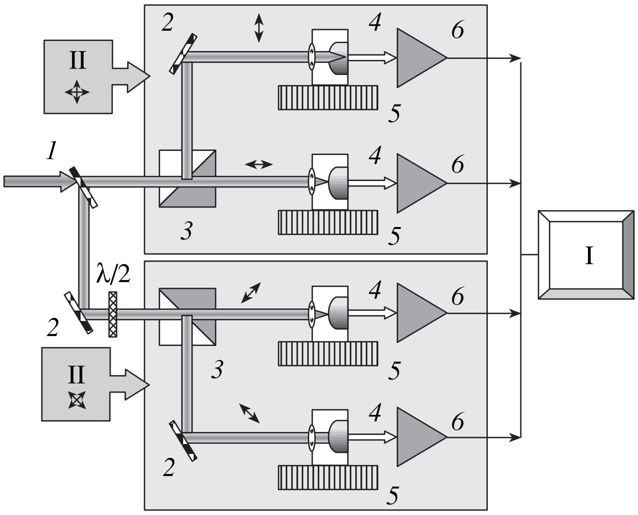


Fig. 2. Receiver of the experimental setup. (*1*) 50% beam splitter, (*2*) mirror, (*3*) Glan polarizing separating prism, (*4*) APD with collecting lens, (*5*) Peltier microcooler, and (*6*) amplifier. I, computer; II, basis.

criterion is $\bar{n} = 0.1$, the fraction of two-photon and three-photon pulses is 5 and 0.16% of the number of single-photon ones. In other words, nine of ten pulses contain no photons in this case. A photon of any polarization sent by Alice may fall on any of three photodetectors: if it has its initial basis, on one of them (the second one is prohibited by the polarizing splitting prism); if it has the foreign basis, equiprobably on two of them. The fraction of multiphoton pulses in the session can be calculated based on the Poison statistic by counting the signals arriving at all four photodetectors and the number of simultaneous records by two or more photodetectors at one time. By appropriately selecting the lasing power of each of the lasers, one can provide a desired number of photons in pulses generated by the transmitter.

Since the data transfer is secured only if each laser pulse has no more than one photon, the photodetectors must meet stringent requirements. Specifically, they must offer a high quantum efficiency of recording, a low noise, and a sufficiently high count rate. Cryptographic systems intended for key transfer over open space [16, 17] operate at a wavelength of ≈0.85 μm, i.e., in the atmospheric transparency window. Today, avalanche photodiodes seem to be the best single-photon detectors in this range. In our setup, we used EG&G C30902S avalanche photodiodes (APDs), which are the most sensitive devices at a wavelength of 0.8 μm among those currently available [7, 19]. Around these diodes, a high-speed single-photon counter with active avalanche quenching was built [20, 21]. To count single photons, the APD must operate as a Geiger counter [19–21], where one photon can trigger a charge carrier avalanche. Three of the four APD detectors (*4* in Fig. 2) operated in the passive avalanche quenching mode; one, in the active mode. The current through the APDs was limited by a 200-kΩ ballast resistance. The signal was picked up from a load resistance of 50 Ω [6]; amplified by amplifier *6*; and applied to a pulse shaper, which generated standard TTL pulses to provide an interface with the computer. If a voltage applied to the APD exceeds some threshold voltage U_{th}, a photon triggers the avalanche multiplication of charge carriers with a gain reaching 10^5–10^6. At a wavelength of 830 nm, the probability of recording a single photon reaches 50%. The diodes were cooled by Peltier semiconductor microcoolers down to –20°C in order to reduce the intrinsic noise. The noise pulse rate in an APD operating in the Geiger mode depends on the temperature and a voltage excess over the threshold value. Figure 3 plots the noise pulse rate against excess $U - U_{\text{th}}$ over the threshold voltage at a diode temperature of –20°C.

Each of photodetectors *4* in Fig. 2 represents a tailored high-speed single-photon counter operating in the active avalanche quenching mode [20, 21] with a count rate of several megahertz. Unlike experiments carried out in [20, 21], our experiments allowed for APD operation in various current-limiting modes in order to protect the device against bright illumination. This feature of our operating scheme is of great important, since, otherwise, the C30902S devices operating as a Geiger counter may fail under a too high illumination level. The avalanche quenching pulse amplitude can be varied from 5 to 25 V. A pulse-delay and pulse-width window discriminator (strobe) can be applied to the device. The minimal duration of the strobe is 30 ns. The output signal of the counter can be displayed in analog form or be

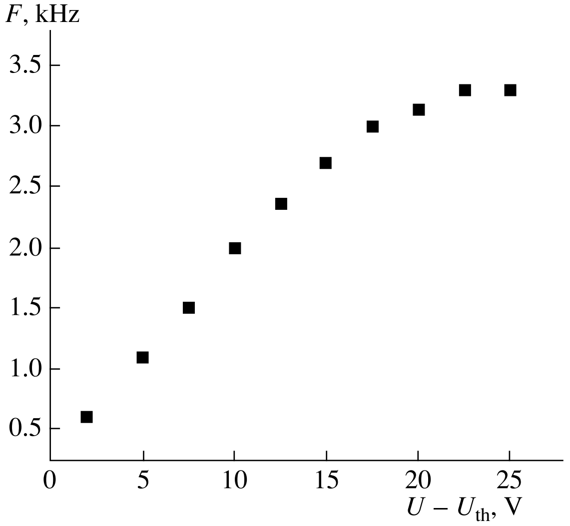


Fig. 3. Noise pulse rate F vs. voltage excess $U - U_{\text{th}}$ for the C30902S APD at –20°C.

converted to a TTL pulse that is then applied to the computer. Figure 4 plots the photon count rate of the counter with active avalanche quenching versus the laser pulse repetition rate for various current-limiting modes and pulse parameters (the count rate is limited from above by (a) 2.50 and (b) 0.25 MHz).

In our experiments, the quantum key was generated as follows. Alice's computer sets the laser pulse repetition rate (clock frequency). For each clock period, a sync (strobe) is generated, which is sent to Bob for transmission–reception synchronization. Simultaneously with the strobe, another pulse is applied at random to one of the four lasers, which generates a 10-ns-wide light pulse. A random number is created by an intelligent random-number generator, although generally it is preferable to use a noise generator to create a random number [7]. Having received the sync, Bob generates his own 20-ns-wide strobe. The pulses from the photodetectors are recorded only during the strobe application. In this case, a major part of the intrinsic noise of the photodetector is eliminated. For example, at –20°C and a voltage excess of 20 V, the rate of noise pulses is about 3×10^3 per second (Fig. 3). At the same time, this value reduces to 100 per 10^6 clock periods when time gating of the signals is employed. After amplification, the duration of the noise pulses and the pulses from the APD was found to be 8–10 ns. Pretuning of the delay between the strobe and the APD pulse (when the diode is triggered by a laser pulse from the transmitter) makes it possible to considerably improve the signal-to-noise ratio and reduce the error rate in the final code. The pulse from the APD is regarded as bearing information only if it coincides in time with the laser pulse. Intrinsic noise pulses not appearing during the strobe did not fall on the counter. The data from the four photodetectors were read out by Bob's computer under the control of the sync. In our setup, we used the same computer. Without loss in generality, such a design makes the experiment less hardware-intensive. When a pulse from any of the APDs came during the strobe, Bob remembered these data and the number of the sync and generated a signal for Alice by which she remembered the number of the sync and which of the lasers operated in that clock period. Since the mean number of photons in a light pulse was much less than unity, there was no need to remember the transfer process as a whole. Fifty percent mirror *1* at the entrance randomly turns the photon to the vertical–horizontal or diagonal basis for recording. If Alice's and Bob's bases coincided, the measurements were assigned a successive serial number and filed to generate the key. Other-

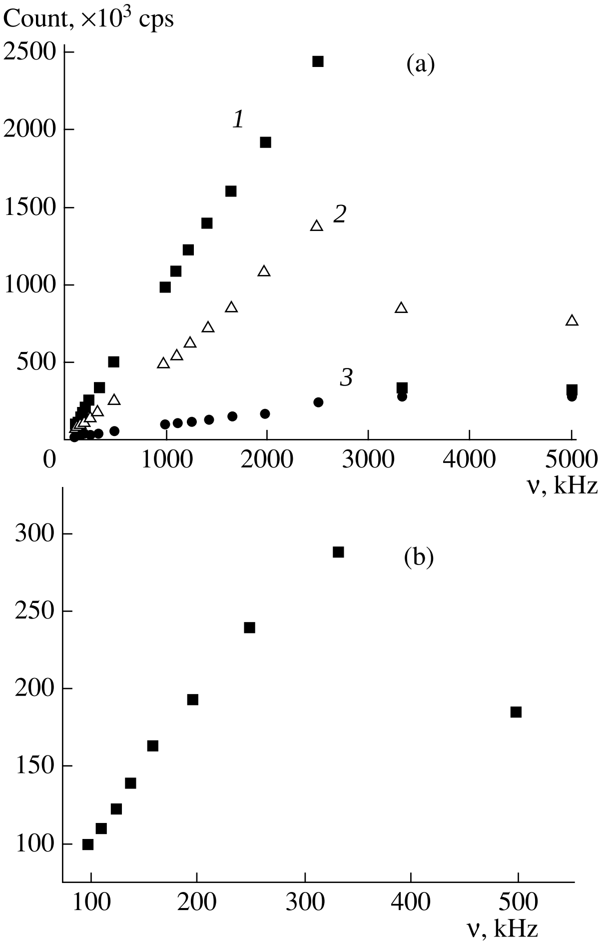


Fig. 4. Count rate of the single-photon counter vs. laser pulse repetition rate ν. (a) The maximal count rate is set at a level of 2.5×10^6 cps: (*1*) a laser pulse is of high intensity and is detected with a probability of 1; (*2*) a laser pulse has an intensity lower than in case *1* and is detected with a probability of 0.5; and (*3*) a laser pulse is still less intense and is detected with a probability of 0.1, which roughly corresponds the mean number of photons per pulse $\bar{n} = 0.2$. (b) The maximal count rate is set at a level of 2.5×10^5 cps. The laser pulse intensity corresponds to a detection probability of 1.

wise, they were disregarded. According to the BB84 protocol, such a procedure generates a matched random secured key.

The key generation rate depends on the laser pulse repetition rate, number n of photons per pulse, and APD speed. In our experiment, the key generation rate was controlled by the rate of data exchange between the computer and transmitter/receiver, which corresponded to a data transfer rate (pulse repetition rate) of 100 kHz.

Following are numerical results on quantum key generation that were obtained in our experiment. In the transfer of $\bar{n} \approx 0.1$ pulses per 10^6 syncs, a key 21303 bit long was formed. Of this number of bit, 209 bit (0.98%) were erroneous (Alice’s and Bob’s bit values diverged). For $\bar{n} \approx 0.2$, the length of the key was 38578 bit (371 bit or 0.96% were invalid). For a clock frequency of 100 kHz used in the experiment, the key generation rate was, respectively, ≈2.1 and 3.8 kbit/s. The low error rate observed in our work (compared with the results in [16]) is explained by the absence of noise and signal losses in the optical data channel. Using the same setup, we simulated an unauthorized capture of all the photons by the detector of an outside listener and his attempts to transfer the data captured to Bob. When analyzing the resulting code, which came over an unsecured channel, we immediately revealed that the error rate in the code had increased and, thus, disclosed the fact of penetration into the quantum transmission line.

Further research in this field seems to be aimed at applying long-range atmospheric and fiber-optic communication lines and increasing the data transfer rate.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, grant no. 04-07-90432).

REFERENCES

1. C. E. Shannon, Bell Syst. Tech. J. **28**, 658 (1949).
2. C. H. Bennet, Phys. Rev. Lett. **68**, 3121 (1992).
3. W. K. Wooters and W. H. Zurek, Nature **299**, 802 (1982).
4. C. H. Bennet and G. Brassard, in *Proceedings of the IEEE International Conference on Computer Systems and Signal Processing, Bangalore, 1984*, pp. 175–179.
5. C. H. Bennet, F. Bessette, G. Brassard, *et al.*, J. Cryptology **5**, 3 (1992).
6. A. K. Ekert, Phys. Rev. Lett. **67**, 661 (1991).
7. N. Gisin, G. M. Ribordy, W. Title, *et al.*, Rev. Mod. Phys. **74**, 145 (2002).
8. S. N. Molotkov, Zh. Éksp. Teor. Fiz. **124**, 1172 (2003) [JETP **97**, 1053 (2003)].
9. S. N. Molotkov, Pis’ma Zh. Éksp. Teor. Fiz. **78**, 194 (2003) [JETP Lett. **78**, 597 (2003)].
10. S. F. Yelin and B. C. Wang, quant-ph/0309105.
11. K. Inoue, E. Waks, and Y. Yamamoto, Phys. Rev. Lett. **89**, 037902 (2002).
12. K. Inoue, E. Waks, and Y. Yamamoto, Phys. Rev. A **68**, 022317 (2003).
13. D. Stucki, N. Gisin, O. Guinnard, *et al.*, New J. Phys. **4**, 41.1 (2002).
14. H. Kosaka, A. Tomita, Y. Nambu, *et al.*, Electron. Lett. **39**, 1119 (2003).
15. T. Kimura, Y. Nambu, T. Hatanaka, *et al.*, quant-ph/0403104.
16. R. J. Hughes, J. E. Nordholt, D. Derkacs, and C. G. Peterson, New J. Phys. **4**, 43.1 (2002).
17. C. Kurtsiefer, P. Zarda, M. Halder, *et al.*, Nature **419**, 450 (2002).
18. J. G. Rarity, P. M. Tapster, P. M. Gorman, and P. Knight, New J. Phys. **4**, 82.1 (2002).
19. EG&G Data Sheet C30902S.
20. M. Ghioni, S. Cova, F. Zappa, *et al.*, Rev. Sci. Instrum. **67**, 3440 (1996).
21. S. Cova, M. Ghioni, and A. Laciata, Appl. Opt. **35**, 1956 (1996).

Translated by V. Isaakyan

Technical Physics, Vol. 50, No. 6, 2005, pp. 732–735. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 6, 2005, pp. 59–62.
Original Russian Text Copyright © 2005 by Shvayko, Krechetov, Aduev, Gudilin, Serov.

OPTICS, QUANTUM ELECTRONICS

Experimental Setup for Studying the Spectral–Kinetic and Space–Dynamic Characteristics of Explosive Decomposition of Energetic Materials

V. N. Shvayko*, A. G. Krechetov*, B. P. Aduev, A. V. Gudilin***, and S. A. Serov*****
** Kemerovo State University, ul. Krasnaya 6, Kemerovo, 650043 Russia*
e-mail: lira@kemsu.ru
*** Institute of Solid-State Chemistry and Mechanochemistry, Kemerovo Branch,*
Russian Academy of Sciences, Kemerovo, 650099 Russia
e-mail: filial@kemnet.ru
**** Unitary State Enterprise All-Russia Research Institute of Opticophysical Measurements,*
Moscow, 119361 Russia
e-mail: ross_ltd@mtu-net.ru

Received July 9, 2004

Abstract—An experimental setup for studying spectral–kinetic and space–dynamic characteristics of explosion-induced luminescence from energetic materials is described. Explosion is initiated by nanosecond- and picosecond-wide electron and laser beams. Explosive luminescence of a single sample is detected in a spectral interval 450 nm wide with a spectral resolution of 2 nm and a spatial resolution of 50 μm. The time resolution, which is determined by a radiation source, is 30 ps. © *2005 Pleiades Publishing, Inc.*

Analysis of the spectral and kinetic characteristics of insulator explosion using time-resolved optical spectroscopy has provided rich information regarding the processes of formation, migration, and recombination of electronic excitations and primary defects produced by pulsed ionizing radiation [1, 2].

The application of similar experimental methods for studying the explosive decomposition of solid energetic materials also seems to be promising, since luminescence is initialized even in the solid state and then is observed in explosion products [3]. The spectral and kinetic characteristics of this luminescence provide important information on the chemical reaction leading to the detonation of explosives. However, the pulsed optical spectroscopy instrumentation used in [1, 2] (including basically a monochromator, photoelectron multiplier, and oscilloscope) needs substantial upgrading when applied to studying the explosive decomposition of solid energetic materials. In this case, one has to extract the spectral and kinetic characteristics from one sample in a spectral interval as wide as possible, since an initializing pulse applied to a single object causes its explosion, while using many samples is inappropriate because of the stochastic nature of the explosion.

The most promising way to solve this problem is the employment of a spectrograph and streak camera as basic elements of the detection scheme. The processes under study impose the following requirements on the detection scheme: a time interval of 10^{-11}–10^{-5} s, a spatial resolution of 20 lines per millimeter or higher, a spectral resolution of 5 nm for the spectrometer maximal aperture ratio, and a spectral interval of detection as wide as possible. In addition, the instrument must be equipped with a computerized data readout and processing system. Here, we describe an experimental setup meeting these requirements.

The basic elements of the setup are a source of excitation, a vacuum chamber with the sample, and a measuring channel including a synchronization system and computerized signal detection system. The signal detection system is built around a specially tailored SKh-1A spectral streak camera, which integrates a Spektr-1A polychromator and Vzglyad-2A streak camera.

The block diagram of the setup is shown in Fig. 1. Crystal holder *2* with sample *3* is mounted in vacuum chamber *1*. Excitation source *4* initiates the explosion of the sample, and lens *L* focuses the resulting radiation (luminescence) on the entrance slit of polychromator *5*, in which the radiation is decomposed into spectral components, converted, amplified, swept in time by streak camera *6*, recorded by SONY ICX039DLA CCD camera *7* (which is a part of streak camera *6*), and memorized in PC *8* for subsequent processing.

The units of the setup are synchronized with CCD camera *7* and G5-56 pulse generator *9*. The synchronization accuracy is ±1 ns. Vacuum chamber *1* is evacuated to 10^{-4} Pa.

Crystal holder *2* can be fixed to the flange of KVO.1575.00.000 microcryogenic setup *10*. In this case, the sample temperature can be varied from 12 to 300 K.

As an excitation source, one can use, depending on the task being solved, a GIN-600 electron accelerator (an effective electron energy of 0.25 MeV, a beam current density of up to 2 kA/cm^2, and a controllable pulse duration of 3–30 ns), a GIN-540 electron accelerator with a shaping gap (an effective electron energy of 0.15 MeV, a beam current density of up to 10 kA/cm^2, and a pulse duration of 50 ps), or a YAG : Nd mode-locked laser (a pulse energy of up to 30 mJ, a radiation wavelength at the fundamental frequency of 1060 nm, and a pulse duration of 30 ps).

The time resolution of the experimental setup as a whole depends on the pulse duration of the excitation source.

The optical scheme of polychromator *5* is the classical Cherny–Turner scheme and ensures the input radiation dispersion in the spectral range from 200 to 1000 nm. When recording wide-band luminescence, the instrument discriminates a 450-nm-wide spectral interval with a spectral resolution of 2 nm and provides tuning within this spectral range. The design of the polychromator allows reconfiguration of the entrance slit in two mutually perpendicular directions, thus making it possible to vary the time and spectral resolutions of the luminescence spectra. The instrument has an aperture ratio of 1 : 5, is compact, and can be mounted directly (without intermediate optics) on the end face of the streak camera so that the image of the luminescence spectrum in the form of a narrow strip is projected onto the photocathode of the electro-optical converter (EOC) of the streak camera.

The electro-optical scheme of streak camera *6* consists of a PV-001 time-analyzing EOC and a PMU-2V image amplifier. Allowance is made for interchangeable EOCs with photocathodes intended for spectral intervals of 250–850 and 400–1200 nm. The image of the spectrum swept on the EOC screen is transferred to a photosensitive CCD camera with a wide-aperture lens.

The circuitry of the streak camera provides time-base sweeping of the luminescence spectra on the screen in the range from 0.5 ns to 20 μs with a maximal resolution of 10^{-12} s.

At the initial stage of explosive decomposition, fast sweeping is required. However, the persisting luminescence of the explosion products may superimpose on the back sweep in this case and distort the spectrum. To avoid this, the streak camera has a shutter with an operating time from 20 ns to 20 μs and operates in the following modes correlating with the power supply operation: "always open" (initially open, no initializing pulse applied to the shutter); "always closed" (initially closed, no initializing pulse applied to the shutter); "normally open" (initially open, closed for the pulse duration); and "normally closed" (initially closed, open for the pulse duration).

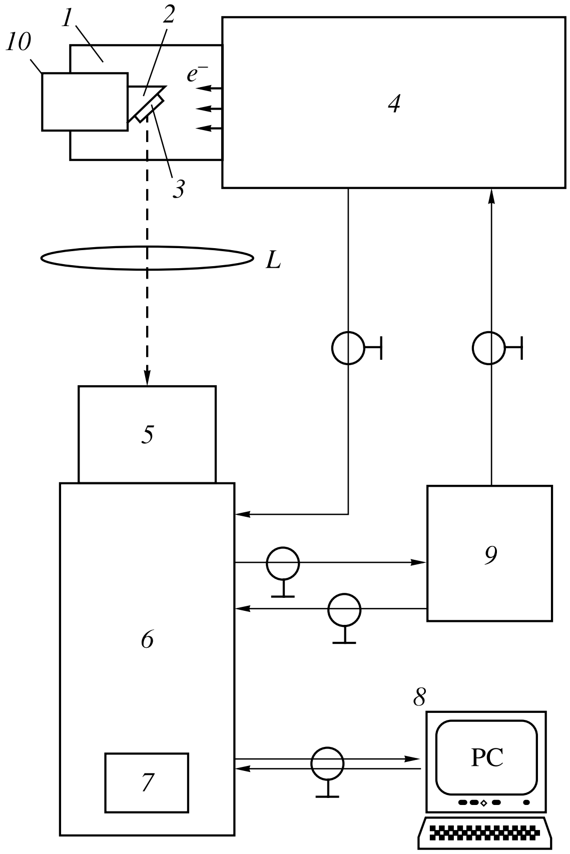


Fig. 1. Block diagram of the setup used to study the spectral and kinetic characteristics of energetic materials.

Switching the sweeping voltage on and off and combining various shutters makes it possible to use the streak camera not only to record the spectral and kinetic characteristics of the luminescence, but also to trace its dynamics in space. Moreover, it also can be applied as a photographing camera with a minimal exposure time of ≈20 ns.

A special software suite was developed for controlling the streak camera and for luminescence spectra processing [4–6]. It allows for remote control of the streak camera through both local-area (Intranet) and global (Internet) networks, which extends the potentialities of spectral–kinetic analysis of energetic materials.

The Sight-2A Server program (i) solves client queries to control the streak camera and provides interactive communication with the streak camera (on the OS Windows 95/98/ME platform), (ii) provides communication with the streak camera via the TVicHW32 generic driver (on the OS Windows 95/98/ME/NT/2000/XP platform), and (iii) transmits images of the spectra from the streak camera to the clients.

The Sight-2A Client program (i) displays the current operating parameters of the streak camera, (ii) solves client queries to change the operating param-

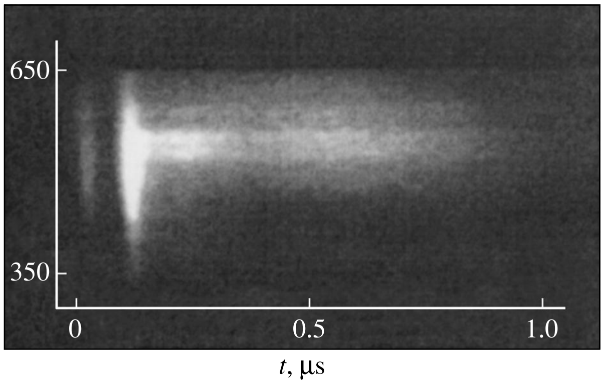


Fig. 2. Explosion luminescence spectrum of the silver azide single crystal. The source of excitation is a GIN-600 electron accelerator. The beam current density is 1 kA/cm^2.

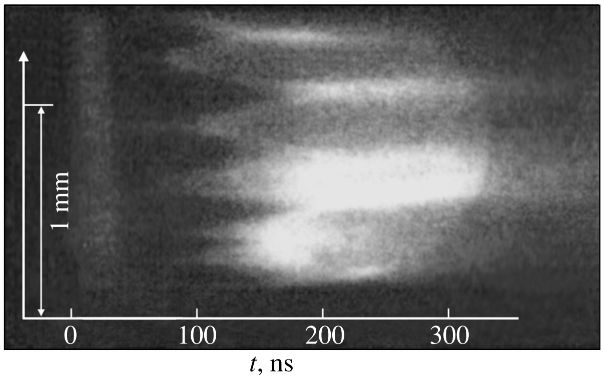


Fig. 3. Spatial dynamics of explosion luminescence from the silver azide single crystal. The source of excitation is a GIN-600 electron accelerator. The beam current density is 1 kA/cm^2.

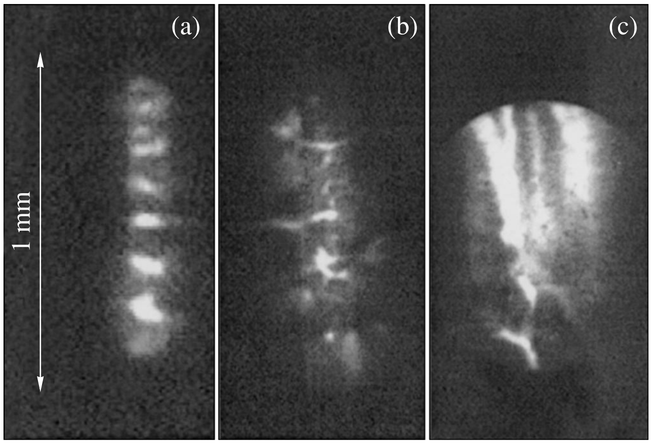


Fig. 4. Instantaneous photos of explosion luminescence from the silver azide single crystal. The source of excitation is a YAG : Nd^{3+} laser. The exposure time is 200 ns. The initializing energy density is (a) 5, (b) 15, and (c) 100 mJ/cm^2.

eters of the streak camera, (iii) takes spectrum images from the streak camera via the service program, (iv) saves the images in standard graphic formats (Bitmap, JPEG Image File, or CompuServe GIF Image), and (v) transmits the images to the handler.

The program Sight-2A intended for luminescence spectrum processing (i) receives spectrum images from the client program or reads out the spectra from the disk, (ii) normalizes the time axis by the sweep coefficients, (iii) normalizes the spectral axis by wavelength calibration (assignment of a wavelength to an image point), (iv) normalizes the spectral intensity by the sensitivity of the streak camera optics, (v) converts the spectra to the energy coordinates, (vi) displays the results of computation in the 3D coordinate system using the OpenGl technology, and (vii) transmits the computation results to Microsoft Excel (using the OLE technology).

The hardware–software environment described above was employed to study the spectral–kinetic characteristics of explosive luminescence from a silver azide single crystal with a GIN-600 electron accelerator as a source of excitation.

The luminescence pattern recorded from the EOC screen is shown in Fig. 2. It can be seen that the initializing pulse gives rise to a continuous luminescence spectrum, which is associated with radioluminescence from the crystal. The radioluminescence is followed by the continuous spectrum second component, which is associated with pre-explosive luminescence due to the onset of the explosive decomposition reaction [3]. Late in the process, the explosion of the sample is accompanied by a plasma glow, on which the lines of explosion product luminescence are imposed.

The experimental setup described above can be used not only for studying the spectral and kinetic characteristics of luminescence induced by explosion of energetic materials but also for tracing the spatial dynamics of the luminescence. Data of this kind are extremely important, since it was shown (at least, for azides of heavy metals) that the kinetics of pre-explosion luminescence is to some extent similar to the process kinetics at the initial stage of the chain reaction of explosive decomposition. Consequently, pre-explosion luminescence may serve as an indicator of this reaction [3].

To perform such experiments, the experimental scheme shown in Fig. 1 should be modified. Sample *3*, in this case, is a filamentary crystal whose axis runs parallel to the EOC photocathode and normally to the sweep direction in streak camera *6*. Polychromator *5* is removed, and the sample image magnified is projected onto the plane of the EOC photocathode. In such a configuration, the time resolution depends on the sample width and the sweep rate of the streak camera. The spatial resolution determined with a reference object (myrrh) is 50 μm.

Figure 3 shows the result for the silver azide crystal. At the instant of initialization, a uniform glow of the sample associated with radioluminescence is observed. Subsequently, luminescence centers expand and ultimately cover the entire sample. The propagation rate of the reaction, ≈1500 m/s, can be estimated from the angle of divergence of the luminescent cones.

Figure 4 demonstrates the potentialities of the experimental setup operating in the instantaneous photography mode. In this case, a picosecond laser is used as a source of excitation. The photographs are taken for three initializing energies at an exposure time of 200 ns. When the initializing energy density is low, the reaction is seen to start at separate points. At a high initializing energy density, the sample explodes (at the given exposure time).

ACKNOWLEDGMENTS

This work was financially supported by the Ministry of Industry and Science of the Russian Federation and by grant no. MK-1348.2004.3 of the President of the Russian Federation.

REFERENCES

1. D. I. Vaĭsburd, *High-Power Solid-State Electronics* (Nauka, Novosibirsk, 1982) [in Russian].
2. É. D. Aluker, V. V. Gavrilov, R. G. Deich, and S. A. Chernov, *Fast Radiation-Induced Processes in Alkali Halide Crystals* (Zinatne, Riga, 1987) [in Russian].
3. Yu. A. Zakharov, É. D. Aluker, B. P. Aduev, *et al.*, *Pre-explosion Phenomena in Azides of Heavy Metals* (TsÉI *Khimmash*, Moscow, 2002) [in Russian].
4. V. N. Shvayko, RF Inventor’s Certificate No. 2,004,610,837 (2004).
5. V. N. Shvayko, RF Inventor’s Certificate No. 2,004,610,835 (2004).
6. V. N. Shvayko, RF Inventor’s Certificate No. 2,004,610,836 (2004).

Translated by N. Wadhwa

Technical Physics, Vol. 50, No. 6, 2005, pp. 736–741. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 6, 2005, pp. 63–68.
Original Russian Text Copyright © 2005 by Lysak, Trofimov.

OPTICS,
QUANTUM ELECTRONICS

Catastrophic Change in the Efficiency of Second Harmonic Generation by a High-Intensity Femtosecond Pulse under Phase Mismatch

T. M. Lysak and V. A. Trofimov
Moscow State University, Vorob'evy Gory, Moscow, 119899 Russia
e-mail: vatro@cs.msu.ru

Received July 19, 2004

Abstract—Earlier, computer simulation discovered the effect of catastrophic and asymmetric change in the efficiency of second harmonic generation by high-intensity femtosecond pulses under phase mismatch. Based on this effect, a simple method for experimentally verifying the existence of the frequency conversion bistable mode predicted earlier is proposed. The results of computer simulation are corroborated by analytical dependences obtained in the long-pulse approximation. © *2005 Pleiades Publishing, Inc.*

INTRODUCTION

The problem of frequency conversion and, in particular, of generating the second harmonic of high-intensity femtosecond pulses remains topical [1–6], because high-efficiency (above 60%) frequency conversion techniques of practical value are lacking. Strong self-action, typical of propagation of high-intensity femtosecond laser pulses, is the cause of low-efficiency frequency conversion under phase and group synchronism. At the same time, it has been shown in our recent works [7–10] that, if certain conditions are satisfied, a 60–80% energy conversion efficiency can be achieved through frequency doubling in optical fibers (or by using wide-aperture beams), with the quality of a second-harmonic pulse being good. In particular, when self-action effects are significant, phase synchronism breaking may increase the conversion efficiency of the high-intensity femtosecond pulse frequency by one order of magnitude [8].

It should be noted that all the methods proposed by us rely on the ambiguous (bistable) dependence of the efficiency of generating the second harmonic by high-intensity femtosecond pulses on the initial conditions of interaction between optical waves. Specifically, along with the usual generation mode of low efficiency, which is typical of physical experiments where radiation, for example, at the double frequency does not enter a nonlinear medium, there exists another, high-efficiency, mode of second harmonic generation. However, to check the theoretical speculations, it is necessary to conduct physical experiments that, in our opinion, are too sophisticated. If experimentalists are not certain that the methods of sufficiently increasing the generation efficiency (see [7–10]) are feasible (and do exist!), relevant experiments can hardly be performed. Therefore, it would be appropriate to devise a simpler (from the standpoint of physical experiment) method to check the conclusions based on the approach being developed by the authors. Tackling this problem in this paper, we describe a method derived from the effect of catastrophic decrease in the efficiency of fundamental-to-second harmonic energy conversion when the system is detuned from phase synchronism.

The term "catastrophic" appearing in the title of this paper stresses that we are dealing with a sharp drop in the frequency conversion efficiency when the phase mismatch changes insignificantly about a certain value—the situation observed under definite conditions of wave interaction. Such conditions are typical of optical bistability problems (and of the theory of catastrophes in general), which consider the system's switching from one state to another.

BASIC EQUATIONS

The system of dimensionless equations that describe the generation of the second harmonic by a femtosecond pulse with allowance for its self-action has the form

$$\frac{\partial A_1}{\partial z} + iD_1\frac{\partial^2 A_1}{\partial \eta^2} + iD_\perp\Delta_\perp A_1 + i\gamma A_1^* A_2 e^{-i\Delta kz}$$

$$+ i\alpha_1 A_1(|A_1|^2 + (1+\beta)|A_2|^2) = 0, \quad 0 < z \le L_z,$$

$$\frac{\partial A_2}{\partial z} + \nu\frac{\partial A_2}{\partial \eta} + iD_2\frac{\partial^2 A_2}{\partial \eta^2} + i\frac{D_\perp}{2}\Delta_\perp A_2 \qquad (1)$$

$$+ i\gamma A_1^2 e^{i\Delta kz} + i\alpha_2 A_2((1+\beta)|A_1|^2 + |A_2|^2) = 0,$$

$$\Delta_\perp = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right), \quad \alpha_2 = 2\alpha_1 = 2\alpha.$$

Here, η is the dimensionless time in the coordinate system related to the fundamental harmonic pulse; z is the normalized longitudinal coordinate;

$$D_j \sim -0.5\frac{\partial^2 \bar{k}_j}{\partial \bar{\omega}_j^2}$$

are the coefficients describing the group velocity dispersion; $\bar{k}_j$ and $\bar{\omega}_j$ are, respectively, the dimensional wavenumber and frequency of a jth wave; $D_\perp$ is the coefficient allowing for diffraction of light (D equals the ratio between the length of the medium and quantity $2k_1a_1^2$, which is four times as large as the diffraction length of the beam at the fundamental frequency); a_1 is the physical dimension of the beam; r is the transverse coordinate normalized by a_1; γ is the coefficient of nonlinear coupling between interacting waves; $\Delta k = k_2 - 2k_1$ is the dimensional phase mismatch; α_j are the coefficients of self-action of the waves; A_j are the complex amplitudes of the harmonics ($j = 1, 2$) normalized by the maximum amplitude of the first harmonic at the initial cross section of the medium ($z = 0$); ν is a parameter proportional to the difference in the inverse group velocities of the second and fundamental harmonics; L_z is the length of the nonlinear medium; and β is a parameter that characterizes the contribution of different permittivity gratings induced by interacting waves to the self-action of a given wave. Parameter β varies from 0 to 1, depending on the lifetime of the gratings induced. It should be remembered that, for example, terms $(A_2A_2^*)A_1$, $(A_2A_1)A_2^*$, and $(A_2^*A_1)A_2$ are the contributors to term $|A_2|^2A_1$. The products of the wave amplitudes in the parentheses correspond to the permittivity gratings at different spatial frequencies. The lifetimes of the gratings may therefore be different. A more general form of the terms in Eqs. (1) that allows for different contributions of the permittivity gratings to the self-action of the wave can be found elsewhere [1].

At the entrance to the nonlinear medium, the initial distribution of the pulse at the fundamental frequency is specified,

$$A_1(z=0,\eta,r) = A_0(\eta,r), \quad 0 \le \eta \le L_t,$$
$$0 \le r \le R,$$

where R is the transverse dimension of the domain under study and L_t is the dimensionless time of process analysis.

The second harmonic amplitude in this initial cross section is zero, $A_2(z = 0, \eta, r) = 0$. It should also be noted that the interacting wave amplitude distributions considered here are finite and the amplitudes are limited on the axis. Therefore, the initial and boundary conditions for Eqs. (1) are written as follows:

$$A_j\big|_{\eta=0,L_t} = r\frac{\partial A_j}{\partial r}\bigg|_{r=0} = A_j\big|_{r=R} = 0; \quad j = 1, 2.$$

In physical experiments, initial distribution $A_0(\eta, r)$ of the fundamental harmonic of the pulse is usually approximated by the Gaussian waveform and beam profile,

$$A_0(\eta,r) = A_{10}\exp(-((\eta - L_t/2)/\tau)^2/2 - r^{m_r}/2).$$

The parameters are normalized in such a way that the dimensionless amplitude at the fundamental frequency equals unity, $A_{10} = 1$. Parameter m_r, characterizing the beam's profile, may take values of 2, 4, 6, and so on.

Generation of the second harmonic under the conditions of wave self-action possesses a number of invariants (see, e.g., [10]), which must be checked in computer simulation. We used conservative difference schemes, which keep these invariants constant.

Efficiency θ of fundamental-to-second harmonic energy conversion is estimated as

$$\theta(z) = \int_0^R\int_0^{L_t} r|A_2(z,\eta,r)|^2 d\eta\, dr \Big/ \int_0^R\int_0^{L_t} r|A_1(z=0,\eta,r)|^2 d\eta\, dr.$$

ANALYTICAL ESTIMATION

In [9], we considered the generation of the second harmonic in the long-pulse approximation and showed that the contribution of self-action to the phase difference between interacting waves depends on the difference in their intensities and the product of α and β,

$$\frac{da_1}{dz} = \gamma a_1 a_2 \sin\varphi, \quad \frac{da_2}{dz} = -\gamma a_1^2 \sin\varphi,$$

$$\frac{d\varphi}{dz} = 2\alpha\beta(a_2^2 - a_1^2) - \gamma\cos\varphi\left(\frac{a_1^2}{a_2} - 2a_2\right) + \Delta k,$$

$$\varphi = \varphi_2 - 2\varphi_1,$$

where a_j and φ_j are, respectively, the amplitudes and phases of the waves, $A_j = a_je^{i\varphi_j}$ ($j = 1, 2$).

If the double-frequency wave is absent at the entrance to the medium, the amplitudes of the interacting waves and their phase difference φ in the medium are related as follows:

$$2(1 - a_2^2)(\alpha\beta a_2 + \gamma\cos\varphi) + \Delta k a_2 = 0,$$

$$a_1^2 + a_2^2 = 1.$$

The first equation implies that the extremal (maximal or minimal) values of amplitude a_2 of the second harmonic are achieved at $\varphi = \pi n$. Importantly, there may be several maxima, because this equation is cubic (in general, the equation for a_2 is quartic). As the waves propagate, the second harmonic amplitude oscillates between the maximal achievable and zero values. The dependence of this maximum, which corresponds to the highest conversion efficiency, $\theta_{\max}$, in the long-pulse

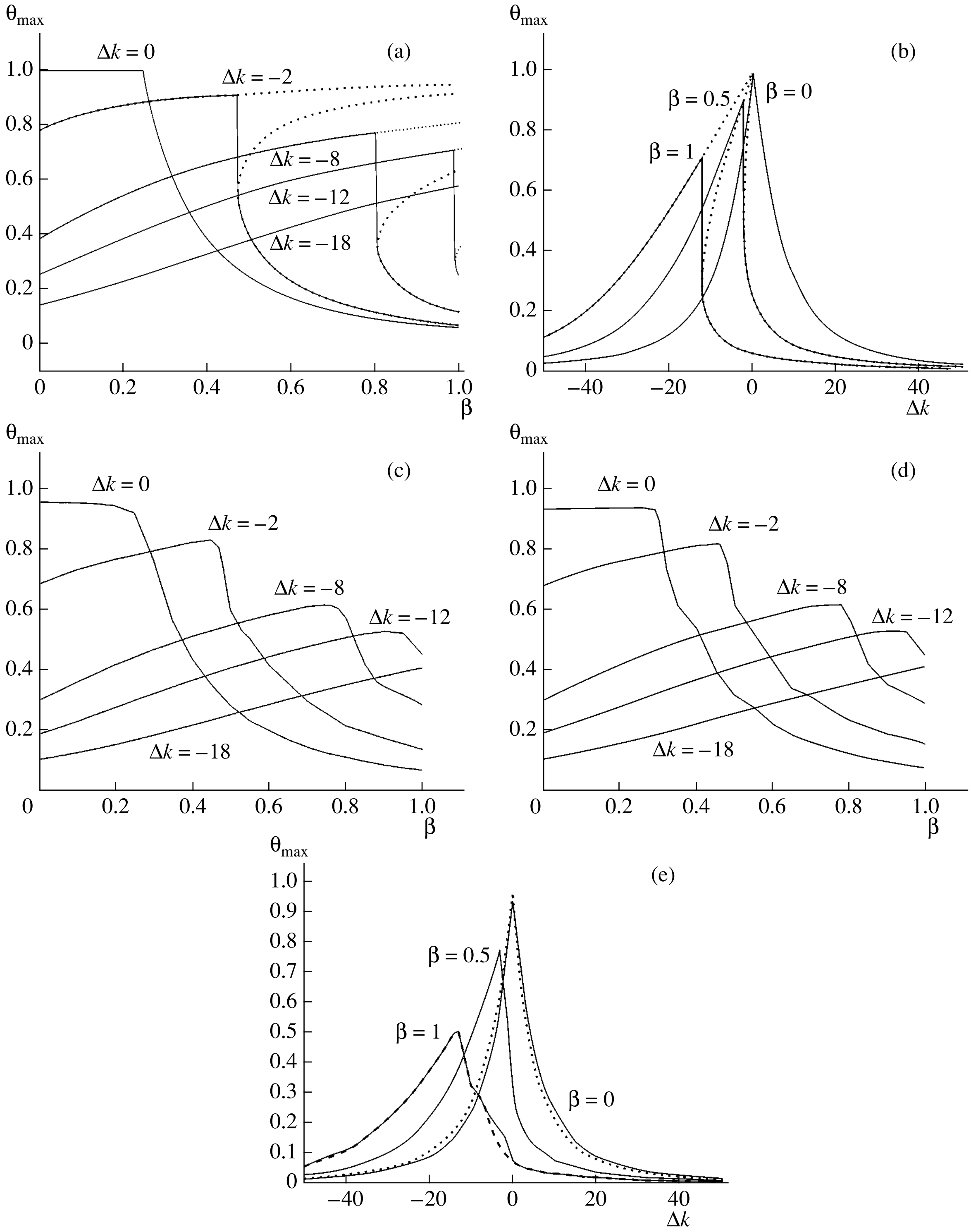


Fig. 1. Maximum efficiency θ_{max} of second harmonic generation vs. (a) β with the wave number detuning as a parameter and (b) wave number detuning with β as a parameter in the long-pulse approximation at $\alpha = 16$ and $\gamma = 4$. Also shown is the generation efficiency in the cross section $z = 1$ vs. (c, d) β and (e) Δk in the case of the Gaussian pumping pulse with $\tau = 4$ and $\nu = 0$ for (c) $D_1 = D_2 = 10^{-6}$ and (d, e) $D_1 = D_2 = 10^{-1}$ (the energy conversion takes place in the optical fiber, $D_\perp = 0$ V in Eqs. (1)). In Fig. 1e, the dotted line corresponds to $\alpha = 1$ and $\beta = 1$; the dashed line, to the cross section $z = 0.5$.

approximation, on β at constant Δk and on phase mismatch Δk at constant β is shown in Fig. 1. The variation of both parameters (β and Δk) is seen to cause sharp steps in the maximum efficiency, which are caused by self-action of the waves. The difference in the lifetimes of the induced permittivity gratings (decreasing parameter β) reduces the effect of cubic nonlinearity on the phase difference between the interacting waves, which eventually fades out at $\beta = 0$ [9]. Therefore, as far as an inserted phase difference is concerned, the decrease in β from 1 to 0 is as effective as a decrease in the peak intensity of the laser pulse fundamental harmonic in the case when all the permittivity gratings make the same contribution ($\beta = 1$).

Let us discuss in more detail the plots in Fig. 1, which are drawn in the long-pulse approximation. The sharp (catastrophic) change in the conversion efficiency in Figs. 1a and 1b stands out. This change is caused by the ambiguous dependence of $\theta_{\max}$ on β (dashed line in Fig. 1a) and Δk (dashed line in Fig. 1b). Note that the dashed lines in Fig. 1a are not closed for the interaction parameters chosen. As soon as β (or Δk) reaches the point where the efficiency is bound to switch to a higher (lower) value, this switching proceeds sharply in response to an insignificant change in the parameter (β or Δk). It is clear that, when the same crystal (medium) is used to generate the second harmonic in a physical experiment, phase mismatch Δk is varied, while β remains constant. Therefore, if the dependence of the energy conversion efficiency on the phase mismatch is asymmetric in a physical experiment, one can argue that the bistable mode of generation due to the effect of cubic nonlinearity is set and that the theoretical speculations are valid.

One should keep in mind several circumstances. Figures 1a and 1b show that, as the effect of self-action increases (i.e., β or the input intensity grows), the variation of the mismatch over a wide range does not have a significant effect on the energy conversion efficiency. In particular, if the contributions of all permittivity gratings induced equal each other ($\beta = 1$) and the model parameters are the same as above, a change in Δk to about -15 raises the frequency doubling efficiency by only about 10%, as under the phase synchronism conditions. Only when $\Delta k = -16$ does the efficiency of energy transfer into the second harmonic increase fivefold. The effect of the pulse waveform and beam profile is discussed in the next section.

PULSE PROPAGATION IN AN OPTICAL FIBER

The propagation of light pulses in a fiber is described by Eq. (1) with the only difference that the diffraction coefficients are zero and the amplitude distribution at the entrance to the nonlinear medium depends on time alone. As a result, for the Gaussian waveform, the β dependence of the frequency doubling (energy conversion) efficiency at a fixed length of wave interaction becomes smoother, i.e., does not exhibit abrupt switching of the system to the state with a low conversion efficiency (Fig. 1c). This is because the propagation path is fixed (in this case, $z = 1$), since the maximum efficiency for different portions of the pulse is achieved on different paths when the second-order dispersion is weak.

The next remark concerns the effect of the second-order dispersion on the shape of the curves. The dispersion also affects the behavior of the curves (Figs. 1d and 1e). The character of the changes depends on the amount of self-action and the wave interaction path. For example, for a medium with $\beta = 0.5$ (this value corresponds to the example considered in [2]), the Δk dependence of the maximum efficiency remains virtually the same (Figs. 1b and 1e), although the maximum value, $\theta_{\max}$, decreased by about 10%. On the contrary, at $\beta = 1$, when the contributions of all the permittivity gratings are equal to each other, the Δk dependence of $\theta_{\max}$ changes significantly: the sharp drop in the fundamental-to-second harmonic energy conversion efficiency in the given cross section of the medium is absent. However, the distinct asymmetry in the generation efficiency versus the phase mismatch dependence persists. Significantly, if the self-action due to cubic nonlinearity ($\beta = 0$) does not influence the phase mismatch or this influence is weak compared with the influence of quadratic nonlinearity (dotted line in Fig. 1e), this dependence becomes symmetric and the maximal efficiency is achieved at phase synchronism. In this case, the absence of the sharp decrease in the energy conversion efficiency in Fig. 1e (which corresponds to the cross section $z = 1$) is attributed, in particular, to backward energy conversion (Fig. 2). Therefore, at the given parameters of interaction, relevant measurements should be made for a shorter medium ($z \leq 0.25$).

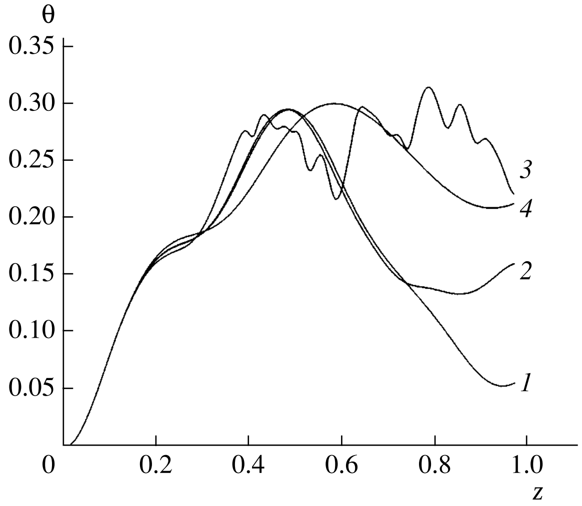


Fig. 2. Evolution of second harmonic generation efficiency θ at (*1*) weak dispersion ($D_1 = D_2 = 10^{-6}$), (*2*) medium dispersion ($D_1 = D_2 = 10^{-1}$), (*3*) strong dispersion ($D_1 = D_2 = 1$), and (*4*) anomalous dispersion ($D_1 = D_2 = -1$) for $\beta = 1$, $a = 16$, $\gamma = 4$, $\nu = 0$, $\Delta k = -8$, and $\tau = 4$.

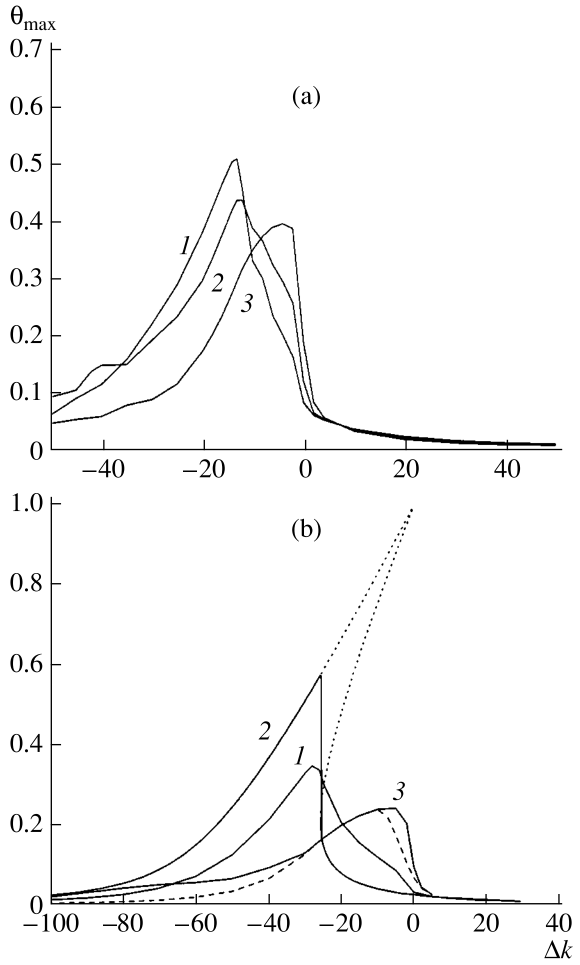


Fig. 3. Maximum efficiency θ_{max} of second harmonic generation in the cross section $z = 1$ at $\beta = 1$, $\gamma = 4$, and $\nu = 0$ vs. parameter Δk. (*1*) Gaussian ($\tau = 4$) pumping pulse in the optical fiber ($D_\perp = 0$), (*2*, Fig. 3a) wide-aperture ($D_\perp = 10^{-6}$) hyper-Gaussian ($m_r = 6$) beam, and (*3*) Gaussian ($m_r = 2$) beam at $D_1 = D_2 = 10^{-1}$ and a = (a) 16 and (b) 25 for the second harmonic generation in the optical fiber at (*1*) $D_1 = D_2 = 10^{-6}$ and (*2*) in the long-pulse approximation. The dotted line in Fig. 3b corresponds to $z = 0.5$.

GENERATION OF THE SECOND HARMONIC BY OPTICAL BEAMS

Of most practical importance is generation of the second harmonic with regard to the beam profile ($D_\perp \neq 0$ in Eqs. (1)). In this case, there arise a variety of nonlinear optical effects, which have been extensively studied experimentally in recent years [1, 2, 4–6]. Considering these effects is beyond the scope of this paper. We will only touch upon the dependence of the frequency conversion efficiency on the wave number detuning for spatially inhomogeneous beams with the emphasis on the practically important case that can be used for checking the dependences being discussed, namely, when the diffraction length for a wave incident on the nonlinear medium far exceeds the length of the nonlinear medium ($D_\perp \ll 1$). Under these conditions, the optimum phase relationship between interacting waves is not disturbed by diffraction. Therefore, the Δk dependences of the energy conversion efficiency in this case are similar to those described above, which is illustrated in Fig. 3.

Note that, for spatially inhomogeneous beams, sharp switching of the system to the state with a low frequency conversion efficiency takes place as Δk increases in the vicinity of $\Delta k = 0$. This effect is observed on sufficiently long wave interaction paths (in our case, $z = 1$) over a wide range of the self-action parameter. For shorter interaction paths, the maximum of the conversion efficiency shifts toward negative Δk. A similar effect is observed when the influence of cubic nonlinearity is enhanced (Fig. 3b).

CONCLUSIONS

Thus, we described feasible physical experiments aimed at verifying the earlier inference [7–10] that generation of the second harmonic by high-intensity femtosecond pulses exhibits a bistability showing up in the presence of high-efficiency and low-efficiency generation modes. Accordingly, the frequency conversion efficiency versus the phase mismatch dependence is asymmetric, this asymmetry arising only when cubic nonlinearity makes unequal contributions to the equations for interacting waves, causing self-action of the waves. Importantly, under certain conditions imposed on the length of the medium and pulse duration that are easy to implement in physical experiments, this dependence exhibits a sharp change in response to an insignificant change in Δk. Such behavior is typical of the effect of optical bistability and is predicted by the theory of catastrophes. That is why (we emphasize once more) the word *catastrophic* has appeared in the title of this paper.

Our calculations also explain why the variation of the phase mismatch over a wide range does not improve the frequency conversion efficiency when the second harmonic is generated by high-intensity femtosecond pulses. The reason is just the establishment of the bistable generation mode.

To conclude, we believe that this paper will be of interest for experimentalists and would be appreciative of receiving information on results obtained in physical experiments. We are also ready to perform computation adequate to a particular physical experiment.

ACKNOWLEDGMENTS

This work was partially supported by the Russian Foundation for Basic Research, project no. 02-01-727.

REFERENCES

1. T. Ditmire, A. M. Rubenchik, L. Eimer, and M. D. Perry, J. Opt. Soc. Am. B **13**, 649 (1996).
2. X. Lui, K. Beckwitt, and F. Wise, Phys. Rev. E **62**, 1328 (2000).
3. A. V. Lukashev, S. A. Magnitskiĭ, and V. I. Pryalkin, Izv. Ross. Akad. Nauk, Ser. Fiz., No. 12, 123 (1995).
4. Z. Zheng, A. M. Weiner, K. R. Paramatswaran, *et al.*, J. Opt. Soc. Am. B **19**, 839 (2002).
5. J. Zhang, J. Y. Huang, H. Wang, *et al.*, J. Opt. Soc. Am. B **15**, 200 (1998).
6. N. Yu, S. Rurimura, K. Kiramura, *et al.*, Appl. Phys. Lett. **82**, 3388 (2003).
7. T. M. Lysak and V. A. Trofimov, Zh. Vychisl. Mat. Mat. Fiz. **41**, 1275 (2001).
8. T. M. Lysak and V. A. Trofimov, Opt. Spektrosk. **92**, 323 (2002) [Opt. Spectrosc. **92**, 103 (2002)].
9. T. M. Lysak and V. A. Trofimov, Opt. Spektrosk. **95**, 489 (2003) [Opt. Spectrosc. **95**, 455 (2003)].
10. T. M. Lysak and V. A. Trofimov, Zh. Tekh. Fiz. **71** (11), 53 (2001) [Tech. Phys. **46**, 1401 (2001)].

Translated by A. Khzmalyan

Technical Physics, Vol. 50, No. 6, 2005, pp. 742–746. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 6, 2005, pp. 69–73.
Original Russian Text Copyright © 2005 by Bilenko, Polyanskaya, Getsman, Gorin, Neveshkin, Yaschenok.

OPTICS, QUANTUM ELECTRONICS

Effect of the Transition Layer on Ellipsometric Measurement of Nanodimensional Layers

D. I. Bilenko, V. P. Polyanskaya, M. A. Getsman, D. A. Gorin, A. A. Neveshkin, and A. M. Yaschenok
Chernyshevsky State University, Saratov, 410012 Russia
e-mail: GorinDA@info.sgu.ru

Received July 27, 2004

Abstract—The results of numerical simulation of a nanodimensional film–transition layer–absorbing substrate structure are presented. It is found that the transition layer affects the accuracy of determining the refractive index and thickness of the nanodimensional coating. It is shown that the introduction of the effective values of the refractive index and absorption coefficient of the substrate improves the accuracy of ellipsometric measurements of the nanodimensional film parameters. Physical (full-scale) and numerical experiments demonstrate that, when the thickness and refractive index of a nanodimensional film on a substrate with an unknown transition layer comparable in thickness with the film are measured, it is appropriate to replace the substrate–transition layer structure by a substrate with effective optical parameters. It is found that a change in the thickness of the transition layer does not noticeably affect the accuracy of determining the thickness and refractive index of the film deposited when the effective values of the refractive index and absorption coefficient of the substrate are used. *© 2005 Pleiades Publishing, Inc.*

INTRODUCTION

At present, extensive research aimed at designing and fabricating nanodimensional electronic and optoelectronic devices is underway [1–3]. This has become possible owing to the development of advanced methods for obtaining nanodimensional structures, such as MBE, MOCVD, the Langmuir–Blodgett method [4, 5], and layer-by-layer absorption of oppositely charged macromolecules (the so-called polyionic assembly) [1]. In these methods, precise measurement of the optical constants and thickness of the structure formed is a critical issue [1–3]. Ellipsometry, offering a high sensitivity and accuracy and also providing *in situ* measurements, is being widely used for this purpose [6–14].

It is well known that ellipsometric accuracy depends on a variety of factors, of which the adequacy of a chosen model to an object being measured is of primary importance [9–12]. In calculating optical constants, one often uses known values of the refractive index of a clean substrate. Such an approach may frequently turn out to be invalid, since substrates almost without exception are covered by a transition layer (in particular, by natural oxide in the case of silicon substrates [7, 10]). The Langmuir–Blodgett technology is most prone to the formation of the transition layer, since (i) monolayers prepared are transferred in air [4], (ii) hydrophobic coatings are often used [4, 5], and (iii) the layer may form as a result of substrate–water interaction when the deposition from the water subphase takes place [4]. These facts may noticeably affect the measurements of the film parameters. In light of this, to estimate and minimize this effect seems to be a topical problem in ellipsometric studies.

MODEL AND COMPUTATION TECHNIQUE

In our computational experiments, the test structure consisted of a silicon substrate, transition layer, and film to be measured (Fig. 1).

We solved the direct and inverse problems of ellipsometry. It is known that the basic equation of ellipsometry relates polarization angles Ψ and Δ to the reflection coefficients of the *s* and *p* components of incident radiation,

$$\rho = \frac{R^p}{R^s} = \tan\Psi\exp(i\Delta). \tag{1}$$

To calculate the reflection coefficients, we applied the matrix method in which zero-thickness spacings with the optical parameters of the environment are inserted between the layers [13, 14]. With this method and the model described above, one can derive general-

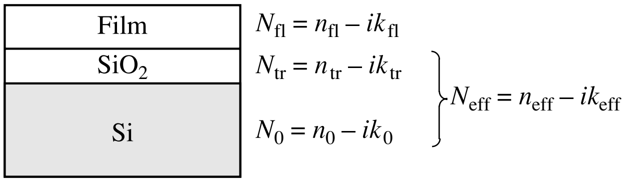


Fig. 1. Model of the structure used in the calculations.

1063-7842/05/5006-0742$26.00 © 2005 Pleiades Publishing, Inc.

ized formulas making it possible to solve the direct problem of ellipsometry for multilayer structures with an infinitely large number of layers,

$$r_m^{s,\,p} = \left[\frac{F_m + \dfrac{r_{m-1} - F_m}{1 - r_{m-1} - F_m}\exp(-2i\chi_m)}{1 + F_m\dfrac{r_{m-1} - F_m}{1 + -r_{m-1}F_m}\exp(-2i\chi_m)}\right]^{s,\,p}, \qquad (2)$$

$$\chi_m = \frac{2\pi}{\lambda}\sqrt{N_m^2 - N_c^2\sin^2(\Theta)}\,d_m, \qquad (3)$$

where N_c is the refractive index of the environment ($N_c = 1$ for air); Θ is the angle of incidence of the radiation on the structure; λ is the wavelength; N_m and d_m are the complex refractive index and the thickness of an mth layer, respectively; and F_m are the Fresnel coefficients at the interface between the mth layer and the environment,

$$F_m^p = \frac{N_m^2\cos(\Theta) - N_c^2\sqrt{N_m^2 - N_c^2\sin^2(\Theta)}}{N_m^2\cos(\Theta) + N_c^2\sqrt{N_m^2 - N_c^2\sin^2(\Theta)}}, \qquad (4)$$

$$F_m^s = \frac{N_c^2\cos(\Theta) - \sqrt{N_m^2 - N_c^2\sin^2(\Theta)}}{N_c^2\cos(\Theta) + \sqrt{N_m^2 - N_c^2\sin^2(\Theta)}}. \qquad (5)$$

In the calculations, the substrate was assumed to be the medium with $m = 0$.

In the direct problem, the polarization angles for the reflection of the radiation from the substrate–surface oxide (Ψ_1, Δ_1) and substrate–transition layer–nanodimensional film (Ψ_2, Δ_2) structures were calculated.

In the simulation, refractive index n_0 and absorption coefficient k_0 of the substrate were taken to be $n_0 = 3.86$ and $k_0 = 0.02$ for $\lambda = 632.8$ nm (this wavelength is typical of an LÉF-3M ellipsometer used in the measurements), the angle of incidence was $\Theta = 70°$ (the value commonly used in ellipsometric measurements on silicon plates), the refractive index of the oxide simulating the transition layer was taken to be $n_{\mathrm{tr}} = 1.46$ and 1.54 [10, 15] and oxide layer thickness d_{tr} was varied from 0 to 23 nm. The smaller the thickness d_{fl} of the film applied, the stronger an effect of the transition layer may be expected. For this reason, the thickness of the film was varied from 10 to 20 nm. Refractive index n_{efl} of the film was varied from 1.05 to 1.8. When selecting the value of refractive index n_{efl} of the film, we took into account the fact that Langmuir–Blodgett films contain primarily molecules of biphyllic compounds, for which the average refractive index equals 1.5, being dependent on the presence of alkyl radicals [4, 16, 17].

The influence of the transition layer on the accuracy of determining the refractive index and thickness of the nanodimensional film, as well as the feasibility of minimizing this influence on the nanodimensional film parameters measured, was determined by comparing a solution to the inverse problem with the parameters set in the direct problem.

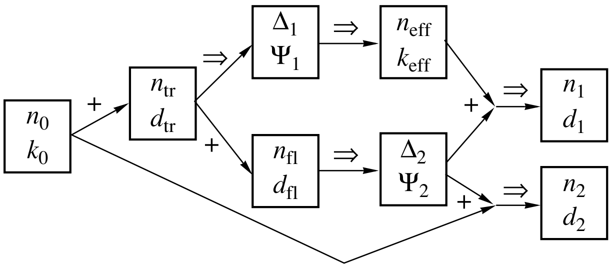


Fig. 2. Solution algorithm for the direct and inverse problem of ellipsometry: n_0 and k_0, the refractive index and absorption coefficient of the clean substrate surface (Fig. 1); n_{tr} and d_{tr}, the refractive index and thickness of the transition layer (Fig. 1); Ψ_1 and Δ_1, the polarization angles obtained from the reflection of radiation from the surface of the clean substrate–transition layer structure; n_{fl} and d_{fl}, the refractive index and thickness of the film deposited (Fig. 1); Ψ_2 and Δ_2, the polarization angles obtained from the reflection of radiation from the surface of the clean substrate–transition layer–film structure; n_{eff} and k_{eff}, the effective refractive index and absorption coefficient of the substrate (the clean substrate–transition layer structure) (Fig. 1); n_1 and d_1, the refractive index and thickness of the film that are calculated using the effective refractive index and absorption coefficient of the substrate; and n_2 and d_2 are the refractive index and thickness of the film that are obtained from the refractive index and absorption coefficient of the clean surface of the substrate.

The inverse problem was solved using the Holmes method [12, 18], in which the properties of the substrate must be preset. We considered two cases (Fig. 1): (i) the parameters of the substrate are the same as those of a clean silicon surface ($n_0 = 3.86$, $k_0 = 0.02$), so that the presence of the transition (surface) layer is disregarded, and (ii) the optical constants of the substrate are replaced by effective parameters (viz., refractive index n_{eff} and absorption coefficient k_{eff}) characterizing the real surface.

The values of n_{eff} and k_{eff} were calculated by the well-known formula [9–12]

$$N_{\mathrm{eff}} = n_{\mathrm{eff}} - ik_{\mathrm{eff}} = N_c = \sin(\Theta)\sqrt{\left(\frac{1-\rho_1}{1+\rho_1}\tan\Theta\right)^2 + 1}, \qquad (6)$$

where Θ is the angle of incidence of radiation, $\rho_1 = \tan\Psi_1\exp(i\Delta)$, and Ψ_1 and Δ_1 are the measured angles of polarization for the substrate–transition layer structure.

The solution algorithm for the direct and inverse problems is given in Fig 2.

RESULTS OF SIMULATION

The results of simulation are illustrated in Figs. 3–6. For comparison, Figs. 5 and 6 also show the results of

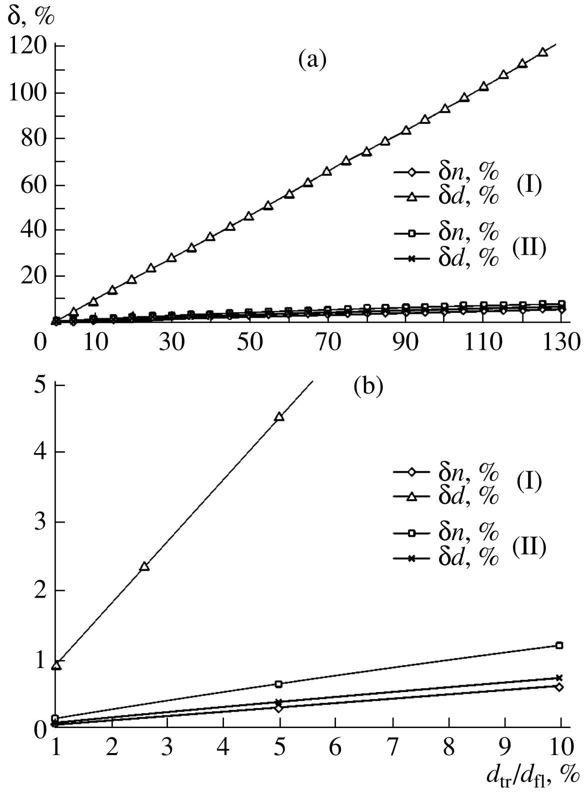


Fig. 3. (a) Effect of transition layer (oxide) thickness d_{tr} on relative accuracy δ in determining the refractive index and thickness of the nanodimensional film (for n_{tr} = 1.46): (I) without oxide and (II) with N_{eff}. (b) Enlarged fragment of Fig. 3a.

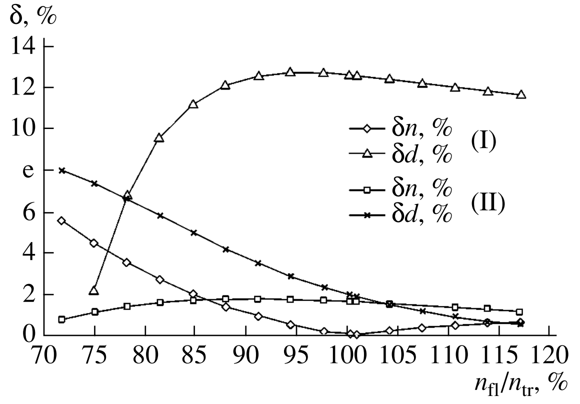


Fig. 4. Effect of the difference in the refractive indices of the transition layer, n_{tr}, and the film, n_{fl}, on relative accuracy δ in determining the refractive index and thickness of the nanodimensional film (for n_{tr} = 1.46): (I) without oxide and (II) with N_{eff}.

calculation for $n_{tr} = 3$. From the above dependences, the following conclusions can be drawn.

(1) Effective refractive index n_{eff} decreases and effective absorption coefficient k_{eff} increases with increasing thickness d_{tr} of the oxide layer simulating the transition layer (Fig. 5). For example, as thickness d_{tr} of the oxide layer grows to 15 nm ($n_{tr} = 1.46$), effective refractive index n_{eff} drops from 3.86 to 3.57, while effective absorption coefficient k_{eff} rises from 0.02 to 0.99. The presence of a transition layer on the surface of the substrate can be judged from the values of the effective refractive index and absorption coefficient.

(2) The error in determining the thickness and refractive index of the film deposited on the oxide-coated substrate depends on the oxide thickness (Fig. 3). As d_{tr} increases, so does the error. If the presence of the oxide layer is disregarded, the error in the thickness is maximal and comparable to the thickness of the oxide layer (see Fig. 3). For example, if the thickness of the oxide is only 5% of the film thickness (d_{tr}/d_{fl} = 5%), the error in the film thickness amounts to 4.5% when the refractive index and absorption coefficient are taken for the clean substrate and drops to 0.4% when the effective parameters are used. For comparable (e.g., equal) thicknesses of the oxide and film, the error in the film thickness exceeds 90% when we ignore the presence of the transition layer and is less than 6% when the effective parameters of the substrate are used.

(3) The error in the thickness and refractive index of the film depends on the difference in the refractive indices of the oxide and film (see Fig. 4). If the refractive index of the film is 1.13 times higher than that of the oxide (n_{fl}/n_{tr} = 1.13), the error in the film thickness is 1.3% if the effective parameters of the substrate are used and grows to more than 11% if the clean-substrate parameters are employed. Use of the effective parameters of the substrate also proves itself if the refractive index of the oxide is smaller than that of the film (Fig. 4). If the refractive index of the film is smaller than that of the transition layer by a factor of 0.85 (n_{fl}/n_{tr} = 0.85), the error in the film thickness approaches 7 against 10% in the case when the refractive index of the clean substrate is used.

(4) For the structures simulated, we calculated the value of Φ, the so-called effective substrate applicability criterion [19]. It was found that the applicability criterion for the effective substrate with an oxide layer 23 nm thick (n_{tr} = 1.54) is Φ = 1.01. The ellipsometric measurement error depends on the oxide-to-film thickness ratio: specifically, for the same thicknesses of the film and oxide, i.e., 23 nm (see Figs. 3 and 4), the error is 6 and 5% for the refractive index and thickness of the film, respectively.

EXPERIMENTAL RESULTS

With the aim of finding the effective parameter ranges, we studied experimentally the effect of standard silicon processing techniques on the effective parameters of wafers. The experiments were carried out on KÉF-5 (phosphorus-doped to a resistivity of 5 Ω cm) Si(111) and Si(100) single crystals.

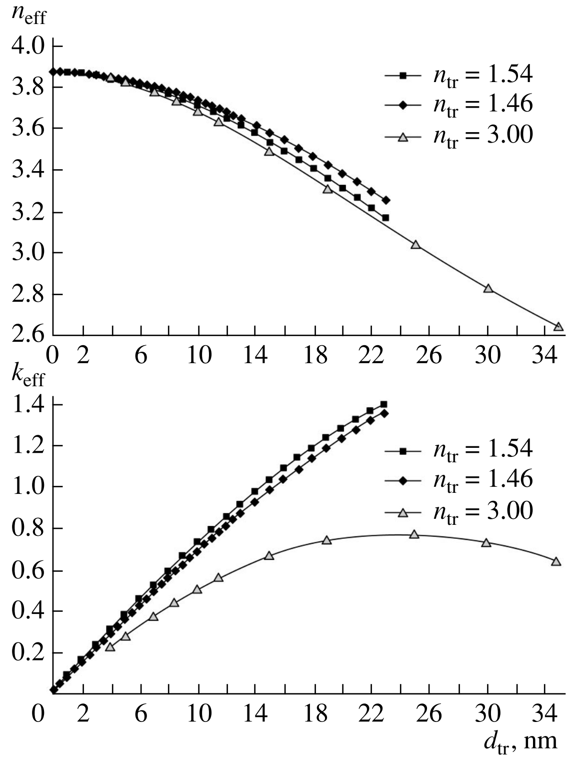


Fig. 5. Dependence of the effective refractive index and absorption coefficient of the substrate on the oxide layer thickness.

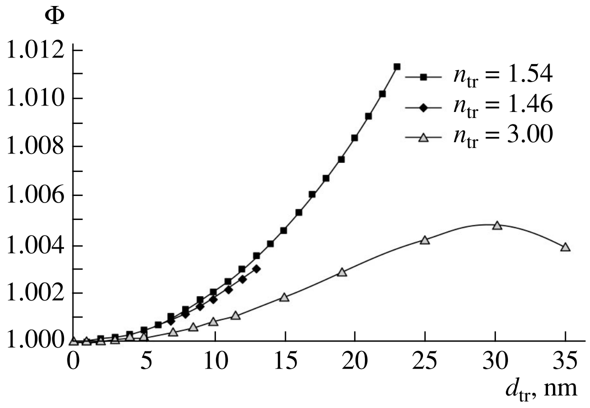


Fig. 6. Dependence of the criterion of substrate effective parameter applicability on the transition layer thickness.

The effective parameters were calculated from measured polarization angles Ψ and Δ for angles of incidence of 60° and 70°.

Prior to the measurements, the wafers were kept in air for a long time. Once the wafers had been boiled in carbon tetrachloride, they were subjected to different types of processing: two successive etchings in an aqueous solution of hydrofluoric acid, ion etching, and keeping for 30 min in distilled water.

The measurements were taken from several points on the surface and then were averaged. The table lists the average values of the refractive index of the substrate after each operation. The (100) wafers are characterized by narrower ranges of the effective refractive indices and absorption coefficients. The values of the effective refractive indices and absorption coefficients vary from 3.56 to 3.86 and from 0.20 to 0.92, respectively, for the (100) substrates and from 3.71 to 3.85 and from 0.24 to 0.72 for the (111) substrates. It is found that the effective refractive indices and absorption coefficients approach those for a clean substrate upon degreasing. In all the cases, effective substrate applicability criterion Φ is close to unity. The maximum deviation of Φ from 1, 0.0035, is observed upon etching in hydrofluoric acid (in particular, upon the second etching of the KÉF-5 (100) wafers in hydrofluoric acid).

The experimental values of n_{eff} and k_{eff} were used in a numerical experiment aimed at determining the properties of Langmuir–Blodgett films deposited on silicon substrates. The sequence of operations was the following: (i) ellipsometric measurement of parameters Ψ_1 and Δ_1 of the substrate before deposition; (ii) calculation of effective refractive index n_{eff} and absorption coefficient k_{eff}; (iii) deposition of the film and ellipsometric measurement of parameters Ψ_2 and Δ_2 of the structure; (iv) calculation of refractive index n_{fl} and thickness d_{fl} of the film deposited using the values of n_{eff}, k_{eff}, Ψ_2, and Δ_2; and (v) error analysis based on the curves obtained in the numerical experiment (Figs. 3–6).

Effect of substrate surface processing on the effective refractive index and absorption coefficient of the substrate

Processing technique	KÉF-5 [100]		KÉF-5 [111]	
	n_{ef}	k_{ef}	n_{ef}	k_{ef}
As-prepared substrate	3.856 ± 0.002	0.210 ± 0.016	3.842 ± 0.016	0.645 ± 0.018
Degreasing in CCl_4	3.858 ± 0.001	0.195 ± 0.004	3.852 ± 0.002	0.242 ± 0.025
First etching in HF	3.630 ± 0.103	0.608 ± 0.204	3.780 ± 0.029	0.464 ± 0.058
Second etching in HF	3.559 ± 0.078	0.923 ± 0.148	3.711 ± 0.015	0.723 ± 0.085
Ion etching	3.759 ± 0.005	0.788 ± 0.014	3.793 ± 0.005	0.584 ± 0.012
Keeping in distilled water	3.760 ± 0.041	0.542 ± 0.060	3.818 ± 0.014	0.372 ± 0.036

It was found that the use of n_{eff} and k_{eff} substantially reduces the errors in determining the parameters of the film deposited. The properties of the transition layer are known to vary noticeably when it is covered by a film. However, even when the layer gets thicker by a factor of 4 (from 10 to 40% of the film thickness; see Fig. 3), the error increases only by ≈1.8% if the effective parameters are used.

CONCLUSIONS

When the thickness and refractive index of a nanodimensional film deposited on a substrate with an unknown transition layer comparable in thickness with the film are to be determined, it is appropriate to replace the substrate–transition layer structure by a substrate with effective optical parameters, as follows from full-scale and numerical experiments. Using the effective parameters determined prior to deposition may cut the measurement accuracy of the thickness and refractive index of the film when the transition layer thickness changes.

REFERENCES

1. G. S. Plotnikov and V. B. Zaĭtsev, *Physical Basis of Molecular Electronics* (Mosk. Gos. Univ., Moscow, 2000) [in Russian].
2. V. P. Evtikhiev, I. V. Kudryashov, E. Yu. Kotel'nikov, *et al.*, Fiz. Tekh. Poluprovodn. (St. Petersburg) **32**, 1482 (1998) [Semiconductors **32**, 1323 (1998)].
3. D. A. Vinokurov, V. A. Kapitonov, O. V. Kovalenkov, *et al.*, Fiz. Tekh. Poluprovodn. (St. Petersburg) **33**, 858 (1999) [Semiconductors **33**, 788 (1999)].
4. M. Petty, *Langmuir–Blodgett Films: An Introduction* (Cambridge Univ. Press, Cambridge, 1996).
5. I. V. Gavrilyuk, V. I. Kazantseva, N. V. Lavrik, *et al.*, Poverkhnost, No. 11, 93 (1991).
6. V. S. Stashchuk and V. I. Shkurat, *Ellipsometry as a Method of Surface Study* (Nauka, Novosibirsk, 1983), pp. 35–37 [in Russian].
7. B. M. Ayupov, N. P. Sysoeva, and E. F. Titova, *Ellipsometry: Theory, Methods, and Application* (Nauka, Novosibirsk, 1987), pp. 136–139 [in Russian].
8. R. I. Lyubinskaya, A. S. Mardezhov, and V. A. Shvets, *Ellipsometry: Theory, Methods, and Application* (Nauka, Novosibirsk, 1987), pp. 59–67 [in Russian].
9. M. M. Gorshkov, *Ellipsometry* (Sov. Radio, Moscow, 1974) [in Russian].
10. V. I. Pshenitsyn, M. I. Abaev, and N. Yu. Lyzlov, *Ellipsometry in Physicochemical Investigation* (Khimiya, Leningrad, 1986) [in Russian].
11. *Basic Concepts of Ellipsometry*, Ed. by A. V. Rzhanov (Nauka, Novosibirsk, 1978) [in Russian].
12. V. K. Gromov, *Introduction to Ellipsometry* (Leningr. Gos. Univ., Leningrad, 1986) [in Russian].
13. *Physics of Thin Films: Advances in Research and Development* (Academic, New York, 1963; Mir, Moscow, 1967).
14. D. I. Bilenko, B. A. Dvorkin, T. Yu. Druzhinina, *et al.*, Opt. Spektrosk. **55**, 885 (1983) [Opt. Spectrosc. **55**, 533 (1983)].
15. G. A. Egorov, N. S. Ivanova, E. V. Potapov, *et al.*, Opt. Spektrosk. **36**, 773 (1974).
16. A. I. Yanklovich, *Advances in Colloid Chemistry* (Khimiya, Moscow, 1991), pp. 263–291.
17. S. N. Shtykov, B. N. Klimov, D. A. Gorin, *et al.*, Zh. Fiz. Khim. **78**, 503 (2004).
18. D. A. Holmes, Appl. Opt. **6**, 168 (1967).
19. V. A. Shvets, Candidate's Dissertation (Novosibirsk, 1988).

Translated by N. Wadhwa

Technical Physics, Vol. 50, No. 6, 2005, pp. 747–753. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 6, 2005, pp. 74–80.
Original Russian Text Copyright © 2005 by Ramazanov, Sotnikov, Tkach.

High-Power Coaxial Microwave Ubitron: Simulation by the Particle-in-Cell Method

R. Z. Ramazanov, G. V. Sotnikov, and Yu. V. Tkach
Institute for Electromagnetic Research, Kharkov, 61022 Ukraine
e-mail: sotnikov@kipt.kharkov.ua

Received July 22, 2004

Abstract—Results of complete numerical simulation of a high-power coaxial microwave ubitron are reported. The simulation uses a 2.5-dimensional numerical electromagnetic code based on the particle-in-cell method and is performed for the amplification and oscillation operating modes. The results of our numerical simulation are compared with analytical results and results following from the nonlinear theory that are obtained by partial computer simulation. *© 2005 Pleiades Publishing, Inc.*

INTRODUCTION

Investigation into electron devices based on the coherent radiation of electrons moving in periodic static electric and magnetic fields started long ago [1, 2]. Theoretical analyses [3, 4] and experimental implementations [5–7] have been devoted mostly to cylindrical ubitron configurations. A new stage in studying these devices is associated with the coaxial configuration [8–11], in which a periodic magnetic field is created by a set of permanent magnets regularly arranged on the inner and outer conductors of the coaxial line. This configuration can produce a wiggler with the longitudinal magnetic field alternating along the radius, whereas the cylindrical ubitron creates a wiggler with a spatially periodic magnetic field.

The coaxial ubitron with the alternating magnetic field offers a number of advantages over the cylindrical configuration. The first is more stable transport of the electron beam in the alternating static magnetic field of the wiggler. If the poles of the magnets that reside on the outer conductor are half-period displaced relative to those of the magnets on the inner conductor, a magnetic well is formed in the coaxial gap. When the beam particles are at the bottom of the magnetic well, the amplitude of their transverse oscillations in the magnetic field of an alternating coaxial wiggler is significantly smaller than when the beam is transported in the spatially periodic field of a cylindrical ubitron. This advantage allows the coaxial ubitron configuration to employ lower magnetic field intensities than those used in the cylindrical configuration at the same electron beam current.

The second important advantage of the coaxial ubitron is that the field of the beam's space charge in its drift chamber is weaker than that in a cylindrical one; i.e., the limiting vacuum current of the electron beam in a coaxial line is higher [12]. Therefore, the device can use a higher electron beam current and, thereby, raise the output power.

The third advantage of the coaxial ubitron is its higher efficiency compared with the cylindrical device. Comparative nonlinear analysis [13] has shown that the maximum power of the amplified microwave signal in a coaxial ubitron is higher than in its cylindrical counterpart. This fact is due to a higher resistance of coupling between the electron beam and the wave being amplified.

Designing a developmental version of a coaxial ubitron needs not only a theoretical substantiation of its operation but also elaboration of a model in terms of which the performance of the device can be estimated numerically. To date, the theory of the ubitron has been thoroughly developed [8–11, 14], unlike numerical approaches. An efficient tool for numerical calculations is the particle-in-cell method, which is widely applied to high-power microwave devices. However, few works are known where this method is applied to the coaxial ubitron. In one of them, it is used to perform complete numerical simulation of oscillation at 2.8 GHz (the MAGIC numerical code) [11]. In the simulations, the Q factor of the resonator was specified by introducing a finite-conductivity region for a required power and efficiency to be achieved.

Below, we study microwave oscillations in the coaxial ubitron under different conditions; namely, it is assumed that energy losses are due to the radiation from the open end of the waveguide.

PHYSICAL STATEMENT OF THE PROBLEM, BASIC EQUATIONS, AND NUMERICAL SIMULATION ALGORITHM

Consider a coaxial waveguide (Fig. 1) of inner radius a and outer radius b into which an annular relativistic electron beam is continuously injected. The

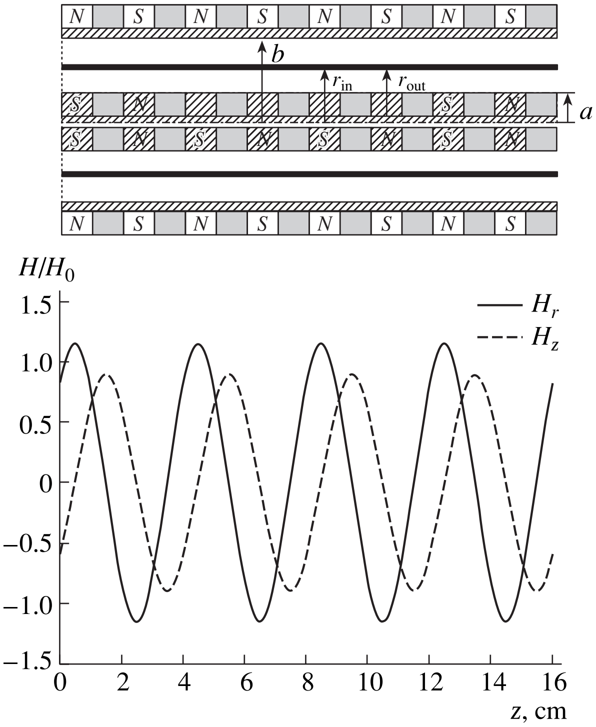


Fig. 1. Coaxial wiggler and the longitudinal distribution of magnetic fields H_r and H_z on the inner conductor of the coaxial waveguide (the first spatial harmonic is taken into account). *N* and *S* are the poles of the magnets; the shadowed regions, pole tips.

beam current is designated as I_b; the energy of beam electrons, as W_b; the inner radius of the beam, as r_{in}; and the outer radius of the beam, as r_{out}. It is assumed that the injected beam is monoenergetic; i.e., the transverse velocity components of electron beam particles are zero at the entrance to the drift chamber ($z = 0$). The left end of the waveguide ($z = 0$) is closed by a metallic wall, which is transparent for the particles and nontransparent for waves excited. On the inner and outer conductors of the coaxial waveguide, permanent magnets are regularly arranged along the axis with period L_w. The poles of the outer magnets are shifted relative to those of the inner magnets by half of the period.

The constant magnetic field produced by such a coaxial wiggler is given by [8, 14]

$$H_{0z}(\bar{r}, \zeta) = -H_0 \sum_{n=1}^{\infty} C_n \cos\left[n\left(\zeta - \frac{\pi}{4}\right)\right] F_n^{(0)}(n\bar{r}), \quad (1)$$

$$H_{0r}(\bar{r}, \zeta) = -H_0 \sum_{n=1}^{\infty} C_n \sin\left[n\left(\zeta - \frac{\pi}{4}\right)\right] F_n^{(1)}(n\bar{r}), \quad (2)$$

where $\bar{r} = k_w r$ and $\zeta = k_w(z + 0.25L_w)$ are the dimensionless coordinates, $F_n^{(0)}(n\bar{r}) = f_n I_0(n\bar{r}) - g_n K_0(n\bar{r})$, $F_n^{(1)}(n\bar{r}) = f_n I_1(n\bar{r}) + g_n K_1(n\bar{r})$, $f_n = [K_0(n\bar{r}_b) + K_0(n\bar{r}_a)]/\Delta_n$, $g_n = [I_0(n\bar{r}_b) + I_0(n\bar{r}_a)]/\Delta_n$,

$$C_n = \frac{4}{\pi n}\sin^2\frac{\pi n}{2}\sin\frac{\pi n}{4};$$

$k_w = 2\pi/L_w$, $\bar{r}_a = k_w a$, $\bar{r}_b = k_w b$, I_0 and I_1 are the Bessel functions, and K_0 and K_1 are the Macdonald functions.

The numerical simulation of the ubitron was based on the original 2.5-dimensional (three-dimensional in velocity and two-dimensional in coordinates) relativistic electromagnetic code, which uses the particle-in-cell method [15]. The trajectory of each macroparticle was calculated by solving a set of equations of motion using a predictor–corrector scheme in cylindrical coordinates. Components **E** and **H** of an electromagnetic field excited by the beam in the waveguide, which enter the expression for the Lorentz force, were calculated by TSC interpolation from the mesh nodes when the set of the Maxwell equations was solved by the finite-difference method on a rectangular grid. The Maxwell equations were solved using an explicit time- and coordinate-centered leapfrog scheme.

The boundary conditions on the metallic surfaces of the drift chamber are as follows: $E_z|_{r=a,b} = 0$, $E_\varphi|_{r=a,b} = 0$, $E_r|_{z=0} = 0$, and $E_\varphi|_{z=0} = 0$. On the right (open) end of the waveguide, the condition of radiation into free space is set: $E_r/H_\varphi|_{z=z_{out}} = 1$ and $E_\varphi/H_r|_{z=z_{out}} = -1$.

The mechanism of charge and current weighting over the discrete grid does not exactly meet the continuity equation, as a result of which numerical results are unstable. Therefore, the electromagnetic field components found were subsequently corrected by the Boris method during the solution of the Laplace equation.

NUMERICAL RESULTS

In the simulation, we used the following parameters of the ubitron: $a = 2$ cm, $b = 4$ cm, $r_{in} = 2.85$ cm, $r_{out} = 3.2$ cm, $I_b = 3$ kA, $W_b = 490$ keV, $H_0 = 3$ kOe, the length of the system $z_{out} = 96$ cm, and wiggler period $L_w = 4$ cm.

Before discussing the numerical results, it is appropriate to qualitatively estimate the range of frequencies and wavelengths generated. Figure 2 shows the dispersion curve for the injected beam and TE_{01} mode in the coaxial waveguide with the above dimensions. The microwave generation occurs when the resonance condition $\omega = (k_w + k_z)v_z$ (k_z is the wavenumber of the high-frequency mode and v_z is the electron average longitudinal velocity) is satisfied. For an initial beam energy of 490 keV, the resonant frequency is $f_1 = \omega/2\pi = 7.75$ GHz and the resonance wavenumber is $k_{z1} \approx 0.32$ cm^{-1}. This point of resonance is indicated in Fig. 2 by figure 1. As was noted above, the injection condi-

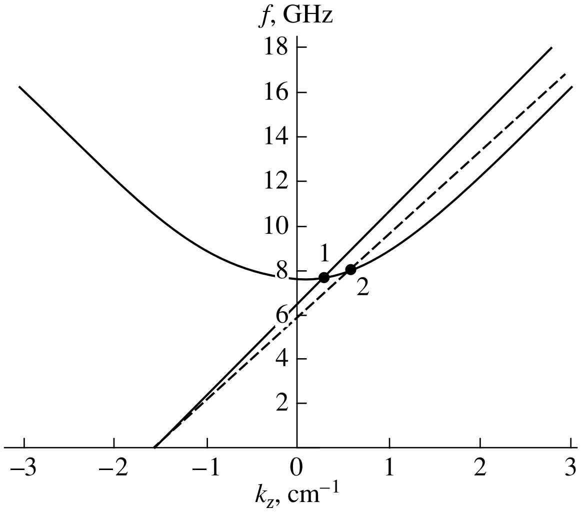


Fig. 2. Dispersion curve for the TE_{01} mode and beam in the coaxial waveguide with inner radius a = 2 cm and outer radius b = 4 cm. Resonance points 1 and 2 correspond to the initial and average electron beam velocity in the wiggler field.

tions in our numerical simulation were such that the electrons had a zero transverse velocity at the entrance to the system. When moving in the wiggler's field, they acquire a transverse velocity at the sacrifice of the longitudinal velocity. As a result, the range of resonant frequencies shifts toward higher frequencies and longitudinal wavenumbers. As follows from Fig. 2, a change in wavenumber k_z may be significant even at a small change in the electron average longitudinal velocity.

Figure 3 shows the configuration space (r, z) of the beam electrons propagating in the drift chamber at three time instants: t = 3.9 ns, the approximate time when the first electrons injected reach the end of the ubitron; t = 10 ns, the time when the amplitude of the electromagnetic field excited reaches a maximal value; and t = 20 ns, the time when transients in the ubitron fade out (further, the dynamics of the particles and fields qualitatively remains the same). As can be seen from Fig. 3, early in the injection, the electrons flow in a laminar manner, describing helical paths. The beam surface is radially modulated in the longitudinal direction with a period much longer than the wiggler period. Such dynamics is typical of particles moving in a periodic magnetic field and can be described by a Mathieu-type equation [16]. The electron azimuth velocity is longitudinally modulated with a period equal to the wiggler period (see Fig. 4 for t = 3.9 ns). As the injection proceeds, smaller scale modulations (perturbations) with a period approximately equal to the wiggler period ($L_w \approx 4$ cm) and a period equal to the spatial period of the excited wave are superimposed on the large-scale modulation with a period of about 20 cm. These perturbations propagate from the exit of the ubitron toward its entrance (Fig. 5b) and finally break the initial large-scale modulation of the beam virtually throughout the system (Fig. 5c). Although the beam transport through the coaxial waveguide remains stable (particle losses at the walls are negligible), the initial laminar flow of the particles breaks. Subsequently, the particle dynamics in the ubitron is qualitatively the same as that illustrated in Fig. 3c for t = 20 ns.

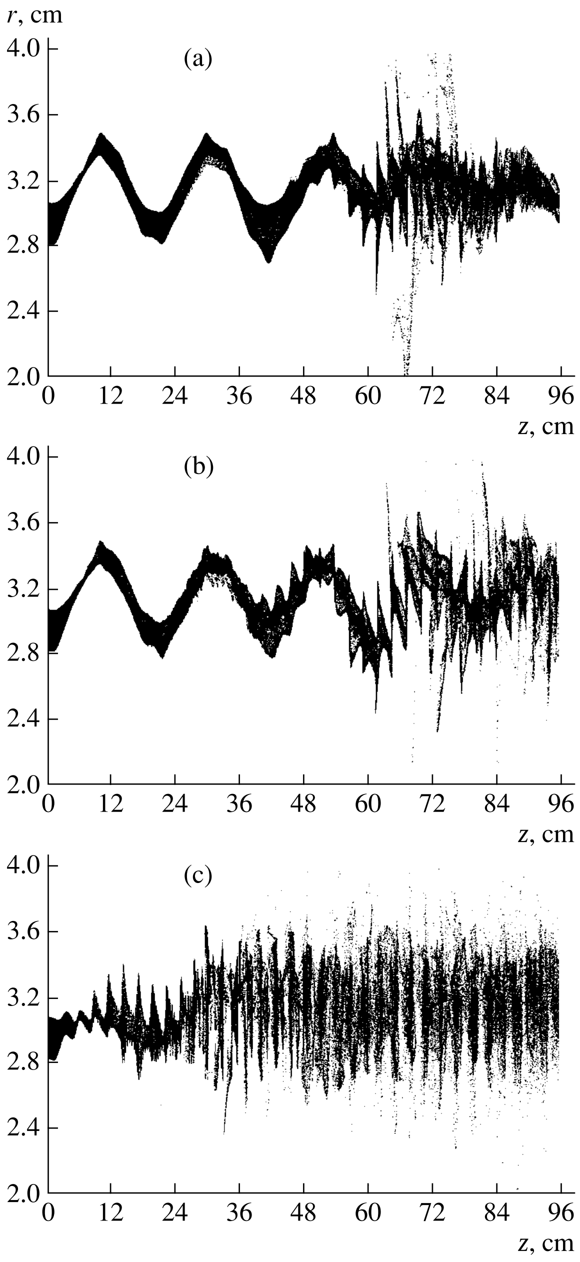


Fig. 3. Configuration space (r, z) of the beam's electrons in the coaxial ubitron at times t = (a) 3.9 (b) 10, and (c) 20 ns.

Figure 5 shows the phase plane (γ, z) of the beam's particles (γ is the relativistic factor) for the same time instants as the configuration space in Fig. 3. We simulated the self-excitation of the system: the initial field in the simulation domain was set equal to zero and the

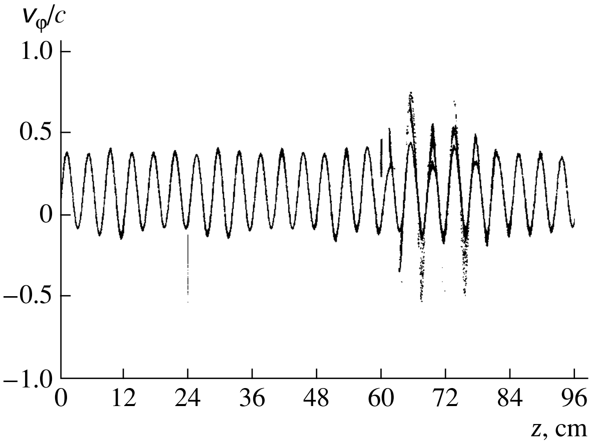


Fig. 4. Phase space (v_φ, z) of the beam's electrons in the coaxial ubitron at t = 3.9 ns.

microwave field at the entrance to the ubitron was also taken to be zero at any time. Therefore, initially, the beam particle energy changes due to only the electrostatic field of the beam. The particles injected into the ubitron first slow down slightly (their kinetic energy decreases) and then traverse a major part of the ubitron with an almost constant energy. Only at the end of the system can a significant energy modulation of the beam be observed (Fig. 5a). With time, the modulation extends along the beam (Fig. 5b). The energy modulation is accompanied by a velocity modulation, bunching the particles in the longitudinal direction. The instability progresses with time and becomes steady when the particles are trapped by the field of the wave excited (Fig. 5c). The effects of instability development that were found in our simulation are numerically identical to those predicted theoretically [3, 4, 13, 17–19] and observed experimentally [5–7]. Thus, our numerical model provides an adequate physical description of resonance wave excitation by an electron beam.

The longitudinal distribution of azimuth electric field E_φ is illustrated in Fig. 6. Initially, E_φ is small almost throughout the system, increasing and reaching a maximum near the exit from the coaxial waveguide (Fig. 6a). With time, the electromagnetic field of the wave excited grows, retaining its longitudinal profile; i.e., instability evolves in the amplification mode. In our numerical model, the boundary condition at the right end of the waveguide corresponds to radiation into free space. In this case, the ubitron is unmatched. As a result, waves reflecting at the exit end of the waveguide and propagating oppositely to the beam inevitably arise. Initially, their amplitude is small, because so is the amplitude of the direct wave. Also, for the reflected wave to affect the beam and substantially contribute to the instability, it must reach the place of injection and reflect from it. Therefore, the effect of the reflected wave on the beam's dynamics and instability development is weak for a time. Initially, the ubitron operates in the amplification mode and the amplitude of the excited wave increases from the place of injection towards the exit from the structure (Fig. 6b). With time, the reflected wave augments the energy of the backward wave and the field amplitude distribution throughout the system becomes uniform. In this case, the longitudinal field pattern is similar to that in a resonator with the only difference that the wavelength depends on the wavevector of the resonance wave (Fig. 6c) rather than on the longitudinal dimension of the resonator. As follows from Fig. 6, the longitudinal structure of the field has a period of ≈10 cm. This period is twice as small as that calculated under the assumption that the

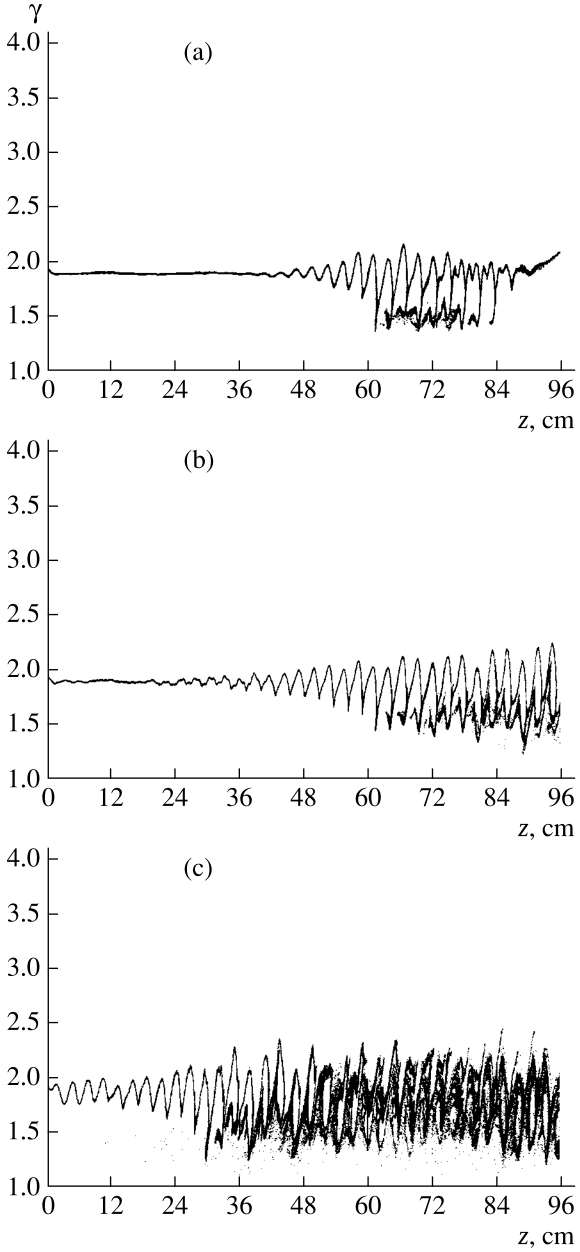


Fig. 5. Same as in Fig. 4 for t = (a) 3.9 (b) 10, and (c) 20 ns.

longitudinal velocity is completely defined by the initial energy of the particles being injected, $v_z = c\sqrt{1 - 1/(1 + W_b/mc^2)^2}$. As we noted above when analyzing the dispersion characteristics shown in Fig. 2, the beam entering the field of the wiggler enhances its transverse velocity upon the sacrifice of the longitudinal velocity. As a result, the average longitudinal velocity of the beam decreases and the region of beam–wave interaction shifts toward higher wavenumbers. A radiation wavelength of ≈10 cm corresponds to longitudinal velocity $v_z \approx 0.78c$. The resonance point for this velocity is indicated by figure 2 in Fig. 2. It can be seen from Fig. 2 that the resonance frequency changed insignificantly. To corroborate this statement, Fig. 7 plots the average longitudinal velocity of the electrons in the ubitron against time. Thus, the estimate of the electron average longitudinal velocity, which is based on the longitudinal structure of the azimuth field, is in fairly good agreement with the result obtained by the numerical simulation.

Figure 8 shows the transverse structure of the azimuth field at $z = 70$ cm for $t = 10$ ns. The wave excited corresponds to the TE_{01} mode of the coaxial waveguide. This result also validates the physical correctness of our numerical model.

The azimuth component of the electric field is the highest. The radial, E_r, and longitudinal, E_z, electric fields are four to five times lower than the azimuth component. The longitudinal structure of E_r is irregular, while the longitudinal distribution of E_z is distinctly regular. For the parameters used in the calculation, the beam is weakly diamagnetic, since the transverse motion of the beam electrons is nonrelativistic (see Fig. 4, where $v_\varphi \approx 0.4c$). At the maximum, $H_r \approx 300$ Oe, $H_z \approx 200$ Oe, and both components of the magnetic field are modulated in the longitudinal direction with the same period as the azimuth electric field.

The frequency spectrum of the electron beam current is shown in Fig. 9. The beam current is modulated with a frequency of ≈8.3 GHz, which corresponds to the resonance frequency (point 2) on the dispersion curve in Fig. 2. Also, significant oscillations are observed at the frequency twice as high as the resonance one.

To calculate the ubitron efficiency and check the accuracy of the numerical model, we used the energy conservation law

$$W + P + S = 0, \tag{3}$$

where

$$W = \frac{1}{4}\iint dzdrr(\mathbf{E}^2 + \mathbf{H}^2), \tag{4}$$

$$P = 2\pi\int_0^t dt'\iint dzdrr\mathbf{j}(t')\cdot\mathbf{E}(t'), \tag{5}$$

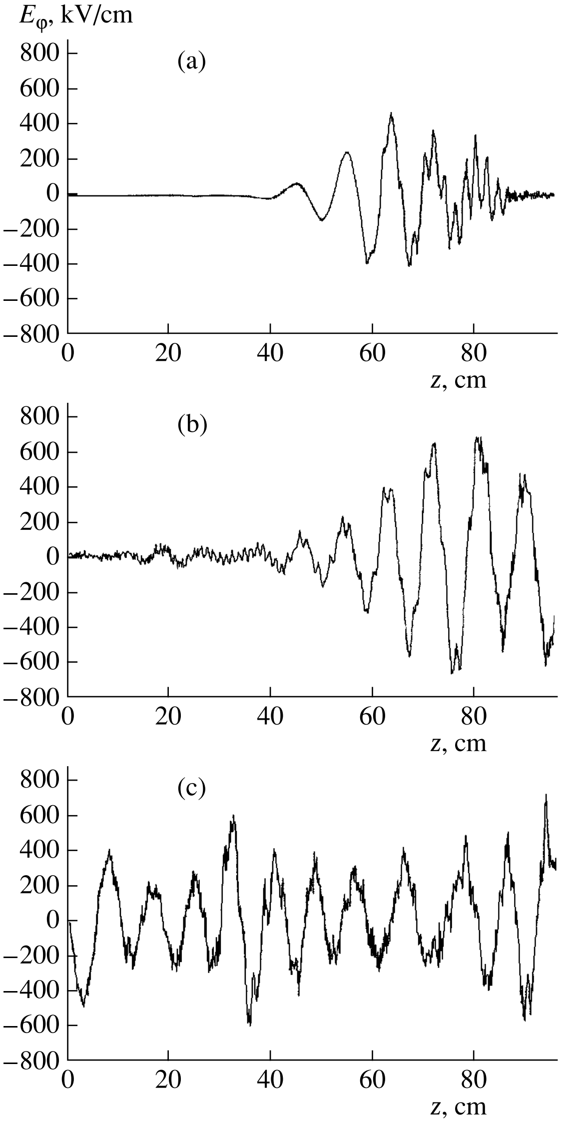


Fig. 6. Longitudinal distribution of the azimuth field in the cross section $r = 3$ cm at times t = (a) 3.9, (b) 10, and (c) 20 ns.

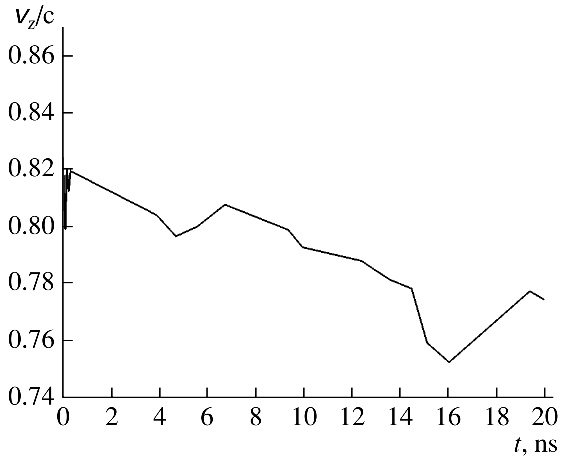


Fig. 7. Average longitudinal velocity of beam’s electrons in the ubitron vs. time.

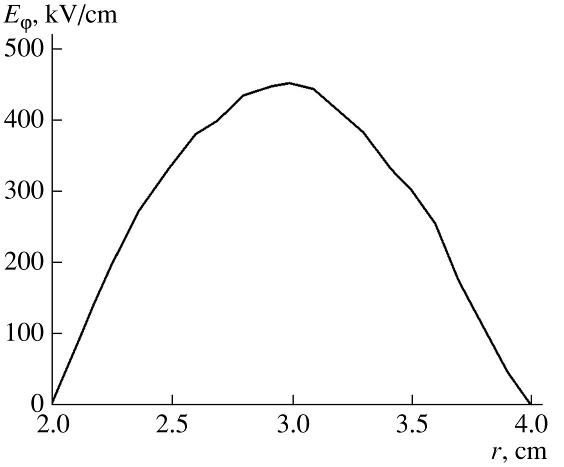


Fig. 8. Transverse structure of the azimuth electric field E_φ in the cross section $z = 70$ cm at $t = 10$ ns.

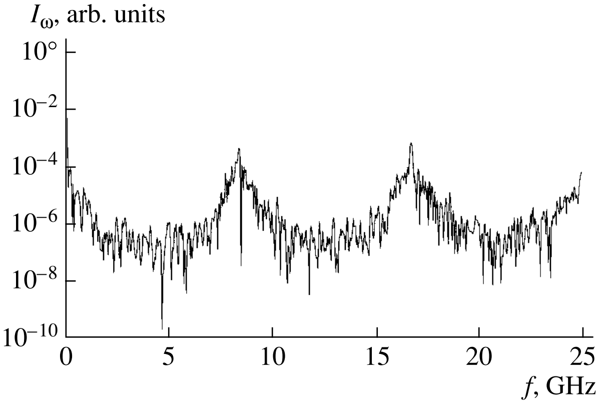


Fig. 9. Frequency spectrum of the electron beam current at the exit from the ubitron.

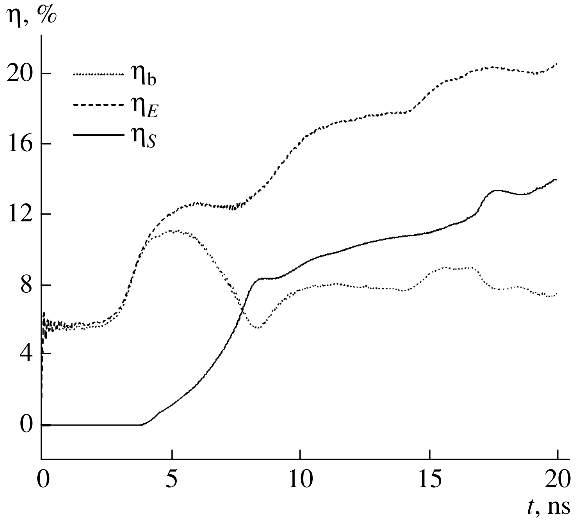


Fig. 10. Energy characteristics of the coaxial ubitron versus time: η_b, the efficiency in terms of beam energy losses; η_E, the efficiency of conversion of the beam energy to the energy stored in the resonator; and η_S, the beam-to-radiation energy conversion efficiency.

$$S = \frac{c}{2}\int_0^t dt' \int dr r(\mathbf{E}(t') \cdot \mathbf{H}(t'))_z \Big|_{z=0}^{z=L}. \qquad (6)$$

In expressions (4)–(6), W is the electromagnetic field energy stored in the system, P is the particle energy loss, and S is the electromagnetic field energy radiated from the entrance and exit ends of the microwave device. In our statement of the problem, the entrance end of the ubitron is nontransparent for the waves; therefore, the energy is radiated from the exit alone. Figure 10 shows the time variation of these energy characteristics normalized by the energy of all particles injected into the system,

$$\eta_E = \frac{W}{I_b U_b t}, \quad \eta_b = \frac{P}{I_b U_b t}, \quad \eta_S = \frac{S}{I_b U_b t}, \qquad (7)$$

where $U_b = W_b/e$ and e is the electron charge.

Until the coaxial ubitron is almost completely filled with the beam, the energy losses increase in time nearly linearly (beam efficiency η_b is almost constant). The particle energy is entirely converted to the energy of the electromagnetic field. Subsequently, the linear oscillation mode changes to the nonlinear mode and the beam energy losses rapidly grow. The electromagnetic field energy stored in the ubitron also grows. The excited electromagnetic waves propagate towards the exit from the waveguide with the group velocity of the resonance wave. The radiation from the system grows with time first nonlinearly and then, after the energy stored in the system becomes constant, almost linearly. Ultimately, the system acquires a quasi-steady state, in which the energy lost by the particles equals to the radiated energy. In this state, the efficiency in terms of the beam energy loss is $\eta_b \approx 20\%$; the efficiency of beam-to-radiation energy conversion, $\eta_S \approx 14\%$; and the efficiency of conversion of the beam energy to the energy stored in the resonator, $\eta_E \approx 6\%$.

CONCLUSIONS

Using the particle-in-cell method, we numerically simulated the transport of and excitation of microwave oscillations by a high-current electron beam in a coaxial drift chamber in the presence of an alternating periodic magnetic field. The shape of the high-current electron beam being transported through the alternating periodic static magnetic field agrees with numerical estimates obtained from the envelope equations within a high accuracy [19].

The analysis of the electromagnetic field dynamics showed that, until the wave reflected from the waveguide's exit comes to its entrance, the ubitron operates as an amplifier. The frequency and wavevector of the resonance wave coincide with those obtained from the linear dispersion relation with allowance for longitudinal–transverse velocity transformation in the field of the wiggler. The reflected wave provides an internal feedback and the microwave device switches to the self-excitation mode. The ubitron efficiency in terms of radiated energy is estimated as ≈14%. For the beam parameters taken in the simulation, this efficiency provides a radiated power of ≈200 MW.

ACKNOWLEDGMENTS

We are grateful to V.A. Balakirev and T.Yu. Yatsenko for submitting analytical expressions for the static magnetic fields of the coaxial wiggler and valuable discussion.

REFERENCES

1. R. M. Phillips, IRE Trans. Electron Devices **7**, 231 (1960).
2. C. E. Ederby and R. M. Phillips, Proc. IRE **53**, 1648 (1965).
3. M. I. Petelin and A. V. Smorgonskiĭ, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **16**, 294 (1973).
4. M. I. Gorshikova and A. V. Smorgonskiĭ, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **18**, 1201 (1975).
5. Yu. V. Tkach, Ya. B. Faĭnberg, N. P. Gadetskiĭ, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **22** (3), 136 (1975) [JETP Lett. **22**, 62 (1975)].
6. Yu. V. Tkach, E. A. Lemberg, Yu. P. Bliokh, *et al.*, Fiz. Plazmy **8**, 75 (1982) [Sov. J. Plasma Phys. **8**, 42 (1982)].
7. R. M. Phillips, Nucl. Instrum. Methods Phys. Res. B **272**, 1 (1988).
8. R. H. Jackson, H. P. Freund, D. E. Pershing, and J. M. Taccetti, Nucl. Instrum. Methods Phys. Res. B **341**, 454 (1994).
9. H. P. Freund, R. H. Jackson, D. E. Pershing, and J. M. Taccetti, Phys. Plasmas **1**, 1046 (1994).
10. D. B. McDermott, A. J. Balkcum, R. M. Phillips, and N. C. Luhmann, Phys. Plasmas **2**, 4332 (1995).
11. A. J. Balkcum, D. B. McDermott, R. M. Phillips, and N. C. Luhmann, IEEE Trans. Plasma Sci. **26**, 548 (1998).
12. G. V. Sotnikov and T. Yu. Yatsenko, Zh. Tekh. Fiz. **72** (5), 22 (2002) [Tech. Phys. **47**, 535 (2002)].
13. V. A. Balakirev, A. V. Borodkin, and T. Yu. Yatsenko, Élektromagnitnye Yavleniya **4** (13), 58 (2001).
14. V. A. Balakirev, Yu. V. Tkach, and T. Yu. Yatsenko, Élektromagnitnye Yavleniya **4** (13), 18 (2001).
15. Yu. A. Berezin and V. A. Vshivkov, *The Method of Particles in the Dynamics of Rarefied Plasma* (Nauka, Novosibirsk, 1986) [in Russian].
16. N. W. McLachlan, *Theory and Application of Mathieu Functions* (Clarendon, Oxford, 1947; Inostrannaya Literatura, Moscow, 1953).
17. J. W. Luginsland, T. A. Antonsen, Jr., J. P. Verboncoeur, *et al.*, in *High-Power Microwave Sources and Technologies*, Ed. by R. J. Barker and Edl Schamiloglu (Wiley-IEEE, New York, 1999).
18. T. M. Antousen, A. A. Mondelli, B. Levush, *et al.*, Proc. IEEE **87**, 804 (1999).
19. V. A. Balakirev, A. V. Borodkin, R. Z. Ramazanov, *et al.*, in *Proceedings of the 13th International Crimean Microwave Conference (CriMiCo'2002), Sevastopol, 2003*, pp. 617–619.

Translated by A. Khzmalyan

Technical Physics, Vol. 50, No. 6, 2005, pp. 754–757. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 6, 2005, pp. 81–84.
Original Russian Text Copyright © 2005 by Aliev, Andreev, Klimov, Lebedev, Nikitin, Terukov, Shadrin.

SURFACE, ELECTRON AND ION EMISSION

Effect of Vacuum Heat Treatment on the Metal–Semiconductor Phase Transition in Thin Vanadium Dioxide Films

R. A. Aliev*, V. N. Andreev, V. A. Klimov**, V. M. Lebedev***, S. E. Nikitin**, E. I. Terukov**, and E. B. Shadrin****

** Amirkhanov Institute of Physics, Dagestan Research Center, Russian Academy of Sciences,*
ul. 26 Bakinskikh Komissarov 94, Makhachkala, 367003 Dagestan, Russia

*** Ioffe Physicotechnical Institute, Russian Academy of Sciences, Politekhnicheskaya*
ul. 26, St. Petersburg, 194021 Russia

e-mail: v.klimov@mail.ioffe.ru

**** Konstantinov Institute of Nuclear Physics, Russian Academy of Sciences, Gatchina,*
Leningrad Oblast, 188300 Russia

Received July 20, 2004

Abstract—The effect of vacuum heat treatment of thin vanadium dioxide films on the parameters of the metal–semiconductor phase transition is studied. The results of heat treatment are compared with those obtained upon irradiation of the synthesized films by medium-energy electrons. The elemental composition of the films that is found by the Rutherford backscattering (RBS) method suggests that an observed change in the hysteresis loop of the films is associated with the reduction of the vanadium dioxide upon heating in a vacuum. *© 2005 Pleiades Publishing, Inc.*

INTRODUCTION

The first-order phase transition in vanadium dioxide (VO_2) single crystals is observed at 67°C. Upon cooling below this temperature, the tetragonal phase transforms into the monoclinic phase, which shows up as a step in the electric conductivity and optical constants of vanadium dioxide [1]. Emerging elastic stresses may damage the single crystals. Therefore, the material of choice in both investigation and application is thin films, and not single crystals, of vanadium dioxide, which remain intact upon repeated cycling about the phase-transition temperature.

When VO_2 is heated or irradiated by a laser or electron beam, the phase transition parameters (the transition temperature, as well as the shape and width of the hysteresis loop) are known to change [2–4]. This seemingly gives a chance to refine the properties of vanadium dioxide films synthesized [5]. However, VO_2-based multilayer devices (e.g., laser-beam limiters [6]) usually involve thin layers of other materials, which are applied on the surface of the VO_2 active film. Application of these layers often implies vacuum heating of the VO_2 thin film, causing an uncontrolled change in its properties. Moreover, devices of this type are normally used under severe thermal and electron irradiation conditions, which may result in degradation of the VO_2 film.

In this work, we compare the effects of medium-energy electron irradiation and vacuum heating on thin vanadium dioxide films.

EXPERIMENTAL

Thin VO_2 films were deposited by laser evaporation [5, 7]. The thickness of the films deposited on Si(001) substrates was 60–80 nm. The absence of vanadium lower oxides in the films was judged by the optical transmission at wavelength $\lambda = 10.6$ μm. We studied the temperature dependences of the dc and ac conductivity (in the frequency range 0.2–200 kHz) and the reflectivity of the films at wavelength $\lambda = 1.54$ μm.

The films were annealed and irradiated by electrons at a pressure of $\sim 10^{-5}$ Torr. The electron energy was ~10 keV; the current density, 120 μA/mm^2. The irradiation dose was estimated from the exposure time, and the annealing temperature was varied from 200 to 350°C.

The elemental composition and concentration profile in the films were determined by nondestructive methods of nuclear physics: RBS and instantaneous deuteron-induced nuclear reactions. The measurements were performed on the analytical installation for materials research based on the electrostatic accelerator at the St. Petersburg Institute of Nuclear Physics [8, 9]. The deuteron energy was $E_d = 0.9$ MeV. The deuterons scattered and the reaction products (protons and alphas) were detected using a planar silicon detector with energy resolution $\Delta E = 10$ keV for $E_\alpha = 2.7$ MeV, which was placed at angle $\theta = 135°$ to the beam direction.

The vanadium concentration in the films was calculated from the peak intensity of backscattered deuterons; the oxygen concentration, from the peak intensity of the alphas produced by the nuclear reaction of deuterons with oxygen-16, $^{16}O(d, \alpha)^{14}N$.

RESULTS AND DISCUSSION

Figure 1a shows the evolution of the hysteresis loop for the metallic phase in the films upon vacuum annealing. Short-term annealing at 250°C shifts the loop toward lower temperatures and affects its shape and width only weakly. An increase in the annealing temperature shifts the loop further and simultaneously makes it wider. After the shift has reached ≈50°C, the widening of the hysteresis loop is accompanied by a decrease in the reflectivity; eventually, the phase transition disappears completely. The heat treatment increases the conductivity of the films in the semiconducting state, whereas the phase-transition-induced jump of the conductivity decreases. For example, for the as-deposited films, this jump is $\sim 10^3$, and after annealing at 300°C, the conductivity at the phase transition changes only several-fold.

Similar results are obtained when the films are irradiated by 10-keV electrons (Fig. 1b). This similarity is thought to indicate that the main cause of the irradiation-induced changes in the properties of the films is heating of the films during irradiation. Indeed, elastic scattering of such low-energy electrons cannot generate a significant amount of Frenkel pairs, so that the nonradiative relaxation of excitations in the electron subsystem merely causes local heating of the films and substrate. Other relaxation mechanisms related to electrostatic instability are characteristic of only wide-gap semiconductors, higher oxides, and alkali-halide crystals [10]. Unlike those compounds, vanadium dioxide, not exhibiting a high bond ionicity, has a comparatively narrow energy gap (≈0.7 eV) and even acquires metallic properties upon heating above 67°C. The free path of 10-keV electrons exceeds the thickness of the films being studied. Moreover, along with these electrons, the vanadium dioxide interacts with a wide spectrum of lower energy secondary electrons appearing in the substrate. Taken together, these factors result in electron-beam heating of the substrate and film.

When treating the results obtained, we used the approach [11] developed for polycrystalline vanadium dioxide films where grains are of different size and degree of nonstoichiometry. According to [11], the conductivity or reflectivity hysteresis loop in such a film is considered as a sum of elementary loops assigned to individual grains. The temperature position of an elementary hysteresis loop is related to the deviation of the grain composition from stoichiometry, the phase transition temperature declining with increasing oxygen deficiency in the grain. The width of an elementary hysteresis loop is inversely proportional to the square root of the grain size [12]. This approach allows us to explain the significant differences between the hysteresis loop for VO_2 single crystals and that for VO_2 thin films, as well as the evolution of the latter loop when the synthesis conditions are changed or the film synthesized is subjected to various physical actions [4, 5].

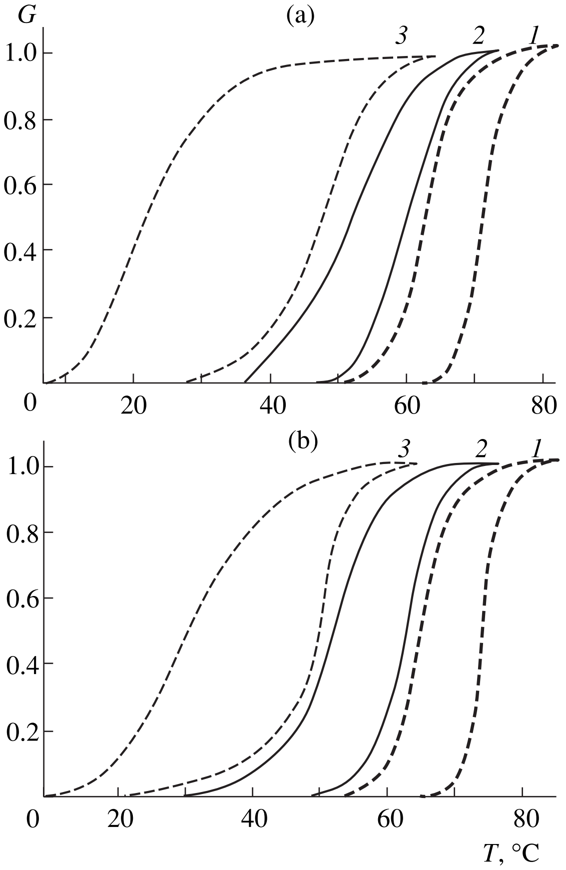


Fig. 1. Temperature dependences of fraction *G* of the metallic phase in the vanadium dioxide film 70 nm thick. (a) The effect of vacuum heat treatment on the shape and temperature position of the hysteresis loop: (*1*) as-deposited, (*2*) annealed at 250°C for 15 min, and (*3*) annealed at 300°C for 15 min. Fraction *G* was calculated from the reflectance of the films measured at wavelength $\lambda = 1.54$ μm. (b) The effect of 10-keV electron irradiation (at a current density of 120 $\mu A/mm^2$) on the hysteresis loop: (*1*) as-deposited, (*2*) irradiated for 5 min, and (*3*) irradiated for 15 min. Fraction *G* was calculated from the conductivity measured at a frequency of 200 kHz.

According to the above approach, the shift of the hysteresis loop without a noticeable change in its width and shape means that moderate-dose irradiation or heating to 250°C (Fig. 1, loops *2*) changes the stoichiometric composition of the grains, leaving their size the same. The decrease in the phase transition temperature is assumed to be related to the generation of valent defects (V^{3+} ions) in the vanadium dioxide. It is the formation of V^{3+} ions, which are necessary for electroneutrality to be provided, that can explain the decrease in the phase transition temperature in the vanadium dioxide containing Nb^{5+} and W^{6+} impurities [1]. This decrease can be expressed as $\Delta T = C_V \cdot 12°$, where

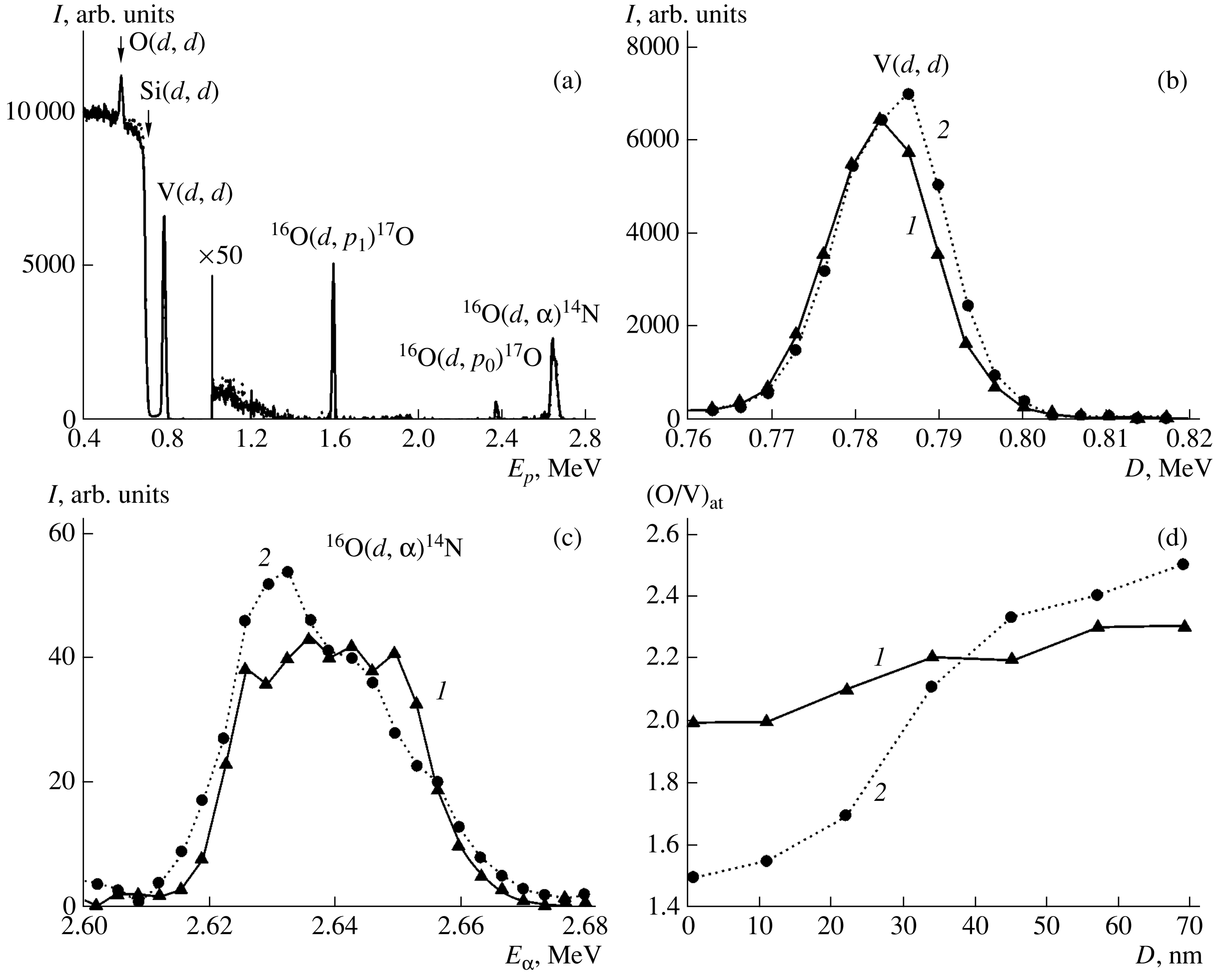


Fig. 2. (a) Experimental spectrum of the particles that form during the irradiation of the vanadium dioxide film on silicon by deuterons with energy $E_d = 0.9$ MeV, (b) spectrum fragment related to deuteron scattering by vanadium, (c) spectrum of alpha particles forming as a result of the nuclear reaction $^{16}O(d, \alpha)^{14}N$, and (d) the depth distribution of the oxygen-to-vanadium atomic ratio. (*1*) As-deposited and (*2*) vacuum-annealed at 300°C for 15 min.

C_V is the V^{3+} ion concentration in at. %. One can write a chain of quasi-chemical reactions [13] that start at the interface and lead to the formation of V^{3+} ions in the vanadium dioxide when oxygen atoms leave it,

$$O^{2-} \longrightarrow V_{O}^{2-} + 1/2O_2,$$

$$V^{4+} + V_{O}^{2-} \longrightarrow V^{3+} + V_{O}^{-},$$

$$V^{4+} + V_{O}^{-} \longrightarrow V^{3+} + V_{O}.$$

Here, O^{2-} is an oxygen ion occupying a site of the VO_2 crystal lattice and V_{O}^{2-}, V_{O}^{-}, and V_{O} are the neutral and singly and doubly ionized oxygen vacancies, respectively.

These reactions describe the case when the oxygen merely escapes into a vacuum when the oxygen pressure in the working chamber is lower than equilibrium. Obviously, the result (i.e., the appearance of V^{3+} ions) would be the same if the vanadium dioxide is reduced by, e.g., hydrogen or carbon monoxide that enter into the residual gases. Accordingly, H_2O or CO_2 is removed to a vacuum in this case.

As was noted above, the maximum shift in the phase transition temperature that was induced by annealing or electron irradiation in a vacuum is ≈50°C. If oxygen vacancies are considered to be doubly ionized, then the lower limit of the homogeneity range estimated by the above equation for ΔT is close to $VO_{1.98}$, which agrees well with the data reported in [14].

Thus, with the approach adopted in [11], the evolution of the hysteresis loops can be explained as follows. During low-temperature annealing, the oxygen leaves the film and a nonstoichiometric oxygen-depleted VO_2 film forms. As long as the film composition is within the homogeneity range, the hysteresis loop merely shifts toward lower temperatures. When the annealing temperature increases, the reaction causing VO_2 reduction proceeds at a higher rate and the initially single-phase VO_2 film becomes two-phase: another vanadium oxide or a mixture of oxides appears. The regions occupied by the vanadium dioxide inside the grains shrink, which widens the hysteresis loop and smoothes out the step in the physical parameters at the phase transition [12]. Eventually, electron irradiation or vacuum heating completely removes the vanadium dioxide from the film.

This conclusion is qualitatively supported by the methods of RBS and instantaneous nuclear reactions. Figure 2 shows the spectra of scattered deuterons and particles generated by the deuteron–oxygen-16 nuclear reactions in the films before and after annealing. At energies below 1 MeV, the peaks of the deuterons scattered by vanadium and oxygen atoms are observed; at higher energies, the peaks of protons and alphas, which are the products of the nuclear reactions, are detected. Figu-res 2b and 2c detail the spectral fragments used to calculate the concentration profiles of the elements in the film.

Figure 2d shows the depth distribution of the oxygen-to-vanadium atomic ratio calculated according to [8]. In the as-deposited films, the ratio O/V (Fig. 2d, solid line) is virtually constant (2.0 ± 0.2) within a calculation error. The small increase in this ratio near the substrate is likely to be due to silicon diffusion from the substrate into the film during synthesis at 600°C. In contrast, in the film annealed at 300°C (Fig. 2d, dotted line), the oxygen fraction in the near-surface layer drops noticeably, whereas the ratio O/V near the silicon substrate remains almost unchanged.

Thus, more than half the film does no longer contain vanadium dioxide. Near the surface, from where the oxygen or reduction products escape in fact, the film composition is closer to V_2O_3 rather than to VO_2. In conductivity measurements, this layer would shunt the remaining part of the vanadium dioxide.

It should be noted that the effects observed are not irreversible. Vacuum heating of the film to 600°C followed by cooling under the same conditions as post-synthesis cooling recovers the step in the reflectivity of the film, as well as the temperature position and shape of the hysteresis loop.

CONCLUSIONS

Both vacuum heating and irradiation of thin VO_2 films by medium-energy electrons damage the films, rather than stabilize the VO_2 metallic phase, as was assumed in [2]. Treated in such a way, the films represent a mixture of different phases with metallic conductivity.

As for the technology of multilayer coatings, one has to avoid vacuum heating of thin vanadium dioxide layers at temperatures above 200°C.

ACKNOWLEDGMENTS

We thank Yu.G. Luk'yanov, V.A. Smolin, and L.M. Nikitin for assistance.

This work was supported by the Department of Physical Sciences at the Russian Academy of Sciences.

REFERENCES

1. A. A. Bugaev, B. P. Zakharchenya, and F. A. Chudnovskiĭ, *The Metal–Semiconductor Phase Transition and Its Application* (Nauka, Leningrad, 1979) [in Russian].
2. A. A. Bugaev, A. I. Gavrilyuk, A. A. Gur'yanov, *et al.*, Pis'ma Zh. Tekh. Fiz. **4** (6), 65 (1978) [Sov. Tech. Phys. Lett. **4**, 27 (1978)].
3. V. G. Mokerov, A. R. Belikov, and L. S. Ignat'ev, Fiz. Tverd. Tela (Leningrad) **21**, 1482 (1979) [Sov. Phys. Solid State **21**, 855 (1979)].
4. E. B. Shadrin and A. V. Il'inskiĭ, Fiz. Tverd. Tela (St. Petersburg) **42**, 1092 (2000) [Phys. Solid State **42**, 1126 (2000)].
5. R. A. Aliev and V. A. Klimov, Fiz. Tverd. Tela (St. Petersburg) **46**, 515 (2004) [Phys. Solid State **46**, 532 (2004)].
6. O. B. Danilov, V. A. Klimov, O. P. Mikheeva, *et al.*, Zh. Tekh. Fiz. **73** (1), 79 (2003) [Tech. Phys. **48**, 73 (2003)].
7. V. N. Andreev, M. A. Gurvich, V. A. Klimov, *et al.*, Pis'ma Zh. Tekh. Fiz. **19** (9), 63 (1993) [Tech. Phys. Lett. **19**, 283 (1993)].
8. G. Gavrilov, A. Krivchitch, and V. Lebedev, Nucl. Instrum. Methods Phys. Res. A **515**, 108 (2003).
9. V. M. Lebedev, Yu. G. Luk'yanov, and V. A. Smolin, in *Proceedings of the 13th International Conference on Electrostatic Accelerators, Obninsk, 2001*, p. 60.
10. Ch. B. Lushchik and A. Ch. Lushchik, *Relaxation of Electron Excitation with Formation of Defects in Solids* (Nauka, Moscow, 1989) [in Russian].
11. T. G. Lanskaya, I. A. Merkulov, and F. A. Chudnovskiĭ, Fiz. Tverd. Tela (Leningrad) **20**, 1201 (1978) [Sov. Phys. Solid State **20**, 193 (1978)].
12. I. A. Khakhaev, F. A. Chudnovskiĭ, and E. B. Shadrin, Fiz. Tverd. Tela (St. Petersburg) **36**, 1643 (1994) [Phys. Solid State **36**, 898 (1994)].
13. P. Kofstad, *Nonstoichiometry, Diffusion, and Electrical Conductivity in Binary Metal Oxides* (Wiley, New York, 1972; Mir, Moscow, 1975).
14. C. H. Griffiths and H. K. Eastwood, J. Appl. Phys. **45**, 2201 (1974).

Translated by K. Shakhlevich

Technical Physics, Vol. 50, No. 6, 2005, pp. 758–765. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 6, 2005, pp. 85–92.
Original Russian Text Copyright © 2005 by Rezinkina.

EXPERIMENTAL INSTRUMENTS AND TECHNIQUES

Growth of Dendrite Branches in Polyethylene Insulation under a High Voltage Versus the Branch Conductivity

M. M. Rezinkina

Molniya Research and Design Institute, National Technical University Kharkov Polytechnical Institute,
Kharkov, 61013 Ukraine
e-mail: marinar@kpi.kharkov.ua

Received July 26, 2004

Abstract—Based on experimental data for electrical aging of cast polyethylene insulation under the action of a high voltage, it is concluded that the branches of dendrites (aging defects) growing in the insulation may be both conductive and nonconductive according to experimental conditions. A method of numerical simulation of the electric field around breakdown channels is described. Potential distributions in the neighborhood of dendrites are calculated for a number of high voltages. It is shown that the growth of dendrites with nonconductive branches slows down gradually as the branch diameter increases and the electric field strength drops. Conversely, the growth of conductive-branch speeds up, since the high potential is localized at their thin ends and the field strength rises. *© 2005 Pleiades Publishing, Inc.*

INTRODUCTION

Polyethylene as a high-voltage insulator is widely used for insulating, e.g., electric cables, which stimulates great interest in studying its electrical aging. Electrical (field) aging, eventually causing breakdown, triggers irreversible processes, one of which is dendritic growth (treeing). Dendrites represent a set of micrometer-diameter cavities with a length of several millimeters [1]. Today, there is no agreement among authors as to whether dendrite channels are conductive and whether they distort the electric field distribution. According to [1, 2], dendrite channels cannot be viewed as a mere extension of a tip. At the same time, Ushakov [3] argues that the potential of the high-voltage electrode is localized near the top of a dendrite channel. According to Kuchinskiĭ [1, p. 134], the surface resistance of a dendrite channel may amount to 10^{10} Ω, while Morshuis [4] indicates that the surface resistance of dendrite cavities in polyethylene initially equals 10^{17} and then drops to 10^9 Ω. Studying the effect of high electric fields on the surface of polymeric insulators, Aleksandrov and Solov’ev [5] concluded that a long-term action of a chemically active environment on the discharge-heated surface of a polymer generates radicals and low-molecular decomposition products, which may considerably reduce the surface resistance of the material.

Studies [4, 6] on electrical aging of artificial cavities consisting of three polyethylene layers, each 0.1–0.2 [4] or 0.4–0.5 mm [6] thick, with a hole of diameter 5 mm in the middle layer revealed three aging stages with different partial discharge (PD) parameters. At the first (unaged) stage, the discharge has the form of a streamer. At the second (aged) stage (which sets in after the sample has been exposed to electric field E = 5 kV/mm for 60 min), a thin oxide layer appears on the sample surface, the surface resistance drastically drops, most electrons are trapped, and breakdown occurs by the Townsend mechanism. At the third stage (which sets in after exposure to E = 5 kV/mm for 100 h), crystals start growing and the discharges are localized on these crystals, becoming strictly periodic.

Champion and Dodd [7] analyzed the growth of dendrites in epoxy resin subjected to a high voltage. Photographing flashes due to PDs, they concluded that dendrite branches may be both conductive and nonconductive. For example, in the case of bush dendrites, a PD flash occupies several branches and originates at the tip. Here, the branches are nonconductive. For tree-like dendrites, the flashes arise only at the very ends of the branches, not occupying them completely and not originating at the tip. These branches are conductive. In the photographs taken in transmitted light [7], the nonconductive branches appear white, whereas the conductive ones are colored black (possibly because of carbonization). Since PDs in the nonconductive branches occupy them completely, such branches are bound to expand due to erosion, as argued in [7]. It was noted [7] that dendrite branches in polyethylene also may pass to the conductive state at a certain stage of dendritic growth [8, 9].

EXPERIMENTAL STUDY OF DENDRITIC GROWTH

Experimental investigation into the growth of dendrites in cast polyethylene [10] confirmed the statement that the diameter of dendrite channels increases during dendritic growth under a low field strength that is close to the treeing-initiating field. For example, if the tip radius is $R_t = 2$ μm, the distance to the grounded electrode is $h = 1.5$ mm, and the applied voltage is $U_0 = 6$–7 kV, the extensive growth of dendrites lasts 25–30 min. In this case, dendrite branches thicken at the base, while remaining thin at the ends. Next is the stagnation phase, when the dendrites almost stop propagating toward the grounded electrode. In this phase, the dendrites branch (lateral growth) and the already existing branches thicken. The maximal thickness of the branch is 10–20 μm.

For the same R_t and h and the voltage applied to the tip $U_0 = 8$–10 kV, the breakdown time equals 2–20 min. Branching is weak: a main branch alone outstands. For $t \geq 3$ min, some of the branches thicken at the base adjacent to the tip. The maximal thickness (1–10 μm) of the branches is attained not long before breakdown, with the branches thickened at the bases remaining narrow at the ends.

According to [11–13], there exists a correlation between a stage of dendritic growth, the shape of dendrites, and the statistic distribution of the difference between the phase of PD initiation and phase φ of applied voltage. It was shown that the distributions of $n^* = n/n_{\max}$ (where n is the PD occurrence frequency at a given phase of the applied voltage and $n_{\max}$ is the maximal PD frequency observed in a given experiment) at the early stage of growth and at the stage when the dendrite ceases to propagate show common features; namely, they are asymmetric on both the positive and negative half-waves of the applied voltage (Fig. 1). At high applied voltages, the PD occurrence frequency distributions change: in the course of dendritic growth, they become similar to each other (right-asymmetric and slightly peaked) on both half-waves (Fig. 2). With regard to the PD photographing data [7], one may argue that, early in the growth of bush dendrites and also at the stage of their intense growth, the branches are nonconductive or poorly conductive, which initiates PDs in the gas-filled dendrite cavities. The result is the asymmetry of the PD occurrence frequency distribution (hereafter, PD frequency distribution) on the negative and positive half-waves of the applied voltage, which arises from the strong dependence of the gas breakdown voltage on electrode polarity in the nonuniform field. With time, dendrite branches developing at high voltages become conductive and the PD frequency distributions change: they become similar to each other on both half-waves of the applied voltages. This is because the polyethylene breakdown voltage is less dependent on electrode polarity [14] and also because of its great scatter.

Thus, during high-field dendritic growth, the branches become conductive. At the same time, fresh dendrites or those growing under an electric field close to the field of growth initiation have nonconductive or poorly conductive branches. At high field strengths, when ponderomotive forces are high, the walls of the cavities are carbonized and their conductivity rises sharply [7]. Since PDs in conductive dendrites are initiated at their ends rather than at the base of the tip, they do not propagate along the whole length of the branches and so do not increase their diameter. If the field and ponderomotive forces are not too high, the cavities widen, because PDs multiply propagate along the whole length of the branches. Thus, high-field dendritic growth exhibits positive feedback: the branches become conductive but remain thin, which further enhances the field and favors the growth toward the counter electrode. At low fields close to the growth-initiating value, the growth proceeds under the negative-feedback conditions: the branches remain nonconductive and their diameter increases. As a result, the growth of such dendrites slows down and may even cease, since the electric field around the plasma-filled nonconductive branches drops when PDs are initiated and little, if any, new areas of the polyethylene experience breakdown. This speculation is corroborated by experimental data. Under the above experimental conditions, dendrites grow slowly at tip voltages $U_0 \leq 7$ kV, while at $U_0 \geq 8$ kV, complete breakdown takes from several minutes to several tens of minutes.

MATHEMATICAL SIMULATION OF THE DENDRITE'S ELECTRIC FIELD

The mathematical simulation of growth of conductive and nonconductive dendrite brunches under different voltages was described in author's earlier work [15]. A mathematical simulation of PD initiation and propagation in dendrite conductive and nonconductive branches was developed in [7]. However, all the models suggested consider too simplified forms of dendrite cavities when calculating the electric field distribution. In [7], the cavity is represented as a set of circles with a diameter of 50 μm (the simulation is made in the 2D statement). In [15], the electric field distribution is calculated in the 3D statement but the dendrite is represented as a sequence of cubic cells of side 10 μm. Actually, however, dendrite branches are a series of microcavities with a near-circular cross section of diameter D varying from 1 to 20 μm. The reason for simplification is quite clear: the dendrite is an elongated object with a high length-to-diameter ratio and analysis of the field generated by such objects is a mathematical challenge.

An electric field in a discharge gap is often calculated with numerical techniques, such as the finite-difference method (FDM) (see, e.g., [16, 17]). This method makes it possible to take into account the electrical nonuniformity of the medium and the variation of the electrical parameters with an extension of a dendrite channel. When using the FDM to calculate the electric field of dendrite-like structures, one should appropriately choose step Δ of grid. To allow for the dependence of the electric field distribution on radius R_0 of the discharge structure, the step must be no smaller than R_0. However, the length of a discharge structure may

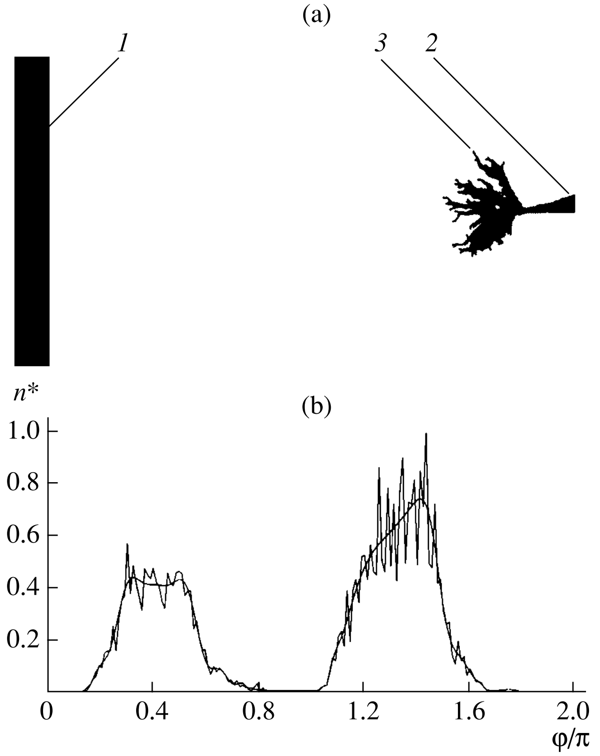


Fig. 1. (a) Dendrite and (b) distribution of the phase differences between the phase of dendrite occurrence and that (φ) of the applied voltage at U_0 = 7 kV (*1,* plane; *2,* tip; and *3,* dendrite.

exceed its diameter by several orders of magnitude. As applied to the 3D problem, the case $\Delta < R_0$ would greatly raise the order of equations to be solved in calculating the electric field and make the problem insolvable even using advanced computing facilities. Therefore, the size of a spatial mesh is usually taken to be much larger than R_0 (for example, equal to the mean step used in simulating discharge structures [16]).

The discharge structures arising at the breakdown of various discharge gaps can be represented as a set of “wires,” conductive circular cylinders with a diameter much smaller than the transverse size of the mesh. In solving the problem, it is taken into consideration that the field of an infinite conducting cylinder decays by the logarithmic law. However, in [18], the problem is solved only for an infinitely long wire and only in the 2D statement. The approach applied in [18] may be extended for three-dimensional calculation of the electric field distribution in a set of finite-length fine wires, including those making an angle with the coordinate axes. Of such elementary wires, dendrites of different spatial configuration, as well as other discharge channels arising at the breakdown of dielectric gaps, can be composed.

To find the electric field distribution using the FDM, an insulating domain containing fine conducting wires was split into parallelepipeds (meshes). Splitting is accomplished in such a way that nodes (i, j, k) of the calculation grid lie at the interfaces and on the axes of the wires (Fig. 3) and nodes (i, j, k), $(i + 1, j, k)$, $(i + 1, j + 1, k)$, $(i, j + 1, k)$, $(i, j + 1, k + 1)$, $(i + 1, j + 1, k + 1)$, $(i + 1, j, k + 1)$, and $(i, j, k + 1)$, at which unknown potentials are determined, are the vertices of an (i, j, k) mesh. Within each of the meshes, except for those adjacent to the wire, the properties of the medium are assumed to be uniform. For each node of the grid, the equation for the scalar electric potential can be written in the form

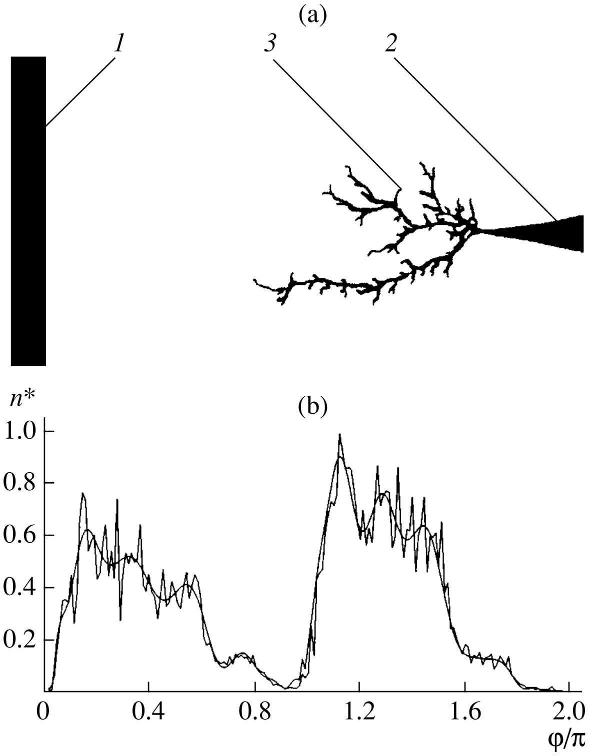


Fig. 2. The same as in Fig. 1 for U_0 = 9 kV.

$$\oint_S \varepsilon\left(-\frac{\partial\varphi}{\partial n}\right)ds = 0, \tag{1}$$

where

$$\varepsilon = \begin{bmatrix} \varepsilon k^x & 0 & 0 \\ 0 & \varepsilon k^y & 0 \\ 0 & 0 & \varepsilon k^z \end{bmatrix}, \tag{2}$$

S is the surface area of the parallelepiped whose faces halve the distances between neighboring nodes (Fig. 3), k^x, k^y, and k^z are coefficients, ε is the relative permittivity, $E_n = -\partial\varphi/\partial n$ is the electric field component normal to elementary surface area ds, and φ is the scalar electric potential.

To solve the problem by the FDM, Eq. (1) was written in the difference form for each the node. The boundary conditions in the calculation were the homogeneous

Neumann conditions $\partial\varphi/\partial n = 0$ at the boundaries X = const and Z = const and $\varphi = 0$ at $Y = 0$; $\partial\varphi/\partial n = E_0$ at $Y = Y_{\max}$; and $\varphi = U_0$ at $X = XT$, $Y = Y_{\max}$, and $Z = ZT$ (here XT, $Y_{\max}$, and ZT are the coordinates of the tip base). The set of equations obtained was solved by the iterative method of variable directions [19] modified for solving three-dimensional elliptic equations (for details of the solution method, see [20]).

To simplify the mathematics, consider a cubic grid with step Δ. The wire segments nonparallel to the coordinate axes may either be directed along the diagonal of the cubic mesh or connect opposite nodes of the faces lying in one of the planes X = const, Y = const, and Z = const. The nodes belonging to the wire will be assigned index w. To be definite, we take the segment connecting nodes (iw, jw, kw) and ($iw - 1, jw + 1, kw$) (Fig. 3). Surfaces S covering each node belonging to the wire cross the wire. Let us express the electric field vector flux across S_1 (the section of the segment connecting these nodes by planes X = const and Y = const) through the potential difference between these nodes, the cross-sectional area of the wire, and its conductivity γ_W as

$$\Lambda\varphi^{\mathrm{W}}_{iw,\,jw,\,kw} = k_{\mathrm{W}}(\varphi_{iw-1,\,jw+1,\,kw} - \varphi_{iw,\,jw,\,kw}), \qquad (3)$$

where $k_{\mathrm{W}} = S_1/d_1k_1 = \gamma_{\mathrm{W}}\pi D_0^2/(8\Delta)$, $S_1 = 0.5\pi(D_0\sqrt{2})^2/4$, D_0 is the wire diameter, $d_1 = \Delta\sqrt{2}$ is the spacing between nodes ($iw - 1, jw + 1, kw$) and (iw, jw, kw), and $k_1 = 1/\sqrt{2}$ is the factor allowing for the fact that the electric field vector flux decreases when projected onto sections X = const and Y = const.

The term on the right of (3) that contains potential $\varphi_{iw-1,\,jw+1,\,kw}$ will be assumed to be known and its value will be defined by the previous iteration.

In difference form, integral (1) representing the electric field induction vector flux across cell surface S is written as

$$\Lambda_x\varphi_{i,\,j,\,k} + \Lambda_y\varphi_{i,\,j,\,k} + \Lambda_z\varphi_{i,\,j,\,k} = f_{i,\,j,\,k},$$

$$\Lambda_x\varphi_{i,\,j,\,k} = \varphi_{i-1,\,j,\,k}AX_{i,\,j,\,k} - \varphi_{i,\,j,\,k}CX_{i,\,j,\,k} + \varphi_{i+1,\,j,\,k}BX_{i,\,j,\,k},$$

$$AX_{i,\,j,\,k} = \Delta(k^x_{i-1,\,j-1,\,k-1}\varepsilon_{i-1,\,j-1,\,k-1} + k^x_{i-1,\,j-1,\,k}\varepsilon_{i-1,\,j-1,\,k} + k^x_{i-1,\,j,\,k-1}\varepsilon_{i-1,\,j,\,k-1} + k^x_{i-1,\,j,\,k}\varepsilon_{i-1,\,j,\,k})/4,$$

$$BX_{i,\,j,\,k} = \Delta(k^x_{i,\,j-1,\,k-1}\varepsilon_{i,\,j-1,\,k-1} + k^x_{i,\,j-1,\,k}\varepsilon_{i,\,j-1,\,k} + k^x_{i,\,j,\,k-1}\varepsilon_{i,\,j,\,k-1} + k^x_{i,\,j,\,k}\varepsilon_{i,\,j,\,k})/4,$$

$$CX_{i,\,j,\,k} = AX_{i,\,j,\,k} + BX_{i,\,j,\,k} + k_{\mathrm{W}X},$$

$$\Lambda_y\varphi_{i,\,j,\,k} = \varphi_{i,\,j-1,\,k}AY_{i,\,j,\,k} - \varphi_{i,\,j,\,k}CY_{i,\,j,\,k} + \varphi_{i,\,j+1,\,k}BY_{i,\,j,\,k},$$

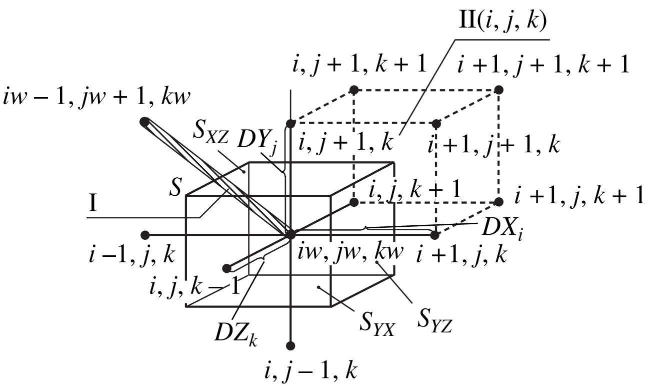


Fig. 3. Calculation cell (I, wire; II, cell).

$$AY_{i,\,j,\,k} = \Delta(k^y_{i-1,\,j,\,k-1}\varepsilon_{i-1,\,j,\,k-1} + k^y_{i-1,\,j,\,k}\varepsilon_{i-1,\,j,\,k} + k^y_{i,\,j,\,k-1}\varepsilon_{i,\,j,\,k-1} + k^y_{i,\,j,\,k}\varepsilon_{i,\,j,\,k})/4,$$

$$BY_{i,\,j,\,k} = \Delta(k^y_{i-1,\,j,\,k-1}\varepsilon_{i-1,\,j,\,k-1} + k^y_{i-1,\,j,\,k}\varepsilon_{i-1,\,j,\,k} + k^y_{i,\,j,\,k-1}\varepsilon_{i,\,j,\,k-1} + k^y_{i,\,j,\,k}\varepsilon_{i,\,j,\,k})/4,$$

$$CY_{i,\,j,\,k} = AY_{i,\,j,\,k} + BY_{i,\,j,\,k} + k_{\mathrm{W}Y},$$

$$\Lambda_z\varphi_{i,\,j,\,k} = \varphi_{i,\,j,\,k-1}AZ_{i,\,j,\,k} - \varphi_{i,\,j,\,k}CZ_{i,\,j,\,k} + \varphi_{i,\,j,\,k+1}BZ_{i,\,j,\,k},$$

$$AZ_{i,\,j,\,k} = \Delta(k^z_{i-1,\,j-1,\,k-1}\varepsilon_{i-1,\,j-1,\,k-1} + k^z_{i,\,j-1,\,k-1}\varepsilon_{i,\,j-1,\,k-1} + k^z_{i-1,\,j,\,k-1}\varepsilon_{i-1,\,j,\,k-1} + k^z_{i,\,j,\,k-1}\varepsilon_{i,\,j,\,k-1})/4,$$

$$BZ_{i,\,j,\,k} = \Delta(k^z_{i-1,\,j-1,\,k}\varepsilon_{i-1,\,j-1,\,k} + k^z_{i,\,j-1,\,k}\varepsilon_{i,\,j-1,\,k} + k^z_{i-1,\,j,\,k}\varepsilon_{i-1,\,j,\,k} + k^z_{i,\,j,\,k}\varepsilon_{i,\,j,\,k})/4,$$

$$CZ_{i,\,j,\,k} = AZ_{i,\,j,\,k} + BZ_{i,\,j,\,k} + k_{\mathrm{W}Z}.$$

Here, $f_{i,\,j,\,k} = -(k_{\mathrm{W}X} + k_{\mathrm{W}Y} + k_{\mathrm{W}Z})\varphi_{iw-1,\,jw+1,\,kw}$; $k_{\mathrm{W}X} = k_{\mathrm{W}Y} = k_{\mathrm{W}}$ for the cells covering the wire (for other cells, $k_{\mathrm{W}X} = k_{\mathrm{W}Y} = 0$, $k_{\mathrm{W}Z} = 0$, since, in our case, the wire lies in plane Z = const and does not cross it); $\varepsilon_{i,\,j,\,k}$ is the relative permittivity of the material of the (i, j, k)th cell; and $k^x_{i,\,j,\,k}$, $k^y_{i,\,j,\,k}$, and $k^z_{i,\,j,\,k}$ are the respective coefficients for the (i, j, k)th cell that enter into expression (2) for permittivity tensor ε.

Let us derive an expression for coefficients k^x, k^y, and k^z. The potential of a fine conductive wire of length $2C$ that is parallel to the OY axis is given by [21]

$$\varphi(x, y, z) = k_Q\ln\frac{\sqrt{(y_1 + C)^2 + x_1^2 + z_1^2} + y_1 + C}{\sqrt{(y_1 - C)^2 + x_1^2 + z_1^2} + y_1 - C}, \qquad (4)$$

where $k_Q = Q/8\pi\varepsilon_0\varepsilon C$; $\varepsilon_0 = 0.885 \times 10^{-11}$ F/m; Q is the charge of the wire; x, y, z are the coordinates of the observation point and x_1, y_1, and z_1 are the distances between the center of the wire and the observation point in the X, Y, and Z directions, respectively.

In the case of a uniform spatial grid with the wire making an angle of $-\pi/4$ with the OX axis, we have

$$x_1 = [(x - x_0) + (y - y_0)]\frac{1}{\sqrt{2}},$$

$$y_1 = [-(x - x_0) + (y - y_0)]\frac{1}{\sqrt{2}}, \quad z_1 = z,$$

where x_0, y_0, and z_0 are the coordinates of the center of the charged wire.

To apply Eq. (1) to the nodes located on or around the wire axis, we will deduce an expression for the derivative of the potential in the X direction,

$$\frac{\partial\varphi}{\partial x} = \frac{1}{\sqrt{2}}\left(\frac{\partial\varphi}{\partial x_1} - \frac{\partial\varphi}{\partial y_1}\right).$$

On rearrangement in view of (4), we eventually come to

$$\frac{\partial\varphi}{\partial x} = \frac{k_Q}{\sqrt{2}}K_x, \tag{5}$$

where

$$K_x = K_1 - K_2,$$

$$K_1 = \frac{x_1}{x_1^2 + z_1^2}\left(\frac{(y_1 - C)}{\sqrt{(y_1 - C)^2 + x_1^2 + z_1^2}} - \frac{(y_1 + C)}{\sqrt{(y_1 + C)^2 + x_1^2 + z_1^2}}\right),$$

$$K_2 = \frac{1}{\sqrt{(y_1 - C)^2 + x_1^2 + z_1^2}} - \frac{1}{\sqrt{(y_1 + C)^2 + x_1^2 + z_1^2}}.$$

For $R_0 \ll \Delta$, we recast k_Q through the potential difference between the node lying on the wire axis and its neighbor in the X or Y direction in view of (4),

$$\left.\frac{\Delta\varphi}{\Delta}\right|_{R_0 \ll \Delta} \approx \frac{k_Q}{\Delta}\ln\left(\frac{4}{3 + 2\sqrt{2}}\frac{\Delta/\sqrt{2}}{R_0}\right).$$

Hence,

$$k_Q \approx \frac{\Delta\varphi}{\Delta}\frac{\Delta}{\ln\left(\frac{4}{3 + 2\sqrt{2}}\frac{\Delta/\sqrt{2}}{R_0}\right)}. \tag{6}$$

Now we integrate (5) over surface area $X = \text{const} \cdot S_{yz}$ (Fig. 3) of dimension $\Delta \times \Delta$,

$$\int_{S_{yz}} -\frac{\partial\varphi}{\partial x}ds = -k_Q K_{xx}, \tag{7}$$

where

$$K_{xx} = \int_{-\Delta/2}^{\Delta/2}\int_{-\Delta/2}^{\Delta/2} K_x dydz. \tag{8}$$

Substituting (6) into (7) yields

$$\int_{S_{yz}} -\frac{\partial\varphi}{\partial x}ds = -\frac{\Delta\varphi_x}{\Delta}\Delta^2 k^x,$$

where

$$k^x = \frac{K_{xx}}{\Delta\sqrt{2}\ln\left(\frac{4}{3 + 2\sqrt{2}}\frac{\Delta/\sqrt{2}}{R_0}\right)}. \tag{9}$$

Expressions for the other coefficients are derived in a similar way,

$$\int_{S_{xz}} -\frac{\partial\varphi}{\partial y}ds = -\frac{\Delta\varphi_y}{\Delta}\Delta^2 k^y,$$

$$k^y = \frac{K_{yy}}{\Delta\sqrt{2}\ln\left(\frac{4}{3 + 2\sqrt{2}}\frac{\Delta/\sqrt{2}}{R_0}\right)}; \tag{10}$$

$$\int_{xy} -\frac{\partial\varphi}{\partial z}ds = -\frac{\Delta\varphi_z}{\Delta}\Delta^2 k^z,$$

$$k^z = \frac{K_{zz}}{\Delta\ln\left(\frac{4}{3 + 2\sqrt{2}}\frac{\Delta/\sqrt{2}}{R_0}\right)}. \tag{11}$$

Here,

$$K_y = K_1 + K_2, \quad K_{yy} = \int_{-\Delta/2}^{\Delta/2}\int_{-\Delta/2}^{\Delta/2} K_y dxdz; \tag{12}$$

$$K_z = \frac{1}{\sqrt{2}}\frac{z_1}{x_1^2 + z_1^2}\left[\frac{y_1 - C}{\sqrt{(y_1 - C)^2 + x_1^2 + z_1^2}} - \frac{y_1 + C}{\sqrt{(y_1 + C)^2 + x_1^2 + z_1^2}}\right],$$

$$K_{zz} = \int_{-\Delta/2}^{\Delta/2}\int_{-\Delta/2}^{\Delta/2} K_z dxdy. \tag{13}$$

When the wire runs parallel to one of the coordinate axes (for example, to Y), coefficients k^x, k^y, and k^z are derived similarly. For nodes on the wire, $-k^y_{iw,\,jw,\,kw} = 10^6$ [20], except for the extreme node, for which

$$k^y = \frac{K_y}{2\Delta \ln(\Delta/R_0)}, \tag{14}$$

where

$$K_y = \int_{-\Delta/2}^{\Delta/2} \ln\left[\frac{\Delta/2 + \sqrt{(\Delta/2+2C)^2 + x^2 + (\Delta/2)^2}}{-\Delta/2 + \sqrt{(\Delta/2+2C)^2 + x^2 + (\Delta/2)^2}} \times \frac{-\Delta/2 + \sqrt{2(\Delta/2)^2 + x^2}}{\Delta/2 + \sqrt{2(\Delta/2)^2 + x^2}}\right] dx. \tag{15}$$

For coefficients $k^x = k^z$, we have

$$k^x = \frac{K_x}{2\Delta \ln(\Delta/R_0)}, \tag{16}$$

where

$$\begin{aligned} K_x = \int_{-\Delta/2}^{\Delta/2} \frac{\Delta/2}{(\Delta/2)^2 + z^2}\Big[&\sqrt{(y+\Delta/2-C)^2 + (\Delta/2)^2 + z^2} \\ &- \sqrt{(y-\Delta/2-C)^2 + (\Delta/2)^2 + z^2} \\ &- \sqrt{(y+\Delta/2+C)^2 + (\Delta/2)^2 + z^2} \\ &+ \sqrt{(y-\Delta/2-C)^2 + (\Delta/2)^2 + z^2}\Big] dz. \end{aligned} \tag{17}$$

Thus, the influence of a fine conductive cylindrical wire and, in particular, its diameter on the electric field distribution is taken into account through coefficients k^x, k^y, and k^z (see expressions (9)–(11), (14), and (16)) multiplying the components of tensor ε for the nodes on or around the wire (for other nodes, $-k^x = k^y = k^z = 1$). Integrals (8), (12), (13), (15), and (17) are taken numerically with any routine.

Potentials φ were calculated using the method described above and also the checking formulas [22] for a conducting prolate spheroid placed in an electric field at an angle of 45° to the OY axis. The step of the grid was set equal to $\Delta = 50$ μm, and radius R_0 of the wire was varied. Comparing the results of numerical and analytical calculations of the potential at the grid nodes showed that the relative error is less than 1.0% for $2R_0/\Delta < 0.02$ and equals 1.4% for $2R_0/\Delta = 0.2$.

CALCULATION OF THE ELECTRIC FIELD OF DENDRITES

The efficiency of the calculation method described above was tested by estimating the electric field of dendrites in the "plane with tip–plane" configuration. This configuration was applied in experimental studies on field aging of solid polymeric insulation. Figure 4 is the sectional view of the system in plane Z = const. A high voltage is applied to the tip through tip carrier *1*, which is a metal plate on which surface tip *2* is mounted. Since the width of the carrier far exceeds the tip diameter and the distance to grounded plane *3*, the system can be represented in the form of two parallel planes. One is grounded (i.e., is under the zero potential), while the other, containing tip *2*, is under high potential U_0. The insulation is polyethylene *4* immersed in transformer oil *5*. Dendrite *6* originates from the tip. The tip–grounded plane distance is $h_0 = 2.5$ mm, the spacing between planes *1* and *3* is $d_{\max} = 10$ mm, and the length of the part of the tip that is introduced into the polyethylene is $b = 3$ mm. Since the length of the system in the X and Z directions, as well as the length of the tip, is several times larger than h_0, we used absorbing boundary conditions [18] at the calculation domain boundaries $X = X_{\max}$, $X = X_{\min}$, $Z = Z_{\max}$, $Z = Z_{\min}$, and $Y = Y_{\max}$. Such an approach made it possible to significantly shrink the calculation domain, bringing it closer to the zone of dendritic growth. For absorbing layers, expression (2) for the relative permittivity tensor takes the form

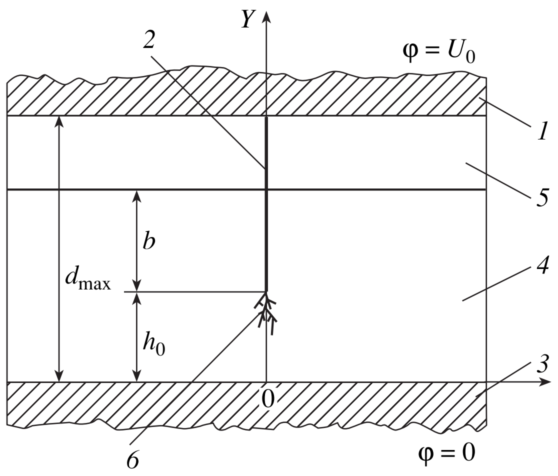


Fig. 4. Model "plane with tip–plane" system.

$$\varepsilon = \begin{bmatrix} \varepsilon/k^x & 0 & 0 \\ 0 & \varepsilon k^x & 0 \\ 0 & 0 & \varepsilon k^x \end{bmatrix} \begin{bmatrix} \varepsilon k^y & 0 & 0 \\ 0 & \varepsilon/k^y & 0 \\ 0 & 0 & \varepsilon k^y \end{bmatrix} \begin{bmatrix} \varepsilon k^z & 0 & 0 \\ 0 & \varepsilon k^z & 0 \\ 0 & 0 & \varepsilon/k^z \end{bmatrix}.$$

It is assumed here that coefficients $k^p = 1 + (k_{\max} - 1)(p/d)^3$ ($p = x$, y, or z) in these matrices vary by the exponential law $k^p = 1 + (k_{\max} - 1)(p/d)^3$. Here, $k_{\max} = 300$ is the maximal value of k^p, $d = N\Delta$ is the absorbing layer thickness, and $N = 5$ is the number of sublayers into which the absorbing layer is split. The given parameter values were taken according to the recommendations made in [18]. The boundary conditions

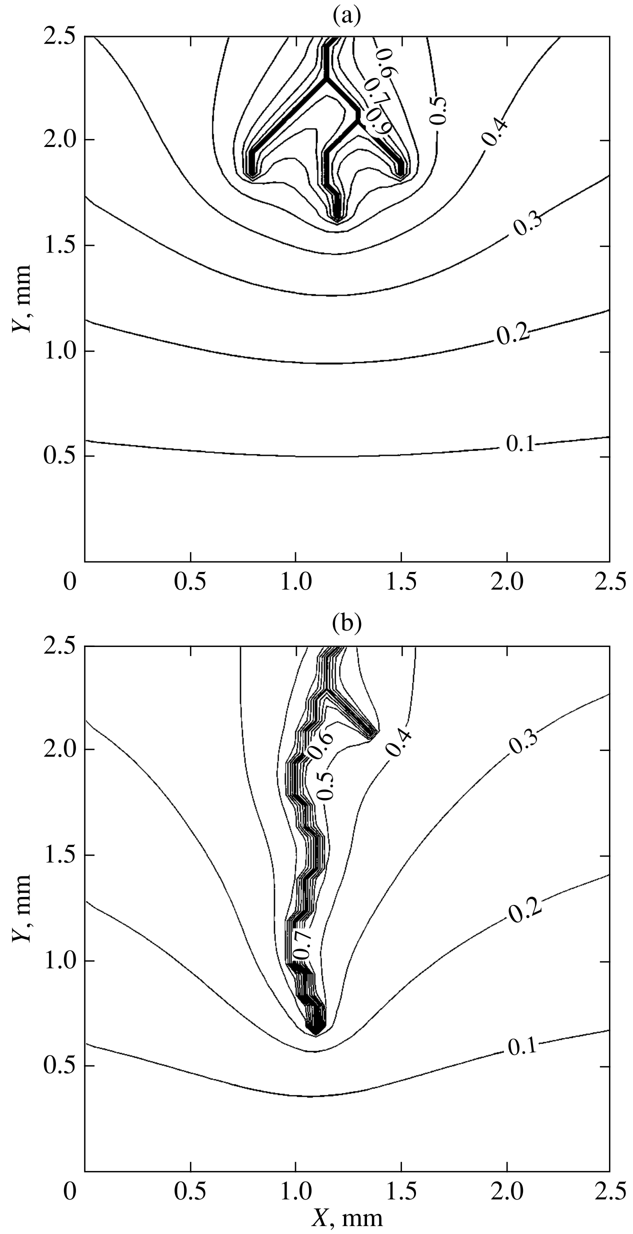


Fig. 5. Calculated isopotentials for U_0 = (a) 6 and (b) 10 kV.

used in the calculation were the following: $\partial\varphi/\partial n = 0$ at boundaries X = const and Z = const, $\varphi = 0$ at $Y = 0$, $\partial\varphi/\partial n = k_{max}E_0$ ($E_0 = U_0/d_{max}$) at $Y = Y_{max}$ (for the details concerning the last boundary condition, see [23]), and $\varphi|_{X = XT,\, Y = Y_{max},\, Z = ZT} = U_0$ (where XT, Y_{max}, ZT are the coordinates of the point where the tip crosses the upper boundary of the calculation domain, Y = const = Y_{max}).

In the calculation, the applied voltage was U_0 = 6 kV; the dendrite channel diameter, D_0 = 10 μm (as evidenced from experimental data gathered at this voltage); the step of the spatial grid used in the FDM, Δ = 50 μm; and $E^*_{max} = E_{max}/E_0 = 55.2$ (where E_{max} is the maximal value of the field strength averaged over the grid volume measuring $\Delta \times \Delta \times \Delta$; the maximal electric field is attained at the ends of the plasma-filled branches). The electric field distributions were also calculated for high-voltage dendritic growth, when the thin basic branch of the dendrite is conductive. In this case, U_0 = 10 kV, D_0 = 1 μm (as evidenced from experimental data gathered at this voltage), and Δ = 50 μm. The maximal electric field, E^*_{max} = 85.3, is attained at the end of the basic branch. Thus, because of the increase in the dendrite branch diameter, the presence of several breakdown channels, and the lower treeing-initiating voltage, the absolute value of grid-volume-averaged maximal field strength E_{max} at U_0 = 6 kV is 2.6 times lower than at U_0 = 10 kV. Note that the maximal strengths at the extreme point of the dendrite differ by a factor of 16.7.

Figure 5 shows isopotentials $\varphi^* = U/U_0$ calculated for U_0 = 6 and 10 kV in section Z = const passing through the dendrite (shown is a part of the calculation region adjacent to the dendrite).

CONCLUSIONS

(1) Experimental data for high-voltage dendritic growth in polyethylene insulation (namely, the evolution of the dendrite shape during growth, localization of PD-related flares, and the distribution of the phase differences between the phase of PD occurrence and that of the applied voltage) strongly suggest that, early in dendritic growth and at voltages close to the growth-initiating voltage, dendrite branches are nonconductive. At high applied voltages (field strengths), some of the branches become conductive and, therefore, propagate at a higher rate. At low fields, when PDs propagate along nonconductive branches, they grow in width and the growth slows down or even ceases.

(2) In this work, we modify the finite-difference method for calculating the 3D electric field distribution when insulating gaps experience breakdown and breakdown channels are represented as a set of conductive cylinders with a high length-to-diameter ratio (wires). The modified approach allows for numerical evaluation of the field distribution for the case when the diameter of the resulting discharge structure is five or more times smaller than the step of the spatial grid used in calculations. Using such a calculation technique, one can take into account the variation in the diameter of a breakdown channel along its length.

(3) Our numerical results for the electric field distribution indicate that, for bush dendrites, which grow at low voltages (6 kV), the maximal grid-volume-averaged field strength is 2.6 times lower than for treelike dendrites growing at high voltages (10 kV).

REFERENCES

1. G. S. Kuchinskiĭ, *Partial Discharge in High-Voltage Equipment* (Énergiya, Leningrad, 1979) [in Russian].
2. Z. Wei and W. Kalner, in *Proceedings of the 7th International Symposium on High Voltage Engineering, Dresden, 1991*, Vol. 2, pp. 249–252.
3. V. Ya. Ushakov, in *Physics and Technology of High-Power Pulsed Systems*, Ed. by E. P. Velikhov (Énergoatomizdat, Moscow, 1987), pp. 295–311.
4. P. Morshuis, in *Proceedings of the 9th International Symposium on High Voltage Engineering, Graz, 1995*, Vol. 5, pp. 5644.1–5644.4.
5. G. N. Aleksandrov and E. P. Solov'ev, Izv. Vyssh. Uchebn. Zaved. Énerg., No. 9, 31 (1971).
6. A. Krivda and E. Gulski, in *Proceedings of the 9th International Symposium on High Voltage Engineering, Graz, 1995*, Vol. 5, pp. 5643.1–5643.4.
7. J. V. Champion and S. J. Dodd, J. Phys. D **34**, 1235 (2001); www.iop.org./Journals/jd.
8. R. J. Densley, IEEE Trans. Electr. Insul. **14**, 148 (1979).
9. C. Laurent and C. Mayoux, IEEE Trans. Electr. Insul. **15**, 33 (1980).
10. M. M. Rezinkina, Vestn. Kharkovsk. Gos. Politekh. Univ., No. 21, 48 (1998).
11. M. M. Rezinkina, O. L. Rezinkin, and M. I. Nosenko, Zh. Tekh. Fiz. **71** (3), 69 (2001) [Tech. Phys. **46**, 339 (2001)].
12. M. M. Rezinkina, O. L. Rezinkin, and M. I. Nosenko, Vestn. Kharkovsk. Gos. Politekh. Univ., No. 66, 107 (1999).
13. M. M. Rezinkina, O. L. Rezinkin, and M. I. Nosenko, Vestn. Kharkovsk. Gos. Politekh. Univ., No. 65, 42 (1999).
14. *Electrical Performance of Polymers*, Ed. by B. I. Sazhin (Khimiya, Leningrad, 1977) [in Russian].
15. M. M. Rezinkina, Pis'ma Zh. Tekh. Fiz. **26** (5), 37 (2000) [Tech. Phys. Lett. **26**, 196 (2000)].
16. A. A. Dul'zon, V. V. Lopatin, M. D. Noskov, *et al.*, Zh. Tekh. Fiz. **69** (4), 48 (1999) [Tech. Phys. **44**, 394 (1999)].
17. L. A. Dissado, J. C. Fothergill, N. Wise, *et al.*, J. Phys. D **33**, L109 (2000).
18. A. Taflove and S. Hagness, *Computational Electrodynamics: The Finite Difference Time Domain Method* (Artech House, Boston, 2000).
19. A. A. Samarskiĭ, *The Theory of Difference Schemes* (Nauka, Moscow, 1989; Dekker, New York, 2001).
20. M. M. Rezinkina, Elektrichestvo, No. 8, 50 (2003).
21. K. A. Krug, *Physical Fundamentals of Electrical Engineering* (Gos. Énergoizdat, Moscow, 1946), Vol. 1.
22. J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill, New York, 1941; OGIZ Gostekhizdat, Moscow, 1948).
23. M. M. Rezinkina and M. M. Shcherba, Tekh. Élektrodinamika, Subject Issue, Part 5, 3 (2004).

Translated by V. Isaakyan

Technical Physics, Vol. 50, No. 6, 2005, pp. 766–770. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 6, 2005, pp. 93–97.
Original Russian Text Copyright © 2005 by Antonyuk, Kovalenko, Moldovan, Petrichuk.

Periodic Structure of Acicular Magnetic Clusters in a Magnetic Liquid

O. A. Antonyuk, V. F. Kovalenko, B. N. Moldovan, and M. V. Petrichuk
Shevchenko National University, Vladimirskaya ul. 64, Kiev, 01033 Ukraine
e-mail: vkovalenko@univ.kiev.ua

Received March 30, 2004

Abstract—The behavior of magnetic clusters in a magnetic liquid placed in a circular capillary is considered. When a uniform magnetic field is applied to the system, acicular clusters grow from the sediment, being aligned with the field. The interaction of the clusters as magnetic dipoles with one another and with an external gradient magnetic field is considered theoretically. In a nonuniform symmetric magnetic field with a peak, the cluster distribution is uniform near the peak. Such a distribution is fairly stable when the magnetic field gradient is varied over certain limits. The ordered (periodic) cluster configuration is realized experimentally, and it is shown that its period can be controlled. As the magnetic field gradient exceeds a certain threshold, the clusters are redistributed, forming a multirow hexagonal array. *© 2005 Pleiades Publishing, Inc.*

INTRODUCTION

Magnetic liquids (MLs), colloidal solutions of fine magnetic particles in a carrier liquid, have recently become an object of extensive research (see, for example, [1]). Interest in MLs is associated, in particular, with their application in photon crystals [2], although the range of application of MLs is actually much wider [3].

Typically, particles in MLs tend toward clustering, which can be regarded as the initial stage of transition into the solid state or as the sedimentation of magnetic particles. Clustering (particle agglomeration), as well as mechanisms of particle–particle interaction in MLs, was studied in detail in [4, 5]. In most applications of MLs (for example, when they are used as lubricants, printing ink, coolants, or thickeners), clustering is an undesirable effect. In all these cases, the hydrodynamic characteristics of MLs must remain stable.

For this reason, agglomeration-preventing techniques have always been the focus of attention. However, one can benefit from clustering in MLs in a number of cases, e.g., visualization of magnetic field lines, magnetic inhomogeneities, and magnetic domains; operation of optical shutters; etc., where such “anomalous” behavior of MLs is used to advantage. Note also that the magnetic clustering effect (more specifically, spatial ordering of the clusters, which is most frequently due to an external magnetic field) is central to the development of ML-based photon crystals [6].

Clusters form via dipole–dipole interaction between magnetic nanoparticles, while the thermal vibration of carrier-liquid molecules hampers this process. Therefore, the coarser the particles, the stronger the clustering effect. The threshold size of nanoparticles at which the probability of clustering becomes significant is about 10 nm [3]. The application of an external magnetic field also favors clustering. A cluster formed in this case is a magnetic dipole where the magnetization vector coincides with the field vector.

Clustering mechanisms and the parameters of spatially ordered clusters in thin films and magnetic liquids have received much attention [7–12]. In most of these works, the magnetic field is applied at right angles to the plane of the film and clusters are distributed among the sites of the 2D hexagonal lattice like magnetic bubbles in thin magnetic films. Such a distribution offers the highest packing density.

In this work, in contrast to those cited above, acicular clusters, which freely grow in MLs, are considered. To facilitate the growth of acicular clusters and their linear arrangement, the ML is placed in circular glass capillaries.

The aim of this study is finding conditions under which magnetic clusters in the capillary are arranged into a given configuration.

THEORY

Clusters in an ML are subjected to dipole–dipole and dipole–magnetic field interaction forces. The gravitational force and the viscosity of the liquid are ignored. We use the point cluster approximation: $h \ll a$, where h is the cluster height and a is the cluster spacing.

It has been noted in the introduction that, when subjected to a moderate magnetic field, clusters in a capillary are arranged linearly; so, we will consider the linear distribution. Let us specify conditions under which the clusters are uniformly distributed along the capillary by taking into account mutual interaction of a limited number of clusters (both nearby and remote). In calculation, the model of periodically arranged magnetic bubbles will be employed [13].

The energy of interaction between two point magnetic dipoles with magnetic moments $\mathbf{m}_1$ and $\mathbf{m}_2$ has the form

$$V_{12} = \frac{\mathbf{m}_1 \cdot \mathbf{m}_2}{|\mathbf{R}_{12}|^3} - \frac{3(\mathbf{m}_1 \cdot \mathbf{R}_{12})(\mathbf{m}_2 \cdot \mathbf{R}_{21})}{|\mathbf{R}_{12}|^5},$$

where R_{12} is the displacement of the dipole with moment $\mathbf{m}_2$ relative to the dipole with moment $\mathbf{m}_1$ and $\mathbf{R}_{21} = -\mathbf{R}_{12}$.

Consider a system of $2N + 1$ magnetic dipoles that are arranged along the x axis at points with coordinates x_i ($i = -N, \ldots, N$) and have the same magnetic moments, $|\mathbf{m}_i| = m$. The dipoles are placed in nonuniform magnetic field $\mathbf{H}(x)$ directed along the y axis. We assume that clusters are saturated; therefore, m is independent of H. The total energy of the system is given by

$$E = \sum_{i=-N}^{N} \left(\sum_{\substack{j=-N \\ j<i}}^{N} \frac{\mathbf{m}_i \cdot \mathbf{m}_j - 3(\mathbf{m}_i \cdot \mathbf{e}_x(\mathbf{m}_j \cdot \mathbf{e}_x))}{|x_i - x_j|^3} \right) - \sum_{i=-N}^{N} \mathbf{m}_i \cdot \mathbf{H}(x_i), \quad (1)$$

where $\mathbf{e}_x$ is the unit vector along the x axis. The first sum on the right-hand side of this relationship is the energy of dipole–dipole interaction between the clusters; the second, the energy of cluster–magnetic field interaction.

Since the major axes of the dipoles are aligned with magnetic field $\mathbf{H} = H\mathbf{e}_y \parallel 0y$, we have $\mathbf{m}_i \cdot \mathbf{e}_x = 0$. The set of equations for dipole coordinates x_i corresponding to the equilibrium state of the system is written as

$$\frac{\partial E}{\partial x_i} = -\sum_{\substack{j=-N \\ i \neq j}}^{N} \left(3\mathbf{m}_i \cdot \mathbf{m}_j \frac{(x_i - x_j)}{|x_i - x_j|^5} \right) - \mathbf{m}_1 \frac{\partial H(x_i)}{\partial x_i} = 0; \quad i = -N, \ldots, N. \quad (2)$$

In the case of a periodic equilibrium structure, $x_i = ai$ and Eq. (2) yields $(\partial \mathbf{H}/\partial x)_{x_i}$ ($i = -N, \ldots, N$),

$$(\partial H/\partial x)_{x_i} = -\frac{3m}{a^4} \sum_{\substack{j=-N \\ j \neq i}}^{N} \frac{i - j}{|i - j|^5},$$
$$|\partial H/\partial x|_{x_i} = \frac{3m}{a^4} \sum_{t=-i+1}^{i} \frac{1}{(N + t)^4}, \quad i > 0. \quad (3)$$

Symmetric terms, which cancel out, are omitted here.

To determine the degree of disordering of the clusters in the case when calculated distribution (3) does not exactly fit the field actual distribution, we will calculate a field gradient at which the cluster distribution is nonuniform. Two cases will be considered: (i) the cluster spacing decreases starting from point $x = 0$, attaining a deviation of 5 (50)% at the ends and (ii) the spacing increases in the same way.

Let $a_j \equiv x_j - x_{j-1}$, where $j = -N + 1, \ldots, N$ (for the periodic distribution, $x_j/a = j$, $a_j \equiv a \equiv (x_N - x_{-N})/(2N + 1)$). The linear dependence of a_j on dipole number j within segment $[0; x_N]$ (with the total length of the segment remaining constant) takes place if

$$a_j/a = 1 + \left(j - \frac{N+1}{2}\right)\Delta \quad (\text{for } j > 0),$$

it follows hence that

$$x_j/a = \sum_{p=1}^{j} a_p = j\left(1 - \frac{N+1}{2}\Delta\right) + \frac{j(j+1)}{2}\Delta$$
$$= j\left[1 + \frac{j - N}{2}\Delta\right] \quad (\text{at } j > 0)$$

or

$$x_j/a = j\left[1 + \frac{|j| - N}{2}\Delta\right], \quad j = -N, \ldots, N$$

due to symmetry. Here, Δ is the parameter of nonuniformity (namely, the difference between adjacent dipole spacings, $D = a_{|j|} - a_{|j|-1}$).

For $N = 15$ ($2N + 1 = 31$), we have

$$a_j/a = 1 + (j - 8)\Delta \quad (\text{for } j > 0);$$

$$x_j/a = j\left[1 + \frac{|j| - 15}{2}\Delta\right], \quad j = -N, \ldots, N.$$

For $n = 5$ ($2N + 1 = 11$),

$$a_j/a = 1 + (j - 3)\Delta;$$

$$x_j/a = j\left[1 + \frac{|j| - 5}{2}\Delta\right], \quad j = -N, \ldots, N.$$

It should be noted that, for $\Delta > 0$, two central spacings, $a_1 = x_1 - x_0$ and $a_0 = x_0 - x_{-1}$, assume the minimal value

$$\left[1 - \frac{N-1}{2}\Delta\right]a,$$

while two extreme distances, a_{-N+1} and a_N, take on maximal value

$$\left[1 + \frac{N-1}{2}\Delta\right]a.$$

For $\Delta < 0$, the situation is reverse: the maximal spacing is at the center of the array and the minimal one at

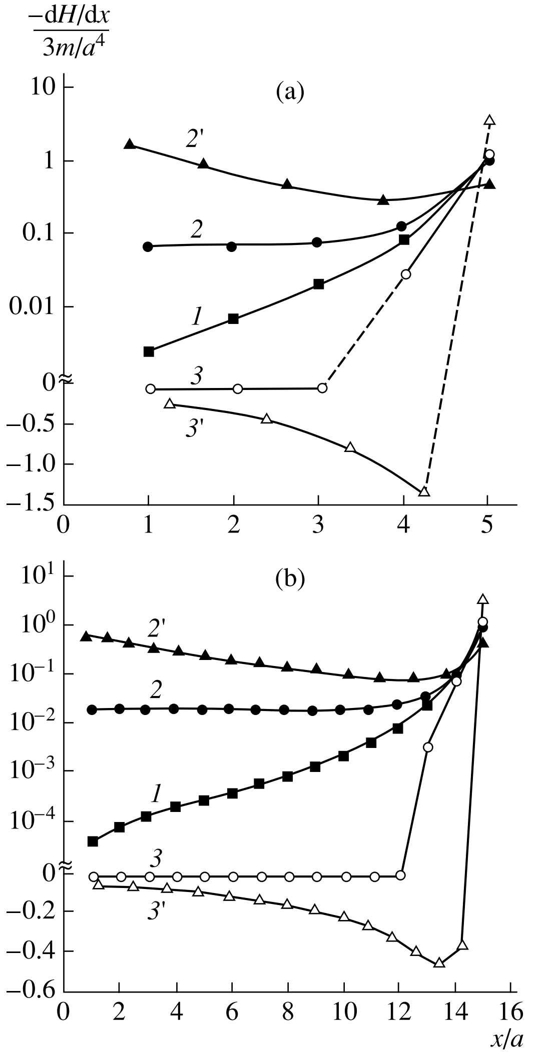


Fig. 1. Magnetic field gradient calculated vs. the coordinate of clusters for N = (a) 11 and (b) 31. (*1*), uniform distribution of the clusters along the y axis; (*2*) 5, (*3*) –5, (*2*') 50, and (*3*') –50% nonuniformity.

the ends. The case $\Delta = 0$ corresponds to the periodic distribution,

$$\delta a = a_N - a_1 = 2\frac{N-1}{2}\Delta a = (N-1)\Delta a;$$

hence,

$$\Delta = \frac{a_N - a_1}{a(N-1)}.$$

From relationship (2), we find the value of the compensating gradient at an ith dipole:

$$(\partial \mathbf{H}/\partial x)_{x_i} = -3m\left[\sum_{j=-N}^{i-1}\frac{1}{(x_j-x_i)^4} - \sum_{j=i+1}^{N}\frac{1}{(x_j-x_i)^4}\right]. \tag{4}$$

The magnetic field gradient versus dipole coordinate dependence calculated by formula (4) is shown in Figs. 1a (N = 5) and 1b (N = 15).

The gradient profile for the cluster nonuniform distribution differs from that corresponding to the uniform distribution both quantitatively and qualitatively. For a 5% nonuniformity ($\delta a/a = 0.05$), the required gradient remains virtually unchanged except at the periphery (curves *2* and *3* in Fig. 1), while for a positive nonuniformity of 50% ($\Delta > 0$), the gradient even increases in absolute value toward the center, unlike the uniform case. In addition, to produce a linear array (row) of clusters with a negative nonuniformity ($\Delta < 0$), the gradients in the central part and at the periphery must be opposite to each other. This means that the uniform structure remains the equilibrium state of the system when the field distribution varies over wide limits; therefore, such a phenomenon may well be called self-organization.

The second conclusion following from the calculations is that a given spatial distribution of magnetic clusters can be provided if the magnetic field is appropriately distributed over the array of clusters; namely, the field must sharply increase in absolute value at the periphery of the array. Obviously, only in this case may the force opposing the repulsion force of magnetic dipoles arise.

The field gradient in Fig. 1 is given in units of $3m/a^4$, and its minimum is the same for N = 5 and 15. This means that the higher the gradient at the extremities of the working interval, the shorter the period (the smaller a). This circumstance makes it possible to control the cluster array period.

The assumption of point clusters weakly affects the calculation results, since the influences of neighboring symmetrically arranged clusters compensate for each other, so that remote clusters can be viewed as points, ($(N + t)a \gg h$), where h is the cluster height.

EXPERIMENTAL

As a model material, we took monodisperse magnetite powder entering, together with polymers and black dye, into the toner of a laser printer. The mean size of magnetic particles was 1.5 μm. Ethyl alcohol was chosen as a carrier liquid. The concentration of the magnetic powder in the ML was 100 mg/cm^3. No anticoagulants were used. In the absence of a magnetic field, the powder settled at the bottom of the container. The ML containers were glass capillaries with inner diameters of 120, 200, and 370 μm.

An external magnetic field was produced by either Helmholtz coils or a permanent magnet. Its strength was varied from 0 to 260 G, and its direction could be arbitrary.

The nonuniform distribution of the magnetic field induction in the capillary plane used in the experiment (the measurements were taken in 0.5-cm steps by a

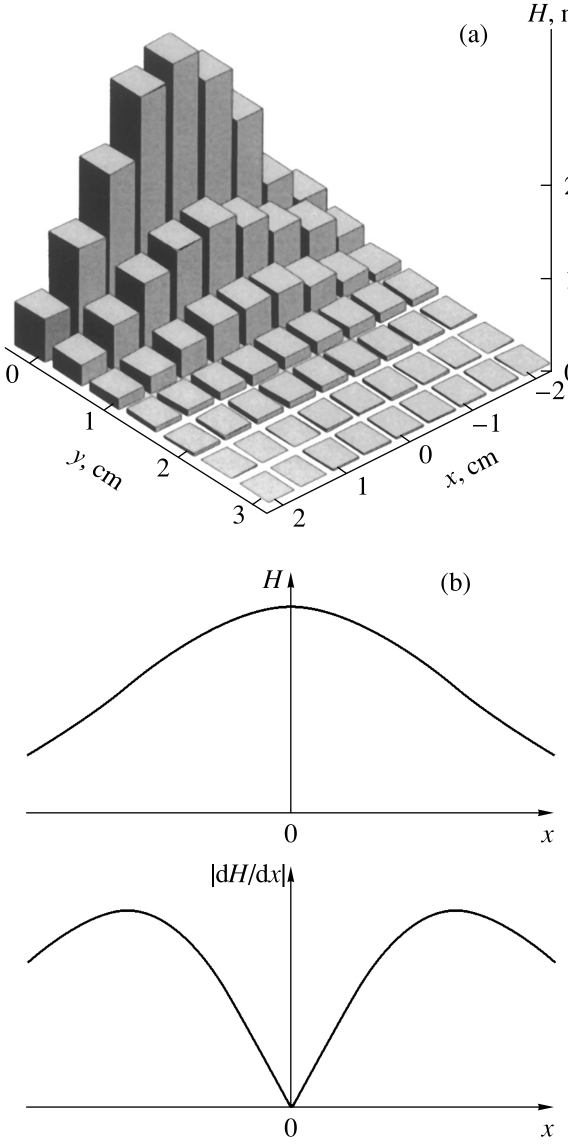


Fig. 2. (a) Magnetic induction distribution measured in the plane of capillary $y = 0$ and (b) field gradient profile along the capillary.

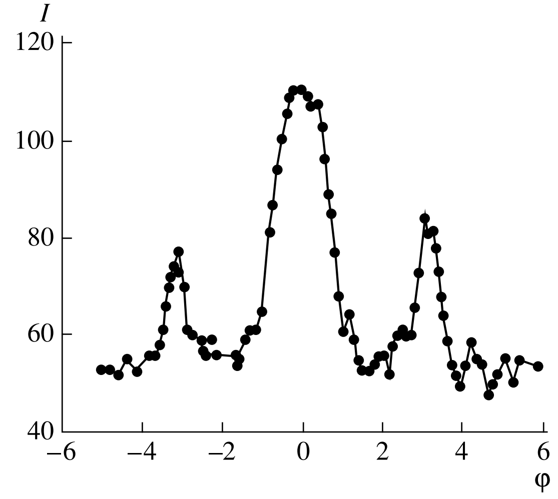


Fig. 3. Light intensity diffracted from the ordered structure of clusters.

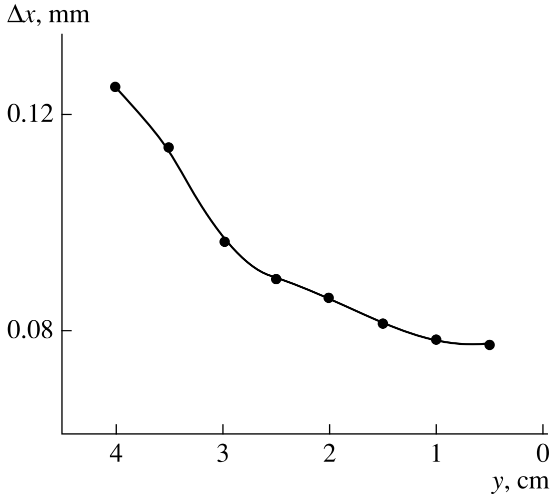


Fig. 4. Period Δx of the linear cluster array vs. the capillary–permanent magnet distance. The period was measured at the zero-gradient point ($x = 0$).

teslameter) is shown in Fig. 2a. Figure 2b shows the distribution of the field gradient.

EXPERIMENTAL RESULTS

Once the magnetic field is applied, acicular clusters start growing from the magnetic sediment in the capillary. By the time the growth terminates, the height of the clusters is approximately the same (110 μm). It is noteworthy that they stop growing even if the capillary diameter is sufficiently large and allows for further growth. The magnetic induction at which the height of the clusters does not change is roughly 15 mT. In the uniform magnetic field, the clusters are randomly spaced and form a linear array, the plane of the clusters passing through the capillary axis. In the presence of a bilateral field gradient, the clusters are arranged near its maximum and their distribution becomes periodic. The degree of periodicity is fairly high, as indicated by the weakly diffuse first maxima of the light diffracted from the ordered structure of clusters obtained in this way (Fig. 3). The diffraction efficiency of the structure is limited by a small ratio of the cluster diameter to the structure period.

As the gradient increases (i.e., as the magnet is approached; see Fig. 2), the period decreases (Fig. 4). When it exceeds a certain threshold, the clusters are rearranged to form two or even more rows. In this case, their structure is close to hexagonal (Fig. 5). The spacing between the clusters in such a structure remains unchanged. In our opinion, here, as in the single row of clusters, the interaction between the magnetized clus-

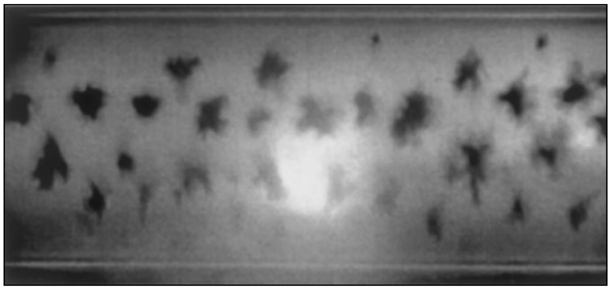


Fig. 5. Acicular clusters (top view) after splitting of the linear array into several rows in the nonuniform external magnetic field. The magnetic field direction is perpendicular to the capillary axis and to the plane of the photo.

ters and the interface between the magnetic liquid and the nonmagnetic material (glass) (i.e., the emergence of forces repelling the clusters from the interface) plays a certain role.

A change in the magnetic field direction causes the corresponding rotation of the cluster axes. In the uniform magnetic field, the clusters turn synchronously about their centers without a displacement along the capillary axis.

Comparison of the experimental gradient profile (Fig. 2) with the theoretical curves (Fig. 1) leads to conclude that the shape of the field used in the experiment qualitatively coincides with the shape required for the uniform distribution (curves *1* in Fig. 1). At the same time, it differs significantly from the shape providing for the distribution with a nonuniformity even as small as ±5% (curves *2* and *3* in Fig. 1). It is therefore not surprising that the resulting cluster distribution exhibits a high degree of periodicity (see Fig. 3).

The possibility of controlling the period is illustrated in Fig. 4, which plots period Δx of the structure against the distance between the magnet and the container with the ML. It follows from Fig. 2 that a decrease in this distance causes an increase in gradient $|dH/dx|$, while a higher gradient corresponds to a smaller period of the linear structure, in accordance with the above calculations (Fig. 1). The change in the period observed in the experiment amounts to 60%. A further monotonic decrease in the period with increasing field gradient is hampered by the formation of additional rows of clusters. This result agrees with the findings in [11], where a similar effect was observed at an increase in the uniform magnetic field.

CONCLUSIONS

(1) Our method of applying an appropriately configured gradient magnetic field to a magnetic liquid contained in a capillary made it possible to produce periodic structures of acicular clusters of magnetic particles.

(2) The distinguishing features of the behavior of a magnetic liquid contained in a capillary and subjected to an external magnetic field with a gradient at the ends of the capillary are as follows: (i) the clusters are arranged into a periodic structure along the capillary axis; (ii) the period of this structure can be varied over certain limits (up to 60% in our case) by changing the magnetic field; (iii) needle-shaped clusters rotate freely in the magnetic field according to its sense; and (iv) beginning from a certain value of the magnetic field, the ordered linear array of clusters splits into several rows with the formation of a hexagonal structure.

(3) The calculated profile of the magnetic field gradient that produces a uniform distribution (ordering) of magnetic clusters in the capillary coincides with the experimental profile, thus confirming the validity of the model proposed.

(4) In the calculated magnetic field profile corresponding to the nonuniform distribution of magnetic clusters, the field gradient grows at the periphery. At the same time, it may be constant, increase, or decrease in the central part of the working region of the capillary, depending on the nonuniformity parameter.

ACKNOWLEDGMENTS

The authors are grateful to V.N. Krivoruchko, I.A. Danilenko, and T.E. Konstantinova (Galkin Physicotechnical Institute, Donetsk, Ukraine) for submitting magnetite nanopowders.

REFERENCES

1. M.-P. Pileni, Adv. Funct. Mater. **11**, 323 (2001).
2. X. Xu, G. Friedman, K. D. Humdeld, *et al.*, Chem. Mater. **14**, 1240 (2002).
3. R. E. Rosensweig, *Ferrohydrodynamics* (Cambridge Univ. Press, Cambridge, 1985; Mir, Moscow, 1989).
4. P. G. de Gennes and P. A. Pincus, Condens. Matter Phys. **11**, 189 (1970).
5. P. C. Jordan, Mol. Phys. **25**, 1412 (1973).
6. X. Xu, G. Friedman, K. D. Humdeld, *et al.*, Chem. Mater. **14**, 1249 (2002).
7. Chin-Yin Houg, J. Magn. Magn. Mater. **201**, 178 (1999).
8. X. Xu, G. Friedman, K. D. Humdeld, *et al.*, J. Am. Chem. Soc. **124**, 13864 (2002).
9. S. Y. Yang, Y. F. Chen, H. E. Horng, *et al.*, Appl. Phys. Lett. **81**, 4931 (2002).
10. H. E. Horng, Chin-Yin Hong, S. Y. Yang, and H. C. Yang, Appl. Phys. Lett. **82**, 2434 (2003).
11. S. Y. Yang, H. E. Horng, Chin-Yin Hong, *et al.*, J. Appl. Phys. **93**, 3457 (2003).
12. V. M. Dubovik, M. A. Martsenyuk, and N. M. Martsenyuk, J. Magn. Magn. Mater. **145**, 211 (1995).
13. V. G. Bar'yakhtar and Yu. I. Gorobets, *Magnetic Bubbles and Their Lattices* (Naukova Dumka, Kiev, 1988) [in Russian].

Translated by N. Wadhwa

Technical Physics, Vol. 50, No. 6, 2005, pp. 771–779. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 6, 2005, pp. 98–106.
Original Russian Text Copyright © 2005 by Khomchenko, Sotsky, Romanenko, Glazunov, Shulga.

EXPERIMENTAL INSTRUMENTS AND TECHNIQUES

Waveguide Technique for Measuring Thin Film Parameters

A. V. Khomchenko, A. B. Sotsky, A. A. Romanenko,
E. V. Glazunov, and A. V. Shulga
Institute of Applied Optics, National Academy of Sciences of Belarus, Mogilev, 212793 Belarus
e-mail: avkh@physics.belpak.mogilev.by

Received July 27, 2004

Abstract—A waveguide technique for measuring the absorption coefficient, refractive index, and thickness of thin films is suggested. It is based on taking the angular dependence of the light beam reflection coefficient in an optical scheme involving a prism coupler. Application of the technique to determining the parameters of thin-film waveguides, insulating coatings, and metal films is considered. *© 2005 Pleiades Publishing, Inc.*

INTRODUCTION

Advances in optics and electronics are stimulating upgrading of existing and development of new procedures for measuring thin film parameters. Promising in this respect are integral optical methods based on taking the angular dependence of light beam reflection coefficient $R(\gamma)$ when thin-film structure modes are excited with a prism [1–5]. From the position of a minimum in the dependence $R(\gamma)$, the real part of the complex propagation constant for a guided mode excited was measured [1–3], and substituting Reh found (h is the complex propagation constant) into known dispersion relations yielded the refractive index and thickness of the films. Similar techniques have been developed for measuring the parameters of films guiding leaky modes [4]. An approach suggested in [5], which is also based on taking and processing curves $R(\gamma)$, makes it possible to find both the real and imaginary parts of the mode propagation constant and, thereby, estimate the refractive index, absorption coefficient, and thickness of a guiding film. In this paper, the whys and wherefores of this approach are presented and its efficiency as applied to thin-film waveguides, insulating coatings, and metal films is considered.

THEORY OF THE METHODS

When guided modes are excited with a prism coupler (Fig. 1), a series of so-called *m* lines can be observed in the reflected light [1]. The typical angular dependence of light beam reflection coefficient $R(\gamma)$ under resonance excitation of guided modes is shown in Fig. 2 (curve *1*). The same dependence is observed for leaky modes in a waveguide where the refractive index of the film is lower than that of the environment (Fig. 2, curves *2* and *3*) and also for plasmon modes (curve *4*) guided by a metal film.

To gain a deeper insight into the dependence $R(\gamma)$, we suppose that a d thick film with complex permittivity ε_w that guides optical modes covers a substrate with complex permittivity ε_s and is in contact with a isosceles prism (see Fig. 1). The permittivities of the prism with base angle θ, prism-surrounding environment, and a g thick gap are known (ε_p, ε_e, and ε_g, respectively). In addition, ε_p and ε_a ($\varepsilon_p > \varepsilon_a$) are assumed to be real. The structure is illuminated by a coherent light beam with its axis making angle γ with the normal to a lateral face of the prism.

Let us take advantage of the results in [6], where the power density of a reflected light beam was calculated for the case of a prism-excited planar optical waveguide in the vector electrodynamic statement. Integrating relationships (9) and (10) in [6] yields

$$R(j) = S_r(j)S_{\mathrm{i}}^{-1}, \tag{1}$$

$$\begin{gathered}
S_{\mathrm{i}} = \int\limits_{-\infty}^{\infty}\!\!\int dq_0 dq_1 z_0 (q_0^2 + q_1^2)^{-1} \\
\times[|q_0\widehat{E}_x + q_1\widehat{E}_y|^2 + \varepsilon_{\mathrm{e}} z_0^{-2}|q_1\widehat{E}_x - q_0\widehat{E}_y|^2],
\end{gathered} \tag{2}$$

$$S_{\mathrm{i}}(\gamma) = \int\limits_{-\infty}^{\infty}\!\!\int dq_0 dq_1 z_a^2 [z_0(q_1^2 + \beta^2)^2 (q_1^2 + q_2^2)^3]^{-1}$$

$$\times(|\varphi_r|^2 + \varepsilon_{\mathrm{e}}|\psi_r|^2),$$

$$\varphi_r = (1 - z_a z_{\mathrm{p}}^{-1})^{-1}\{\varphi_{\mathrm{i}}(r_p\varepsilon_{\mathrm{p}} q_1^2 \sin^2\theta - r_s s^2)(1 + z_p z_a^{-1})^{-1}$$

$$-\psi_{\mathrm{i}}\varepsilon_{\mathrm{e}} q_1 a(r_s + r_p)[z_a(1 + z_{\mathrm{p}}\varepsilon_{\mathrm{e}} z_a^{-1}\varepsilon_{\mathrm{p}}^{-1})]^{-1}\sin\theta\},$$

$$\psi_r = (1 + z_a\varepsilon_{\mathrm{p}} z_{\mathrm{p}}^{-1}\varepsilon_{\mathrm{e}}^{-1})^{-1}$$

$$\times\{\varphi_{\mathrm{i}}\varepsilon_{\mathrm{p}}\varepsilon_{\mathrm{e}}^{-1} q_1 s(r_s + r_p)(1 + z_{\mathrm{p}} z_a^{-1})^{-1}\sin\theta$$

1063-7842/05/5006-0771$26.00 © 2005 Pleiades Publishing, Inc.

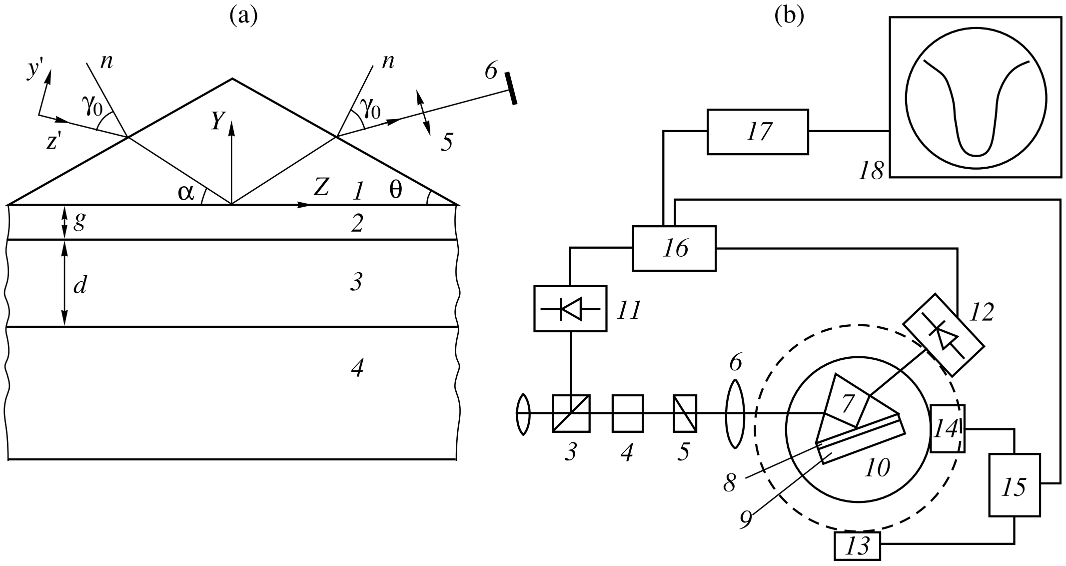


Fig. 1. (a) Prism coupler and (b) setup for taking the angular dependence of the light beam reflection. (*1*) Laser, (*2*) collimator, (*3*) beam splitter, (*4*) attenuator, (*5*) polarizer, (*6*) lens, (*7*) prism, (*8*) gap, (*9*) thin-film structure, (*10*) rotary table, (*11*, *12*) photodetectors, (*13*, *14*) stepping motors, (*15*) motor synchronizer, (*16*) comparator, (*17*) ADC, and (*18*) PC.

$$+\psi_{\mathrm{i}}(\varepsilon_{\mathrm{p}}a_1^2 r_s \sin^2\theta - s^2 r_p)[z_a(1 + z_{\mathrm{p}}\varepsilon_{\mathrm{e}}z_a^{-1}\varepsilon_{\mathrm{p}})]^{-1}\}, \quad (3)$$

$$\varphi_{\mathrm{i}} = (q_1^2\sin^2\gamma - q_2 z_0)z_a^{-1}\widehat{E}_x - q_1\widehat{E}_y,$$

$$\psi_{\mathrm{i}} = q_1(z_0 - q_2\sin\gamma)z_a^{-1}\widehat{E}_x - q_2\widehat{E}_y,$$

$$s = q_1^2\cos\theta - q_2\beta, \quad \beta = z_{\mathrm{p}}\sin\theta - q_2\cos\theta,$$

$$q_2 = z_0\sin\gamma - q_0\cos\gamma,$$

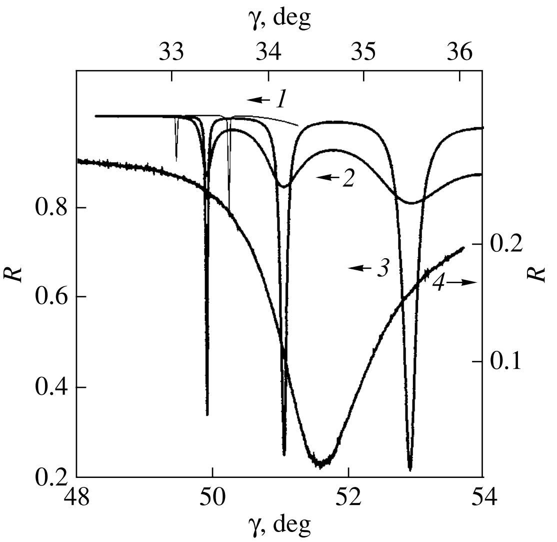


Fig. 2. Angular dependence of the light beam reflection coefficient for the (*1*) SiO$_x$ film/quartz glass, (*2*) SiO$_x$/K8 glass, (*3*) SiO$_x$/Si, and (*4*) air/aluminum waveguide structures. w_0 = 490 μm.

$$z_0 = \sqrt{\varepsilon_{\mathrm{e}} - q_0^2 - q_1^2}, \quad z_{a,\,\mathrm{p}} = \sqrt{\varepsilon_{\mathrm{e,\,p}} - q_1^2 - q_2^2},$$

$$\widehat{E}_{x,\,y} = \int\limits_{-\infty}^{\infty}\!\!\int dxdyE_{x,\,y}\big|_{z\,=\,0}\exp(iq_0k_0y + iq_1k_0x).$$

Here, E_x and E_y are the electric field components of the exciting beam, $k_0 = 2\pi\lambda_0^{-1}$ is the wavenumber of vacuum, and r_s and r_p are the coefficients of reflection of *s*- and *p*-polarized plane waves from the base of the prism for angle of incidence $\theta_{\mathrm{i}} = \arctan\sqrt{(q_1^2 + q_2^2)(\varepsilon_{\mathrm{p}} - q_1^2 - a_2^2)^{-1}}$. The infinite limits of integration in (2) and (3) are taken on the assumption that functions $\widehat{E}_x(q_0, q_1)$ and $\widehat{E}_y(q_0, q_1)$ acquire infinitesimal values outside the domain $q_0^2 + q_1^2 \le \varepsilon_{\mathrm{e}}$.

Expressions (1)–(3) are very awkward and can hardly be used for solving the inverse problem of reconstructing the complex propagation constant of a mode excited. They can be simplified if we assume that the waveguide is excited by a polarized beam; that is,

$$E_x\big|_{z\,=\,0} = (1 - \chi)\psi(x, y), \quad E_y\big|_{z\,=\,0} = \chi\psi(x, y),$$

where χ, which describes the transverse distribution of the beam field, equals either 0 or 1.

Let function $\psi(x, y)$ be even in x, $\psi(x, y) = \psi(-x, y)$, and the characteristic scale of its variation, $w_0(|\underline{\nabla}\psi| \sim |\psi|w_0^{-1})$, meet the condition $k_0w_0 \gg 1$. In this case, the

Fourier transform of function $\psi(x, y)$,

$$\widehat{\psi}(q_0, q_1) = \int\int_{-\infty}^{\infty} dxdy\psi(x, y)\exp(iq_0k_0y + iq_1k_0x) \quad (4)$$

is far from zero in the domain $q_0^2 + q_1^2 \le (k_0w_0)^{-2}$ [7], so that quantities of order q_0^2, q_1^2, q_1 and q_0 in (2) and (3) can be neglected.

Eventually, expression (1) can be recast as

$$R(\gamma) = 16k^2(1+k)^{-4}\bar{R}(\gamma), \quad (5)$$

where

$$k = (\varepsilon_{\mathrm{p}}\varepsilon_{\mathrm{e}}^{-1})^{\chi}(\varepsilon_{\mathrm{p}}\varepsilon_{\mathrm{e}}^{-1} - \sin\gamma)^{-0.5}\cos\gamma,$$

$$\bar{R}(\gamma) = \left(\int\int_{-\infty}^{\infty} dq_0dq_1|\widehat{\psi}|^2\right)^{-1}\int\int_{-\infty}^{\infty} dq_0dq_1|r_\chi|^2|\widehat{\psi}|^2 \quad (6)$$

and $r_\chi = (1-\chi)r_s + \chi r_p$ is calculated at $q_1 = 0$. Applying the resonance approximation to $|r_\chi|^2$ in (6) [6], we get

$$\bar{R}(\gamma) = \frac{|(1-\delta)(1+\delta)^{-1}|^2}{\int\int_{-\infty}^{\infty} d\tau dq_1|\widehat{\psi}|^2}\int\int_{-\infty}^{\infty} d\tau dq_1 \quad (7)$$

$$\times\frac{|\widehat{\psi}|^2[(p_1 + 2|p_2|\cos\sigma)^2 + (\tau\sqrt{2} - p_4 + 2|p_2|\sin\sigma)^2]}{p_1^2 + (\tau\sqrt{2} - p_4)^2}.$$

Here,

$$p_1 = w\mathrm{Im}\bar{h}, \quad |p_2|\exp(i\sigma) = -2iw\Delta h\delta(1-\delta)^2,$$
$$p_4 = k_0\sqrt{\varepsilon_{\mathrm{e}}}w_0(\gamma - \bar{\gamma}); \quad (8)$$

$$\delta = -i(\varepsilon_{\mathrm{p}}\varepsilon_q^{-1})^{\chi}\sqrt{(k_0^{-2}h^2 - \varepsilon_{\mathrm{g}})(\varepsilon_{\mathrm{p}} - k_0^{-2}h^2)^{-1}}; \quad (9)$$

$$w = \frac{w_0(\sqrt{1-\alpha^2}\cos\theta + \alpha\sin\theta)}{\sqrt{(1-\alpha^2)[1 - \varepsilon_{\mathrm{p}}\varepsilon_{\mathrm{e}}^{-1}(\sqrt{1-\alpha^2}\sin\theta - \alpha\cos\theta)^2]}}; \quad (10)$$

$\tau = q_0k_0w_0(\sqrt{2})^{-1}$; $\alpha = k_0^{-2}(\mathrm{Re}h)^2\varepsilon_{\mathrm{p}}^{-1}$; $\Delta h = \bar{h} - h$; $\bar{h}$ is the propagation constant of the leaky mode in the film structure in the presence of the prism coupler; and $\bar{\gamma}$ is a root of the equation $\beta(\bar{\gamma}) = k_0^{-1}\mathrm{Re}\bar{h}$ (where $\beta(\gamma) = \sin\theta\sqrt{\varepsilon_{\mathrm{p}} - \varepsilon_{\mathrm{e}}\sin^2\gamma} - \sin\gamma\cos\theta\sqrt{\varepsilon_{\mathrm{p}}}$).

Expression (7) is valid for an arbitrary planar waveguide excited by a prism. In this expression, the properties of a specific waveguide are embodied through factor Δh. An explicit expression for Δh is given in [6], where it was shown, in particular, that σ is a small parameter, which has a considerable effect only when leaky or plasmon modes are considered.

Further analysis of expression (7) requires that function $\widehat{\psi}(q_0, q_1)$ be specified. Let oscillations in the film be excited by a Gaussian beam, which approximates the radiation field of a single-mode gas laser. Putting $\psi \sim \exp[-(xw_0^{-1})^2 - (yw_0^{-1})^2]$ in (4) and ignoring quantities of order σ^2 in (7), we arrive at

$$\bar{R}(\gamma) = \left|\frac{1-\delta}{1+\delta}\right|^2\left\{1 - 4\sqrt{\frac{2}{\pi}}|p_2|\right.$$
$$\left.\times\left[\left(\frac{|p_2|}{p_1} + 1\right)\mathrm{Re}G(\xi) - \sigma\mathrm{Im}G(\xi)\right]\right\}, \quad (11)$$

where $\xi = -p_1 + ip_4$ and

$$G(\xi) = i(\sqrt{2})^{-1}\int_{-\infty}^{\infty}\exp(-\tau^2)(i\xi - \tau\sqrt{2})^{-1}d\tau.$$

To reconstruct mode propagation constant h, we assume that the dependence $\bar{R}(\gamma)$ is taken within angular range $|\gamma - \gamma_0| \le a$, where γ_0 is the angle at which $\bar{R}(\gamma)$ reaches minimal value $\bar{R}_0$. According to (11), $\bar{R}(\gamma)$ reaches a minimum at $p_4 = p_4^{(0)}$, where

$$p_4^{(0)} = \sigma N_0,$$

$$N_0 = [p_1G(-p_1) + \sqrt{0.5\pi}](1 + |p_2|p_1^{-1})^{-1} \quad (12)$$
$$\times[p_1\sqrt{0.5\pi} + (1 + p_1^2)G(-p_1)]^{-1}.$$

Here,

$$\bar{R}_0 = |(1-\delta)(1+\delta)^{-1}|^2$$
$$\times[1 - 4|p_2|\sqrt{2\pi^{-1}}(|p_2|p_1^{-1} + 1)G(-p_1)]. \quad (13)$$

From (8) and (13), we have

$$h = k_0\beta(\gamma_0)$$
$$+ w^{-1}\{p_4^{(0)} + i[p_1 - p_2(1-\delta)^2(2\delta)^{-1}]\}, \quad (14)$$

where $p_2 = |p_2|\exp(i\sigma)$,

$$|p_2| = 0.5[-p_1 + (-1)^{\rho}$$
$$\times\sqrt{p_1^2 + \sqrt{0.5\pi}p_1(1 - \bar{R}_0|(1+\delta)(1-\delta)^{-1}|^2)[G(-p)]^{-1}}], \quad (15)$$

and ρ equals either 0 or 1.

Expressions (12)–(15) involve unknown parameters p_1 and σ. To find them, we note that the real part and imaginary part of function $G(\xi)$ are, respectively, an even and odd function of parameter p_4. With this in

Table 1. Mode propagation constants and film parameters restored by numerical simulation

w_0, μm	Mode type	Re (hk_0^{-1})	Im (hk_0^{-1})	Re ε	Im ε	d, μm
∞	TE_0	1.5126356	-9.8182×10^{-6}	2.295226	-3.0298×10^{-5}	3.1636
	TM_0	1.5125604	-9.8036×10^{-6}			
500	TE_0	1.5126356	-9.8186×10^{-6}	2.295221	-3.0313×10^{-5}	3.1651
	TM_0	1.5125605	-9.8099×10^{-6}			
250	TE_0	1.5126356	-9.8184×10^{-6}	2.295221	-3.0298×10^{-5}	3.1651
	TM_0	1.5125605	-9.8074×10^{-6}			
100	TE_0	1.5126359	-9.8144×10^{-6}	2.295227	-3.0287×10^{-5}	3.1635
	TM_0	1.5125607	-9.8003×10^{-6}			
50	TE_0	1.5126354	-9.4941×10^{-6}	2.295223	-2.9296×10^{-5}	3.1652
	TM_0	1.5125613	-9.2525×10^{-6}			

mind, we integrate relationship (11) to find

$$\sigma = k_0\sqrt{\varepsilon_e}w_0\left|(1+\delta)(1-\delta)^{-1}\right|^2 \times N_1^{-1}\left(\int_{\gamma_0}^{\gamma_0+a}\bar{R}(\gamma)d\gamma - \int_{\gamma_0-a}^{\gamma_0}\bar{R}(\gamma)d\gamma\right), \tag{16}$$

$$N_1 = 8\sqrt{2\pi^{-1}}|p_2|\left\{\mathrm{Im}\int_0^{\bar{a}}Gdp_4 - (|p_2|p_1^{-1}+1) \times [\mathrm{Re}G(-p_1+i\bar{a}) - G(-p_1)]N_0\right\},$$

$$\frac{1}{G(-p_1)}\mathrm{Re}\int_0^{\bar{a}}Gdp_4 = \frac{0.5k_0\sqrt{\varepsilon_e}w_0}{1-\bar{R}_0\left|(1+\delta)(1-\delta)^{-1}\right|^2} \times\left[2a - \left|(1+\delta)(1-\delta)^{-1}\right|^2\int_{\gamma_0-a}^{\gamma_0+a}\bar{R}(\gamma)d\gamma\right], \tag{17}$$

where $\bar{a} = k_0\sqrt{\varepsilon_e}w_0a$.

Expressions (12) and (14)–(17) make it possible to solve the inverse problem of reconstructing complex propagation constant h from distribution $\bar{R}(\gamma)$ provided that quantities w_0, ε_e, ε_p, ε_g, and k_0 are known. Here, central is the solution of transcendental equation (17) for parameter p_1 (it can be shown that this equation always has a unique negative root). Then, h is calculated directly using (12) and (14)–(16). Note that δ and w_0 in (12)–(17) depend on h (see (9) and (10)). Since quantity $(k_0w)^{-1}$ in (14) is small, these parameters can be calculated for $h = k_0\beta(\gamma_0)$ and then refined by iterations if necessary.

The obtained solution to the inverse problem is valid for modes propagating near the surface of a waveguide with an arbitrary refractive index. It is determined only by integration of function $\bar{R}(\gamma)$ found experimentally, which renders the results statistically stable. Moreover, expressions (12) and (14)–(17) lack difficult-to-control parameter g, which is responsible for prism-induced measurement errors. On the other hand, filtering out of the prism interference makes ρ in (15) double-valued. With the equality $|p_2| = -0.5p_1$ fulfilled, the degree of prism–waveguide coupling (or thickness g) is such that one of three possible values of $\bar{R}_0$ is achieved, $\bar{R}_{0\min} = |(1-\delta^2)(1+\delta^2)^{-1}|^2[1+\sqrt{2/\pi}\,p_1G(-p_1)]$. Accordingly, at $\rho = 0$ or 1 in (15), the coupling is stronger or weaker than this value. Practically, this means that one should decide between two values of the propagation constant that follow from experimental data processing, either may turn out to be true and so cannot be eliminated beforehand. In this situation, the curves $\bar{R}(\gamma)$ should be constructed for several values of g, which is varied by loosening or tightening the prism–waveguide contact. The value of h corresponding to the correct value of ρ will remain unchanged, while the other value will vary. Such an expedient was used in the experiments described below.

To determine complex permittivity ε_w and thickness d of the film, the values of h for two modes are substituted into known dispersion relations. The resulting system of two complex nonlinear equations is solved by contour integration as described in [8].

When deducing relationships (11)–(17), we made a number of assumptions that were validated by estimating the accuracy of numerical solutions to the inverse problem. As “experimental” data, we used distributions $\bar{R}(\gamma)$ calculated from (1)–(3) for different w_0. Table 1 demonstrates a typical example of calculation for a waveguide film with $\varepsilon_w = 2.295225 - i3.03 \times 10^{-5}$ and $d = 3.164$ μm covering a substrate with permittivity $\varepsilon_s = 2.25 - i3.0 \times 10^{-6}$ ($\lambda_0 = 0.6326$ μm). This film supports two *TE* and two *TM* polarization modes. The exact values of the propagation constants are $h_{TE_0} = 1.512635657 - i9.8187 \times 10^{-6}$ and $h_{TM_0} = 1.512560485 - i9.80548 \times 10^{-6}$. As follows from these data, the propagation constants of the modes and the film parameters are recovered most exactly in the plane-wave approximation ($w_0 \longrightarrow \infty$). As w_0 decreases, the error in film parameter determination grows, since the vector character of the problem of beam reflection, which is ignored in approximation (5), is highlighted. Nevertheless, the data presented indicate the applicability of the given approach to thin film parameter measurement.

EXPERIMENTAL

To measure the distribution $R(\gamma)$, we designed a computerized setup the schematic of which is shown in Fig. 1. A Gaussian beam from a He–Ne laser with $\lambda_0 = 0.633$ μm and cross-sectional radius $w_0 = 490$ μm is incident on prism *7* mounted on rotary table *10*. The angle of incidence of the beam on the prism was varied in 20" steps with stepping motor *15*. The prism was made of TF12 optical glass with a refraction index of 1.77905 (at $\lambda_0 = 0.633$ μm). The radiation polarization and wavelength can be varied. Also, the light beam radius can also be varied from 70 to 500 μm by reconfiguring the optical scheme. The radius was measured from the intensity level $I = I_0/e$, where I_0 is the intensity at the center of the beam. Test sample *9* is pressed against the measuring prism so as to provide optimal conditions for optical mode excitation. The curve $R(\gamma)$ was recorded by means of photodetector *12*, which measures the intensity of the beam reflected from the prism coupler and moves under the control of another stepping motor *14*, and photodetector *11*, which measures the incident beam intensity (the unit controlling motor *16* is synchronized with comparator unit *17*). After digitization, the signal entered the PC RAM.

The basic points in taking an angular dependence of the light beam reflection coefficient when oscillations were excited in a thin-film structure are as follows. As the origin of the angles, we took the angular position of photodetector *12* at the instant it records the incident (unreflected) light beam (in this case, the prism is located outside the beam). After the angular position of the normal to one of the faces of the prism has been determined with regard to the prism coupler geometry and optical parameters of the prism and substrate, the processor tentatively finds a range of angles to be measured and sets the stage for measurements. Statistical data gathered for a given number of samples and a given number of measurements made on each sample are then averaged, and the resulting distribution of the reflection coefficient is memorized. A computer program built around this algorithm processes the reflection coefficient distribution and finds the real and imaginary parts of the mode propagation constant. When determining the absorption coefficient of the film material, we supposed that the attenuation of the beam due to scattering is much smaller than the attenuation due to absorption.

Table 2. Parameters of the thin waveguiding film that are determined by various methods

Parameters	Guided mode spectroscopy		Angular dependence of reflection	
	$m = 0$	$m = 1$	$m = 0$	$m = 1$
$h'k_0^{-1}$	1.46755	1.45814	1.46748	1.45810
$h''k_0^{-1}$	9.98×10^{-6}	6.51×10^{-6}	1.02×10^{-5}	6.71×10^{-6}*
n	1.47104		1.47099	
k	1.03×10^{-5}		1.08×10^{-5}	
d, μm	2.49		2.53	

* Optical losses 5.5 dB/cm.

DETERMINATION OF THE GUIDING FILM PARAMETERS

Thin guiding films studied in this work were prepared by rf magnetron scattering of KV quartz glass on a substrate made of the same glass in the $Ar : O_2 = 4 : 1$ atmosphere [8]. The parameters of such a waveguide supporting two modes of *TE* polarization are listed in Table 2. In this case, the errors in determining the real and imaginary parts of propagation constant were 2×10^{-5} and 0.015×10^{-5}, respectively. The correctness of the data obtained was checked by other techniques. The optical losses for the second mode were measured by scanning a fiber along the waveguide [9] were found to be 5.5 ± 0.1 dB/cm. The thickness of the film measured with a profilograph was found to be 2.51 ± 0.02 μm. These values are seen to be in satisfactory agreement with the data listed in Table 2.

In measuring thin film parameters to take an angular dependence of reflection coefficient, it is necessary to properly choose the probing beam diameter. Fairly correct results can be obtained only if the beam is rather wide. When the reflection coefficient distribution was measured by exciting a thin-film waveguide with Gaussian beams with diameters $w_0 = 90$, 145, and 490 μm, the optical parameters of the film were in best agree-

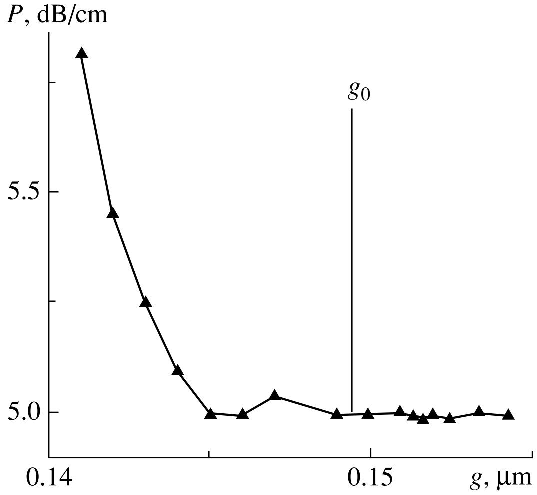


Fig. 3. Optical losses vs. the gap between the prism and thin-film structure.

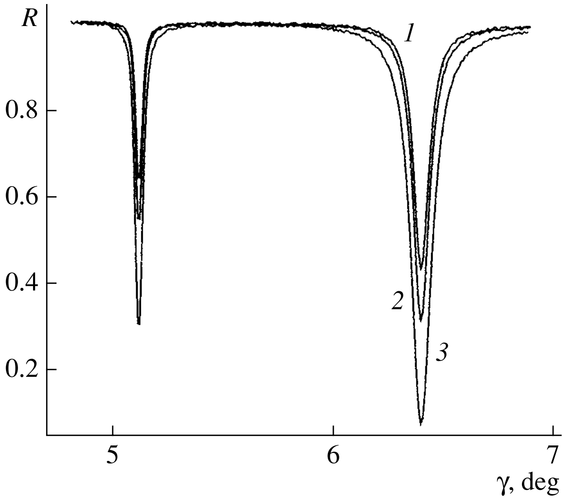


Fig. 4. Angular dependence of the beam reflection coefficient for the SiO_x film on the K8 glass substrate for different g (the smaller g, the larger the curve number).

ment with the results obtained by other techniques for $w_0 = 490$ μm. A possible explanation for such an effect has been given in the previous section (Table 1). It should also be noted that, when measuring low optical losses ($k < 10^{-5}$), we ran into the problem of energy leakage from the prism coupler. This is because our technique measures the total radiation attenuation in the film. If optical losses are low, the radiation escapes from the measuring prism and, thereby, "overestimates" the absorption coefficient being measured. Generally, the measurement of low losses is correct if light energy leakage from the prism is prevented, e.g., by shrinking the gap between the prism and film. As follows from Fig. 3, the measured energy losses remain unchanged if the gap thickness is smaller than some critical value g_0. Unfortunately, parallelism between the base of the prism and the sample surface breaks in this case, introducing an additional error (up to 20%) into the measurements of optical losses in the waveguide. Thus, special experimental conditions must be provided to obtain correct data for low absorption. At a moderate absorption, no difficulties arise as a rule.

Table 3. Results of processing the angular dependences of the reflection coefficient shown in Fig. 4 (the data for the mode $m = 1$)

Curve	h'/k_0		h''/k_0	
1	1.46512	1.46483	1.77×10^{-3}	1.22×10^{-4}
2	1.46512	1.46493	1.76×10^{-3}	7.97×10^{-5}
3	1.46512	1.46492	1.76×10^{-3}	3.39×10^{-5}

Table 4. Parameters of the SiO_x films deposited on various substrates

	n	$k \times 10^{-5}$	d, μm
SiO_x/SiO_2	1.47095	3.39	2.51
SiO_x/Si	1.47091	3.34	2.53
SiO_x/K8	1.47024	2.5	2.69

DETERMINATION OF THE NONGUIDING FILM PARAMETERS

Thin nonguiding films were grown by rf sputtering of quartz glass on substrates with a higher refractive index. The substrates were made of K8 glass and single-crystal silicon. At a wavelength of 0.633 μm, the refractive indices of the glass and silicon were, respectively, 1.5146 and 3.515. Such structures support only leaky modes. These structures and those discussed in the previous section were prepared simultaneously in a single process; therefore, it was expected that the optical parameters and thicknesses of both are close to each other. The technique for measuring the angular dependence $R(\gamma)$ and the processing algorithm were the same as in the case of the guiding films. Here, as before, we took into account the influence of the prism coupler on the guided mode parameters when measuring the properties of the reflected beam. It is well documented that, when the prism tunnel excitation technique is applied for measuring film parameters, the accuracy of the final results depends on the degree of prism–waveguide coupling (i.e., on the gap between the film and prism base) [10]. The angular dependence of the reflection coefficient also varies with the gap (Fig. 4). However, the results obtained with the technique suggested are independent of experimental conditions [5]. As was noted above, the true value of h remains constant irrespective of the pressure the prism exerts on the sample (i.e., irre-

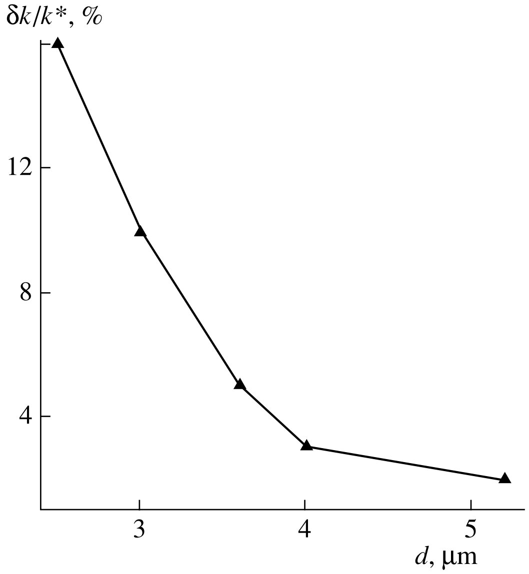


Fig. 5. Relative error in the absorption coefficient for the SiO_x film on the K8 glass substrate vs. the thickness of the film.

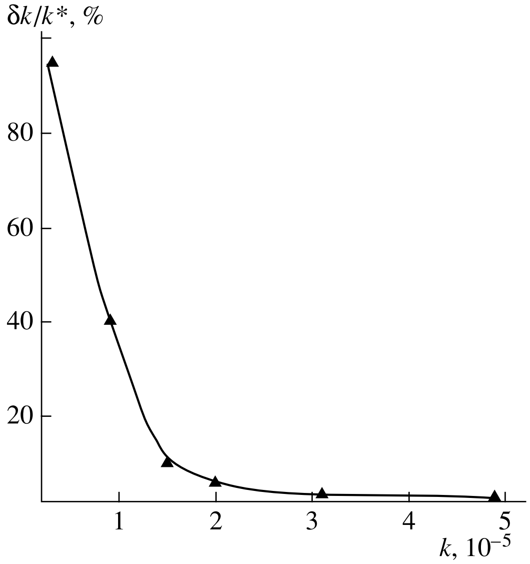


Fig. 6. Relative error in the absorption coefficient for the SiO_x film on the silicon substrate vs. the absorption coefficient value. The measurements were made for guided modes excited in reference SiO_x/SiO_2 structures.

spective of the gap between the prism and sample), while other values vary. This effect is dramatized in measuring the imaginary part of the propagation constant (Table 3). The parameters of SiO_x films deposited on different substrates that are derived from the angular dependences of the reflection coefficient (Fig. 2) are given in Table 4.

It is seen that the parameters of the films deposited on the K8 glass differ considerably from those of the films deposited on the other substrates. Presumably, because of a small difference between the refractive indices of the film and substrate ($\Delta n = n - n_s$), the leaky mode is weakly localized in the film. Measurements made on SiO_x films of roughly equal thickness (≈2.5 μm) deposited on various substrates corroborate this conjecture. In this case, for a refractive index of the substrates varying between 1.9 and 2.0, the absorption coefficient of the film coincides (within 3%) with the values obtained by other techniques. Obviously, the degree of localization of the mode field in the film and, hence, the reliability of results depend on the film thickness. Figure 5 shows the thickness dependence of the relative error in measuring absorption coefficient k. All the films were deposited under identical conditions and had nearly the same values of n (1.4701) and k (3×10^{-5}). Simultaneously, the error in determining the thickness and refractive index decreased from 6 to 1% and from 5×10^{-5} to 1×10^{-5}, respectively, with increasing thickness.

Our technique to measure the parameters of weakly absorbing nonguiding thin films has limitations. By way of example, consider the results for the SiO_x films on the silicon substrate. The films had roughly the same thickness (≈2.5 μm) but differ in composition, since they were grown by rf sputtering in atmospheres with a different oxygen concentration. Therefore, the absorption coefficient was also different. Figure 6 demonstrates how the relative error in determining the absorption coefficient depends on the absorption coefficient value in such structures. At $k < 10^{-5}$, the error exceeds 30%. This may be related to the fact that, at low k ($<10^{-5}$), the error in measuring $h''k_0^{-1}$ becomes comparable to, or even exceeds, the value of k itself. In this situation, determination of the absorption in the film becomes a challenge. As for the refractive index and thickness, the respective errors are no higher than 5×10^{-5} and 2–3%.

Thus, the measuring technique suggested in this work is applicable to structures where the refractive indices of the films and substrate differ considerably (Δn must be at least >0.5). If Δn is small, this approach provides valid results for $d \geq 5$ μm. If the absorption in the film is not to be measured, this technique can be used for estimating the refractive index and thickness of the film with an accuracy of $\sim 10^{-5}$ and $\sim 10^{-2}$ μm, respectively, provided that the refractive indices of the film and substrate differ insignificantly.

DETERMINATION OF THE PARAMETERS OF METAL FILMS SUPPORTING PLASMON MODES

The approach discussed above (determination of the complex propagation constant from the angular dependence of the reflection coefficient) also applies to plasmon modes propagating over the surface of metal films

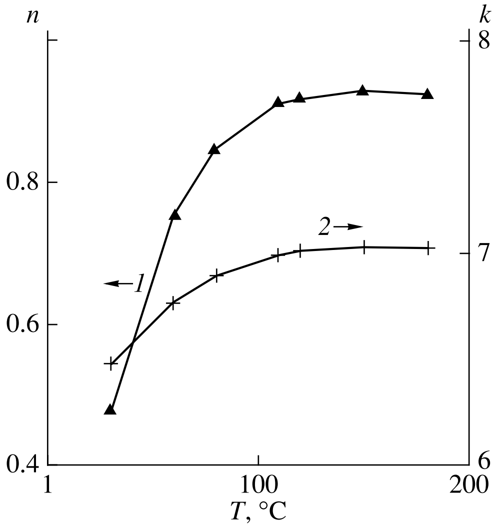


Fig. 7. Dependences of (*1*) refractive index n and (*2*) absorption coefficient k on the substrate temperature for the Al films on the quartz glass.

placed in an insulating environment. However, the thickness of a metal film can be measured if plasmon modes are excited at both its boundaries [11]. It is evident, however, that plasmon modes at the outer (relative to the prism coupler) boundary can be excited only if the film thickness lies in the range 300–500 Å because of a high absorption of visible radiation in a metal. Hence, our approach makes it possible to estimate the absorption coefficient, refractive index, and thickness of only very thin films. The measuring procedure is the same as that described in the previous sections.

For metal films more than 800 Å thick, the parameters of a plasmon mode excited at one interface are not affected by the second interface. Then, from the angular dependence of the reflected light beam intensity, one can determine only the optical parameters of a thick film or surface layers in massive metal samples. Since the propagation constant of a plasmon mode is related to the permittivity of a metal film (or a surface layer of a metal), $\varepsilon = \varepsilon' + i\varepsilon''$, and that of the environment, ε_e, as $h^2 = h' + ih'' = k_0^2\varepsilon_e\varepsilon/(\varepsilon_e + \varepsilon)$, we have [11]

$$\varepsilon' = [(h'^2 - h''^2)k_0^2\varepsilon_e - (h'^2 - h''^2)^2]z^{-1},$$

$$\varepsilon'' = 2k_0^2 h'h''\varepsilon_e/z,$$

where

$$z = (\varepsilon_e k_0^2 - h'^2 + h''^2)^2 + (2h'h'')^2.$$

Having measured propagation constant h, one can find the permittivity and, hence, refractive index and absorption coefficient of the metal. Figure 7 shows the temperature dependences of the optical parameters for aluminum films deposited on the quartz glass by cathode sputtering at different substrate temperatures. It is known that the films deposited at substrate temperature $T_0 = 120$ K are of good quality and exhibit a high adhesion to the substrate [12, 13]. From Fig. 7, it follows that the film has a high refractive index at temperatures $T > T_0$, which indicates that it is close-packed. Consequently, our technique allows one to estimate the parameters of thin metallic coatings and surface layers of massive metal samples, as well as to judge their quality.

CONCLUSIONS

In this paper, we substantiate a method of measuring the absorption coefficient, thickness, and refractive index of thin films that is based on taking the angular dependence of the light beam reflection coefficient for the case when guided (leaky or plasmon) modes are excited with a prism. The potential and domain of applicability of the method are considered. Whether the parameters of the films supporting leaky modes are measured correctly or not depends on the degree of localization of the mode fields in the film and on film thickness d. In turn, degree of localization depends on difference $\Delta n = n - n_s$ in refractive indices between the film and substrate. For $\Delta n \geq 0.5$, the measured values of k ($\delta k/k = 0.03$) coincide with those measured on a similar waveguiding film. For $\Delta n < 0.5$, relative error $\delta k/k$ depends on the film thickness: at $d \geq 3.5$ μm, it does not exceed 0.05 for a SiO_x film on a K8 glass substrate. The least value of k that can be measured by this method with a reasonable accuracy depends on losses due to leakage. For a 2.5-μm-thick SiO_x film deposited on a Si substrate, $\delta k/k = 0.1$ for $k = 10^{-5}$ and $\delta k/k = 0.03$ for $k = 3 \times 10^{-5}$. The refractive index and thickness of the films are measured accurate to 5×10^{-5} and 3%, respectively. Therefore, application of this method to measure the absorption of thin films supporting leaky modes alone is appropriate only if the refractive indices of the film and substrate differ considerably (at least by $\Delta n > 0.5$). When Δn is small, this method works well at film thicknesses $d \geq 5$ μm. The method suggested in this paper allows one to estimate the parameters of thin metallic coatings and surface layers of massive metal samples, as well as to judge their quality.

Thus, we developed a technique for measuring the optical properties of thin films that is based on taking the angular dependence of the light beam reflection coefficient when visible-range modes are excited in the films. It can be used for characterization of various thin-film structures in optics and microelectronics.

REFERENCES

1. P. R. Tien, R. Ulrich, and R. J. Martin, Appl. Phys. Lett. **14**, 291 (1969).
2. W. Lukosz and P. Pliska, Opt. Commun. **17**, 1 (1995).
3. R. Ulrich and R. Torge, Appl. Opt. **12**, 2901 (1973).
4. T. W. Hou and C. J. Mogab, Appl. Opt. **20**, 3184 (1981).

5. A. V. Khomchenko, A. B. Sotskiĭ, A. A. Romanenko, *et al.*, Pis'ma Zh. Tekh. Fiz. **28** (11), 51 (2002) [Tech. Phys. Lett. **28**, 467 (2002)].
6. A. B. Sotskiĭ, A. A. Romanenko, A. V. Khomchenko, and I. U. Primak, Radiotekh. Élektron. (Moscow) **44**, 687 (1997).
7. P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953; Inostrannaya Literatura, Moscow, 1995), Vol. 1.
8. A. A. Romanenko A. B. Sotskiĭ, Zh. Tekh. Fiz. **68** (4), 88 (1998) [Tech. Phys. **43**, 427 (1998)].
9. G. T. Petrovskiĭ, V. P. Red'ko, and A. V. Khomchenko, Zh. Tekh. Fiz. **54**, 2045 (1984) [Sov. Phys. Tech. Phys. **29**, 1199 (1984)].
10. R. Th. Kersten, Vak.-Tech. **23**, 16 (1973).
11. S. Monneret, P. Huguet-Chantome, and F. Flory, J. Opt. A **2**, 188 (2000).
12. H.-P. Chiang, Y.-C. Wang, and P. T. Leung, Thin Solid Films **425**, 135 (2003).
13. *Handbook of Thin-Film Technology,* Ed. by L. I. Maissel and R. Glang (McGraw-Hill, New York, 1970; Sov. Radio, Moscow, 1977), Vol. 1.

Translated by V. Isaakyan

Technical Physics, Vol. 50, No. 6, 2005, pp. 780–786. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 6, 2005, pp. 107–113.
Original Russian Text Copyright © 2005 by Rapis.

EXPERIMENTAL INSTRUMENTS AND TECHNIQUES

On the Problem of Nucleation (Cell Formation) in Self-Organization of Protein Nanostructures *in Vitro* and *in Vivo*

E. Rapis
Laboratory of Applied Physics, Tel-Aviv University, Ramat-Aviv, Tel Aviv, 64230 Israel

Received September 27, 2004

Abstract—No convincing theory or hypothesis concerning the origin of biological cells exists today. Insight into the problem is difficult, because an empiric model of cell origination and division at the crucial phase of life, self-organization of protein nanostructures, is lacking. It has been shown experimentally that protein nanostructures exhibit signs of self-organization when an open far-from-equilibrium protein–water system condenses *in vitro*. In other words, to be active, protein must be in the nonequilibrium state. Such a form of self-organization is accompanied by nucleation and the formation of defects, which divide the protein film into domains (“cells”) with nuclei. This type of structuring in the nonequilibrium (active) protein may be viewed as a crude empiric model of protein nucleation, since it includes the formation and division (self-organization) of biological cells, the origination of which, in turn, is intimately related to the self-organization of protein at the nanolevel. The reason for the similarity of the basic processes is identical conditions of protein condensation *in vitro* and *in vivo*. In both cases, when water evaporates rapidly from an open water–protein system that is far from thermodynamic equilibrium, the conditions necessary for protein nonequilibrium nanostructures be self-organized with nucleation in the form of nucleus-containing “cells” are set. *© 2005 Pleiades Publishing, Inc.*

INTRODUCTION

The origin of biological cells, the building blocks of organized matter, still remains a mystery. This problem occupies a prominent place in a rapidly emerging field of biology, proteomics (science of protein). A specific task of proteomics is to trace the evolution of structure information from an atom to a cell. The ultimate goal in this direction is to discover a basic principle of self-organization that is central to all biological processes and primarily to the origination and division of macrocells.

Despite recent advances in the development of analytic instruments, X-ray analysis remains the most popular technique for protein investigation [1]. With X-ray equipment, equilibrium energy-inactive protein crystals are studied with an angstrom resolution.

However, much evidence indicates that visualization of the protein molecule structure at the angstrom level bears little information on the functioning of protein. Efforts are therefore made to analyze the structure of larger protein complexes. Instead of studying the folding of protein molecular chains at the atomic level, it seems to be more appropriate to gain insight into the structure of noncovalent macromolecular complexes featuring a unique structure and biological function on the nanoscale. It has turned out that nature starts creating molecular communities that govern key processes in the living matter (division, multiplication, functioning, etc.) just from this scale, and so it is this scale that is of scientific interest.

It becomes evident that gaining insight into cell origination is hindered by the lack of a technique making it possible to develop an empiric model that covers the self-organization of nonequilibrium protein *in vitro* from the nano- to macrolevel.

A complex investigation performed by the author [2–7] has shown that the *in vitro* condensation of an open protein–water system that is far from thermodynamic equilibrium does reproduce the process of self-organization of the protein nanostructures. Specifically, it has been established that protein is active (i.e., capable of functioning) only when in the nonequilibrium state. Most interestingly, this process is accompanied by the emergence of defects dividing the protein film into nucleus-containing domains, which are morphologically similar to living cells (Fig. 1).

The question arises as to why structures morphologically similar to living cells appear in a film of pure protein drying on a solid substrate in open air. It is well documented that biological cells contain not only protein but also a variety of other ingredients, such as DNA, adenosine triphosphoric acid (ATPA), and others. What is the role of protein in cell origination? New impressive findings have been obtained in this respect. Experimentally, it has been convincingly proved [8] that protein in living matter may function independently. Unique experiments have made it possible to visualize (up to a second) the cytokinesis (cell division) of the protein cytoplasm. In this case, cytokinesis proceeds in the complete absence of the genetic apparatus and, hence, without nucleus division (i.e., without karyomitosis).

However, the mechanism underlying this process remains unknown. In particular, it is unclear why and how the protein of the cell protoplasm plays the leading functional role or whence come signals controlling the protein fibers functioning as constrictors in divisible cells of plants and animals [9, 10]. In other words, biologists do not know whence cracks dividing the protein protoplasm in a cell come and, accordingly, do not know the origin of living cells themselves.

To gain a better understanding of the problem, the author carried out special experiments on protein solidification under conditions that thermodynamically and kinetically approach those under which protein functions in living matter, that is, in an open and thermodynamically nonequilibrium (rather than in a closed and equilibrium) water–protein system. The experiments are in many ways similar to those performed earlier [5–7] but the emphasis is on other things, specifically, on nucleation in protein films and complexation with the formation of nucleus-containing domains (cells) upon condensation.

EXPERIMENTAL

The object under study is an open (exposed to air) water–protein (egg lysozyme, egg protein, or human blood serum with erythrocyte hemolysis) system. The blood serum with pronounced erythrocyte hemolysis (i.e., containing a complex of plasma proteins together with a high concentration of hemoglobin) was prepared by taking 5.0 ml of venous blood and keeping it in an open test tube placed in a cooler at 1–2°C for 10 days. Drops of the material were placed on a microscope slide and dried at room temperature and atmospheric pressure. Then, the condensation kinetics was observed under optical, polarizing, scanning electron, and laser confocal microscopes (series 1 of tests). In series 2, the same system was dried under a cap, so that water evaporation and, accordingly, protein condensation and polymerization proceeded more slowly (for details of the experimental procedure, see [1–7]). A total of three thousand runs were carried out.

RESULTS

The series-1 experiments showed that the condensation of egg protein, egg lysozyme, and blood serum with erythrocyte hemolysis in the open system, which is far from thermodynamic equilibrium, is accompanied by the self-organization of the protein solution.

When the condensation of the proteins or protein complexes reaches a certain critical point, the homogeneous film consolidates and the nucleation process starts. The onset of nucleation is manifested by a characteristic click and the appearance of geometrically related regularly shaped large-scale defects (straight lines, spirals, etc.). They divide the 3D film of the protein or its complex (in the blood serum) into separate domains (cells). Each of the cells has distinct regular boundaries and central nucleus-containing areas with nano- and large-scale opposite vorticities. Various forms of discontinuous symmetry can also be distinguished: helical (ellipsoidal), superspiral (rotational), translational radial, pure rotational, chiral, “hedgehog” or “bird’s wing” symmetry with signs of fractal geometry, and others.

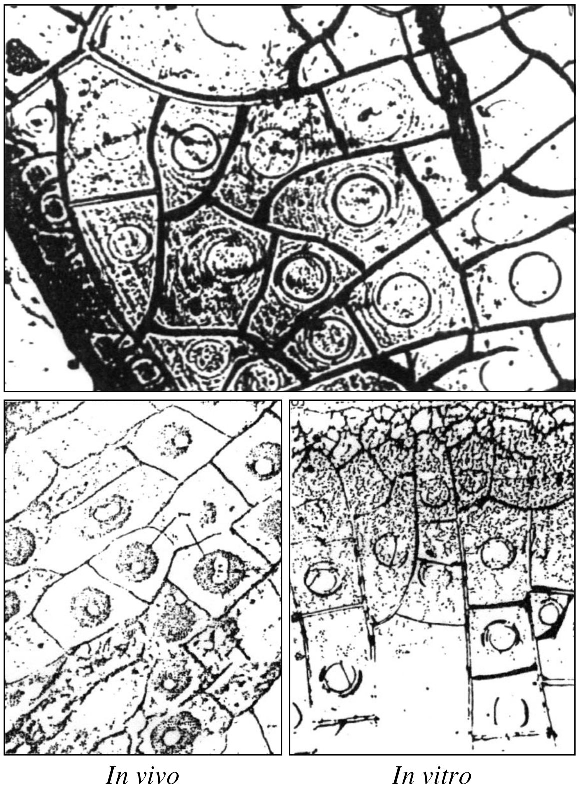


Fig. 1. Nonequilibrium nanofilm of the protein “protos” (lysozyme–water system). The blocks (cells) with nuclei inside *in vitro* are geometrically similar to living cells. Shown also are biological cells *in vivo*. Optical microscope, ×200.

In the series-2 experiments (the condensation of the egg protein or pure lysozyme in the closed equilibrium system), 2D reticular crystals appear and the above types of nucleation with different kinds of discontinuous symmetry at the nano- and macrolevel are absent.

The sharp difference in the two states of the proteins (egg protein and lysozyme) was confirmed by X-ray diffraction analysis. Namely, it was shown that, in the nonequilibrium state of the protein (series 1), the spontaneous formation of nanostructures takes place (signs of ordering are observed starting with 4.33–10.00 Å) and the long-range lattice is absent (it is certain to arise in the experiments of series 2).

It was also demonstrated that, in series 1 of experiments (the highly nonequilibrium system), the condensation of the proteins and blood serum results in the formation of an anisotropic liquid-crystal phase with large-scale defects and electrical property pulsation. Such properties are typical of many suspensions upon solidification [11] and different liquid-crystal materials [11–14]. At the same time, when the protein suspension solidifies in the closed near-equilibrium system (series 2), the liquid-crystal phase of the film does not form [15]; instead, the solid crystalline phase appears irrespective of the film size [12]. The experiments demonstrated that, in pure protein (egg protein or lysozyme without DNA, ATPA, and other ingredients of living matter), as well as in the blood serum with erythrocyte hemolysis in the presence of a variety of the ingredients, distinct supermolecular self-organization of the protein nanostructure takes place. This process is accompanied by nucleation and the appearance of 3D cluster films in the liquid-crystal phase at the meso- and macrolevel. Under these conditions, a new nonequilibrium state of protein that has unusual structural, electrical, optical, and magnetic properties was revealed [2–7].

DISCUSSION

The data obtained allow one to trace common features in the *in vitro* and *in vivo* self-organization of protein in the nonequilibrium state. The experiments with pure protein suggest that the bulk of consolidating protein free of other ingredients of living matter contains an independent program of activity that is responsible for nucleation (or the formation of domains) in protein with time. It also becomes clear from these experiments why the divisibility of protein into domains (cells) with circular nuclei at their centers is retained in the absence of the genetic apparatus.

It is obvious that, in this case, we are facing an important role of the protein component in the division of living cells. However, the only fact revealed so far is the similar behavior of protein (nucleation) under both the biotic (*in vivo*) and abiotic (*in vitro*) conditions. To gain a better insight into this phenomenon in living matter, it is necessary to consider in greater detail the experiments *in vitro* and discover the true reasons for nucleation, i.e., division of the protein film by cracks under nonequilibrium conditions.

The role of the process kinetics has been convincingly proved by simple experiments. Varying the rate of evaporation of water from the water–protein system, we radically change its character, i.e., made it nonequilibrium under the open conditions and equilibrium under the closed conditions. In the first case (open system, series 1), we observe the nonlinear oscillating dynamics of protein nanostructure self-organization in the nonequilibrium state. The texture in the form of domains with nuclei arises, which closely resembles the protein nucleation process. In the other case (closed equilibrium system, series 2), no nucleation is observed.

The aforesaid impressively supports the well-known theoretical and experimental data that the nucleation and growth of neighboring grains (blocks, cells, etc.) in the course of matter self-organization proceed only in a nonlinear unstable system and texture is a derivative of self-organization. As early as in 1952, Sternling and Scriven [16] put forward the idea that, in nonequilibrium nonlinear dynamic systems, the surface current produces cellular domains similar to Benard cells. Benard [17] described nonequilibrium processes attendant to heat convection that result in the formation of cellular structures with clear-cut centers and striation, which appears roughly within 72 h.

These data coincide with the behavior of other film structures occurring during condensation and acquiring the nonequilibrium state [18–20].

It is obvious that the emergence of domains (cells) in all the above examples, including in protein, is of general universal character and is associated with the self-organization of matter in the nonequilibrium state. This is totally consistent with the concept [21] that, in the course of spontaneous self-organization, the mere division of condensing matter switches to more sophisticated forms of organization, since self-organization is the “apex” of condensation and executes the program hold in a given system [21]. According to the program, this process generates an ordered (in a certain way) algorithmically repeated functional supermolecular architecture from the nano- to macrolevel. As follows from published data, the process is under temporal kinetic control. Nonequilibrium, in turn, governs the appearance of ordered coherent dissipative nanostructures with nonlinear dynamics.

Eventually, the system takes a universal functionality and structure with domains, blocks, cells, etc., all playing a major part in the further self-organization of various kinds of matter and, in particular, in evolution of living cells [21].

Nevertheless, the appearance of domains (cells) in protein *in vitro*, unlike cells in oil (the Benard phenomenon) and blocks observed upon drying of a takyr soil in a clay-surfaced desert [22] or of ordinary mud (which also exemplifies the nonequilibrium state of matter), is a phenomenon of special importance, since it seems to be intimately related to the production of cells in a living organism.

All our experiments without exception revealed not only generation of individual domains (cells), the process inherent in other nonequilibrium systems as well, but also the previously unknown specific feature of the cells, namely, a geometrically regular clear-cut central region (nucleus). Each of the cells has only one nucleus whose size correlates with the size of the cell. A strict geometric similarity is observed between the cells and nuclei. Taken together, they constitute a highly ordered structure, which bears a striking phenomenological resemblance to cells and cell layers in living biological systems. Of particular interest is the fact that, in the samples of the drying hemolyzed serum, which contain protein complexes (serum and hemoglobin) together with many other organic and inorganic blood components, the pattern becomes still more ordered, very expressive, and typical of pure protein, differing from the pattern of pure protein only by color. Presumably, this is the result of independent activity of protein, where active competing autowave processes have a very high oscillation frequency and, thereby, suppress the morphology due to other substances [7]. However, these very important observations remain poorly understood. It is hoped that detailed investigation of the nonequilibrium mechanism in terms of the empiric model suggested will provide additional insight into the protein self-organization process under both abiotic and biotic conditions.

To understand the nature of the process, it is necessary to work out a theoretical model of protein self-organization and symmetry that takes into consideration the specific features of the equilibrium and nonequilibrium states of protein. Therein lies the goal of subsequent studies.

However, a number of points can be highlighted even today based on the documented data currently available. The fact that the liquid–solid phase transition occurring in nonequilibrium protein *in vitro* is accompanied by the formation of vortices is apparently associated with hydrodynamic turbulence. Considering the low rate of evaporation of the water from the water–protein system exposed to air, this effect can be explained based on the experiments by Groisman and Steinberg [23]. They showed that a specific type of turbulence, elastic turbulence, occurs even if the rate of solidification is low. This is especially true for high-viscosity (elastic) polymer systems, which a water–protein colloid belongs to.

At the same time, turbulence is known to be a periodic source of defects; that is, turbulence-generated defects possess discontinuous symmetry [24]. This elucidates the origin of the discontinuous symmetry in the protein film upon condensation *in vitro* (helical or orbital, radial, rotational, hedgehog-like, etc.), which is similar to the symmetry of protein *in vivo* (Petukhov, private communication, 1999).

The discontinuous helical (orbital) symmetry of protein has been recently discovered in biological experiments made in outer space. This symmetry is attributed to the orbital interaction of electromagnetic fields, while the nucleus-related asymmetry is associated with the action of nuclear forces (Sakina-Weave, private communication, 2003).

In our experiments *in vitro* [5–7], the formation of cells (more specifically, their boundaries) in the protein film was due to large-scale defects appearing in the liquid-crystal phase of the protein, which is consistent with the properties of various liquid crystals [9]. This allows us to understand the origination of large defects in the protein of a living organism, where protein is most frequently just in the liquid-crystal state. (It is remembered that the defects inevitably cause nucleation in protein with the formation of structural components of the cells with the symmetry similar to that in protein films and blocks observed in experiments *in vivo* and *in vitro* at the nano- and macrolevel). The above reasoning checks well with Mints and Kononenko's opinion [13] that cell division as applied to living organism morphology can be considered as the motion of a linear defect (dividing groove) through the body and membrane of a cell. It is believed that such a motion is possible due to the presence of liquid-crystal phases performing biological functions [11–13].

The experiments carried out by the author support the above hypothesis. It can be argued that the liquid-crystal phase of protein behaves in accordance to the elasticity theory. Then, electric pulses passing through a protein film can be regarded as a manifestation of the well-known property of periodic elasticity variation [25, 26]. In terms of the elasticity theory, the free energy of the system varies depending on nonequilibrium elastic forces, the direction of which is responsible for its dissipation and minimization.

The above speculations concerning the process of nucleation in the protein films, as well as the feasibility of observing the morphological pattern in real experiments *in vitro*, make it possible to draw some preliminary inferences. Specifically, structural ordering showing up in nucleation with the formation of nucleus-containing domains (cells) upon the condensation of liquid-crystal protein films *in vitro* at the nano- and macrolevel and the same effect observed in cells and cell layers in a living organism at the same levels are basically similar. This similarity seems to be associated with the relaxation and minimization of the energy of the protein nonequilibrium state in both cases. The same symmetry types in the protein films upon self-organization *in vitro* and *in vivo* strongly suggest the same reasons for nucleation under the biotic and abiotic conditions; eventually, these reasons outline the energy stabilization route in the course of self-organization of protein nanostructures.

This inference can be strengthened by a number of well-documented experimental (morphological), experimental, and functional data.

(1) When observing morphological changes inside a living egg, embryologists, without giving any explanation, usually describe the following periodically repeating pattern. At the vegetative pole of the egg, the cortical layer in the surface disk of the cytoplasm consolidates. It becomes very viscous, dense, and homogeneous (i.e., free of chromosomes, DNA, and other inclusions). Prior to division, the stress in the protein increases sharply and an initial dividing groove, a defect, arises. Such is the onset of defect formation, which gives rise to 2, 4, 8, 16, etc., blocks (Fig. 2). These blocks are called cells, although the fourth side distinctly separating the so-called cells from the bulk of

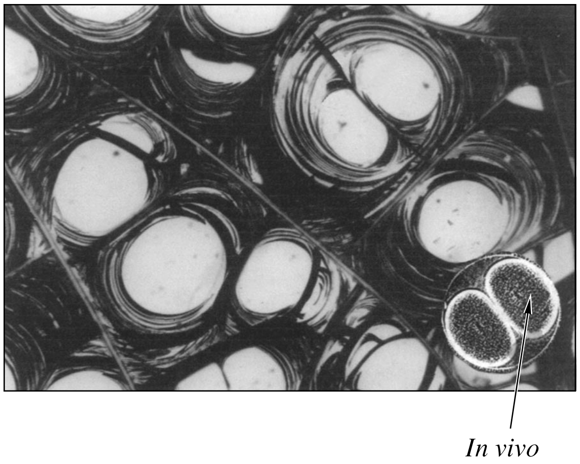


Fig. 2. Division at the center of the mother spiral (nucleus) of a cell in the drying protein film *in vitro* (×120). At the bottom, the division of a snake's egg *in vivo* is shown (×185). Optical microscope.

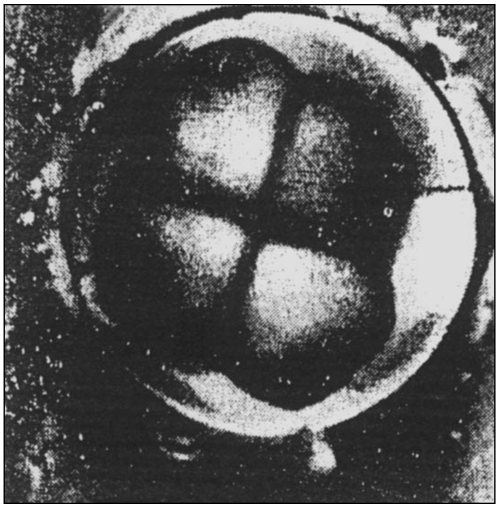


Fig. 3. Division of an egg *in vivo.* The first dividing groove passes through the animal pole. Four cells are being formed.

the homogeneous structureless protein is absent at the early stage. That is all on this issue the author has managed to find in relevant publications on embryology [27, 28].

Reading this description, one might think that the processes observed in the experiments (namely, the emergence of defects (cracks) in the internally strained protein film when its viscous bulk consolidates under the nonequilibrium conditions typical of a living system) suffice to produce cells in the egg (*in vivo*). Surprisingly, the only difference is that the egg is under biotic conditions (*in vivo*), while the experimental conditions are abiotic (*in vitro*).

(2) It is well known that the protein in a living organism is self-organized and performs a variety of functions, which are sometimes of explosive character. Today, ATPA is considered to be the basic energy source of protein. Our experiments *in vitro* have shown that, when protein is self-organized without ATPA (lysozyme or egg protein), it passes to the high-energy nonequilibrium state, which promotes the restructuring and, hence, functioning of protein [2–7].

In these experiments, we managed to discover the nonequilibrium state of protein at the same conformation transitions *in vivo* and *in vitro* and, thus, vividly demonstrate the energy contribution of this state to the process of protein self-organization *in vivo* and *in vitro* when ATPA concurrently enters into the phosphorylation reaction.

The analogy (as yet qualitative) consists in the correspondence between the outer configurations of the cell and nucleus in both cases (Fig. 1). The films with helical (orbital) symmetry surrounding the centers of the cells *in vitro* resemble helical proteinaceous microtubules of the cytoskeleton, which are arranged around the nuclei in living cells [29, 30]. In addition, 3D films appearing in the form of stacks *in vitro* are akin to stacks of membranes in the living cell organelles (mitochondria, Golgi apparatus, etc.) [29].

It is known that the processes of structuring and forming (e.g., membranes, layers, films, fibers, tissues, cells with nuclei, organelles, and so on) proceeding in a living organism seemingly follow the unified construction program, which is associated primarily with proteinaceous complexes existing in these components. These complexes inevitably arise, as in the experiments *in vitro* described above, upon the condensation of various liquid or liquid-like protein–water systems and then turn into denser or solid biological objects. However, the reasons for the geometric similarity between various proteins, cells, and tissues irrespective of their nature and features of a living organism still remain unclear.

Our experimental data made it possible to explain this phenomenon by the ability of various proteins, when in the nonequilibrium state, to trigger the epitaxial growth of nanofilms (membranes) in the form of stacks. Such behavior in the course of self-organization is typical of organic polymers. However, the author is unaware of works where the self-organization of polymers causing nucleus-containing cells that are similar to living cells in shape is described.

The experiments allowed us to observe not only the statics but also the dynamics late in the protein self-organization, when the blocks and their nuclei start spontaneously dividing at the center of the field (similarly to the division of living cells and their nuclei) to form superspirals (Fig. 3). The onset of the process in the central zone in both cases may indicate that this zone is energetically the most active.

Next, the appearance of defects and extremely flat double-wall "channels" between the blocks *in vitro,* as well as *in vivo,* is noteworthy. In terms of morphology, these channels are called intercellular spaces, gaps, or grooves, in which the intercellular liquid is contained. This observation is consistent with observations of "twins" when the gelatinous bulk of the protein divides into blocks (Fig. 1).

From the data listed, it follows that the nucleation of nonequilibrium protein under the conditions of nanostructure film self-organization *in vitro* geometrically coincides with the formation of cells with nuclei *in vivo*. It is obvious, however, that "cells" *in vitro* are far from being identical to living cells. Yet, the similarity of structures (frameworks), symmetries, and scales *in vitro* and *in vivo*, as well as the functional activity of protein in both cases, is a valid one and deserves further investigation.

CONCLUSIONS

The structuring of nonequilibrium (active) protein *in vitro* described in this paper may be viewed as a crude empiric model of nucleation with the formation and division of biologic cells, since the basis for the cell activity *in vivo* is also self-organization at the nanolevel. Even the qualitative data presented here give an idea of a source of energy needed for nucleation and formation of defects in protein *in vitro* and *in vivo*. The structuring and general behavior of the nonequilibrium liquid-crystal protein film during condensation suggest universal properties of protein, which is in agreement with the theory of cluster nanofilms [29, 30] and the theory of anisotropic nematic films and elastic films [31].

We for the first time managed to visualize the dynamics of the spontaneous phase transition in protein at its condensation and self-organization *in vitro* under conditions that closely approach those in a living organism at a cellular level. The *in vitro* conditions used in this work and *in vivo* conditions are similar (i) in kinetics (the rate of dehydration is sufficient that the process is nonequilibrium; *in vivo*, nonequilibrium is related to the rapid hydrolysis of ATPA), (ii) in thermodynamic parameters (the open system far from thermodynamic equilibrium), (iii) in scale (nanolevel, cellular level, and macrolevel; X-ray diffraction analysis, optical and electron microscopes), and (iv) in solvent (water).

Our experiments showed that, when protein nanostructures are self-organized at the mesolevel and macrolevel (cellular level), the emergence of a new liquid-crystal phase with elastic properties is accompanied by the generation of large-scale defects promoting nucleation and the formation of multilayer films. These films divide into blocks (cells) with nuclei at their centers under both the biotic and abiotic conditions.

Significantly, while the process *in vitro* goes in pure protein free of additional energy sources (such as ATPA, DNA, and others), the resulting types and scales of symmetry are similar to those *in vivo*. This may indicate the same underlying mechanisms in both cases and suggests the independent activity of protein in a living organism. It appears that a more comprehensive quantitative correlation between the self-organization dynamics of pure protein *in vitro* and the self-organization dynamics of protein *in vivo* will allow us to estimate the "competitiveness" of its activity and, hence, to see whether it can efficiently "operate" in a complex system with a variety of ingredients. The experiments with the hemolyzed serum seemingly exclude the assumption that protein, when in a complex system like a living organism and operating in cooperation with many other ingredients, may change the route of self-organization and structuring. This is consistent with the concept in [21], which states that supramolecular chemistry in the course of self-organization is the result of executing a certain conformation molecular code existing both *in vitro* and *in vivo*. This fact forbids the systems from evolving in directions other than the given one, which generate products of another essence [21].

It is hoped that detailed analysis of the empiric model of protein self-organization under nonequilibrium conditions, which is accompanied by nucleation from the nano- to the macrolevel, will provide a deeper insight into the origination and division of living cells.

ACKNOWLEDGMENTS

I thank M. Amus'ya, A. Arel', V. Buravtsev, V. Volkov, A. Zaikin, M. Klinger, L. Manevich, Yu. Neeman, and I. Prigogine for encouragement, assistance, and fruitful discussions.

REFERENCES

1. A. Sali *et al.*, Nature **422**, 216 (2003).
2. E. Rapis, Pis'ma Zh. Tekh. Fiz. **14**, 1561 (1988) [Sov. Tech. Phys. Lett. **14**, 679 (1988)].
3. E. Rapis and G. Gasanova, Zh. Tekh. Fiz. **61** (4), 62 (1991) [Sov. Phys. Tech. Phys. **36**, 406 (1991)].
4. E. Rapis, Pis'ma Zh. Tekh. Fiz. **21** (5), 13 (1995) [Tech. Phys. Lett. **21**, 321 (1995)].
5. E. Rapis, Pis'ma Zh. Tekh. Fiz. **23** (4), 28 (1997) [Tech. Phys. Lett. **23**, 263 (1997)].
6. E. Rapis, Zh. Tekh. Fiz. **70** (1), 122 (2000) [Tech. Phys. **45**, 121 (2000)].
7. E. Rapis, *Protein and Life (Self-Assembling and Symmetry of Protein Nanostructures)* (MiltaPKPTIT, Moscow, 2003) [in Russian].
8. D. Zhang and B. Nicklas, Nature **382**, 466 (1996).
9. R. Heald *et al.*, Nature **382**, 420 (1996).
10. M. Buckingham, Nature **408**, 773 (2000).
11. Ph. Poulin *et al.*, Science **275**, 1770 (1997).
12. M. Zapotocky *et al.*, Science **283**, 209 (1999).
13. R. Mints and E. Kononenko, Priroda, No. 6, 36 (1984).
14. V. Indenbom, V. Nikitenko, and V. Strunin, Priroda, No. 4, 74 (1974); No. 6, 82 (1974).
15. Xiao Bing *et al.*, Nature **376**, 188 (1995).
16. C. Sternling and L. Scriven, AIChE J. **5**, 514 (1959).

17. H. Benard, Rev. Gen. Sci. Pures Appl., No. 11, 1261 (1990).
18. D. Avnir, D. Darin, and P. Pfeifer, Nature **308**, 261 (1984).
19. B. Parkinson, Science **270**, 1157 (1995).
20. J. Preskill, Science **272**, 966 (1996).
21. J. M. Lehn, Proc. Natl. Acad. Sci. USA **99**, 16 (2002).
22. A. Groisman and E. Kaplan, Europhys. Lett. **25**, 415 (1994).
23. A. Groisman and V. Steinberg, Nature **405**, 53 (2000).
24. R. Service, Science **276**, 356 (1997).
25. A. Aggeli *et al.*, Nature **386**, 259 (1997).
26. T. Maruyama *et al.*, Science **274**, 233 (1996).
27. R. Nicklas, Chromosoma **21**, 1 (1967).
28. C. W. Bodemer, *Modern Embryology* (Holt, Rinehart and Winston, New York, 1968; Mir, Moscow, 1971).
29. L. Kjer Nielson *et al.*, Curr. Biol. **9**, 385 (1999).
30. A. Linstedt, Curr. Biol. **9**, 893 (1999).
31. M. Trau *et al.*, Science **272**, 706 (1996).

Translated by V. Isaakyan

Technical Physics, Vol. 50, No. 6, 2005, pp. 787–789. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 6, 2005, pp. 114–116.
Original Russian Text Copyright © 2005 by S. Drapak, I. Drapak, Kovalyuk.

SHORT COMMUNICATIONS

Optical and Electrical Properties of Turpentine Films

S. I. Drapak*, I. T. Drapak, and Z. D. Kovalyuk***
** Frantsevich Institute of Materials Science Problems, Chernovtsy Branch,*
National Academy of Sciences of Ukraine, Chernovtsy, 58000 Ukraine
*** Fed'kovich State University, ul. Kotsubinskogo 2, Chernovtsy, 58012 Ukraine*
e-mail: chimsp@unicom.cv.ua

Received September 7, 2004

Abstract—The basic optical and electrical properties of turpentine, a biologically active natural organic compound, are studied. At wavelengths from 350 to 550 nm, the room-temperature photoluminescence spectrum of turpentine shows a number of peaks. In the temperature range 280–300 K, the activation energy of conduction is found to be 0.53 eV, which is much lower than the optical bandgap and indicates the hopping mechanism of conduction with a variable length of hops between states near the Fermi level. *© 2005 Pleiades Publishing, Inc.*

Advances in semiconductor optoelectronics have led to the development of a variety of heterophotocells with different functionality [1–3]. Together with routine studies of conventional solid-state semiconductor–semiconductor, semiconductor–metal, and semiconductor–insulator structures, investigations into hybrid semiconductor–organic substance heterojunctions are becoming more and more extensive [4–7]. When appropriately designed, these structures may offer photoelectric parameters that are comparable, or even much superior, to those of commercial devices [8]. Structures consisting of a semiconductor and a biologically active organic substance are of special interest [9–11]. In our recent work [12], we demonstrated that turpentine, a natural organic substance, when incorporated into photosensitive structures, behaves as a *p*-semiconductor. A feature of turpentine-based structures is that photovoltaic conversion takes place mainly in the organic substance.

Turpentine, a gummy substance that exudes from incisions in the trunks of pine trees, is a valuable raw material for the production of various technical and biologically active compounds: lacquers; camphore; immersion oil, which is applied in microscopy; rosin (a mixture of resin acids with the general formula $C_{19}H_{29}OOH$); and turpentine oil (a mixture of unsaturated terpene hydrocarbons of general formula $C_{10}H_{16}$). Reversibility of chemical processes taking place in the turpentine constituents with time and under illumination was reported [13]. However, despite progressively growing interest in percolation composites (insulator–conductor and insulator–semiconductor systems) [14, 15], data on the physical properties of turpentine are currently lacking.

In this work, we for the first time determined the basic optical and electrical properties of films of this natural biologically active organic material. Namely, we recorded the transmission and luminescence spectra of turpentine in the wavelength range λ = 300–900 nm, measured its conductivity at room temperature, and took the temperature dependence of the conductivity.

To prepare a turpentine film uniform in thickness, a drop of an alcoholic solution (96% ethyl alcohol) of turpentine was applied on a sapphire substrate mounted on a centrifuge. The centrifuge was rotated until the alcohol completely evaporated. The thickness of the test films thus obtained was 20–30 μm; the density, 0.714 g/cm^3. The conductivity of the films was measured by the standard two-point probe method [16] using silver contacts.

The optical measurements are shown in Fig. 1. It is seen that (i) the turpentine films are transparent in the range 500–900 nm ($T \approx 90\%$ at thickness $d \approx 20$ μm, curve *1*); (ii) the long-wave part of the optical absorption edge is highly diffuse; and (iii) absorption coefficient α changes by one order of magnitude (from 10^3 to 10^2 cm^{-1}) in the range 300–400 nm, i.e., at photon ener-

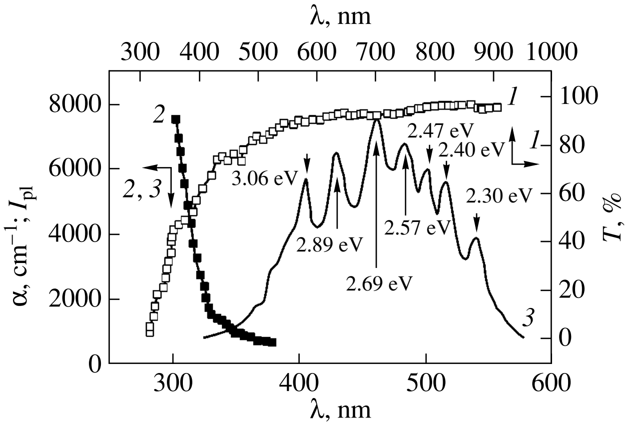


Fig. 1. (*1*) Transmission, (*2*) absorption, and (*3*) photoluminescence spectra taken of turpentine films at T = 288 K.

1063-7842/05/5006-0787$26.00 © 2005 Pleiades Publishing, Inc.

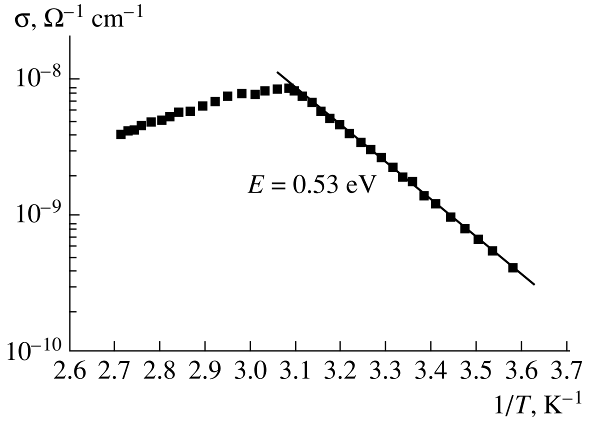


Fig. 2. Temperature dependence of the conductivity of the turpentine films.

gies varying between 4.13 and 3.1 eV (curves *2*). Note that rosin films obtained by the same technique are transparent ($T \approx 90\%$ at the same film thickness) throughout the spectral range studied.

Luminescence was excited by 337-nm radiation. The luminescence spectrum (Fig. 1, curve *3*) exhibits many peaks in the range 350–550 nm, indicating several types of luminescence centers present in turpentine, which is characteristic both of compounds with a complicated energy band structure [17] and of complex molecular mixtures [18].

Figure 2 shows the temperature dependence of the turpentine conductivity, $\sigma(T)$. In the temperature range 291–325 K, the activation energy of conduction is constant, ≈0.53 eV. This value is much lower than the energy bandgap of the material that might be expected from the absorption spectrum. Presumably, we are dealing with either the impurity mechanism of conduction or with the hopping mechanism with a variable length of hops between states localized near the Fermi level. A more detailed study of the temperature dependence of the turpentine conductivity will be carried out in our subsequent works. Here, we only note that the linear run of $\ln\sigma$ as a function of $T^{-1/4}$ at T = 280–325 K, as well as the presence of the portion where $\sigma \sim \omega^{0.8}$ (where ω is the frequency of the electric field used in the measurements), counts in favor of the hopping mechanism of conduction in turpentine films [19]. At $T > 325$ K, the conductivity of the organic material starts decreasing. Such behavior of the curve $\sigma(T)$ may be accounted for by several reasons. First, the decline in the conductivity with rising temperature can be attributed to the saturation of the rate of increase of the charge carrier mobility after all holes (electrons) have passed from localized centers to the valence (conductivity) band under the action of temperature. This scenario is typical of both amorphous inorganic semiconductors and a number of organic compounds [19, 20]. Second, such behavior of $\sigma(T)$ may be associated with a decrease in the carrier mobility due to phonon scattering. Bearing in mind that organic substances have usually a crystal structure [21] and that even amorphous films have crystalline areas [19], the second supposition does not seem meaningless, while it needs additional substantiation.

Thus, the results of the pioneering investigation into the optical performance and electrical conductivity of turpentine indicate that this biologically active natural organic material is a semi-insulator with an optical bandgap of ≈3.3 eV and so can be used in designing optoelectronic devices. It should also be emphasized that the resistivity of turpentine does not exceed that of a number of artificial (synthesized) wider bandgap organic materials that are to be used or even are being used in various commercial electronic components [22]. One more advantage of turpentine over its biological counterparts mentioned in the literature (skin, green leaves, protein, and others [9, 10]) and a number of organic materials is its high environmental resistance.

REFERENCES

1. A. G. Milns and J. J. Feucht, *Heterojunctions and Metal–Semiconductor Junctions* (Academic, New York, 1972; Mir, Moscow, 1975).
2. Yu. V. Gurevich and F. V. Pleskov, *Semiconductor Photoelectrochemistry* (Nauka, Moscow, 1983; Consultants Bureau, New York, 1986).
3. Zh. I. Alferov, Fiz. Tekh. Poluprovodn. (St. Petersburg) **32** (2), 3 (1998) [Semiconductors **32**, 1 (1998)].
4. S. Naka, K. Shinko, H. Okada, *et al.*, Jpn. J. Appl. Phys., Part 2 **33**, L1772 (1994).
5. G. Ferrari, D. Natali, M. Sampietro, *et al.*, Org. Electron. **3**, 33 (2002).
6. B. Abay, Y. Onganer, M. Saglam, *et al.*, Microelectron. Eng. **51–52**, 659 (2000).
7. N. V. Blinova, E. L. Krasnopeeva, Yu. A. Nikolvaev, *et al.*, Fiz. Tekh. Poluprovodn. (St. Petersburg) **37**, 53 (2003) [Semiconductors **37**, 53 (2003)].
8. R. Rinaldi, E. Branca, R. Cingolani, *et al.*, Appl. Phys. Lett. **78**, 3541 (2001).
9. V. Yu. Rud’, Yu. V. Rud’, and V. Kh. Shpunt, Fiz. Tekh. Poluprovodn. (St. Petersburg) **31**, 129 (1997) [Semiconductors **31**, 97 (1997)].
10. V. Yu. Rud’, Yu. V. Rud’, and V. Kh. Shpunt, Zh. Tekh. Fiz. **70** (2), 114 (2000) [Tech. Phys. **45**, 255 (2000)].
11. S. I. Drapak, V. B. Orletskiĭ, V. D. Fotiĭ, and Z. D. Kovalyuk, Prikl. Fiz. **11** (2), 73 (2004).
12. S. I. Drapak and Z. D. Kovalyuk, Pis’ma Zh. Tekh. Fiz. **30** (6), 73 (2004) [Tech. Phys. Lett. **30**, 250 (2004)].
13. V. A. Raldugin, V. V. Grishko, Yu. V. Gatilov, *et al.*, Sib. Khim. Zh., No. 5, 66 (1992).
14. V. A. Sotskov and S. V. Karpenko, Zh. Tekh. Fiz. **73** (1), 106 (2003) [Tech. Phys. **48**, 100 (2003)].
15. V. A. Sotskov, Zh. Tekh. Fiz. **74** (11), 107 (2004) [Tech. Phys. **49**, 1501 (2004)].

16. N. Ya. Karkhanina, *Technology of Semiconductor Materials* (GITL, Kiev, 1961) [in Russian].

17. V. V. Serdyuk and Yu. F. Vaksman, *Luminescence in Semiconductors* (Vyshcha Shkola, Kiev, 1988) [in Russian].

18. C. A. Parker, *Photoluminescence of Solutions* (Elsevier, Amsterdam, 1968; Mir, Moscow, 1972).

19. N. F. Mott and E. A. Davis, *Electronic Processes in Non-Crystalline Materials* (Clarendon, Oxford, 1979; Mir, Moscow, 1982).

20. F. Yakuphanoglu, M. Aydin, N. Arsu, and M. Sekersi, Fiz. Tekh. Poluprovodn. (St. Petersburg) **38**, 486 (2004) [Semiconductors **38**, 468 (2004)].

21. É. A. Silinsh and L. F. Taure, *Organic Semiconductors* (Znanie, Moscow, 1980) [in Russian].

22. É. A. Lebedev, M. Ya. Goĭkhman, M. E. Kopman, *et al.*, Fiz. Tekh. Poluprovodn. (St. Petersburg) **37** (7), 82 (2003) [Semiconductors **37**, 816 (2003)].

Translated by V. Isaakyan

Technical Physics, Vol. 50, No. 6, 2005, pp. 790–793. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 6, 2005, pp. 117–120.
Original Russian Text Copyright © 2005 by Petrakovskaya, Isakova, Bayukov, Velikanov.

SHORT
COMMUNICATIONS

Superparamagnetism of Magnetite Particles in C_{60} Fullerite Powder

É. A. Petrakovskaya, V. G. Isakova, O. A. Bayukov, and D. A. Velikanov

*Kirenskiĭ Institute of Physics, Siberian Division, Russian Academy of Sciences,
Akademgorodok, Krasnoyarsk, 660036 Russia*

e-mail: rsa@iph.krasn.ru

Received August 23, 2004

Abstract—A new magnetic material, C_{60} fullerite powder doped by magnetite (Fe_3O_4) nanoparticles, is obtained by heating a mixture of fullerite and iron(III) acetylacetonate. It is shown that the material offers superparamagnetic properties. Surface bonding between the nanoparticles and the fullerite is established. © *2005 Pleiades Publishing, Inc.*

The physical properties of nanocomposites and, hence, their functionality depend on the size of the nanoparticles and the properties of the matrix. A challenging problem in this field is preparing nanoparticles of a given composition by one-stage synthesis. Fullerite seems to be an appropriate matrix for magnetic nanocomposites, which are today viewed as very promising. In this work, we study the resonance and magnetic properties of C_{60} fullerite powder with Fe_3O_4 magnetite nanoinclusions. The material is prepared by the chemical reaction between C_{60} and volatile iron triacetylacetonate $Fe(acac)_3$ ($Fe(acac)_3 = Fe(CH_3COCHCOCH_3)_3$), during which the magnetite nanoparticles enter the magnetically neutral but chemically active matrix.

The samples of magnetic materials MC_{60} and M^*C_{60} (M stands for magnetite) were prepared by the original method [1], according to which a solid mixture $Fe(acac)_3 : C_{60} = 1 : 2$ is heated to 300°C until the burning of $Fe(acac)_3$ is seen visually. Then, the process goes in air spontaneously without heating (MC_{60} samples). When M^*C_{60} samples were synthesized, the crucible with the reagents was cooled by water of temperature 10–15°C.

The heating process was studied by the method of differential thermal analysis (DTA). It was found that, in the presence of C_{60}, $Fe(acac)_3$ decomposes with the formation of $Fe(acac)_2$ and the radical (acac) followed by pyrolysis of $Fe(acac)_2$ without substantial gas evolution from C_{60}. The weight loss meets the conversion of $Fe(acac)_3$ to Fe_3O_4.

The state of the iron was determined with Mössbauer measurements using a $^{57}Co(Cr)$ source. The electron magnetic resonance (EMR) spectra were taken with an SE/X-2544 electron paramagnetic resonance X-ray spectrometer equipped with a thermal liquid-nitrogen trap. The spectra were characterized by effective (peak-to-peak) resonance linewidth ΔH, which was found from the derivative of the absorption line, and position H_0 of the peak, which is found by integrating the derivative. Magnetic measurements were made with a SQUID magnetometer. The X-ray fluorescence spectrum taken from the C_{60} fullerite synthesized by Eletskiĭ exhibits a standard set or peaks typical of the face-centered cubic lattice of C_{60} at 300 K. For this fullerite, the electron paramagnetic resonance (EPR) of the radical is characterized by parameters $g = 2.0017$ and $\Delta H = 0.08$ mT. In the X-ray spectra taken from the products of the C_{60}–$Fe(acac)_3$ reaction, the reflections due to C_{60} persist.

Early Mössbauer measurements revealed magnetite particles in the material synthesized [2–4]. The parameters of the Mössbauer spectrum taken from the MC_{60} sample are listed in the table. The parameters of the hyperfine structure were obtained by fitting the model spectrum to the experimental one under the assumption that the absorption line is of Lorentz shape. Three distinct nonequivalent positions of the iron were distinguished in the spectrum that are represented by the sextet, doublet, and singlet. As judged from the hyperfine field on the nucleus and the isomer shift value, the sextet can be assigned to defects in the magnetite, whose sublattice remained unresolved because of the poor statistic of the spectrum. The particle size did not exceed 18 nm [3].

The isomer shift of the doublet is typical of trivalent iron oxide, which can be assumed to consist of finer, ≈13 nm, superparamagnetic magnetite particles. The isomer shift of the singlet is typical of iron-based alloys, including iron–carbon alloys [2]. In this case, the iron may cover the magnetite particle surface and chemically bonded to the fullerite.

Figure 1a shows the temperature dependences of magnetization for the composite samples cooled in a magnetic field of 20 mT (field cooling, FC) and in the absence of the field (zero-field cooling, ZFC). The

Mössbauer data for MC_{60}

Sextet	δ, mm/s	H, 10^5 A/m	ε, mm/s	FWHM, mm/s	S
1	0.45 ± 0.05	490 ± 10	0.33 ± 0.08	1.66 ± 0.08	0.45 ± 0.1
2	0.36 ± 0.05	0	0.79 ± 0.08	0.58 ± 0.08	0.41 ± 0.1
3	0.08 ± 0.05	0	0	1.72 ± 0.1	0.13 ± 0.1

Notes: δ, the isomer chemical shift relative to α-Fe; ε, the quadrupole splitting; H, the hyperfine field on the iron nucleus; and S, the fraction of iron-occupied sites.

curves are typical of superparamagnets. In the M^*C_{60} sample, the magnetization decreases with increasing temperature by roughly one order of magnitude, the difference in the FC and ZFC curves persisting throughout the temperature range (Fig. 1a shows only the behavior of the magnetization under the FC conditions). Based on the data of DTA and quantitative analysis (the latter show that both the iron and the fullerite do not diminish in weight upon temperature synthesis), we calculated the magnetization per gram of iron. In the MC_{60} sample at 300 K, this value equals 17.6 emu/g, which correlates with the data for magnetite particles in the carbon matrix, 19–45 emu/g [4].

The EMR spectra taken in the temperature range 80–500 K are presented in Fig. 2. The spectrum for M^*C_{60} contains the line of the C_{60} radical. In the MC_{60} sample, this line is resolved only in a narrow magnetic field range and is highly broadened, indicating that the magnetization grows. The absorption line intensity for M^*C_{60} is roughly one order of magnitude lower than for MC_{60}, which is also consistent with the magnetization data.

The temperature dependence of the EMR spectra (Fig. 3) is typical of iron oxides in the superparamagnetic state [5–7]. It is known that, because of the first-order phase transition in the temperature range 250–256 K, the C_{60} fullerite lattice becomes primitive cubic. At these temperatures, the resonance characteristics of the material exhibit a specific feature: the slope of the temperature curves for the EMR linewidth and resonance field changes (Fig. 3a). Such a run confirms the presence of particle–fullerite bonding.

As the temperature grows, the EMR line narrows, which is associated with the averaging effect of magnetization thermal fluctuations. According to the calculation made in [5], the EMR linewidth broadening factor for like particles is proportional to the ratio of their volumes. The experimental spectra show both broad and narrow absorption lines, which differ in position in the magnetic field (this fact is more pronounced for M^*C_{60}, Fig. 2b). Presumably, there exists a statistical double-

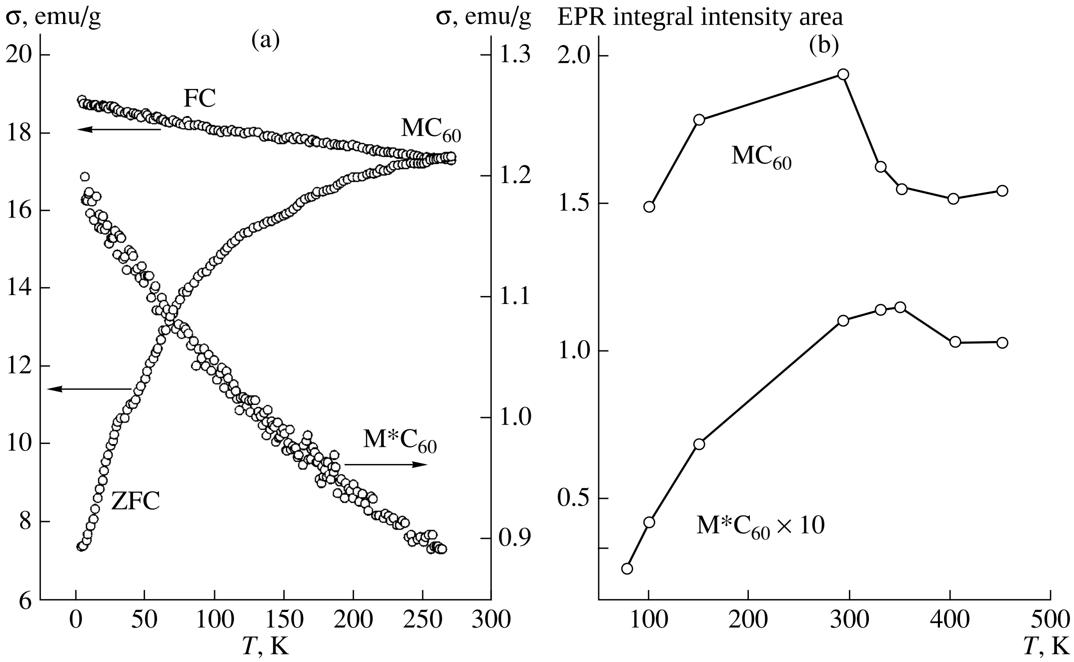


Fig. 1. Temperature dependence of the magnetization.

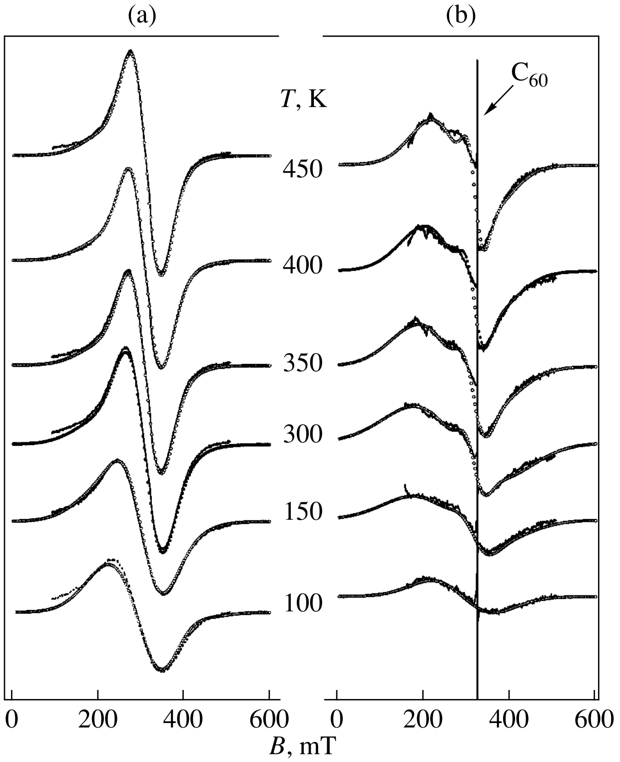


Fig. 2. Temperature dependence of the EMR spectra for (a) MC_{60} and (b) M^*C_{60}. The symbol "○" refers to the computer simulation of the spectrum by using two Gaussian curves. The EPR spectrum of the C_{60} radical is shown by the arrow.

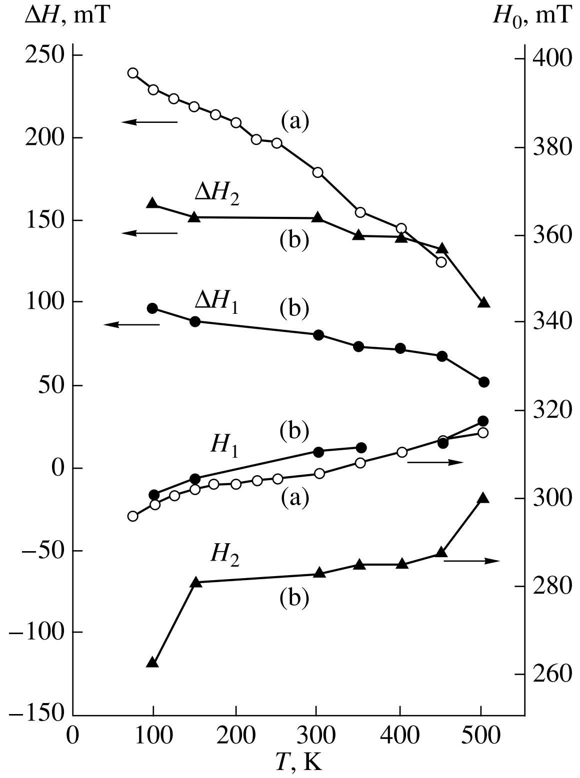


Fig. 3. Temperature dependence of the EMR spectrum parameters for MC_{60}: (a) experimental and (b) simulated spectra.

hump particle size distribution similar to that found in [6]. The same takes place in the Mössbauer spectra. Computer-aided simulation of the EMR spectra by using two Gaussian resonance lines gave the best fit to the experimental data (Fig. 2). Such a shape is likely to reflect a considerable influence of nonuniform magnetic anisotropy in the magnetite nanoparticles embedded in the fullerite. This adversely affects the resolution of the Mössbauer spectra. Throughout the temperature range, the value of ΔH_1 of the broader spectral component for the MC_{60} sample is less than ΔH_1^* for M^*C_{60} and resonance fields obey the inequality $H_1 > H_1^*$. For the narrower components, conversely, $\Delta H_2 > \Delta H_2^*$ and $H_2 > H_2^*$. Therefore, the doublet in the spectra for M^*C_{60} is better resolved. The temperature dependences of the parameters of the resonance lines simulated are shown in Fig. 3 for MC_{60}. For MC_{60} at 300 K, the linewidths and their associated resonance fields are ΔH_1 = 81.6 mT, H_1 = 310.4 mT and ΔH_2 = 151.5 mT, H_2 = 281.7 mT. For M^*C_{60}, ΔH_1^* = 59 mT, H_1^* = 316 mT and ΔH_2^* = 224 mT, H_2^* = 265 mT. For sized magnetite particles, the EMR linewidths at 300 K are given in [8]. Comparing those values and ours, we obtained two characteristic linear sizes of magnetite particles in MC_{60}: 9 and 15 nm. These sizes are comparable to those estimated from the Mössbauer data. In M^*C_{60}, the magnetite particles sizes are 6 and 21 nm.

The difference in particle size between MC_{60} and M^*C_{60} reflects the different synthesis conditions. In the latter case, the process allowed for heat removal. This decreased the rates of vaporization and decomposition of $Fe(acac)_3$ with the formation of the radical. As a result, some iron complexes might stay in the molten state. Apparently, coarser particles are synthesized under such conditions. The decrease in the magnetization and the coarsening of the particles may be attributed to the agglomeration of nanocrystallites and a decrease in their density, as well as to their adsorption on the surface due to the organic products of ligand decomposition. Fine particles form during the gas-phase decomposition of $Fe(acac)_3$ on the surface of the melt. In the case of M^*C_{60}, they became still finer, because vaporization is less vigorous under such conditions.

The temperature run of the magnetization for MC_{60} and the dependence of the magnetization on the cooling conditions lead us to assume that the magnetization in the nanoparticles is suppressed. Because of the particle size distribution, the process covers a wide temperature range. The resonance methods strengthen this assumption. Figure 1b shows the temperature dependences of the integral intensity of the EMP spectra. Its value varies in proportion to the magnetization of the material. According to the general concept [9], the run of these curves clearly indicates magnetization suppression in MC_{60} and in M^*C_{60}. In the latter case, the suppression is observed at a higher temperature, because here coarse particles grow.

In summary, a new magnetic composite (Fe_3O_4 nanoparticles in C_{60} fullerite powder) offers superparamagnetic properties. The suppression of the nanoparticle magnetization starts near room temperature. The magnetic resonant properties of the nanoparticles are sensitive to structural changes in the fullerite. The effects observed are of applied interest.

REFERENCES

1. V. G. Isakova, E. A. Petrakovskaya, A. D. Balaev, *et al.*, Zh. Prikl. Khim. (St. Petersburg) **76**, 597 (2003).
2. V. Chekin, *Mössbauer Spectroscopy for Iron, Gold, and Tin Alloys* (Énergoizdat, Moscow, 1981) [in Russian].
3. I. P. Suzdalev, *Dynamic Effects in Gamma Resonance Spectroscopy* (Atomizdat, Moscow, 1979) [in Russian].
4. S. M. Aharoni and H. M. Litt, J. Appl. Phys. **42**, 352 (1971).
5. V. K. Sharma and F. Waldner, J. Appl. Phys. **48**, 4298 (1977).
6. J. Kliava and R. Berger, J. Magn. Magn. Mater. **205**, 328 (1999).
7. K. Nagara and A. Ishihara, J. Magn. Magn. Mater. **104–107**, 1571 (1992).
8. A. Aharoni, Phys. Rev. **177**, 793 (1969).
9. R. Berger, J.-C. Bissey, J. Kliava, *et al.*, J. Magn. Magn. Mater. **234**, 535 (2001).

Translated by V. Isaakyan

Technical Physics, Vol. 50, No. 6, 2005, pp. 794–798. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 6, 2005, pp. 121–125.

SHORT COMMUNICATIONS

Ion–Optical Scheme for a Portable High-Sensitivity Mass Spectrograph

V. T. Kogan, S. A. Manninen, D. S. Lebedev, O. S. Viktorova, and I. T. Amanbaev

Ioffe Physicotechnical Institute, Russian Academy of Sciences,
ul. Politekhnicheskaya 26, St. Petersburg, 195021 Russia

Received September 20, 2004

Abstract—An ion–optical scheme that allows for a considerable improvement of the sensitivity and speed of a mass spectrograph is suggested. These ends are accomplished by appropriately choosing the parameters of an electrostatic capacitor and the shape and location of a magnetic lens, as well as by using an original multichannel linear sectional detector. The numerical simulation of the detector is carried out. It is shown that rapid analysis of multicomponent mixtures in a wide mass range, along with a substantial extension of the dynamic range of a signal recorded, is a possibility. The basic parameters of a mass analyzer based on the scheme suggested are verified experimentally, and the results obtained are discussed. *© 2005 Pleiades Publishing, Inc.*

INTRODUCTION

Portable mass spectrometers are effective instruments of physicochemical analysis. Their advantages are highlighted in performing *in situ* real-time measurements in environmental engineering, geophysics, space research, and process control [1]. Being capable of operating independently, these devices are indispensable in carrying out long-term routine measurements under extreme or hazardous conditions. Static sectoral mass spectrometers offer a number of advantages that are of importance for portable instruments: a low power consumption, the feasibility of continuous analysis of multicomponent mixtures, and a simple interface between the ion mass separator and any ion source currently available.

Most today’s portable static analyzers operate in the scanning mode [2–4]. This mode cannot provide a high sensitivity and speed of measurement, since peaks in the ion mass spectrum are recorded in turn. To remedy the situation at least partially, several sensors, channel electron multipliers, are usually used. They are arranged so as to provide an optimal partitioning of the total ion mass range into subranges. The drawback of such an approach is that scannings of the ion mass spectrum within a subrange are interdependent and proceed in parallel when an electric or magnetic field in the mass analyzer is varied.

At present, the Mattauch–Hertzog scheme using a coordinate-sensitive detector [5] is becoming the workhorse in static mass spectrometers intended for simultaneous analysis of mixture components. This scheme incorporates chevron assemblies of microchannel plates or assemblies of microchannel plates combined with CCD arrays. Mass spectrometers equipped with such detectors provide a high resolution and sensitivity, as well as a short resolving time (several milliseconds). However, in sectoral mass spectrometers, the ions are spatially separated in the focal plane; therefore, a large amount of charge is accumulated somewhere in the detector when the conventional schemes (including the Mattauch–Hertzog one) are used even if a spread in ion mass is moderate. That is why the dynamic range where the instrument provides a linear relationship between the component concentration and signal amplitude at the output of the detector is limited (10^4–10^5). Such a dynamic range prevents the portable devices from being efficient in serving a variety of research needs. In addition, the portable devices impose restrictions on the dimension and weight of the 90° magnetic system employed in the Mattauch–Hertzog scheme.

In [6], an original mass spectrograph using a magnet with a sector angle of 50° was described. Its characteristics meet the portability requirements to a greater extent than those of the devices based on the Mattauch–Hertzog scheme. At the same tine, the design of that mass spectrograph suffers from disadvantages, such as the absence of focusing in the direction normal to the pole pieces and, accordingly, a low sensitivity.

ION–OPTICAL SCHEME OF THE MASS SPECTROGRAPH

In this work, we consider the ion–optical scheme of a mass spectrograph (Fig. 1a) with angular and energy focusing of ions that provides rapid recording of ion mass spectra over a wide mass range with the separating electric and magnetic fields of the mass analyzer remaining constant. First-order angular and energy focusing is provided in the plane perpendicular to the z axis over a wide range of ion masses. With such focusing, losses of particles that have a velocity component in the z direction are virtually excluded owing to the presence of a spherical electrostatic capacitor and also by properly selecting the angle of incidence of ions on the entrance boundary of the magnet. Figure 1b is the cross-sectional view of an ion beam. We managed to find system parameters such that focusing in the z direction does not impair the resolution of the system. In doing so, we applied the algorithm [6] for optimizing the design of a portable mass spectrometer with an electrostatic cylindrical capacitor and the curved exit boundary of the magnet. In [7], expressions were derived for deviation δ of the ions from the equilibrium trajectory after they have passed through cylindrical and spherical electrostatic systems (the source size is assumed to be small, and the radii of the equilibrium trajectories are set equal to each other). For a cylindrical deflector,

$$\delta = \alpha\left[(l_c^1 + l_c^2)\cos(\sqrt{2}\varphi_c) + \left(\frac{1}{\sqrt{2}} - \sqrt{2}l_c^1 l_c^2\right)\sin(\sqrt{2}\varphi_c)\right] + \sigma(1 - \cos(\sqrt{2}\varphi_c) + \sqrt{2}l_c^2\sin(\sqrt{2}\varphi_c)),$$

for a spherical deflector,

$$\delta = \alpha[(l_s^1 + l_s^2)\cos\varphi_s + (1 - l_s^1 l_s^2)\sin\varphi_s] + \sigma(1 - \cos\varphi_s + l_s^2\varphi_s).$$

From these expressions, we arrive at a relationship that provides the equality of the relevant distances to the region of first-order angular and energy focusing in these systems,

$$\frac{\cos(\sqrt{2}\varphi_c) - \sqrt{2}l_c^1\sin(\sqrt{2}\varphi_c) - 1}{\sqrt{2}\left(\sqrt{2}l_c^1\sin^2(\sqrt{2}\varphi_c) - \frac{1}{2}\sin(\sqrt{8}\varphi_c)\right)} = \frac{\cos\varphi_s - l_s^1\sin\varphi_s}{l_s^1\sin^2\varphi_s - \frac{1}{2}\sin(2\varphi_s)}. \tag{1}$$

Here, l^1 is the ratio of the distance between the ion source and entrance boundary of the deflector to the mean radius of the deflector; l^2 is the ratio of the distance between the exit boundary of the deflector and the region of focusing to the mean radius of the deflector; φ_c and φ_s are the sector angles in the cylindrical and spherical deflectors, respectively; α is the relative angle spread; and σ is the relative energy spread.

With the parameter values taken in accordance with relationship (1), the resolution of the mass spectrograph remains unaffected in switching from one electrostatic deflector to the other.

The analyzer shown in Fig. 1 generates an ion flux with a small velocity component in the z direction at its exit. This allows us to apply an original linear sectional detector as a spectrum recorder.

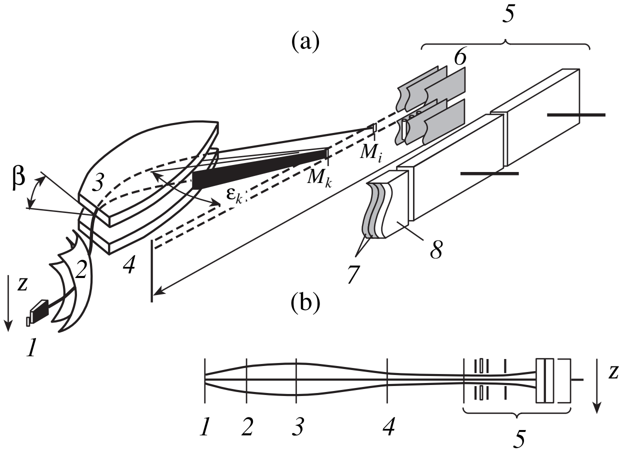


Fig. 1. Mass spectrometer with z focusing: (a) general view and (b) cross-sectional view of the ions in the z direction. (*1*) Ion source; (*2*) electrostatic capacitor (mean radius 60 mm, sector angle 60°); (*3*, *4*) magnets (angle of arrival $\beta = 45°$, sector angle $\gamma = 55°–90°$); and (*5*) detector including electrostatic filter (*6*), chevron assembly of microchannel plates (*7*), and selector (*8*) (angle of arrival $\varepsilon = 30°–90°$).

DESIGN OF THE MULTICHANNEL LINEAR SECTIONAL DETECTOR

The detector consists of an electrostatic filter, a chevron assembly of microchannel plates (MPs), and a collector. The entire ion mass spectrum is displayed on the focal plane of the mass spectrograph. The electrostatic filter extracts ions of a certain mass passing through the MPs (Fig. 1). It has a small extent in the direction of ion propagation and may be viewed as a component of the detector that is placed in the focal plane of the long-focus mass-spectroscopic system. The recording part of the detector is partitioned into segments (sections), each containing its own sensor that comprises a chevron MP assembly and a collector. Each of the segments covers a certain mass subrange. Within each subrange, the detectors operate independently, and each of the mass subranges is scanned in an arbitrary sequence by applying scattering potentials to the electrodes of the electrostatic filter. The desired operating mode is controlled by software that selects ions or ion bunches and directs them toward the detector through the focal plane of the mass spectrometer. Such an operating mode is feasible because of a high dispersion in the heavy ion range. Whether or not the ions may pass toward the detector depends on the presence or absence of a potential on the corresponding electrode of the electrostatic filter. The electrode spacing in the focal plane must be smaller than the peak spacing in the mass spectrum in order to keep the resolution of the instrument at a desired level.

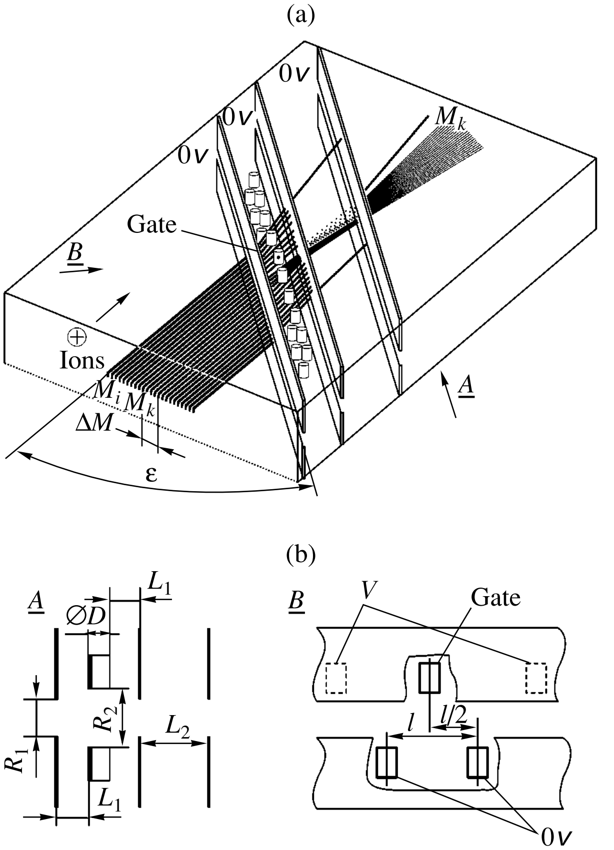


Fig. 2. Segment of the electrostatic filter: (a) general view and ion trajectories (simulation with the SIMION 7 software) and (b) design viewed from *A* and *B*. The lower row of cylindrical electrodes and gate electrode are unbiased; the upper electrodes, except for the gate, are under potential *V* depending on ion energy *E* and the angles of incidence of the ions on the detector.

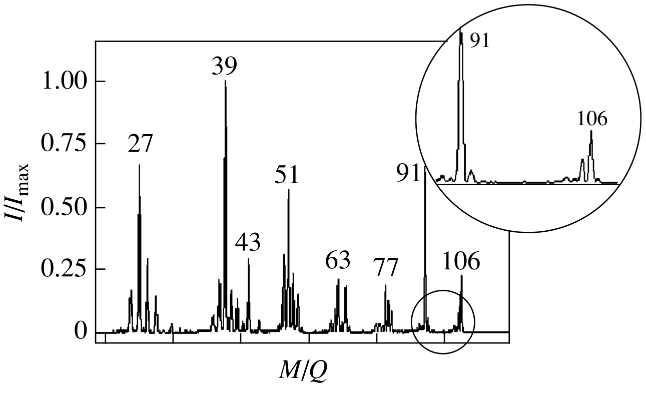


Fig. 3. Mass spectrum of orthoxylene obtained with the prototype of the mass spectrograph operating in the scanning mode. The prototype has the spherical electrostatic capacitor and sectional magnet with the curved exit boundary.

ION–OPTICAL SCHEME OF THE LINEAR SECTIONAL DETECTOR

Figure 2 schematically shows a segment of the electrostatic filter. When the cylindrical electrodes in the upper row are under a negative potential and the electrodes in the lower row remain unbiased, the ions entering the electrostatic filter of the recording unit in the *x* direction and being deflected in the *z* and *y* directions remain in the filter. When one or several upper electrodes are under the zero potential, a weakly scattering region forms near these electrodes, which provides selective passage of the ions toward the MPs. In other words, one can examine a desired part of the mass spectrum (i.e., can allow selected ions to reach the MPs) by varying the potentials on the filter's electrodes. Such an approach provides for rapid analysis owing to a low value of the control potential, on the one hand, and because scanning of part of the spectrum is carried out independently within each subrange totally covering an ion peak or several peaks, on the other.

For a given time interval, the sensor of the detector's segment records one or several ion peaks within a desired ion mass subrange. The electrostatic filter has a weak effect on the ions passing toward the MP chevron assembly, scattering them in the *xy* plane (Fig. 2a). Since the sensitive element (MP) is at a distance from the focal plane, the dynamic range of the detector may reach a level of ~10^7 ions/s. This value is two orders of magnitude higher than that provided by MPs incorporated into the conventional coordinate-sensitive detector.

NUMERICAL AND EXPERIMENTAL SIMULATION OF THE MASS SPECTROGRAPH

The parameters of the electrostatic and magnetic lenses of the portable mass spectrograph were taken based on the results of numerical simulation. For the electrostatic spherical capacitor, the mean radius and the sector angle were taken to be 60 mm and 60°, respectively. For the magnetic system, the angle of arrival of the ions at the system was taken to be $\beta = 45°$, the sector angles were varied between 55° and 90°, the radius of the exit boundary of the magnetic lens was taken to be equal to 130 mm, and the radii of the pole pieces of the magnets (14–100 mm) depended on the ion trajectory radius. Due to *z* focusing, the sensitivity of such a mass spectrograph is roughly one order of magnitude higher than that of the scanning mass spectrometer with the same dimensions (150 × 250 × 400 mm) described in [6].

A developmental version of the mass analyzer was tested experimentally. In the experiments, the linear sectional detector was not used and the parameters of the prototype in different mass ranges were determined in the scanning mode: the voltages on the spherical capacitor plates were varied simultaneously with and in proportion to the accelerating voltage in the ion source.

Figure 3 shows the ion mass spectrum of orthoxylene obtained with a VÉU-6 channel detector placed at a distance of 0.4 mm behind the collimator. The collimator, in turn, was placed at the point where the ions traveling along the largest radius trajectory in the magnetic field meet the focal plane of the mass analyzer. It is in this region (corresponding to the heavy ion range) that the instrument must provide the highest resolution. As follows from Fig. 3, the resolution of the prototype is close to the rated value, ≈130 at 5%. The prototype allows for measurements in the mass range from 4 to 200 u. However, in the absence of the multichannel detector, the time to record the total spectrum is the same as in the scanning instruments, ≈1 min.

SIMULATION OF THE DETECTOR OPERATION

The numerical simulation and the selection of the electrostatic filter parameters were performed using the SIMION 7 program. The transmission functions of the system are presented in Fig. 4. Here, the responses of the system are shown for the case when the angular spread of the ions arriving at the entrance to the filter is absent and the ions are uniformly distributed in the space. The angular spread can be neglected, since the magnet–filter distance far exceeds the ion path in the field of the filter. The response was calculated for different angles of incidence of ions on the entrance window of the detector and for the numbers of electrodes that form a window in the potential barrier of the multichannel deflector varying from one to three, since the ion directions and the resolution differ in various parts of the spectrum.

The simulation results show that an electrostatic filter equipped with a set of electrodes the parameters and location of which are determined for each subrange can provide the controllable spatial separation of ions with different masses and effective collection of the ions at the exit from the mass analyzer.

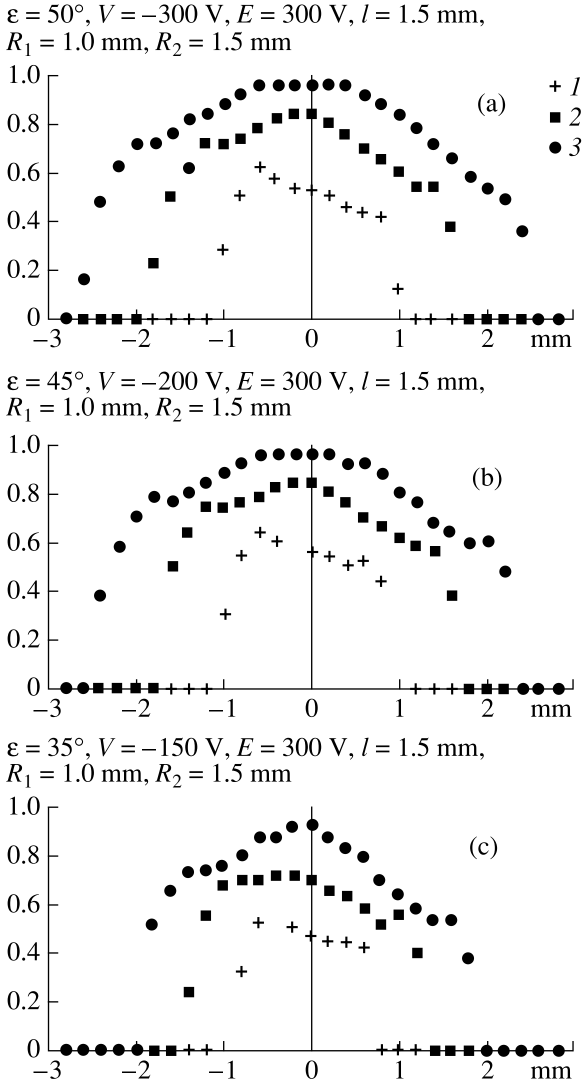


Fig. 4. Response of the multichannel system normalized by the number of ions falling into its entrance window as a uniformly distributed unidirectional flux. The angle of incidence is (a) 50°, (b) 45°, and (c) 35°. The number of electrodes forming a window in the potential barrier of the multichannel deflector is (*1*) one, (*2*) two, and (*3*) three.

RESULTS AND DISCUSSION

The ion–optical scheme presented in this work, on the one hand, makes it possible to improve the sensitivity of the mass spectrograph and, on the other hand, matches the instrument to the multichannel sectional detector providing more effective measurements. In the mass spectrograph under consideration, selective scanning within each subrange of the detector is carried out independently by applying control potentials to the electrodes of the filter. The time it takes for the detector to record each ion peak depends on the intensity of the peak (the higher the intensity, the shorter the time). Therefore, the ion flux intensity may be estimated from the time interval over which the signal reaches a preset level. In this operating mode, the recording time may be varied according to the concentration of a component.

In the operating mode suggested, the speed of analysis is little inferior to the speed when ion spectra are recorded with a coordinate-sensitive detector, since the total measurement time is controlled by the time necessary to record low-concentration components. At the same time, the mode suggested offers a much wider dynamic range of the detecting unit. Moreover, unlike the Mattauch–Hertzog scheme, which alone is used in combination with coordinate-sensitive detectors, our scheme allows the designers to place the MP assembly away from the magnetic field of the mass analyzer.

In comparison with sweeping the mass spectra by means of separating fields of the mass analyzer, the measuring mode suggested in this paper raises the speed of analysis by several tens of times.

CONCLUSIONS

We considered an ion–optical scheme for a mass spectrograph featuring a spherical electrostatic capacitor, as well as an inclined entrance and curved exit boundaries of the magnetic lens, that allows a significant rise in the sensitivity of the instrument. An original multichannel linear sectional detector intended for rapid parallel and/or sequential spectral measurements of many-component chemical compounds is designed.

The high sensitivity, speed, and dynamic range of the instrument combined with the possibility of simultaneously recording partial (componentwise) spectra in a wide ion mass range renders it superior not only to static but also to dynamic mass spectrometers. Due to the advantages mentioned above, the portable mass spectrometer can be applied for tackling a variety of scientific and applied problems needing express measurement. The instrument may also be helpful in monitoring the composition of gases and volatiles in space research, environmental engineering, medicine, geophysics, seismology, and other fields of science and technology.

REFERENCES

1. E. R. Badman and R. G. Cooks, J. Mass Spectrom. **35**, 659 (2000).
2. A. O. Nier, Int. J. Mass Spectrom. Ion Processes **66**, 55 (1985).
3. J. H. Hoffman *et al.*, IEEE Trans. Geosci. Remote Sens. **GE-18**, 80 (1980).
4. V. T. Kogan, A. K. Pavlov, Y. V. Chichagov, *et al.*, Field Anal. Chem. Technol. **1**, 331 (1997).
5. M. P. Sinca *et al.*, Int. J. Mass. Spectrom. **176**, 99 (1998).
6. V. T. Kogan, G. Yu. Gladkov, and O. S. Viktorova, Zh. Tekh. Fiz. **71** (4), 130 (2001) [Tech. Phys. **46**, 492 (2001)].
7. V. P. Afanas'ev and S. Ya. Yavor, *Electrostatic Energy Analyzers for Charged Particle Beams* (Nauka, Moscow, 1978).

Translated by V. Isaakyan

Technical Physics, Vol. 50, No. 6, 2005, pp. 799–801. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 6, 2005, pp. 126–128.
Original Russian Text Copyright © 2005 by Dolov, Kuznetsov.

SHORT COMMUNICATIONS

Modification of the Large Particle Method Applied in Numerical Solution of Problems of Microwave Electronics

A. M. Dolov and S. P. Kuznetsov

Institute of Radio Engineering and Electronics, Saratov Branch, Russian Academy of Sciences, ul. Zelenaya 38, Saratov, 410019 Russia

e-mail: Kuznetsov@sgu.ru

Received September 21, 2004

Abstract—The method of large particles, which is applied in one-dimensional numerical simulation of processes in electron devices, such as traveling-wave tubes, is suggested. With the number of large particles fixed, the amount of computation remains virtually the same, while the accuracy attained under strong nonlinearity rises significantly. *© 2005 Pleiades Publishing, Inc.*

The approach that consists in replacing the electron beam with a set of “large particles” which are few in number compared to the electrons in a real beam is being widely used in numerical simulation of processes taking place in microwave electron devices, such as traveling-wave tubes, backward-wave tubes, gyrotrons, etc. The simplest version of this method dates back to the early works of Nordsieck and Vaĭnshteĭn on numerical solution of the equations used in the one-dimensional nonlinear theory of traveling-wave tubes [1–3]. In the dimensionless form without regard to space-charge and energy losses in the slow-wave structure, these equations are written as [4, 5]

$$\frac{\partial^2\theta}{\partial\xi^2} = -\mathrm{Re}Fe^{i\theta}, \quad \theta|_{\xi=0} = \theta_0, \quad \left.\frac{\partial\theta}{\partial\xi}\right|_{\xi=0} = 0, \tag{1}$$

$$\frac{\partial F}{\partial\xi} + ibF = J, \quad F|_{\xi=0} = F_0, \tag{2}$$

$$J = -\frac{1}{\pi}\int_0^{2\pi} e^{i\theta}d\theta_0. \tag{3}$$

The first relationship here is the equation of motion of electrons, where θ is the particle’s phase relative to the wave, F is the normalized complex amplitude of the wave, and ξ is the dimensionless coordinate. The second relationship is the equation of wave excitation, where parameter b characterizes the asynchronism between the wave and electron beam and right-hand side J is the normalized complex amplitude of the first harmonic of the high-frequency current. These equations are supplemented by initial conditions for the particles and field at the entrance to the space of interaction. In the expression for the current harmonic amplitude in terms of quantity θ, which is associated with beam particles, integration is over initial phase θ_0 of the particles. The dimensionless quantities and their relation to the related dimensional quantities are discussed in more detail elsewhere [4].

The essence of the method of large particles is in replacing the continuous distribution of electrons over their initial phase by finite set K of particles that have phases $\theta_k^0 = 2\pi k/K$ at the entrance to the space of interaction. Then, integration in expression (3) for the complex amplitude of the first harmonic of the bunched beam current may be replaced by summation,

$$J = -\frac{2}{K}\sum_{k=0}^{K-1} e^{i\theta_k}. \tag{4}$$

Formally, calculation by formula (4) corresponds to rectangular integration, which is mentioned in calculus classes as the simplest method of numerical integration and is generally deprecated due to a large computational error. However, when a periodic function is integrated over an interval whose length is exactly the same as the period (the case under study), the error turns out to be relatively low and comparable to that produced by seemingly more accurate trapezoidal integration, which is based on the linear interpolation of the integrand between equidistant points on the initial phase axis. In effect, integration by formula (4) fails under strong nonlinearity, when the relative displacement of neighboring large particles is of order 1. Formally, it is necessary to decrease the step of integration over the initial phase in this case, i.e., to increase the number of particles (the higher the nonlinearity, the larger the number of particles). In practice, this circumstance is often disregarded, because the results obtained with a small number of particles are still acceptable. Actually, however, a relatively small number of particles (say, three) is the major source of the error in the current harmonic, which varies as $1/K$. When fine points, such as the type of dynamic regime in numerical solution of nonstationary problems, stability of stationary regimes, etc., are

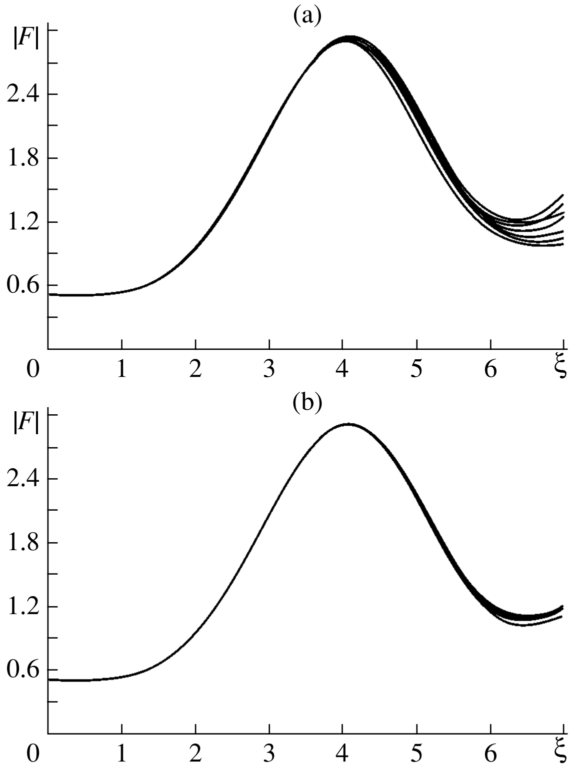


Fig. 1. Wave field amplitude distribution along the system for various initial phases of the input signal. The conventional method of large particles for K = (a) 20 and (b) 60.

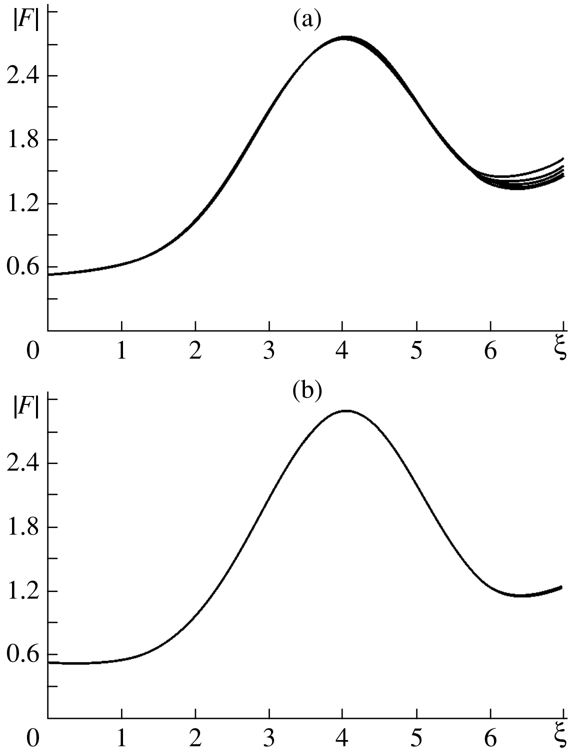


Fig. 2. The same as in Fig. 1 for the modified method at K = (a) 20 and (b) 60.

analyzed, such errors may become crucial. Accordingly, one has to drastically increase the number of particles, which is unfortunately scarcely efficient because of a slow decrease of $1/K$.

Our idea of modifying the method consists in abandoning the conventional approach to numerical integration and applying linear interpolation of the exponent rather than the whole integrand. Physically, this means the following. Instead of being discretely concentrated on the particles, the charge is assumed to be uniformly distributed between any pair of particles whose coordinates are involved in the calculation. Within each interval from $2\pi/K$ to $2\pi(k + 1)/K$ of θ_0, the exponential whose exponent is a linear function of argument is integrated analytically. As a result, the problem reduces to finding a solution to the following system of differential equations:

$$\frac{\partial^2\theta_k}{\partial\xi^2} = -\mathrm{Re}(F\exp(i\theta_k)),\quad \theta_k|_{\xi=0} = \frac{2\pi k}{K},\qquad \left.\frac{\partial\theta_k}{\partial\xi}\right|_{\xi=0} = 0, \tag{5}$$

$$\frac{\partial F}{\partial\xi} + ibF = J,\quad F|_{\xi=0} = F_0, \tag{6}$$

$$J = -\frac{2}{K}\sum_{k=0}^{K-1}\frac{\sin((\theta_k-\theta_{k+1})/2)}{(\theta_k-\theta_{k+1})/2}\exp(i(\theta_k+\theta_{k+1})/2). \tag{7}$$

It should be taken into account in summation that $\theta_K \equiv \theta_0 + 2\pi$ according to the equations of motion.

The error involved in such a scheme can hardly be formally (mathematically) estimated, since it depends on function $\theta(\theta_0)$ appearing in the equations, which itself is determined in the course of nonlinear calculation. Therefore, to demonstrate the improved efficiency of the method, we turn to a numerical experiment.

Figures 1 and 2 show numerical solutions to the equation of the one-dimensional nonlinear theory of traveling-wave tubes: the dependences of the wave field dimensionless amplitude $|F|$ on the dimensionless coordinate ξ. Figure 1 corresponds to the conventional method of large particles; Fig. 2, to its modified version. In either figure, the set of curves corresponds to different phases of input signal $F_0 = A_0e^{i\varphi}$ with the maximal amplitude fixed, $A_0 = 0.5$, and argument φ varying in the range from 0 to $2\pi/K$. The number K of particles is set equal to 20 in Figs. 1a and 2a and 60 in Figs. 1b and 2b.

It can be analytically demonstrated that, in the continuous case (see Eqs. (1)–(3), where integration, rather than summation, is used), the dependence on the initial phase is absent. Then, the phase-dependent discrepancy between the curves obtained numerically gives an idea of the methodical error due to beam discretization. If the degree of nonlinearity corresponds to the first minimum of the field amplitude, the error introduced by the conventional method is rather high, while decreasing with an increase in the number of large particles. As for the method modified, even at $K = 20$, it yields the results close to those yielded by the original method at $K = 60$.

If only Eqs. (1)–(3) of the elementary nonlinear theory of traveling-wave tubes were to be solved, it would not be a problem to increase the number of particles many times in order to ensure a high accuracy using advanced computing facilities. However, in many cases (e.g., in solving nonstationary problems of microwave electronics [6–8], especially with regard to the space charge), the modification of the method may be of great value. It would be of particular benefit in investigation into microwave electron devices by the methods of nonlinear dynamics, where numerical calculations imply a comprehensive analysis of oscillatory modes corresponding to various parameters, bifurcation analysis, and stability analysis of stationary modes.

ACKNOWLEDGMENTS

The work was supported by the Russian Foundation for Basic Research, project no. 03-02-16192.

REFERENCES

1. A. Nordsieck, Proc. IRE **41**, 630 (1953).
2. L. A. Vaĭnshteĭn, Radiotekh. Élektron. (Moscow) **2**, 883 (1957).
3. L. A. Vaĭnshteĭn, Radiotekh. Élektron. (Moscow) **2**, 1027 (1957).
4. L. A. Vaĭnshteĭn and V. A. Solntsev, *Lectures on Microwave Electronics* (Sov. Radio, Moscow, 1973) [in Russian].
5. A. M. Kats, E. M. Il'ina, and I. A. Man'kin, *Nonlinear Phenomena in O-Type Microwave Devices with Long-Term Interaction* (Sov. Radio, Moscow, 1975) [in Russian].
6. N. S. Ginzburg, S. P. Kuznetsov, and T. N. Fedoseeva, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **21**, 1037 (1978).
7. A. P. Kuznetsov and S. P. Kuznetsov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **27**, 1575 (1984).
8. N. S. Ginzburg, N. A. Zavol'skiĭ, and G. S. Nusinovich, Radiotekh. Élektron. (Moscow) **32**, 1031 (1987).

Translated by A. Pankrat'ev

Technical Physics, Vol. 50, No. 6, 2005, pp. 802–804. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 6, 2005, pp. 129–131.
Original Russian Text Copyright © 2005 by Khodenkov.

SHORT COMMUNICATIONS

Sense Anisotropy of the Kink Dynamics in an Evolutionary Equation with Broken Space Symmetry

G. E. Khodenkov

Magnitooptoélektronika Laboratory, General Physics Institute, Ogarev Mordovian State University,
Bol'shevistskaya ul. 68, Saransk, 430000 Russia
e-mail: angeline@mtu-net.ru

Received October 11, 2004

Abstract—A nonlinear diffusion-type evolutionary equation that describes the effect of sense anisotropy of velocity when the front of a solitary wave (kink) moves with unequal velocities in the forward and (when an external field reverses sign) backward directions is suggested. A prerequisite for such behavior is space symmetry breaking in the system. The equation is applied to describing the dynamics of 180° domain walls in ferromagnets with negative cubic crystalline anisotropy. The amount of the sense anisotropy effect in YIG may reach 20%. *© 2005 Pleiades Publishing, Inc.*

It is known that nonequilibrium Langevin random forces (with a zero mean) may cause directional motion (transport) of particles in systems with broken space symmetry of the potential in oscillatory fields. Such a mechanism acting like a ratchet is considered to be common in the dynamics of various objects with “Brownian engines,” such as bacteria, nanoparticles, etc. [1]. Before passing to description of another mechanism, we note that space symmetry breaking, a prerequisite for the transport, may be related not only to the external potential but also to the structure of the particle, where a preferential direction is thus specified. This situation is described well by the model of umbrella [2], where the viscous force involves both odd and even powers of velocity. In terms of this model, sense asymmetry (anisotropy) of velocity (unequal velocities in the forward and backward directions) is a straightforward effect even if external fields are invariable.

The above speculation is obviously equally valid for continuous phenomenological nonlinear evolutionary equations. For example, the author showed [3, 4] that 180° domain walls (DWs) with a kink in magnets with negative crystalline anisotropy (Ni, YIG, etc.) exhibit sense anisotropy of velocity when constant driving field H reverses. Analytical estimates accurate to H^2 (weak field approximation) were made using the Landau–Lifshitz equations.

With the aim to considerably extend the domain of nonlinearity, the author suggests a simple one-component evolutionary equation similar to the nonlinear equation of diffusion that applies to a wide class of systems (including nonmagnetic systems). Results accurate to the second order of smallness in field that follow from this equation are totally coincident with those obtained previously [3, 4]. A numerical extrapolation toward higher order fields discovers a noticeable sense anisotropy of velocity in YIGs, which may amount to 20%, as follows from estimates.

To illustrate the aforesaid, consider a nonlinear dimensionless equation for some field-related quantity $\psi(x, t)$ that depends on spatial coordinate x and time t,

$$\partial_t \psi + H \sin\psi = \partial_{yy}\psi - 1/2\, dw/d\psi, \tag{1}$$

where H is the driving field and $w(\psi)$ is the energy density.

It is supposed that the right of (1) has two saddle points for $d\psi(y \longrightarrow \pm\infty)/dy \longrightarrow 0$, ($\psi = 0$, $\psi' = 0$) and ($\psi = \pi$, $\psi' = 0$), which are connected by two separatrices (180° kinks) with spatial derivatives of opposite sign. The assumption of key importance is that energy density $w(\psi)$ is noninvariant under substitution $\psi \longrightarrow -\psi$. Then, one can show that velocity V in the local frame of reference $x = y - Vt$ where $\partial_t\psi = -V\partial_x\psi$ involves terms that are even in H.

A solution to (1) is sought as series in H with subscripts indicating the order of relevant quantities. In the zeroth-order approximation, expression (1) yields the equation $\psi_0''(x) - (1/2)dw/d\psi_0 = 0$, the first integral of which is $\psi_0'^2 = w(\psi_0)$ by virtue of autonomy. Let us show that there is no need for calculating corrections $\psi_{1,\,2}$ in order to discover the desired effect in the second order of smallness: it suffices to know only V_1 (ψ_0 is known from the first integral) for this purpose. Multiplying (1) by $d\psi/dx$ (putting for the moment $d\psi/dx > 0$ for definiteness) and integrating the result in infinite terms yields the exact relationship $V = 2H/\langle\psi'^2\rangle$ for the kink, from which we derive the expansion

$$\begin{gathered} V = V_1 + V_2 + \ldots, \quad V_1 = 2H/\langle\psi'^2\rangle, \\ V_2 = -4H\langle\psi_0'\psi_1'\rangle/\langle\psi_0'^2\rangle^2, \end{gathered} \tag{2}$$

where $\langle \ldots \rangle$ is the integral over dx in infinite terms.

Note that $\langle \psi'^2 \rangle$ is proportional to the dissipation function for (1). Next, it will be demonstrated that quantity $\langle \psi'_0 \psi'_1 \rangle$ can also be expressed in quadratures and does not require knowledge of ψ_1.

In the first order of smallness, (1) yields the inhomogeneous equation

$$\hat{L}\psi_1 = f(\psi_0(x)), \tag{3}$$

where $\hat{L} = -d^2/dx^2 + \psi_0'''/\psi_0'$ is the Hermitean operator and $f(\psi_0) = -\psi_0' V_1 + H\sin\psi_0$. Since $\hat{L}\psi_0' = 0$, the right of (3) must be orthogonal to ψ_0'; that is, $\langle \psi_0' f(\psi_0) \rangle = 0$ (from which V_1 is determined). Multiplying both sides of (3) by $d\psi_0/dx$ and integrating first over dx from $-\infty$ to y and then over dy by parts in infinite terms, we get for $\langle \psi_0' \psi_1' \rangle$ the expression

$$-\frac{1}{2}\int\limits_{-\infty}^{\infty} dy \int\limits_{-\infty}^{y} dx f(x)\psi_0'(x) = \frac{1}{2}\int\limits_{-\infty}^{\infty} dx x f(x)\psi_0'(x)$$
$$\equiv \frac{1}{2}\int\limits_{\psi_0(-\infty)}^{\psi_0(\infty)} d\psi_0 x(\psi_0) f(\psi_0), \tag{4}$$

where $f(x) \sim H$ is the right of (3) and function $x(\psi_0)$ is found from the first integral.

In space symmetry is lacking, integral (4) is other than zero and velocity (2) has even-in-field components.

As applied to magnetic systems, consider the nonlinear motion of a 180° DW in cubic ferromagnets with negative magnetic anisotropy (this problem was studied in terms of the Landau–Lifshitz equations earlier [3, 4]). Here, magnetostatic interactions prevail, so that the DW structure is similar to the Bloch structure. Then, when allowing for dynamic deviations of the magnetic moment, we can restrict analysis to only one angular variable (instead of two, as in [3, 4]) lying in the DW plane. The anisotropy energy in this plane can be expressed through angular variable ψ alone in the form (see, e.g., [5])

$$w = \sin^2\psi\left[b + \frac{1}{12}(\sin\psi - \sqrt{8}\cos\psi)^2\right], \tag{5}$$

where $b > 0$ is a small constant (in YIGs, $b \sim 0.01$). In this case, relationship (5) in [1] applies. The two complementary separatrices mentioned above (180° DWs) with $\psi_0' < 0$ and $\psi_0' > 0$ have the form

$$\cot\psi_0 = [A \mp B\sinh Bx]\left(b + \frac{2}{3}\right)^{-1}, \tag{6}$$

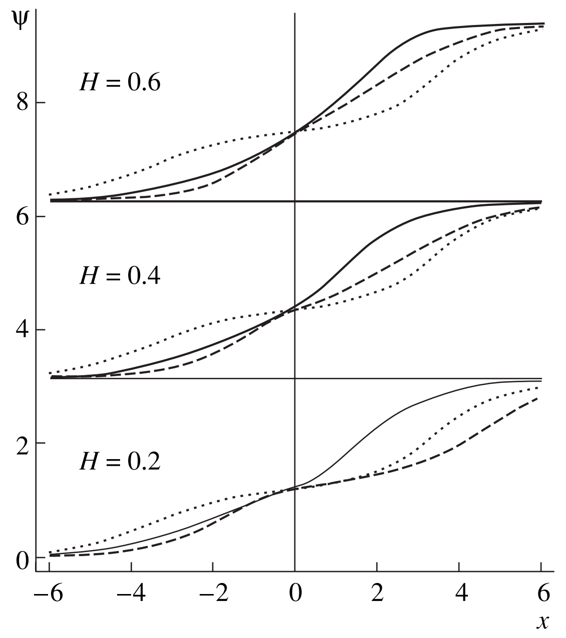


Fig. 1. Structures $\psi(x)$ of π links for $H = 0$ (dotted lines), $H < 0$ (dashed lines), and $H > 0$ shown by the curves (continuous lines).

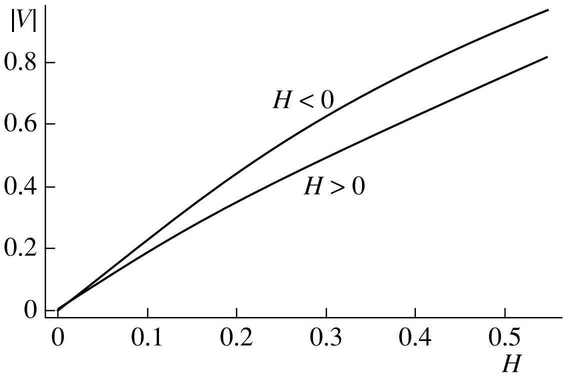


Fig. 2. Field dependence of the velocity of the kink moving in the forward and backward directions.

where $A = 1/(3\sqrt{2}(b + 2/3))$, $B = \sqrt{b\left(b + \frac{3}{4}\right)}/(b + 2/3)$, and $C = b + 2/3$.

In view of (6), dynamic results (2) and (4) are totally coincident with those obtained in [3, 4] in the second order of smallness. In formulas (2) and (4), $b = 0.01$, $\langle \psi_0'^2 \rangle \approx 0.95$, and $\langle \psi_0' \psi_1' \rangle \approx \pm 0.6H$. Here, dimensionless applied magnetic field H is given in units of magnetic anisotropy field H_a and velocity V is measured in units of $\gamma H_a \Delta/\alpha$ (where $\gamma > 0$ is the gyromagnetic ratio, Δ is a width-related parameter of the DW, and α is the Gilbert dimensionless damping parameter).

The results of numerical calculations for YIGs performed at fields $H < 2/3 + b$, at which an unfavorably magnetized domain ($\psi = \pi$) is stable (i.e., $d^2w(\psi)/d\psi^2 > 0$), are shown in Figs. 1 and 2 for the upper sign in (6). It is seen that the DW structure changes considerably not only when the magnitude of the field varies but also when its direction (sense of velocity) reverses. The dependence of the structure on the sense of the field (see (3)) in the case when the symmetry of the initial structure is broken (see (6)) causes a difference in dissipation losses at $\pm H$ and, hence, a difference in the velocities. In general terms, sense symmetry is recovered if both structures described by (6) are taken into account.

To evaluate the effect of sense asymmetry, consider, instead of energy density (5), the expression

$$\sin^2\psi\left(b+\frac{3}{4}\sin^2(\psi-\psi_{00})\right), \qquad (7)$$

which coincides with (6) if $\tan\psi_{00} = \sqrt{8}$. While the structure of 180° DWs coincides with (6), the parameters A, B, and C change:

$$A = \frac{3}{4}\sin\psi_{00}\cos\psi_{00}\Big/\left(b+\frac{3}{4}\sin^2\psi_{00}\right),$$

$$B = \sqrt{b\left(b+\frac{3}{2}\sin^2\psi_{00}\right)}\Big/\left(b+\frac{3}{4}\sin^2\psi_{00}\right),$$

$$C = \sqrt{b+\frac{3}{4}\sin^2\psi_{00}}.$$

Introducing sense asymmetry coefficient $\chi(b, \psi_{00}) = \langle\psi'_0\psi'_1\rangle/H$, we find, in view of (4), (6), and (7), that $\chi(b, \psi_{00}) = \chi(b, \chi_{00} + n\pi)$, where $n = \pm1, \pm2\dots$. Within the interval $0 < \psi_{00} < \pi$, we have $\chi(b, 0) = \chi(b, \pi) = 0$ and a single zero, $\chi(b, \pi/2) = 0$. Numerical estimates give the following extrema for this coefficient: $\max\chi(b = 0.1, \psi_{00} \approx 0.77) = 0.9$ and $\min\chi(b = 0.1, \psi_{00} \approx 2.37) \approx -0.9$ (note that $|\chi(b \longrightarrow 0, \psi_{00})| \sim \ln(1/b)$). These extrema determine the amount (extrema) of sense anisotropy.

REFERENCES

1. R. D. Astunnan and P. Haenggi, Phys. Today **55** (11), 33 (2002).
2. Yu. L. Bolotin, A. V. Tur, and V. V. Yanovskiĭ, Zh. Tekh. Fiz. **72** (7), 9 (2002) [Tech. Phys. **47**, 803 (2002)].
3. G. E. Khodenkov, Pis'ma Zh. Éksp. Teor. Fiz. **68**, 823 (1998) [JETP Lett. **68**, 864 (1998)].
4. G. E. Khodenkov, Fiz. Met. Metalloved. **89** (2), 11 (2000).
5. A. Hubert, *Theorie der Domanenwande in Geordneten Medien* (Springer-Verlag, Berlin, 1974; Mir, Moscow, 1977).

Translated by V. Isaakyan

Technical Physics, Vol. 50, No. 6, 2005, pp. 805–806. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 6, 2005, pp. 132–133.

SHORT COMMUNICATIONS

On Optimization of the Charge Carrier Concentration in a Cooling Thermoelectric Branch

O. I. Markov
Orel State University, Orel, 302015 Russia
e-mail: Markov@e-mail.ru
Received October 11, 2004

Abstract—Computer-aided simulation of the temperature field in a thermoelectric branch is performed under the conditions of a maximal temperature difference and a maximal refrigerating capacity. A differential equation of stationary heat conduction with temperature-dependent kinetic coefficients is used in the simulation. The optimal concentration of charge carriers in the branch is determined in terms of the classical statistics. © *2005 Pleiades Publishing, Inc.*

Wide application of thermoelectric coolers to a great extent is hindered by their low efficiency. Prospects in this field are associated largely with the search for new, more efficient thermoelectric materials. However, it is also worth thoroughly revising the potential of the materials currently available. It is common knowledge that the key quality index of semiconductors used in thermoelectric coolers is the thermoelectric figure of merit [1]

$$Z = \frac{\alpha^2\sigma}{\chi}, \tag{1}$$

where α is the thermopower, σ is the conductivity, and χ is the thermal conductivity.

The temperature dependence of the kinetic coefficients is taken into account by averaging them over the working temperature interval [2],

$$\bar{f} = \frac{1}{\Delta T}\int_{T_0}^{T_1} f\,dT. \tag{2}$$

The fact that the kinetic coefficients depend on the free carrier concentration makes it possible to reach a maximal value of the thermoelectric figure of merit by optimizing the concentration. As was justly noted in [3], optimization is usually carried out separately for either thermoelectric branch, which poorly aids in attaining a maximal efficiency (figure of merit) of the thermocouple as a whole. Parameter Z in p- and n-semiconductors reaches the highest value at a certain carrier concentration. It was shown [3] that a relevant calculation may be brought to an end only if the kinetic coefficients are simple functions of the concentration. However, in that work, the concentration dependence of the thermal conductivity is disregarded and the procedure of averaging the coefficients over the temperature interval remains somewhat unclear.

In this work, a numerical technique of thermoelement optimization is suggested. Specifically, it allows for optimization of the carrier concentration in a thermoelectric branch intended for a given temperature interval. Obviously, first of all, one should know the band structure of the semiconductor and choose relevant statistics. Consider a model of a one-valley nondegenerate impurity semiconductor with a spherical Fermi surface in terms of charge carrier classical statistics. For scattering by acoustic phonons, the kinetic coefficients have the following form [4]:

the differential thermopower,

$$\alpha = \frac{k}{e}\left(2 + \ln\frac{2(2\pi mkT)^{3/2}}{nh^3}\right), \tag{3}$$

the electric conductivity,

$$\sigma = enu, \tag{4}$$

the thermal conductivity,

$$\chi = \chi_{\mathrm{ph}} + 2\left(\frac{k}{e}\right)^2 T\sigma. \tag{5}$$

Here, $\chi_{\mathrm{ph}} = a/\sqrt{T}$ is the lattice component of the thermal conductivity and $u = bT^{-3/2}$ is the charge carrier mobility.

Coefficients a and b in formulas (4) and (5) were taken such that the temperature dependences closely fit the respective experimental data for compound $(\mathrm{Bi, Sb})_2\mathrm{Te}_3$. The effective mass was set equal to $0.7m_0$. At 300 K, Z was estimated as 3×10^{-3} K^{-1}. The maximal temperature difference and the optimal carrier concentration, both defined by classical methods, were 75.8 K and $n_0 = 7.5 \times 10^{24}$ m^{-3}, respectively.

Along with this calculation, the boundary-value problem for a thermoelectric branch that is similar to the problem considered in [5] was solved. In our approach, the carrier concentration in either branch may be optimized separately, since the thermal effect of

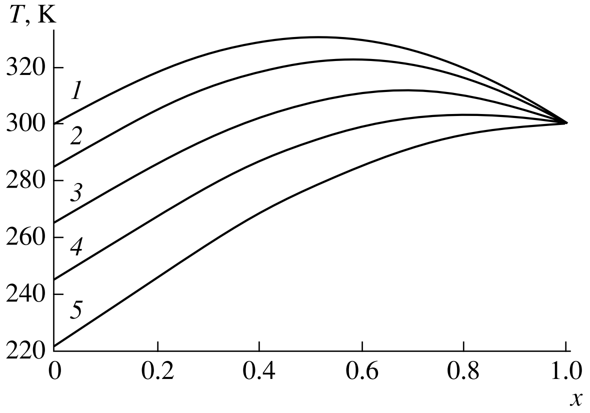


Fig. 1. Temperature distribution along the thermoelectric branch for temperature differences of (*1*) 0, (*2*) 15, (*3*) 35, and (*4*) 55 K. Curve *5* corresponds to the maximal temperature difference.

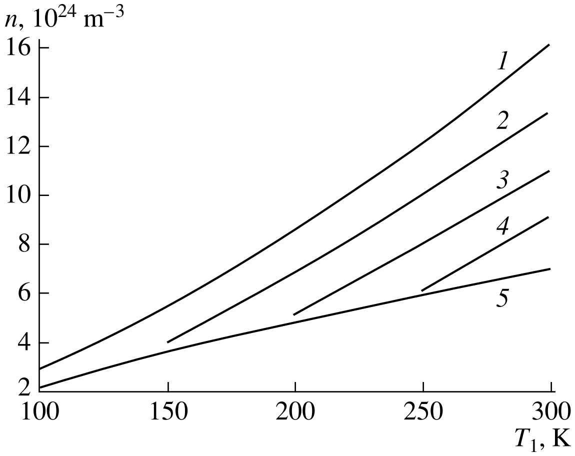


Fig. 2. Charge carrier optimal concentration vs. the hot-junction temperature at different temperature drops. (*1*–*5*) the same as in Fig. 1.

the other is allowed for as load q_0. For the active branch, $q_0 > 0$; for the passive one, $q_0 < 0$ [6]. In the steady-state conditions with allowance for the Thomson effect, the temperature field of a one-dimensional adiabatically insulated homogeneous thermoelectric branch is described by the equation for stationary heat conduction

$$\frac{d}{dx}\left(\chi\frac{dT}{dx}\right)+\frac{Y^2}{\sigma}-YT\frac{d\alpha}{dT}\frac{dT}{dx} = 0 \qquad (6)$$

with the boundary conditions

$$\chi\frac{dT}{dx}\Big|_{x=0} = \alpha YT|_{x=0}-q_0, \quad T|_{x=1} = T_1, \qquad (7)$$

where $q_0 = Q_0 l/S$ is the specific thermal loss of the branch and $Y = Jl/S$ is the specific current in the branch.

The boundary-value problem was solved numerically with simultaneous optimization of the solution in terms of specific current and carrier concentration. The calculation was carried out for both a maximal temperature drop and a maximal refrigerating capacity in the temperature range 100–300 K. The calculation results are shown in Figs. 1 and 2. The former shows the temperature fields in the branch for a maximal temperature difference (curve *5*) and a maximal refrigerating capacity (curves *1*–*4*) at a hot-junction temperature of 300 K. The latter plots the optimal carrier concentration against the hot-junction temperature for a maximal temperature difference (curve *5*) and a maximal refrigerating capacity (curves *1*–*4*). From Fig. 2, it follows that, at a certain temperature difference, the optimal concentration depends on the operating conditions. According to the solution to the boundary-value optimization problem, the optimal concentration at a zero temperature difference equals 16.1×10^{24} m^{-3}, which is more than twice n_0. Since the temperature difference in this case is zero, no averaging over the temperature interval is required and the optimal value should be $n_0 = 7.5 \times 10^{24}$ m^{-3}. This result can be accounted for by the fact that the branch is overheated in its central part (Fig. 1, curve *1*). The optimal concentration is higher than n_0, since an appreciable part of the branch is heated above 300 K. At the zero difference, the specific refrigerating capacity calculated for $n_0 = 7.5 \times 10^{24}$ m^{-3} turns out to be 12.5% lower than that calculated for an optimal concentration of 16.1×10^{24} m^{-3}. Consequently, the optimal concentration determined by the conventional method of average parameters for the active branch will contain a systematic error.

Hence, it follows that the conventional method of designing thermoelectric coolers may fail in determining the charge carrier optimal concentration in the branches. When using the method of average parameters, one should average the kinetic coefficients over the temperature distribution along the branch rather than over the working temperature interval.

REFERENCES

1. A. F. Ioffe, *Semiconductor Thermoelements and Thermoelectric Cooling* (Izd. Akad. Nauk SSSR, Moscow, 1956; London, Infosearch, 1957).
2. A. I. Burshteyn, *Semiconductor Thermoelectric Devices* (Fizmatgiz, Moscow, 1962; Heywood, London, 1964).
3. N. M. Zhukova, A. P. Kashin, M. Z. Maksimov, and O. V. Marchenko, Fiz. Tekh. Poluprovodn. (Leningrad) **23**, 1908 (1989) [Sov. Phys. Semicond. **23**, 1183 (1989)].
4. B. M. Askerov, *Kinetic Effects in Semiconductors* (Nauka, Leningrad, 1970) [in Russian].
5. O. I. Markov, Zh. Tekh. Fiz. **74** (2), 138 (2004) [Tech. Phys. **49**, 277 (2004)].
6. O. I. Markov, Pis'ma Zh. Tekh. Fiz. **30** (13), 11 (2004) [Tech. Phys. Lett. **30**, 532 (2004)].

Translated by V. Isaakyan

Technical Physics, Vol. 50, No. 6, 2005, pp. 807–808. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 6, 2005, pp. 134–135.
Original Russian Text Copyright © 2005 by Sotskov.

SHORT COMMUNICATIONS

Experimental Conductivity Dependence of the Real Part of the Permittivity in a Disordered Paraffin–Graphite Macrosystem

V. A. Sotskov
Kabardino-Balkar State University, ul. Chernyshevskogo 173, Nalchik, 360004 Russia
e-mail:sozkov_va@rambler.ru

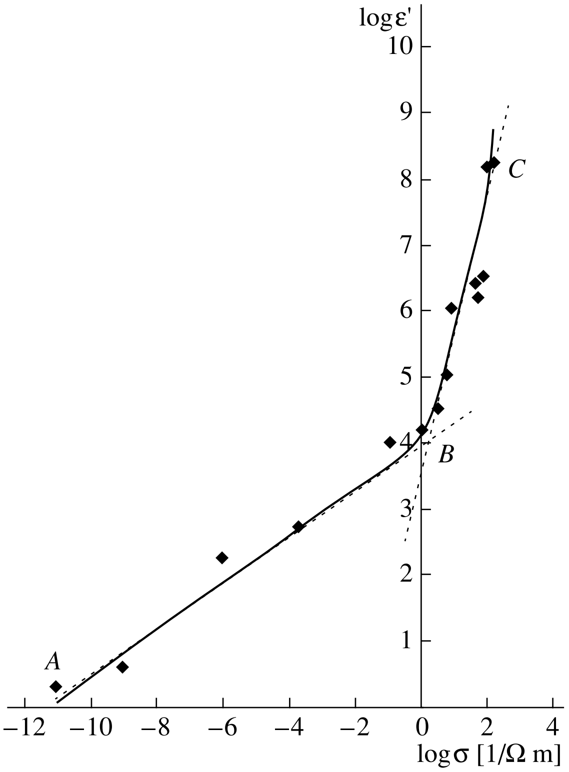
Received October 11, 2004

Abstract—The conductivity dependence of the real part of the permittivity of a paraffin–graphite macrosystem with a variable graphite concentration is studied experimentally. *© 2005 Pleiades Publishing, Inc.*

The dependence of the real part of permittivity, ε', on conductivity σ in microsystems has been extensively studied both theoretically [1, 2] and experimentally [2, 3]. However, as applied to macrosystems, this issue has received only cursory attention [2, 3]. By macrosystems are meant those in which the quantum effects in transfer processes may be ignored and, hence, the conductivity of the system as a whole depends on the geometry of the system, the conductivity of conducting phase particles, and the contact resistance between them. It seems that knowing the dependence $\varepsilon' = f(\sigma)$ may be of value in designing capacitors, cermet resistors, and other electronic components, as well as in preparing conducting pastes.

In this communication, the author reports the experimentally found dependence of the real part of the permittivity on the volume conductivity of a macroscopically disordered paraffin–graphite system.

The reason why this system was taken as a model was discussed at length in [4]. The sample preparation and measuring techniques are described in [4–6]. The test samples were capacitors with plates made of electrolytic copper that were filled with a paraffin–graphite mixture with a given concentration of the conducting phase. The experiments were carried out with plane and cylindrical capacitors. The resistance, capacitance, and Q factor were measured by the standard techniques [7, 8]. The resistance was measured at a constant voltage. The capacitance and Q factor were measured at a frequency of 1 kHz. The final result of this study is shown in the figure. It is seen that the curve $\log\varepsilon' = f(\log\sigma)$ can be represented as consisting of two portions: *AB* and *BC*. The kink *B* corresponds to volume graphite concentration $x \approx 0.15$; i.e., point *B* is near the percolation threshold $x_{\text{per}} \approx 0.16$ [4]. Portion *AB* can be approximated by the power-type relationship $\varepsilon' = 10^4\sigma^{0.35}$; portion *BC*, by $\varepsilon' = 10^{3.5}\sigma^{2.4}$. Generally, such a behavior of the dependence $\varepsilon' = f(\sigma)$ is predictable, since the system's conductivity changes abruptly at the percolation threshold, which, according to currently existing ideas [1, 2], is bound to change the real part of the permittivity. However, the experimental data reported here introduce important refinements; namely, (i) the dependence $\varepsilon' = f(\sigma)$ for the macrosystem is nonlinear throughout the range of conducting phase volume concentrations and (ii) this nonlinearity grows considerably as the conducting phase concentration exceeds the percolation threshold, i.e., at $x > x_{\text{per}}$.



Log–log dependence of the real part of the permittivity on the volume concentration for the paraffin–graphite macrosystem.

1063-7842/05/5006-0807$26.00 © 2005 Pleiades Publishing, Inc.

REFERENCES

1. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media* (Nauka, Moscow, 1982; Pergamon, New York, 1984).
2. V. F. Gantmakher, *Electrons in Disordered Media* (Fizmatlit, Moscow, 2003) [in Russian].
3. V. V. Pasynkov, *Electronic Materials* (Vysshaya Shkola, Moscow, 1980), pp. 155–160 [in Russian].
4. V. A. Sotskov and S. V. Karpenko, Zh. Tekh. Fiz. **73** (1), 107 (2003) [Tech. Phys. **48**, 100 (2003)].
5. V. A. Sotskov, Pis'ma Zh. Tekh. Fiz. **30** (11), 38 (2004) [Tech. Phys. Lett. **30**, 461 (2004)].
6. V. A. Sotskov, Pis'ma Zh. Tekh. Fiz. **30** (12), 1 (2004) [Tech. Phys. Lett. **30**, 487 (2004)].
7. A. R. von Hippel, *Dielectric Materials and Applications* (Wiley, New York, 1954; Gosénergoizdat, Moscow, 1959).
8. V. P. Berzan *et al.*, in *Handbook of Electrical Capacitors* (Shtiintsa, Kishinev, 1982), p. 27 [in Russian].

Translated by M. Astrov

Technical Physics, Vol. 50, No. 6, 2005, pp. 809–811. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 6, 2005, pp. 136–139.
Original Russian Text Copyright © 2005 by Obraztsov, Volkov, Mikheev, Chakhovskoi, Roddatis, Garshev.

SHORT COMMUNICATIONS

Effect of Laser Radiation on the Morphology and Emissivity of Nanodimensional Carbon Films

A. N. Obraztsov*, A. P. Volkov*, G. M. Mikheev, A. G. Chakhovskoi***, V. V. Roddatis*, and A. V. Garshev***

** Moscow State University, Vorob'evy Gory, Moscow, 119992 Russia*
*** Institute of Applied Mechanics, Ural Division, Russian Academy of Sciences, Izhevsk, 426067 Russia*
e-mail: gmmikheev@udmnet.ru
**** University of California, Davis CA, 95616 USA*

Received October 13, 2004

Abstract—The effect of high-power pulsed laser radiation on the morphology and field emission of electrons from nanocrystalline graphite films is studied. The films are prepared by dc-discharge-activated CVD from a hydrogen–methane mixture. The material of the films partly evaporates under the action of the laser pulse. Threshold laser power densities at which the emissivity and geometry of the cathode surface change are determined. A direct correlation between the morphology and emission characteristics is shown to be absent. © *2005 Pleiades Publishing, Inc.*

Application of carbon nanotubes (CNTs) as field-emission cathodes has recently attracted considerable interest. Field (or cold) emission results from quantum electron tunneling from a cathode into a vacuum through a potential surface barrier. The emission efficiency depends mainly on the electric field strength at the cathode surface and on the work function of the cathode material [1, 2]. A large body of experimental data indicates that CNT cathodes provide a high emission current density at comparatively low cathode voltages [2]. This property of CNT cathodes is extremely promising for vacuum devices [3]. In [4], we showed that electrons could also be effectively emitted from nanodimensional graphite crystallites. The similarity of the behavior and basic parameters of CNT emitters and nanocrystalline graphite (NCG) emitters suggests that the same physical mechanisms govern the emission characteristics of nanocrystalline carbon materials [5]. However, this assumption needs further experimental and theoretical support.

The aim of this work is to experimentally find a correlation between the electron emission efficiency and morphology of NCG cathodes.

NCG films were prepared by dc-discharge-activated CVD of carbon on conducting substrates. The substrates were 25 × 25-mm silicon platelets. The gas mixture consisted of hydrogen and methane taken in the ratio 10 : 1 [4, 6]. The carbon films deposited were porous and consisted of graphite crystallites, each including 5–50 parallel atomic layers, which corresponds to a thickness of 2–20 nm, respectively. The idea that these crystallites have a graphite-like atomic structure is based on experimental data obtained by various methods, including Raman scattering, high-resolution electron microscopy, electron diffraction, etc. [4–6]. Using these methods, it has also been found that atomic layers in the graphite nanocrystallites are oriented mostly along the normal to the substrate surface. Other dimensions (i.e., the length and height) of the crystallites are typically about 1 μm. Figure 1a shows the typical image of such a NCG film taken with a LEO 1550 scanning electron microscope (SEM).

The NCG films were irradiated by linearly polarized pulsed radiation from a YAG : Nd laser ($\lambda = 1064$ nm). The pulse maximum energy was 50 mJ at a pulse duration of about 22 ns. The laser beam was directed normally to the film surface. To determine the threshold powers of the laser radiation that variously influence the emissivity and morphology of the film, we divided the sample surface into 5 × 5-mm squares (Fig. 2a). Then, the sample was scanned relative to the laser beam so that some of the squares were irradiated by a single pulse. In so doing, a given square was subjected to a pulse of given power density. The irradiation was carried out in air with the substrate kept at room temperature. After the irradiation, the emissivity and morphology of the sample were analyzed.

The electron emission characteristics were measured by the techniques described in [4, 5]. The sample was placed in an evacuated measuring cell (at a residual pressure of about 10^{-4} Pa) parallel to a glass plate coated by ITO and phosphor layers. The sample–glass plate distance was 0.5 mm. A voltage of 1 kV, which is sufficient to initiate electron emission, was applied between the cathode (sample) and anode (ITO layer between the phosphor and glass plate). The changes in the emission from different squares after the laser irradiation were estimated from the changes in the phos-

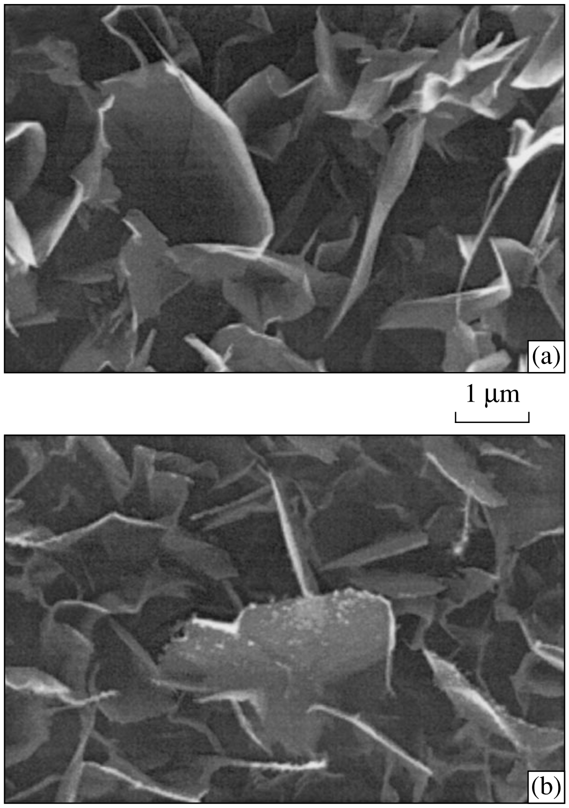


Fig. 1. SEM images of the NCG film surface (a) before and (b) after pulsed laser irradiation at a power density of 7 MW/cm^2.

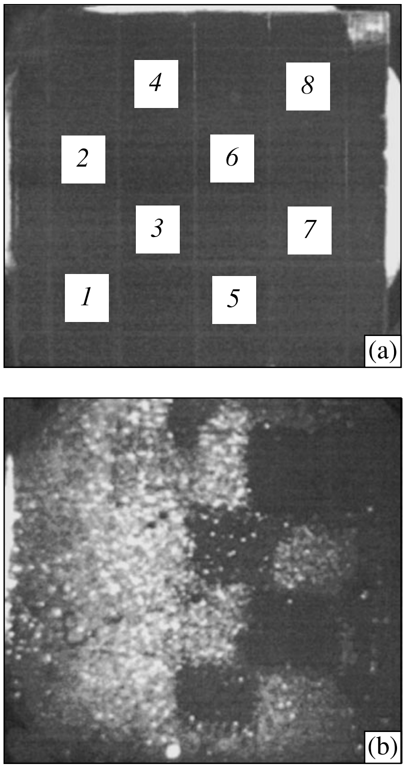


Fig. 2. (a) NCG film subdivided into squares on the 25 × 25-mm silicon platelet and (b) the corresponding cathodoluminescent anode plate (shown are the emission center distribution and efficiency in various regions on the surface of the sample shown in (a)). The radiation power density in these regions is (*1*) 1.6, (*2*) 2.9, (*3*) 3.3, (*4*) 6.9, (*5*) 9.9, (*6*) 14.4, (*7*) 21, and (*8*) 31.7 MW/cm^2. The unmarked squares were not irradiated.

phor glow intensity that were induced by incident electrons.

The laser radiation causes rapid heating of the film due to light absorption followed by evaporation of the film. In particular, we found that a radiation power density above 18 MW/cm^2 induces substantial changes in the initial morphology of the films. Their initially velvety black surfaces become light gray with a metallic luster, which is characteristic of normal graphite. SEM examination shows that irradiation at such a power density causes graphite nanocrystallites to be arranged along the substrate surface rather than normally to it. A similar change in the morphology of the film surface is observed when the films are subjected to a mechanical action. For comparison, Fig. 3 shows the SEM images of the NCG film surface after the laser irradiation (Fig. 3a) and after scratching by tweezers with Teflon tips (Fig. 3b). In both cases, the changes in the crystallite orientation were irreversible. Such behavior can be explained by anisotropic evaporation of graphite nanocrystallites under the action of linearly polarized laser radiation (see [7]). Due to this effect, only those crystallites (or their fragments) remain on the film surface subjected to a high-power laser pulse whose atomic layers are arranged normally to the plane of polarization. Without mechanical support on their sides, these crystallites "fall" on the substrate surface and are held in this position by van der Waals forces, as it takes place under the mechanical action. In both cases (after the high-power irradiation and mechanical action), the emission from the NCG films virtually ceases (an emission-initiating voltage increases by one or two orders of magnitude, and the densities of emission centers and emission current decrease by several orders).

The effect of pulsed laser irradiation on the emissivity of the films was also observed at much lower power densities. As follows from the anode image (Fig. 2b), the emissivity starts changing at a power density of about 7 MW/cm^2. Irradiation with such a density also causes morphological changes in the initial surface, as follows from the SEM images in Fig. 1b. These morphological changes are likely to be associated with the redeposition of the carbon material evaporated by a laser pulse. It is natural to assume that the redeposited carbon is amorphous. The characteristic dimensions of such amorphous carbon aggregates do not exceed several nanometers.

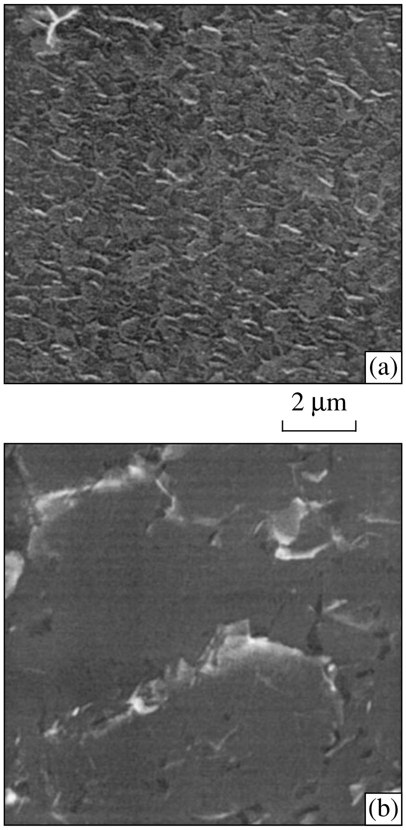


Fig. 3. SEM images taken of the NCG film surface subjected to (a) linearly polarized pulsed irradiation with a power density of 30 MW/cm^2 and (b) mechanical action.

Low-power laser irradiation does not have any noticeable effect on the emission characteristics. However, as the radiation power density increases, the emissivity decreases gradually and disappears at 20 MW/cm^2. The initial surface morphology also changes; specifically, the density of the amorphous carbon aggregates gradually grows. It should be noted that, up to the maximum radiation power density used in the experiments, the dimensions (height and thickness) of graphite nanocrystallites on the film surfaces remain virtually the same; only their amount changes. This fact indicates that the geometry of the NCG films, which is specified by the dimensions of individual emitters (crystallites), is not a decisive factor here. Indeed, the constancy of the geometrical dimensions of the crystallites implies the constancy of the field-enhancing geometrical factor, which is estimated from the ratio between the height of the emitter and the radius of its top. Moreover, extra conducting amorphous carbon aggregates present on the tops of the crystallites are also bound to contribute to the geometrical factor [8]. Similarly, the field-enhancing geometrical factor is expected to grow as the crystallite spacing increases and, accordingly, the electrostatic screening efficiency drops [2].

Thus, it seems natural to suppose that the emission from graphite nanocrystallites depends mainly on the structure and electronic properties of the material rather than on their geometrical characteristics. This conclusion fully agrees with our model of emission centers and the mechanism of low-voltage electron emission, which were proposed for nanocrystalline carbon materials [4, 5]. Results following from these model and mechanism suggest that the ordered graphite-like atomic structure of graphite nanocrystallites, carbon nanotubes, and other similar materials provides a good electric conductivity and mechanical stability of nanocrystalline carbon emitters. Significantly, these graphite-like nanocrystalline carbon materials allow for formation of surface clusters with modified electrical properties, the result of which is a decrease in the effective thickness of the surface potential barrier and a corresponding increase in the emissivity [9].

ACKNOWLEDGMENTS

We thank R.G. Zonov for assistance.

This work was supported by the INTAS (project no. 01-0254) and CRDF (grant no. RP2-2559-MO-03).

REFERENCES

1. R. Gomer, *Field Emission and Field Ionization* (Harvard Univ. Press, Cambridge, 1961; AIP, New York, 1993), p. 195.
2. J.-M. Bonard, H. Kind, Th. Stoeckli, and L.-O. Nilsson, Solid-State Electron. **45**, 893 (2000).
3. D. Temple, Mater. Sci. Eng., R. **24**, 185 (1999).
4. A. N. Obraztsov, I. Yu. Pavlovskiĭ, and A. P. Volkov, Zh. Tekh. Fiz. **71** (11), 89 (2001) [Tech. Phys. **46**, 1437 (2001)].
5. A. N. Obraztsov, A. P. Volkov, A. I. Boronin, and S. V. Koshcheev, Zh. Éksp. Teor. Fiz. **120**, 970 (2001) [JETP **93**, 846 (2001)].
6. A. A. Zolotukhin, A. N. Obraztsov, A. O. Ustinov, and A. P. Volkov, Zh. Éksp. Teor. Fiz. **124**, 1291 (2003) [JETP **97**, 1154 (2003)].
7. G. M. Mikheev, R. G. Zonov, A. N. Obratsov, and A. P. Volkov, Zh. Éksp. Teor. Fiz. **125**, 548 (2004) [JETP **98**, 483 (2004)].
8. E. I. Givargizov, V. V. Zhirnov, A. N. Stepanova, *et al.*, Appl. Surf. Sci. **87**–**88**, 24 (1995).
9. A. N. Obraztsov and A. A. Zakhidov, Diamond Relat. Mater. **13**, 1044 (2004).

Translated by K. Shakhlevich

Technical Physics, Vol. 50, No. 6, 2005, pp. 812–814. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 6, 2005, pp. 140–142.
Original Russian Text Copyright © 2005 by Vandakurov.

SHORT COMMUNICATIONS

New Approach to the Problem of Solar Wind Formation

Yu. V. Vandakurov

Ioffe Physicotechnical Institute, Russian Academy of Sciences,
Politekhnicheskaya ul. 26, St. Petersburg, 194021 Russia

Received October 21, 2004

Abstract—A hypothesis is discussed according to which the variation of the solar activity in 11-year intervals stems from the need for convective heat transfer regularization. It is not improbable that such regularization is disturbed during the Mounder minimum of solar activity, when heat transfer is most likely chaotic. The obstacles resulting from a sharp decrease in the density of the medium in going to the uppermost strata of the convective zone are considered. The hypothesis for solar wind formation in these strata seems to be the only valid one. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

In his recent work [1], the author touched upon the problem of finding the distribution of the rate of rotation of the solar convective zone. The rate of rotation was represented as a series in full orthogonal basis. The coefficients of the series were determined by minimizing the mean discrepancy between the rate predicted theoretically and the rate that can be calculated based on helioseismic data currently available [2]. It is essential that those calculations were carried out in terms of a hypothetical stationary model that generally disregarded the condition of symmetry of rotation about the equatorial plane.

Eventually, it turned out that, in terms of the standard model, which assumes symmetric rotation of the medium about the equator, agreement between theory and experiment is possible only in particular cases, for example, at relative latitudes (depths) of the convective zone roughly equal to 0.840 or 0.925, as well as at all depths if the absolute latitude is roughly equal to 42°. This value is very close to that (41°) at which torsional waves, which manifest themselves as periodic small variations of the rate of rotation (see, e.g., [3]), are not generated on the Sun. Since these waves are observed at any other latitude throughout the convective zone, the hypothesis that nonstationary processes on the Sun play an important part seems the most plausible. We believe that they are responsible for convective heat transfer regularization.

The fact that, in the nonmagnetic convective zone, any radial movement of a convection element generates an unbalanced azimuth force was discussed earlier [4]. If longitude-dependent (longitudinal) modes can be excited, this force can be counterbalanced; however, longitudinal modes can be excited only in the presence of an appropriate axisymmetric toroidal magnetic field. If we are dealing with the lowest mode of this field, its value at the bottom of the solar convective zone must be equal to 110 kG.

Thus, we assume that the rotation of the medium meets the condition of self-excitation of motions that are necessary for convective heat transfer regularization, with the rotation distribution being near-symmetric about the equator. Regular convective heat transfer may be disturbed with the transition to chaotic heat transfer. Chaotic processes are likely to take place during the Mounder minimum of the solar activity, when the solar activity level is known [5] to be greatly depressed.

The problem of determining the state of the rotating magnetized convective zone with allowance for small time variations of the rate of rotation and other parameters implies the solution of a set of complicated equations. However, in a number of limiting cases, the equations simplify. For example, at the upper boundary of the solar convective zone, the density of the medium changes drastically and the problem becomes nearly one-dimensional. This situation is considered in Sect. 2. Concise conclusions are drawn in the final section.

2. EQUILIBRIUM CONDITIONS AT THE UPPER BOUNDARY OF THE CONVECTIVE ZONE

In general form, the equilibrium equations represented as expansions in the full basis of orthogonal vector spherical harmonics were presented in [6]. For an arbitrary vector **U**, such a representation is written as

$$\mathbf{U} = \sum_{\lambda JM} U^{(\lambda)}_{JM}\mathbf{Y}^{(\lambda)}_{JM}, \qquad (1)$$

where $\mathbf{Y}^{(-1)}_{JM} = i_r Y_{JM}$, $\mathbf{Y}^{(+1)}_{JM} = (r/d_J)\nabla_\perp Y_{JM}$, and $\mathbf{Y}^{(0)}_{JM} = -i(r/d_J)\mathbf{i}_r \times \nabla_\perp Y_{JM}$. Here, $\nabla_\perp$ is the horizontal component of the gradient operator; Y_{JM} is the spherical function; $d_J = [J(J+1)]^{1/2}$; and $\lambda = 0, 1$, or -1. Subscripts J and M are integer numbers: $J \geq 0$ and $M = -J, -J+1, \ldots, J$. Summation in formula (1) is over λ, J, and M.

In our case with axial symmetry, $M = 0$ and, according to the results obtained in [4, 7], the azimuth equilibrium equations have the form

$$r\frac{\partial}{\partial t}v_{J0}^{(0)} = -\frac{1}{(2J+1)^{1/2}}\sum_{J_1J_2} C_{J_10J_20}^{J0}E_{J_1J_2}^{J}$$

$$\times\left[I_{J_2}Z_{JJ_1}^{J_2}\left(v_{J_20}^{(-1)}\frac{\partial}{\partial r}rv_{J_10}^{(0)} - \frac{1}{4\pi\rho}B_{J_20}^{(-1)}\frac{\partial}{\partial r}rB_{J_10}^{(0)}\right)\right. \qquad (2)$$

$$\left.-J_1(J_1+1)Z_{JJ_2}^{J_1}\left(v_{J_10}^{(0)}v_{J_20}^{(+1)} - \frac{1}{4\pi\rho}B_{J_10}^{(0)}B_{J_20}^{(+1)}\right)\right],$$

$$J = 1, 3, \ldots;$$

$$r\frac{\partial}{\partial t}B_{J0}^{(0)} = -\frac{1}{(2J+1)^{1/2}}\sum_{J_1J_2} C_{J_10J_20}^{J0}E_{J_1J_2}^{J}$$

$$\times\left[\frac{\partial}{\partial r}r(I_{J_1}Z_{JJ_2}^{J_1}v_{J_10}^{(-1)}B_{J_20}^{(0)} - I_{J_2}Z_{JJ_1}^{J_2}v_{J_10}^{(0)}B_{J_20}^{(-1)})\right. \qquad (3)$$

$$\left.+J(J+1)Z_{J_1J_2}^{J}(v_{J_10}^{(+1)}B_{J_20}^{(0)} - v_{J_10}^{(0)}B_{J_20}^{(+1)})\right],$$

$$J = 2, 4, \ldots.$$

The continuity condition and the equation for the magnetic induction flux are written as

$$v_{J0}^{(+1)} = \frac{1}{r\rho[J(J+1)]^{1/2}}\frac{\partial}{\partial r}r^2\rho v_{J0}^{(-1)}, \qquad (4)$$

$$B_{J0}^{(+1)} = \frac{1}{r\rho[J(J+1)]^{1/2}}\frac{\partial}{\partial r}r^2B_{J0}^{(-1)}, \qquad (5)$$

where

$$E_{J_1J_2}^{J} = \left[\frac{(2J_1+1)(2J_2+1)}{4\pi}\right]^{(1/2)}\frac{1}{2d_Jd_{J_1}d_{J_2}}C_{J_10J_20}^{J0}, \qquad (6)$$

$C_{J_1M_1J_2M_2}^{JM}$ is the Clebsch–Gordan coefficient, $Z_{ab}^{c} = [a(a+1) + b(b+1) - c(c+1)]$, and $I_a = [a(a+1)]^{1/2}$.

Quantities $v_{J_k0}^{(J0)}$ and $B_{J_k0}^{(J0)}$ specify the velocity and magnetic field distributions. Note that possible time variations of the meridian circulation velocity are disregarded in continuity condition (4). In the state symmetric about the equator, coefficients J in Eqs. (2) and (3) are odd and even, respectively. In the right-hand sides of these equations, the sum $J + J_1 + J_2$ is even.

Earlier [1], we performed numerical calculations to see whether it is possible to derive a stationary nonmagnetic model. As was mentioned above, the stationary conditions and the conditions of symmetry of rotation about the equator are consistent with the helioseismic data by Schou *et al.* [2] in exceptional cases. In general, a model of the solar convective zone must include transient processes. Below, we will consider the upper region of this zone, $r/R \geq 0.99$ (where $R = R_{\odot}$), which was not studied in [1]. At the very boundary ($r/R = 0.99$), provided that the rotation is symmetric about the equator, the velocity ratios at latitudes of 0, 30°, 45°, and 60° equal, respectively, 1.02, 0.974, 0.903, and 0.810, which are close to the data reported in [2]. In this work, we refine the results of calculations that were made using the figure in [1].

In the near-surface region, $r/R \geq 0.99$, the density of the medium drops sharply with increasing r. Essentially, matter effluxion due to the rotation of the medium takes place in this region. In a crude approximation, only the radial dependence of the equilibrium quantities is taken into account, so that the term containing $v_{00}^{(-1)}$ will dominate in Eq. (2). Then, we have $v_{00}^{(+1)} = 0$ and

$$r^2\rho v_{00}^{(-1)} = \text{const}, \qquad (7)$$

$$r\frac{\partial}{\partial t}v_{J0}^{(0)} + \frac{1}{(4\pi)^{1/2}}v_{00}^{(-1)}\frac{\partial}{\partial r}rv_{J0}^{(0)} = 0, \qquad (8)$$

$$r\frac{\partial}{\partial t}\left(\frac{1}{r\rho}B_{J0}^{(0)}\right) + \frac{1}{(4\pi)^2}v_{00}^{(-1)}\frac{\partial}{\partial r}\left[\frac{1}{r\rho}B_{J0}^{(0)}\right] = 0. \qquad (9)$$

Equation (8) also follows from one-dimensional equilibrium equations.

Guenther *et. al.* [8, Table 3B], analyzing the uppermost strata of the solar convective zone, found that the relative density gradient equals 4000 and decreases twice when the relative radius decreases by as little as 0.1%. Because of such a drastic drop in the density with increasing r, variations in the rate of rotation in the uppermost strata are expected to be considerable and, therefore, extremely difficult to describe.

According to Eq. (8), the rate of rotation in these strata might be roughly constant if the condition $rv_{J0}^{(0)} \approx$ const were satisfied. However, detailed helioseismic investigations [9–11] performed at altitudes of up to 20000 km show that a complicated radius- and time-dependent asymmetric-about-equator state of rotation sets in at these altitudes. Clearly, such a complex equilibrium state cannot be described by simple equation (8). Any solutions to Eqs. (7) and (8) will be in considerable disagreement with the observations.

In light of the aforesaid, the hypothesis that the uppermost region of the medium stratifies, causing fluxes moving with different velocities, seems to be the most plausible. To put it differently, it is assumed that Eq. (8) describes the mean rate of rotation of the medium as a whole but bunches of fast charged particles (ions) whose mean radial velocity is specified by the heat transfer equation are always present in these strata. The energy needed for acceleration of these fast ions may be essential. The helioseismic data [9–11] also indicate the formation of fluxes in the medium that are asymmetric about the equatorial plane. Therein lies a fundamental difference between the stellar convection and that in the laboratory conditions.

The above hypothesis might explain both heating of the solar chromosphere and the appearance of fast ions in the solar corona. These phenomena were discussed, e.g., in [5, 12, 13] on the assumption that a heat source is located at high altitudes. Our calculations testify that the fast ions are generated in the uppermost strata of the convective zone.

It is worth noting that the effluxion of fast ions may slow down the rate of rotation of the Sun. It is common knowledge that stars like the Sun rotate more slowly than predicted from the simple theory (this problem was discussed, e.g., in [14]). Our calculations count in favor of this effect.

3. CONCLUSIONS

Leaving aside the problem of the solar tachocline, one can conventionally divide solar activity phenomena taking place in the magnetized differentially rotating zone into three groups. The first one covers those where instability is self-excited in the presence of both an axisymmetric toroidal magnetic field and meridian circulation of matter. The typical time of instability development depends on the radial gradient of the angular rate of rotation [15, 16]. Instability of this type seems to be the reason for emerging magnetic fields appearing on the Sun.

The other group of solar activity phenomena is related to the need for convective heat transfer regularization. As was already mentioned, the radial displacement of convection elements (which move along the axis of rotation) is only possible in the presence of longitude-dependent (longitudinal) modes, which are neutrally stable if an appropriate toroidal magnetic field (on the order of 10^5 G) is present in the medium. Torsional waves sustaining a necessary distribution of the mean rate of rotation of the medium are also of great importance. Theoretical calculations aimed at predicting the distribution of the rate of rotation are on the agenda.

The specific situation occurs in the uppermost strata of the convective zone (third group). The difficulty here is that the variation of the rate of rotation that follows from the heat balance condition turns out to be unacceptably high. This bottleneck can be remedied by taking into consideration the formation of fast ion beams. If such beams do form, problems such as heating of the chromosphere, solar wind formation, and rapid slowing down of the rate of rotation of stars with a convective sheath can be tackled.

REFERENCES

1. Yu. V. Vandakurov, Zh. Tekh. Fiz. **74** (10), 35 (2004) [Tech. Phys. **49**, 1280 (2004)].
2. J. Schou, R. Howe, S. Basu, *et al.*, Astrophys. J. **567**, 1234 (2002).
3. S. V. Vorontsov, J. Christensen-Dalsgaard, J. Schou, *et al.*, Science **296**, 101 (2002).
4. Yu. V. Vandakurov, Zh. Tekh. Fiz. **73** (3), 23 (2003) [Tech. Phys. **48**, 298 (2003)].
5. Yu. I. Vitinskiĭ, *Solar Activity* (Nauka, Moscow, 1983) [in Russian].
6. D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskii, *Quantum Theory of Angular Momentum* (Nauka, Leningrad, 1975; World Sci., Singapore, 1988).
7. Yu. V. Vandakurov, Astron. Zh. **76**, 29 (1999) [Astron. Rep. **43**, 24 (1999)].
8. D. B. Guenther, P. Demarque, Y.-C. Kim, and M. H. Pinsonneault, Astrophys. J. **387**, 372 (1992).
9. S. Basu, H. M. Antia, and S. C. Tripathy, Astrophys. J. **512**, 458 (1999).
10. D. A. Haber, B. W. Hindman, J. Toomre, *et al.*, Astrophys. J. **570**, 855 (2002).
11. J. Zhao and A. G. Kosovichev, Astrophys. J. **603**, 776 (2004).
12. E. N. Parker, *Interplanetary Dynamical Processes* (Interscience, New York, 1963; Mir, Moscow, 1965).
13. J. C. Brant, *Introduction to the Solar Wind* (Freeman, San Francisco, 1970; Mir, Moscow, 1973).
14. J. L. Tassoul, *Theory of Rotating Stars* (Princeton Univ. Press, Princeton, 1978; Mir, Moscow, 1982).
15. Yu. V. Vandakurov, Pis'ma Astron. Zh. **27**, 700 (2001) [Astron. Lett. **27**, 596 (2001)].
16. Yu. V. Vandakurov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **44**, 735 (2001).

Translated by V. Isaakyan