Technical Physics, Vol. 48, No. 4, 2003, pp. 377–384. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 4, 2003, pp. 1–8.

THEORETICAL AND MATHEMATICAL PHYSICS

Application of the Method of Nonstationary Waveguide Equations to Simulating Pulsed Processes in Irregular Transmission Lines

V. N. Mikhaĭlov, V. I. Koroza, and M. N. Golikov

Institute for Strategic Stability, Moscow, 115304 Russia

e-mail: iss@ript.in.ru

Received July 2, 2002

Abstract—The potentialities of the model of nonstationary waveguide equations for describing pulsed processes in irregular transmission lines are studied using planar lines as an example. The problem of high-accuracy controllable-error numerical simulation is discussed. Typical examples of simulating ultra-wide-band electromagnetic pulses with an initial TEM structure in terms of time-domain representation are presented with emphasis on the interaction of the pulses with irregularities, including their transformation into longitudinal waves. Both lumped and distributed irregularities are addressed: deep corrugations that cover 90% of the transmission line's aperture (distributed irregularities) and these corrugations in combination with sharp kinks at the boundary surfaces and permittivity steps at the boundaries of the dielectric filling (lumped irregularities). It is shown that a relative rms error involved in the calculated field intensity of no higher than 10^{-4} is easy to achieve.

INTRODUCTION

The growing interest in the electrodynamics of nonstationary and pulsed processes in irregular transmission lines and, in particular, in the propagation of ultra-wide-band (UWB) electromagnetic pulses (EMPs) in them [1] poses new problems in waveguide electrodynamics. The models developed for the numerical simulation of monochromatic (i.e., ~exp(–*i*ω*t*)) processes are inapplicable to UWB EMPs for the following reason. The numerical inversion of Fourier integrals of monochromatic UWB EMP field components (in irregular waveguides, they are also found numerically) encounters insurmountable difficulties in attaining a sufficient and controllable accuracy, because the frequency spectra are very wide. Due to this circumstance, one has to give up frequency representations and develop new more general simulation methods that provide a higher and controllable accuracy.

At the same time, the propagation of nanosecond and subnanosecond pulses in devices that contain irregular transmission lines is accompanied by effects that can be explained neither quantitatively nor qualitatively within the conventional theory of transmission lines, because the theory is approximate and not wholly adequate for a rigorous electrodynamic approach. In particular, according to the generally accepted single-mode model, a TEM pulse in a matched inhomogeneous line (i.e., when the wave impedance is constant over the line length) usually with a low loss per unit length can only slowly change its waveform. In a lossless line, the waveform should remain unchanged. However, experiments [2] and simulations [4] based on the more rigorous model of coupled strings [3] have shown significant qualitative deviations from these predictions. The reason for this discrepancy is that the single-mode model of telegraph equations disregards mode mutual transformation and, in particular, the transformation of the dispersion-free TEM mode into dispersive modes (and vice versa) when the cross section varies continuously. The variational method [3] applied below is more adequate for a rigorous electrodynamic approach and allows for the real processes.

Note also that a combination of the analytical methods that represent transient fields in irregular waveguides as the superposition of monochromatic fields and methods for numerically calculating these fields are associated with significant difficulties where short or steep-edge pulses are concerned. The practical implementation of this approach produces unpredictable errors when calculating the inverse Fourier transform, because the frequency spectra are very wide. This circumstance is another important reason for the application of the method mentioned above [3], which does not use expansions in frequency spectra.

Although interest in transient electromagnetic processes in waveguides is growing, the number of theoretical works on this subject is very limited. Monograph [5] considers transient waves in waveguides. Although the method using the separation of variables [5] is free of the above disadvantages of frequency-domain analysis, it is applicable only to irregular waveguides of certain special shapes.

Below, we consider the propagation of a pulse of a finite initial shape in planar irregular transmission lines

1063-7842/03/4804-0377$24.00

and study the feasibility of providing high-accuracy controlled-error numerical results by the variational method of nonstationary waveguide equations.

1. STATEMENT OF THE PROBLEM AND NONSTATIONARY WAVEGUIDE EQUATIONS

Let the y axis of the Cartesian coordinates be directed transversely to the waveguide under study, i.e., an irregular layer sandwiched in two perfectly conducting surfaces. These surfaces are defined by the equations $y = -a_1(z)$ and $y = a_2(z)$. Here, z is the longitudinal coordinate, which specifies the direction of wave propagation in the layer, and $a_1(z)$ and $a_2(z)$ are continuous functions of z such that $a_1(z) \geq 0$ and $a_2(z) \geq 0$. These functions do not turn to zero simultaneously; hence, the layer width $\Delta(z) = a_1(z) + a_2(z) > 0$ for any z. An example of the profile (normal to the plane $X0Z$ and passing through the Z axis) is shown in Fig. 1. The profile here is a planar layer confined by the lines $y = a_1(z)$ and $y = -a_2(z)$. We assume that the boundaries are regular at $z < 0$: $a_1(z) = a_{10} = \text{const}$ and $a_2(z) = a_{20} = \text{const}$; hence, the width of the half-layer $z < 0$ is $\Delta(z) = a_{10} + a_{20} = \Delta_0 = \text{const}$. We also assume that the electrophysical parameters of the medium filling the line are independent of z and that the layer profile, electrical parameters of the medium, and field components are independent of the x coordinate.

In the calculations that follow, we use dimensionless variables and parameters. To pass to the dimensionless variables, we normalize all quantities that have dimensions of length (spatial coordinates, the functions $a_1(z)$ and $a_2(z)$, etc.) to a certain appropriate linear scale L and normalize the time t by dividing by L/c (c is the velocity of light in free space). Throughout the paper, $L = \Delta_0$. To designate all the dimensionless variables, we use the same notation for convenience. In particular, as a result of the normalization, we have $\Delta_0 = 1$ and $c = 1$.

Let a finite TEM pulse whose waveform is limited in space and time propagate from the regular half-layer ($z < 0$) in the positive z direction. As the pulse reaches the irregular part of the layer because of the interaction with the irregularities, it partially reflects, retaining its TEM structure, and partially transforms into TM modes propagating in both forward and backward directions.

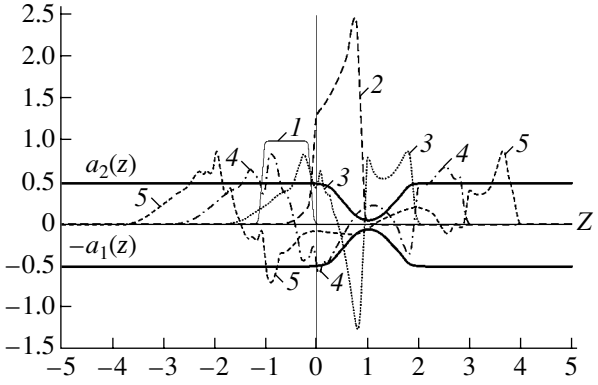

Fig. 1.

Following the variational method [3] to derive nonstationary waveguide equations for our irregular planar layer in view of the above assumptions, we use the bilinear functional

$$J(\mathbf{H}, \mathbf{H}_0) = \int_\Omega \{\varepsilon^{-1}([\nabla \times \mathbf{H}] \cdot [\nabla \times \mathbf{H}_0]) - \mu(\partial \mathbf{H}/\partial t \cdot \partial \mathbf{H}_0/\partial t)\} d^4\Omega, \quad (1)$$

which depends on the vector magnetic and auxiliary fields $\mathbf{H}(y, z, t)$ and $\mathbf{H}_0(y, z, t)$. Here, $\varepsilon = \varepsilon(y, z)$ and $\mu = \mu(y, z)$ are the relative permittivity and permeability of the medium. The domain of integration Ω in (1) is the interior of a fragment of a four-dimensional cylinder whose generatrix is the t axis ($0 \leq t \leq T$) and the directrix is the boundary of the interval $z_1 \leq z \leq z_2$ of the strip line of unit width in the x direction. The values T, z_1, and z_2 will be specified below. Because the integrand in (1) is independent of x, integration over x is reduced to multiplication by unity. As a result, the four-dimensional integral becomes three-dimensional.

The stationarity conditions

$$\delta_{\mathbf{H}_0} J(\mathbf{H} \cdot \mathbf{H}_0) = 0 \quad (2)$$

with $\delta\mathbf{H}_0(y, z, t_0) = \delta\mathbf{H}_0(y, z, t_0 + T) = \delta\mathbf{H}_0(y, z_{1,2}, t) = \mathbf{0}$ are equivalent to the Maxwell equations inside Ω with the boundary conditions $[(\nabla \times \mathbf{H}) \times \mathbf{n}] = \mathbf{0}$ on metal surfaces. For finite pulses, this corresponds to a perfectly conducting metal.

In our case, the magnetic field intensity has only one component $H(y, z, t)$ in the x direction. Let us expand this component in modes of reference waveguides:

$$H(y, z, t) = \sum e_j(y, z) f_j(z, t), \quad (3)$$

where the amplitudes $f_j(z, t)$ ($j = 1, 2, 3, \ldots$) are unknown. As the basis functions $\{e_j(y, z)\}$ in expansion (3), we use the field distributions in the reference waveguides for the planar case studied:

$$e_j(y, z) = \cos[\pi(j-1)(a_1(z) + y)/\Delta(z)]. \quad (4)$$

The amplitude $f_1(z, t)$ defines the magnetic field of the TEM mode; the remaining amplitudes $f_i(z, t)$ ($j \geq 2$), the magnetic fields of the $TM_{0(j-1)}$ modes (E waves).

In matrix form, a system of waveguide equations for transient and pulsed processes in the irregular line under study is similar to system (7) in [3]:

$$\partial/\partial z[G(z)\partial \mathbf{f}/\partial z + Q(z)\mathbf{f}] - Q^\tau(z)\partial \mathbf{f}/\partial z - P(z)\mathbf{f} - T(z)\partial^2 \mathbf{f}/\partial t^2 = \mathbf{0}. \quad (5)$$

The unknown column vector $\mathbf{f}(z, t)$ is formed by its coordinates $f_j(z, t)$, and its dimension N is equal to the number of terms involved in sum (1). The coefficients in system (5) are the $[N \times N]$-dimensional matrix functions $G(z)$, $Q(z)$, $P(z)$, and $T(z)$ with elements given by

$$G_{ns}(z) = \int_{-a_1(z)}^{a_2(z)} \varepsilon^{-1} e_s e_n dy,$$

$$T_{ns}(z) = \int_{-a_1(z)}^{a_2(z)} \mu e_s e_n dy,$$

$$Q_{ns}(z) = \int_{-a_1(z)}^{a_2(z)} \varepsilon^{-1} (e_s)_z e_n dy, \quad (6)$$

$$P_{ns}(z) = \int_{-a_1(z)}^{a_2(z)} \varepsilon^{-1} \{(e_s)_z (e_n)_z + (e_s)_y (e_n)_y\} dy.$$

In the above formulas, $Q^{\tau}(z)$ is the transpose of $Q(z)$ and the subscripts z and y mean differentiation with respect to these variables.

Note two features of system (5) that follow from the properties of functional (1), which was used to derive this system, and from conditions (2).

(1) As N grows, sums (3) provide, in the limit, the perfect conductivity condition on metal boundaries for finite pulses after substituting exact solutions (5) into (3).

(2) Because model (5) is adequate for the rigorous electrodynamic approach, the accuracy of numerical results obtained with this model is the same as the accuracy of its numerical realization. This accuracy can be checked by performing (i) a numerical experiment (by varying the lengths *hz* and *ht* of the intervals used in the finite-difference approximation of the derivatives and also the number N of terms involved in (3)) and (ii) additional energy-balance calculations.

When ε and μ are independent of the transverse coordinate y, the factors $\varepsilon^{-1} = \varepsilon^{-1}(z)$ and $\mu = \mu(z)$ in expressions (6) can be factored out from the integral sign:

$$G_{ns}(z) = \varepsilon^{-1}(z) g_{ns}(z), \quad T_{ns}(z) = \mu(z) t_{ns}(z),$$
$$Q_{ns}(z) = \varepsilon^{-1}(z) q_{ns}(z), \quad P_{ns}(z) = \varepsilon^{-1}(z) p_{ns}(z). \quad (6')$$

The integrals $g_{ns}(z)$, $t_{ns}(z)$, $q_{ns}(z)$, and $p_{ns}(z)$ can be taken analytically using expressions (4) (see Appendix).

We will study system (5) subject to the initial conditions

$$f_1(z, t)|_{t=0} = \varphi(z), \quad f_1'(z, t)_t|_{t=0} = -(\varepsilon\mu)^{-1/2} \varphi'(z),$$
$$f_j(z, t)|_{t=0} = 0, \quad f_j'(z, t)_t|_{t=0} = 0, \quad j = 2, 3, 4, \ldots, \quad (7)$$

which correspond to a given initial waveform $\varphi(z)$ of the TEM pulse. The function $\varphi(z)$ is assumed to be finite and lie in the interval $\alpha < z < \beta$ in the regular part of the layer (i.e., $\alpha < \beta < 0$). Since we are always interested in a certain finite range Λ of dimensionless z and t ($\Lambda = \{z_1 \le z \le z_2, 0 \le t \le T\}$, where $z_1 < \alpha$, $z_2 > 0$ and $T \le \min\{|z_1 + \alpha|, |z_2 - \beta|\}$), formulas (7) should be supplemented by auxiliary boundary conditions

$$\mathbf{f}(z_1, t) = \mathbf{f}(z_2, t) = 0, \quad 0 \le t \le T. \quad (8)$$

With conditions (7) and (8), system (5) is closed and can be used to calculate $\mathbf{f}(z, t)$ on a discrete mesh in the rectangle Λ.

Below, we present an expression for the pulse energy per unit width of the layer in the x direction, $W(t)$:

$$W(t) = \int_{z_1}^{z_2} dz \int_{-a_1(z)}^{a_2(z)} dy \Bigg\{ \mu \Big(\sum e_j f_j\Big)^2 + \varepsilon^{-1} \Bigg[\Bigg(\int_0^t \sum e_j f_j d\eta\Bigg)_y\Bigg]^2 + \varepsilon^{-1} \Bigg[(\varepsilon\mu)^{1/2} e_1 \varphi(z) - \int_0^t \Big(\sum e_j f_j\Big)_z d\eta\Bigg]^2 \Bigg\}. \quad (9)$$

The exact solution to the problem corresponds to $W(t) = W(0) = W_0 = \text{const}$; therefore, the spread of $W(t)$ will be used about W_0 to estimate the accuracy of the numerical solution.

Below are examples of numerical solutions for a continuously distributed smooth irregularity (Section 2) and for combinations of this irregularity with a number of typical lumped obstacles that produce discontinuities in matrix functions (6) and (6') (sharp kinks at the boundary surfaces and steps of $\varepsilon(z)$) (Section 3).

In our calculations, we specify the pulse waveform (the function $\varphi(z)$ in conditions (7)) by the following trapezoidal functions with possibly smooth edges: (a) a geometrical trapezoid of unit height with the base on the z axis and linear edges, where the length of the median of the trapezoid is taken as the pulse width z_p and the projections of the corresponding sides onto the z axis are taken as the durations z_+ and z_- of the leading and trailing edges; (b) a trapezoid-like figure obtained by replacing the sides of the trapezoid in (a) by quarter-wavelength sections of cubed sinusoids; and (c) a trapezoid-like figure obtained by replacing the sides of the trapezoid in (a) with sections of fifth-degree polynomials uniquely defined from the requirement that they, together with their first and second derivatives, pass continuously into the z axis and upper plateau at the end points.

Trapezoid-like pulses (b) and (c), which are obtained from trapezoid (a), will be characterized by

the parameters z_p, z_+, and z_- equal to their values for the original trapezoid. Note that, at $\varepsilon = \mu = 1$, for the durations of the pulse (t_p) and of its leading (t_+) and trailing (t_-) edges, we have $t_p = z_p$, $t_+ = z_-$, and $t_- = z_+$.

To solve the problem stated by (3), (5), and (6) numerically, we apply a uniform mesh with steps *hz* and *ht* along the coordinates *z* and *t*, respectively, to the rectangle Λ. The explicit three-layer scheme gives a chance to calculate $\mathbf{f}$ on an $(n+1)$th layer in *t* from its values on *n*th and $(n-1)$th layers. The standard approximation

$$(\partial^2\mathbf{f}/\partial t^2)_{i,j} = (\mathbf{f}_{i,j+1} - 2\mathbf{f}_{i,j} + \mathbf{f}_{i,j-1})/(ht)^2,$$
$$\mathbf{f}_{i,j} = \mathbf{f}(z_i, t_j)$$

for the derivative $\mathbf{f}_{tt}$ at the point (z_i, t_j) allows us to find from (5) the basic calculational relationship

$$\mathbf{f}_{i,j+1} = 2\mathbf{f}_{i,j} - \mathbf{f}_{i,j-1} + (ht)^2\{T^{-1}\partial/\partial z[G\partial\mathbf{f}/\partial z + Q\mathbf{f}] - T^{-1}Q^{\tau}\partial\mathbf{f}/\partial z - T^{-1}P\mathbf{f}\}_{i,j},$$

in which we use the formulas for five-point and seven-point differentiation to calculate, respectively, the first and second derivatives with respect to *z*.

Unless specified otherwise in each particular case, the numerical results presented below refer to $N = 7$, mesh step sizes $hz = 0.01$ and $ht = 0.004$, and pulse parameters $z_p = 1$ and $z_+ = z_- = 0.2$.

2. PROPAGATION OF THE PULSE THROUGH A CORRUGATION

Consider a transmission line with a distributed smooth irregularity in the form of a symmetric corrugation

$$a_1(z) = a_2(z) = \begin{cases} 0.5 & \text{at} \quad z < 0 \quad \text{and} \quad z > 2 \\ 0.5 - a_0\sin^2(\pi z/2) & \text{at} \quad 0 \le z \le 2 \end{cases} \qquad (10)$$

of height a_0 ($0 < a_0 < 0.5$). Then, $\Delta_0 = 1$ and the waveguide aperture in the narrowest cross section (at $z = 1$) is $\Delta_{min} = (1 - 2a_0)$. In our calculations, we put $\varepsilon = \mu = 1$.

Figure 1 shows the profile of the transmission line containing distributed irregularity (10) with $\Delta_{min} = 0.1$ ($a_0 = 0.45$) and plots the magnetic field intensity *H* in the midplane ($y = 0$) versus *z* at $t = 0$, 1, 2, 3, and 4 (curves *1–5*, respectively). These curves illustrate the waveform variation and the reflection picture for a pulse that initially has the TEM structure and waveform (b) described above (curve *1*). As follows from curves *4* and *5*, which refer to the time instants when the front part of the pulse again enters the regular region ($z > 2$) of the transmission line having passed through the corrugation, the deformation of the front part continues. This is because by then the nondispersed TEM mode, which existed alone at $t = 0$, coexists with dispersed TM modes, which emerged in this region later.

Figure 2 plots the amplitudes $f_1(z, t)$ of the TEM component of the pulse versus *z* for the same values of *t* and for the same initial waveform (b) as in Fig. 1. The curve numbers in Fig. 2 correspond to the same values of *t* as in Fig. 1. As could be expected, after the front part of the pulse has entered the second regular region of the transmission line (curves *4* and *5* at $z > 2$), the deformation of the TEM component of the front part stops.

For the same initial pulse, Fig. 3 shows the distributions of the E_{02} mode (curve *1*), which is absent at $t = 0$ ($f_3(z, 0) \equiv 0$) and arises later ($t = 2$ (curve *1*) and 4 (curve *2*)) as the result of transformation of the TEM mode as the pulse passes through the irregular region of the transmission line. Note that, unlike the situation with an asymmetrical aperture [6], $f_2(z, t) \equiv 0$ in our case, because the geometry is symmetric about the plane $y = 0$ and the E_{01} mode is not generated.

Figure 4 plots *H* versus time in the midpoint of the narrowest cross section ($z = 1$, $y = 0$) at $\Delta_{min} = 0.10$ (curve *1*), 0.25 (curve *2*), and 0.50 (curve *3*) for the same initial waveform and parameters of the pulse. For

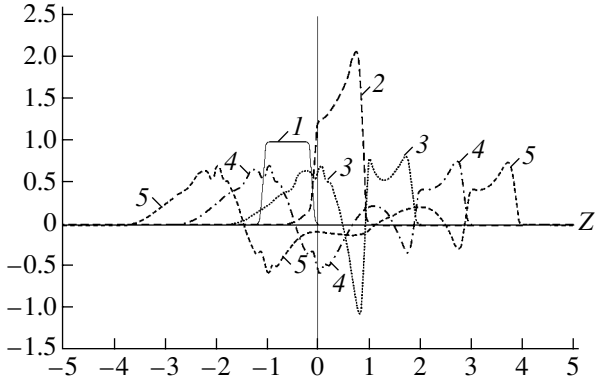

Fig. 2.

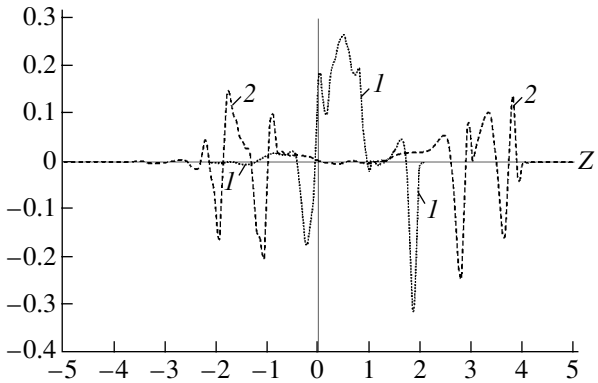

Fig. 3.

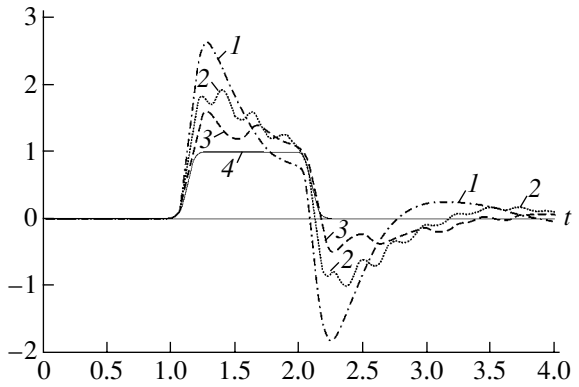

Fig. 4.

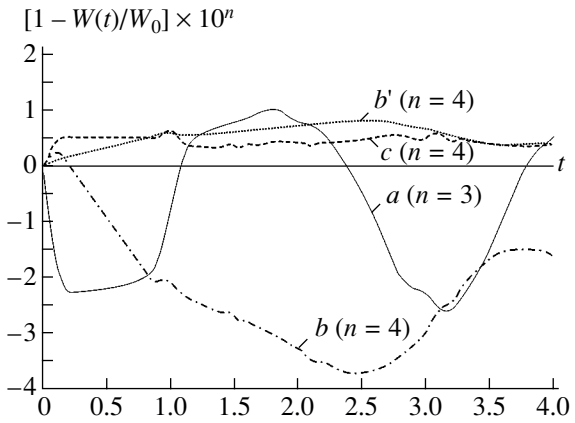

Fig. 5.

comparison, the dependence $H(t)$ at the same point is also shown in the absence of corrugation ($a_0 = 0$ in (10)), when the transmission line becomes regular (curve *4*).

The accuracy of the calculations can be estimated by comparing the time dependence of $[1 - W(t)/W_0]$ in Fig. 5 (where $W_0 = W(0)$) with the data summarized in Table 1. Curve *a* ($n = 3$ on the vertical axis) refers to the initial pulse in the form of a trapezoid (case (a)). Curves *b* ($n = 4$) and *c* ($n = 4$) refer to the initial pulses in the form (b) and (c), respectively. Figure 5 also demonstrates results for the pulse with the initial waveform (b) (curve *b*', $n = 4$) obtained on a $[2 \times 2]$ times finer mesh ($hz = 0.005$, $ht = 0.002$). Curves *b* and *b*' show that the use of the finer mesh decreases the error by a factor of 5. It is also seen that, as the initial waveform changes from (a) to (b) and then from (b) to (c) (with the pulse parameters z_p, z_+, and z_- and mesh steps *hz* and *ht* remaining unchanged), the accuracy of the calculations improves each time by approximately one order of magnitude, because the quality of the differential properties of the finite function $\varphi(z)$ in initial conditions (5) improves in the same order. For instance, when form (a) is changed into (b), the function $\varphi(z)$ that is everywhere continuous but has discontinuities (at four vertices of the trapezoid) is replaced with the function $\varphi(z)$ that is everywhere smooth but has a discontinuous (at two points of the junction with the upper plateau) first derivative. In turn, this function changes into $\varphi(z)$ with a continuous second derivative when form (b) is changed into (c).

Table 1 illustrates the error arising when the infinite sum in formula (3) is replaced by its first $N = 7$ terms. It lists relative partial energies W_j/W_0 of the TEM mode ($j = 1$) and also the E_{01} ($j = 2$), …, E_{014} ($j = 15$) modes. The energies are calculated at $t = 7$ for the pulses with initial waveforms (a), (b), and (c) with $N = 15$. For odd *j*, the calculated values of W_j/W_0 are no higher than $\sim 10^{-24}$ (rather than exact zeroes); therefore, the corresponding rows are omitted in the table (except for the row $j = 2$). The last row of Table 1 gives estimates for the relative residual energy W^r/W_0 of all terms neglected in (3) ($N = 7$). The residual energy W^r is calculated by formula (9) with infinite sum (3) replaced by a finite sum ($8 \le j \le 15$). Note that the calculation of W^r as the sum of its partial components W_j might prove to be incorrect, since modes of the reference waveguides for the irregular lines are not orthogonal.

From Fig. 5 and Table 1, it follows that, when the relative rms error of the field intensity is specified at a level of no higher than $\sim 1/2\max\{[1 - W(t)/W_0], W^r/W_0\} \sim 10^{-4}$ (as in our example), the choice $N = 7$ is justified, because it meets the condition $\max[1 - W(t)/W_0] \sim W^r/W_0$. Note that, for shallower corrugations, the same calculation accuracy can be reached at smaller *N*.

3. OTHER EXAMPLES

Consider combinations of distributed and lumped irregularities by taking as an example a nonsmooth junction between a regular planar line and a corrugation

Table 1

j	(a)	(b)	(c)
1	9.8729×10^{-1}	9.8132×10^{-1}	9.8067×10^{-1}
2	1.1780×10^{-24}	1.4727×10^{-24}	1.4284×10^{-24}
3	1.2406×10^{-2}	1.6568×10^{-2}	1.6901×10^{-2}
5	9.2914×10^{-4}	2.5102×10^{-3}	2.5583×10^{-3}
7	1.0327×10^{-4}	4.0804×10^{-4}	4.3313×10^{-4}
9	5.0128×10^{-5}	7.1098×10^{-5}	7.9569×10^{-5}
11	2.6025×10^{-5}	1.5548×10^{-5}	1.6966×10^{-5}
13	1.1362×10^{-5}	6.4971×10^{-6}	6.4337×10^{-6}
15	6.3585×10^{-6}	4.5141×10^{-6}	4.5382×10^{-6}
W^r/W_0	9.6848×10^{-5}	1.0049×10^{-4}	1.1042×10^{-4}

Table 2

j	(5, 0)	(5, 7)	(5, 11)	(0, 11)
1	9.7905×10^{-1}	9.7610×10^{-1}	9.7588×10^{-1}	9.7588×10^{-1}
2	2.8608×10^{-24}	2.9541×10^{-24}	3.0585×10^{-24}	3.0925×10^{-24}
3	2.1332×10^{-2}	2.1753×10^{-2}	2.1884×10^{-2}	2.1891×10^{-2}
5	3.2702×10^{-3}	3.3279×10^{-3}	3.3750×10^{-3}	3.3746×10^{-3}
7	1.3490×10^{-3}	1.3639×10^{-3}	1.3700×10^{-3}	1.3658×10^{-3}
9	2.9793×10^{-4}	2.7792×10^{-4}	2.7650×10^{-4}	2.7842×10^{-4}
11	7.8100×10^{-5}	6.9677×10^{-5}	6.7895×10^{-5}	6.7055×10^{-5}
13	2.5970×10^{-5}	2.1262×10^{-5}	2.0937×10^{-5}	2.1023×10^{-5}
15	1.6246×10^{-5}	1.3926×10^{-5}	1.3874×10^{-5}	1.3861×10^{-5}
W^r/W_0	4.2455×10^{-4}	3.8984×10^{-4}	3.8565×10^{-4}	3.8667×10^{-4}

in the presence of a dielectric plug with permittivity jumps at its boundaries. The corrugation shape is different from that given by formula (10). Here,

$$a_1(z) = a_2(z) = \begin{cases} 0.5 \quad \text{at} \quad z < 0 \quad \text{and} \quad z > 2 \\ 0.5 - a_0 \sin(\pi z/2) \quad \text{at} \quad 0 \le z \le 2 \end{cases} \tag{11}$$

with $\mu = 1$, $\varepsilon(z) = \varepsilon_- = \text{const}$ at $z < 0$ and $z > 2$, and $\varepsilon(z) = \varepsilon_+ = \text{const}$ at $0 < z < 2$. Figure 6 illustrates the profile of this transmission line for $a_0 = 0.45$, $\Delta = 1$, and $\Delta_{\min} = 0.1$.

As the pulse that satisfies initial conditions (7) propagates from the first regular half-layer ($z < 0$) to the second ($z > 2$), it has to pass through the smooth distributed irregularity (corrugation) and also through the lumped irregularities: (i) sharp kinks at the boundary surfaces that are given by the functions $a_1(z)$ and $a_2(z)$ in (11) at $z = 0$ and 2 and (ii) steps of $\varepsilon(z)$ also at $z = 0$ and 2. As follows directly from formulas (6) and (6'), the lumped irregularities break the continuity (or smoothness) of the matrix coefficients in the system of waveguide equations (5). In all cases of nonsmooth behavior of the matrix coefficients at isolated points on the z axis, system (5) must be completed with joining conditions, which require the transverse magnetic and electric field components to be continuous.

Note, however, that joining conditions greatly complicate the computational algorithms, as the use of implicit schemes becomes necessary. As a result, the body of computation increases substantially and, all other factors being the same, the simulation accuracy degrades because of the accumulation of roundoff errors. Below, we show that, for the lumped irregularities considered in this paper, these difficulties can be avoided by applying an approximate smoothing-out technique that makes it possible to obviate the need for joining conditions and extend system (5) to the entire interior of the rectangle Λ defined above, thereby preserving the possibility of applying the explicit procedure with all its advantages. We will also show that, in

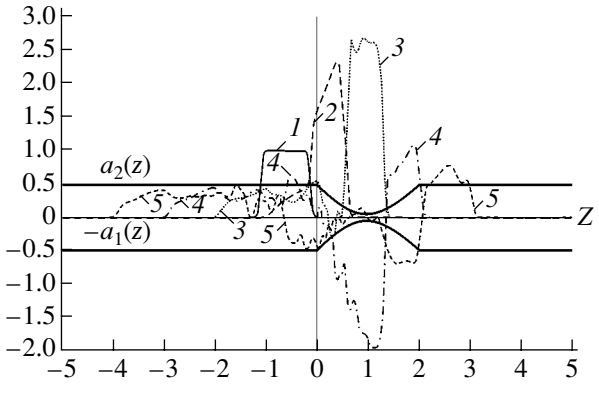

Fig. 6.

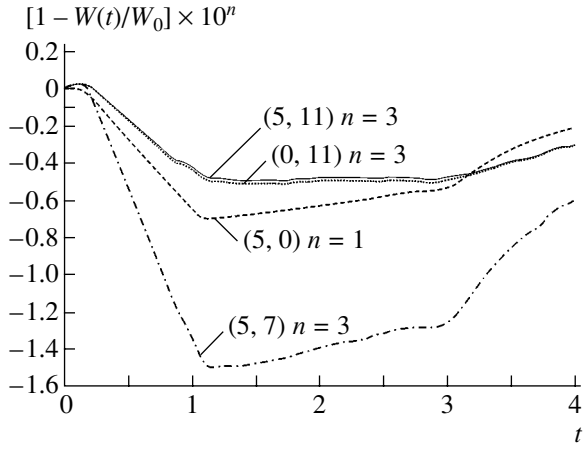

Fig. 7.

the examples that follow, additional errors arising from this approximation can be minimized.

In further calculations, we use $a_0 = 0.45$, $\varepsilon_- = 1$, and $\varepsilon_+ = 2$.

In addition to the profile of the transmission line, Fig. 6 plots the magnetic field intensity H in the median plane ($y = 0$) as a function of z at $t = 0$, 1, 2, 3, and 4 (curves *1–5*, respectively). The curves illustrate the time variation of the waveform for the initial pulse (b) (curve *1*). To smooth out the function $\varepsilon(z)$, we replace its values at 11 points of the mesh in the z direction near each of the jump points (at $z = 0$ and 2 and at five points adjacent to each of them on the left and on the right) by the values of a fifth-order polynomial. This polynomial is uniquely defined from the conditions that it passes continuously into the constants ε_- and ε_+ at the end points and that its first and second derivatives remain continuous at these points. To smooth out the boundary surfaces, we replace the values of the functions $a_1(z)$ and $a_2(z)$ at five points of the mesh in the z direction near each of the kinks (at $z = 0$ and 2 and at two points adjacent to each of them on the left and right) by the values of a third-order polynomial that is uniquely defined from the continuity and smoothness conditions.

The curves show that, as the pulse passes from the position $t = 0$ (curve *1*) to the position $t = 4$ (curve *5*) through the positions $t = 1$, 2, and 3 (curves *2–4*), the front part of the pulse first narrows and then restores its width, because the pulse travels from the optically lower density medium to the higher density medium and then again to the former.

Table 2 lists the relative partial energies W_j/W_0 of the TEM mode ($j = 1$) and for the E_{01} ($j = 2$), …, E_{014} ($j = 15$) modes. The calculation was made at $t = 7$ and $N = 15$ for pulses with initial waveform (b). The columns of the table refer to various pairs (p, q), where p and q are the numbers of mesh nodes in the z direction used to smooth out each of the kinks (p) and each of the jumps in $\varepsilon(q)$ by fifth-order polynomials. The zero values of p or q mean that smoothing-out was absent. The relative energies W_j/W_0 obtained for even j are no higher than $\sim 10^{-24}$ (rather than exact zeroes) and are not included in the table (except for the row $j = 2$). The last row of Table 2 contains estimates of the relative energy W^r/W_0 of the residual in sum (3), which is neglected when $N = 7$.

Figure 7 plots $[1 - W(t)/W_0]$ versus t for $0 \le t \le 4$ and $N = 7$ for the same initial pulse and the same values (p, q) as in Table 2. From Table 2 and Fig. 7, we can draw the following conclusions. The smoothing-out of the boundary kinks in the above examples does not yield a noticeable gain in accuracy, as follows from the results obtained for variants (5, 11) and (0, 11); hence, smoothing-out may be omitted. At the same time, the smoothing-out of the steps in ε gives a significant gain: the upper estimate $\sim 1/2\max\{[1 - W(t)/W_0], W^r/W_0\}$ of the relative rms error of the field intensity in variants (5, 0), (5, 7), and (5, 11) is $\sim 3 \times 10^{-2}$, 7×10^{-4}, and 2×10^{-4}, respectively. Such a significant difference in the effect of smoothing arises because the jumps in ε cause discontinuities in the matrix coefficient $G(z)$ by the higher derivative with respect to z in (5), while the boundary kinks do not affect the continuity of $G(z)$.

CONCLUSIONS

Thus, calculations based on nonstationary waveguide equations provide a relative rms error in the field intensity of the pulse that does not exceed $\sim 10^{-4}$, which is much smaller than the experimental error. This conclusion applies both to distributed irregularities (such as deep corrugation) and to combinations of distributed and lumped irregularities (i.e., sharp kinks at the boundary surfaces and steps of the permittivity). This opens up new possibilities for studies concerned with UWB EMPs and for the development of reference field-shaping systems and devices for calibrating UWB EMP detectors.

APPENDIX

Expressions for the functions $g_{ns}(z)$, $t_{ns}(z)$, $q_{ns}(z)$, and $p_{ns}(z)$.

(1) Functions $g_{ns}(z)$ and $t_{ns}(z)$ have the following form:

$$g_{11} = \Delta(z);$$

$$g_{1s} = g_{n1} = 0, \quad n > 1 \quad \text{and} \quad s > 1;$$

$$g_{ns} = \frac{1}{2}\Delta(z)\delta_{ns}, \quad n > 1 \quad \text{and} \quad s > 1,$$

δ_{ns} is the Kronecker symbol;

$$t_{ns} = g_{ns} \quad \text{for any} \quad n \quad \text{and} \quad s \ge 1.$$

(2) Functions $q_{ns}(z)$ are as follows:

$$q_{n1} = 0, \quad n \ge 1;$$

$$q_{1s} = [(-1)^s a_2'(z) - a_1'(z)], \quad s > 1,$$

(') is the derivative symbol;

$$q_{ns} = \frac{(s-1)^2}{(n-1)^2 - (s-1)^2}$$

$$\times [a_1'(z) + (-1)^{n+s} a_2'(z)], \quad n > 1 \quad \text{and} \quad s > 1, \; n \ne s;$$

$$q_{nn} = -\frac{1}{4}\Delta'(z), \quad n > 1.$$

(3) Functions $p_{ns}(z)$ are

$$p_{1s} = p_{n1} = 0, \quad n \ge 1 \quad \text{and} \quad s \ge 1;$$

$$p_{ns} = \frac{4(s-1)^2(n-1)^2}{[(s-1)^2-(n-1)^2]^2}\frac{\Delta'(z)}{\Delta(z)}$$

$$\times [a_1'(z) + (-1)^{n+s}a_2'(z)], \quad n > 1 \quad \text{and} \quad s > 1,\ n \neq s;$$

$$p_{nn} = \frac{\pi^2(n-1)^2}{6\Delta(z)}$$

$$\times \{3 - 3a_1'(z)a_2'(z) + [\Delta'(z)]^2\} - \frac{[\Delta'(z)]^2}{4\Delta(z)}, \quad n > 1.$$

REFERENCES

1. *Ultra-Wideband, Short-Pulse Electromagnetics 3*, Ed. by C. E. Baum, L. Carin, and A. P. Stone (Plenum, New York, 1997).
2. V. I. Koroza, M. N. Nechaev, and S. A. Tsvetkov, Pis'ma Zh. Tekh. Fiz. **21** (11), 1 (1995) [Tech. Phys. Lett. **21**, 395 (1995)].
3. V. I. Koroza, Pis'ma Zh. Tekh. Fiz. **22** (21), 6 (1996) [Tech. Phys. Lett. **22**, 865 (1996)].
4. V. I. Koroza and V. E. Kondrashov, Pis'ma Zh. Tekh. Fiz. **22** (17), 91 (1996) [Tech. Phys. Lett. **22**, 729 (1996)].
5. V. V. Borisov, *Transients Fields in Waveguides* (Leningr. Gos. Univ., Leningrad, 1991).
6. M. N. Golikov, V. I. Koroza, and V. N. Mikhaĭlov, Dokl. Akad. Nauk **383** (1), 51 (2002) [Dokl. Phys. **47**, 504 (2002)].

Translated by A. Khzmalyan

Technical Physics, Vol. 48, No. 4, 2003, pp. 385–392. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 4, 2003, pp. 9–16.

THEORETICAL AND MATHEMATICAL PHYSICS

Surface Electromagnetic Waves in Faraday Media

A. N. Furs and L. M. Barkovsky

Belarussian State University, ul. F. Skanina 4, Minsk, 220050 Belarus

e-mail: Barkovsky@bsu.by

Received July 10, 2002

Abstract—The effect of unidirectional propagation of surface electromagnetic waves on the interface between an isotropic Faraday medium and isotropic optically inactive medium is predicted. Such waves can be excited when the intensity of an external magnetic field exceeds a certain threshold. Solutions to the dispersion relation are analyzed, and existence conditions for the surface waves are established. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

In recent years, along with surface electromagnetic waves in surface-active negative permittivity media (metals, plasmas, etc. [1]), surface waves of a new type, dispersionless polaritons, have been the subject of much investigation. They appear when one or both contacting media are anisotropic [2–5] and can be observed at the interface between a positive uniaxial crystal and an isotropic medium, between identical positive uniaxial crystals with crossed axes, or between a biaxial crystal and an isotropic medium. General conditions for the existence of these surface waves at the interface between anisotropic and isotropic dielectric media with positively definite permittivity tensors have been found in [6], and conditions for the resonance excitation of them in isotropic layer–anisotropic substrate structures have been discussed in [7]. A typical feature of dispersionless surface waves is that they can propagate only in certain directions along the interface. The angular width of allowed directions increases with the degree of anisotropy of contacting media.

The use of contacting media as parametric materials whose anisotropy can be induced and varied by external fields opens up wide possibilities for controlling dispersionless surface polaritons. In this paper, we touch upon the problem of magnetooptically controlling surface waves at the interface between a Faraday medium and an optically inactive isotropic medium with an external magnetic field applied along the interface. Using the surface-impedance tensor formalism, we derive a dispersion relation, study the symmetry of its solutions, and find the necessary and sufficient conditions for the existence of surface electromagnetic waves. These conditions imply that the interface between optically active and passive media may serve as an isolator. We also consider the particular case where surface waves propagate perpendicularly to an external magnetic field. Expressions for the surface impedance tensors entering the dispersion relation are derived in the Appendix.

DISPERSION RELATION FOR SURFACE WAVES AT THE INTERFACE BETWEEN FARADAY AND INACTIVE ISOTROPIC MEDIA

Constitutive equations for monochromatic waves with a frequency ω in transparent Faraday optically active media have the form [8]

$$\mathbf{E} = \varepsilon^{-1}\mathbf{D}, \quad \mathbf{H} = \mathbf{B},$$

where the inverse tensor of the permittivity ε^{-1} is a linear function of the external magnetic field intensity H:

$$\varepsilon_{ik}^{-1} = (\varepsilon^0)_{ik}^{-1} + ie_{ikl}F_{lm}H_m. \tag{1}$$

In Eq. (1), $F = (F_{ik})$ is the Faraday tensor and e_{ikl} is the completely antisymmetric third-rank pseudotensor (Levi-Civita pseudotensor). Summation over repeating indices is performed from 1 to 3. Below, we consider optically isotropic Faraday media, for which the tensors $(\varepsilon^0)^{-1}$ and F are proportional to the unit tensor: $(\varepsilon^0)_{ik}^{-1} = a\delta_{ik}$ and $F_{ik} = f\delta_{ik}$, where δ_{ik} is the Kronecker delta and a and f are scalar quantities. For such media, Eq. (1) takes the form

$$\varepsilon^{-1} = a + ib\mathbf{c}^{\times}, \tag{2}$$

where the scalar quantity b is a linear function of the magnitude of H and $\mathbf{c}^{\times}$ is the tensor dual [9, 10] to the unit vector $\mathbf{c}$ that is parallel to H.

Artificial optical activity shows up as the rotation of plane of polarization of linearly polarized light (the Faraday effect). The parameters a and b entering formula (2) are expressed through the refractive index n_0 of a linearly polarized wave propagating in the direction of $\mathbf{c}$, the wavelength $\lambda_0 = 2\pi c/\omega$ of this wave in free space, and the Verdet constant V as [8]

$$a = \frac{1}{n_0^2}, \quad b = \frac{V\chi_0 H}{\pi n_0^3}. \tag{3}$$

1063-7842/03/4804-0385$24.00 © 2003 MAIK "Nauka/Interperiodica"

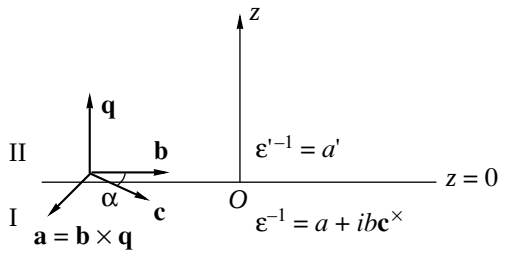

Fig. 1. Interface between Faraday medium and isotropic optically inactive media.

Without loss of generality, we further assume that $0 \le b < a$ and that the unit vector **c** is codirected to the vector **H** for Faraday media with positive rotation ($V > 0$) and counterdirected to the vector **H** for those with negative rotation ($V < 0$).

Let isotropic Faraday medium I ($z < 0$) characterized by inverse permittivity tensor ε^{-1} (2) border on isotropic optically inactive medium II ($z > 0$, Fig. 1) characterized by the inverse permittivity $a' = 1/\varepsilon' > 0$, and let an external magnetic field be applied parallel to the interface (the vector **c** is perpendicular to the unit normal vector **q** to the interface, which is parallel to the z axis). Below, we show that, when

$$b > a - a' \quad (a' < a), \quad b > \sqrt{a'(a' - a)} \quad (a' > a),$$

surface waves may be excited in such a structure that propagate in certain directions **b** ($\mathbf{b}^2 = 1$) relative to the vector **c**.

Equations for a surface electromagnetic wave in medium I can be written as

$$\mathbf{H}(\mathbf{r}, t) = \sum_{s=1}^{2} C_s \mathbf{H}_s^0 \exp[ik(\mathbf{b} - i\eta_s \mathbf{q}) \cdot \mathbf{r} - i\omega t],$$
$$\mathbf{E}(\mathbf{r}, t) = \sum_{s=1}^{2} C_s \mathbf{E}_s^0 \exp[ik(\mathbf{b} - i\eta_s \mathbf{q}) \cdot \mathbf{r} - i\omega t], \quad (4)$$

where k is the projection of the wavevector onto the direction **b**, $\mathbf{H}_0^s$, and $\mathbf{E}_0^s$ are the amplitudes of inhomogeneous partial waves at the interface, η_s are complex coefficients that characterize the decay of these waves away from the interface ($\mathrm{Re}\,\eta_s > 0$), and C_s are weighting factors.

The fields in medium II are described by similar equations involving the amplitudes $\mathbf{H}_s'^0$ and $\mathbf{E}_s'^0$ and the coefficients C_s' and η_s' (with $\mathrm{Re}\,\eta_s' < 0$). The vectors $\mathbf{H}_s^0$, $\mathbf{E}_s^0$, $\mathbf{H}_s'^0$, and $\mathbf{E}_s'^0$ and coefficients η_s and η_s' can be found by substituting (4) into the Maxwell equations and constitutive relations, while the projection k of the wavevector and the weighting factors C_s and C_s' are found from the continuity conditions for the tangential components of the fields at the interface (boundary conditions). The tangential components of the resultant magnetic and electric fields at the interface,

$$\mathbf{H}_\tau^0 = \sum_{s=1}^{2} C_s \mathbf{H}_{s\tau}^0, \quad \mathbf{H}_\tau'^0 = \sum_{s=1}^{2} C_s' \mathbf{H}_{s\tau}'^0,$$
$$\mathbf{q} \times \mathbf{E}^0 = \sum_{s=1}^{2} C_s \mathbf{q} \times \mathbf{E}_s^0, \quad \mathbf{q} \times \mathbf{E}'^0 = \sum_{s=1}^{2} C_s' \mathbf{q} \times \mathbf{E}_s'^0,$$

are related to each other through the surface impedance tensors γ and γ' of the contacting media [11]:

$$\mathbf{q} \times \mathbf{E}^0 = \gamma \mathbf{H}_\tau^0, \quad \mathbf{q} \times \mathbf{E}'^0 = \gamma' \mathbf{H}_\tau'^0. \quad (5)$$

From Eqs. (5) and boundary conditions $\mathbf{H}_\tau^0 = \mathbf{H}_\tau'^0$ and $\mathbf{q} \times \mathbf{E}^0 = \mathbf{q} \times \mathbf{E}'^0$, it follows that

$$(\Gamma - \Gamma')\mathbf{H}_\tau^0 = 0, \quad (6)$$

where $\Gamma = -i\nu\gamma$, $\Gamma' = -i\nu\gamma'$, and $\nu = \omega/(ck)$ is the dimensionless reduced frequency (the phase velocity in terms of the velocity of light in free space).

Assuming that Eq. (6) has nontrivial solutions $\mathbf{H}_\tau^0 \neq 0$, we obtain the dispersion relation

$$\mathrm{Tr}\overline{(\Gamma - \Gamma')} = 0. \quad (7)$$

Here, the bar means the tensor adjugate to $\Gamma - \Gamma'$ and Tr means the trace of this tensor [9, 10].

For surface electromagnetic waves in transparent media described by Eqs. (4), the tensors Γ and Γ' are Hermitean tensors and can be calculated from general formulas obtained in [5, 6]. For a Faraday medium (see Appendix),

$$\Gamma = \frac{1}{a - \nu^2 + a\eta_1\eta_2}$$
$$\times \{\nu^2 a(\eta_1 + \eta_2)\mathbf{b} \otimes \mathbf{b} - \nu^2 b\cos\alpha(\mathbf{b} \otimes \mathbf{a} + \mathbf{a} \otimes \mathbf{b}) \quad (8)$$
$$+ [b\sin\alpha(a - \nu^2 + a\eta_1\eta_2) - a^2\eta_1\eta_2(\eta_1 + \eta_2)]\mathbf{a} \otimes \mathbf{a}\},$$

where $\mathbf{a} = [\mathbf{b} \times \mathbf{q}]$ ($\mathbf{a}^2 = 1$), α is the angle between the vectors **b** and **c** ($\mathbf{c} = \mathbf{b}\cos\alpha + \mathbf{a}\sin\alpha$), and

$$\eta_{1,2} = \frac{1}{a}\left[a(a - \nu^2) - \frac{1}{2}b^2\cos^2\alpha \pm \frac{1}{2}b\cos\alpha\sqrt{4a\nu^2 + b^2\cos^2\alpha}\right]^{1/2} \quad (9)$$

are the decay coefficients of the partial waves.

For an isotropic medium,

$$\Gamma' = -\nu^2\sqrt{\frac{a'}{a' - \nu^2}}\mathbf{b} \otimes \mathbf{b} + \sqrt{a'(a' - \nu^2)}\mathbf{a} \otimes \mathbf{a} \quad (10)$$

and

$$\eta'_{1,2} = -\sqrt{\frac{a' - \nu^2}{a'}}. \quad (11)$$

By substituting expressions (8) and (10) into Eq. (7), we obtain a dispersion relation for surface waves propagating over the interface between Faraday and isotropic media:

$$G(\nu, \alpha) = 0, \quad (12)$$

where

$$G(\nu, \alpha) = \frac{(\eta_1 + \eta_2)ab\sin\alpha}{a - \nu^2 + a\eta_1\eta_2} - \frac{a^3\eta_1\eta_2(\eta_1 + \eta_2)^2 + \nu^2 b^2\cos^2\alpha}{(a - \nu^2 + a\eta_1\eta_2)^2} + \left[b\sin\alpha - \frac{a^2\eta_1\eta_2(\eta_1 - \eta_2)}{a - \nu^2 + a\eta_1\eta_2}\right]\sqrt{\frac{a'}{a' - \nu^2}} - \frac{a(\eta_1 + \eta_2)}{a - \nu^2 + a\eta_1\eta_2}\sqrt{a'(a' - \nu^2)} - a'. \quad (13)$$

Given the parameters a, b, a', and α, relations (12), (13), and (9) can be used to determine the reduced frequency $\nu = \nu_s$ and, consequently, the projection $k = k_s$ of the surface wave's wavevector onto the direction **b**. In these formulas, we assume that

$$0 \le \nu_s < \hat{\nu}_{\lim} = \min(\nu_{\lim}, \nu'_{\lim}), \quad (14)$$

where $\nu_{\lim} = \sqrt{a - b|\cos\alpha|}$ and $\nu'_{\lim} = \sqrt{a'}$ are the limiting frequencies of the waves in the Faraday and isotropic media (see also Eq. (A8)).

Condition (14) means that the electromagnetic wave energy is localized near the interface in both media; i.e., the wave is actually a surface wave. The absence of solutions to Eq. (12) in the interval $[0, \hat{\nu}_{\lim})$, the so-called sublight interval [6], means that it is impossible to excite a surface wave propagating in the given direction **b**.

ANALYSIS OF SOLUTIONS TO THE DISPERSION RELATION AND EXISTENCE CONDITIONS FOR SURFACE WAVES

Dispersion relation (12) does not change its form when the angle α is replaced by $\pi - \alpha$, the equalities $\eta_1(\pi - \alpha) = \eta_2(\alpha)$ and $\eta_2(\pi - \alpha) = \eta_1(\alpha)$ being valid in this case. Therefore, if a surface wave can propagate in the direction specified by a certain vector **b**, it can also propagate in the direction mirror-reflected in the plane to which **c** is normal.

At the same time, the dispersion relation is not invariant under the substitution $\alpha \longrightarrow \alpha - \pi$ (the reversal of the propagation direction: $\mathbf{b} \longrightarrow -\mathbf{b}$) if the coefficients η_1 and η_2 remain positive and η'_1 and η'_2, negative. This means that, in general, the propagation direction of surface electromagnetic waves at the interface between Faraday and inactive isotropic media is irreversible and that such an interface exhibits isolating properties. If $\nu = \nu_s$ is a solution to Eq. (12), $\nu = -\nu_s$ is also a solution. However, the substitution $\nu_s \longrightarrow -\nu_s$ (or $k_s \longrightarrow -k_s$) is not equivalent to the substitution $\mathbf{b} \longrightarrow -\mathbf{b}$, because the corresponding wave not only travels in the opposite direction but also its amplitude grows exponentially with distance from the interface (see Eq. (4)). Such a wave is physically unrealistic.

At the left boundary $\nu = 0$ of the sublight interval, the coefficients η_1 and η_2 (9) are

$$\eta_1(0) = 1, \quad \eta_2(0) = \frac{1}{a}\sqrt{a^2 - b^2\cos^2\alpha}$$

and the function $G(\nu, \alpha)$, which appears in the dispersion relation, is negative:

$$G(0, \alpha) = 2(b\sin\alpha - \sqrt{a^2 - b^2\cos^2\alpha} - a') < 0.$$

If the dispersion relation has a solution, it is unique [6]. Consequently, Eq. (12) has a solution if the function G is positive at the right boundary $\nu = \hat{\nu}_{\lim}$ of the sublight interval:

$$G_{\lim}(\alpha) = \lim_{\nu \to \hat{\nu}_{\lim}} G(\nu, \alpha) > 0. \quad (15)$$

Condition (15) is a necessary and sufficient condition for the existence of surface electromagnetic waves at the interface between Faraday and inactive isotropic media.

Function $G(\nu, \alpha)$ in (15) can be replaced by the function $R(\nu, \alpha) = \sqrt{a' - \nu^2}\,G(\nu, \alpha)$, which is finite at $\nu = \sqrt{a'}$. Figure 2 plots the function $R_{\lim}(\alpha) = R(\hat{\nu}_{\lim}, \alpha)$ for various values of the constitutive parameters a, b, and a'. The angular intervals marked by thick segments on the abscissa are those where dispersion relation (12) has solutions. They define the propagation directions of surface electromagnetic waves. Depending on the parameters, one of the three following cases is realized: (i) no directions exist on the interface in which the surface waves can propagate (Fig. 2a); (ii) the wave propagation is possible if α belongs to either the interval (α_1, α_2) or the interval $(\pi - \alpha_2, \pi - \alpha_1)$, where $0 < \alpha_1 < \alpha_2 < \pi/2$ (Fig. 2b); and (iii) allowed propagation directions are defined by α lying in the interval $(\alpha_1, \pi - \alpha_1)$, where $0 < \alpha_1 < \pi/2$ (Fig. 2c). In other words, these directions fall into the angle $\pi - 2\alpha_1$ with the bisectrix perpendicular to **c**.

In cases (ii) and (iii), the unidirectional propagation mentioned above is observed. Mathematically, this means that there are no intervals with negative α, which correspond to waves traveling in the opposite direction.

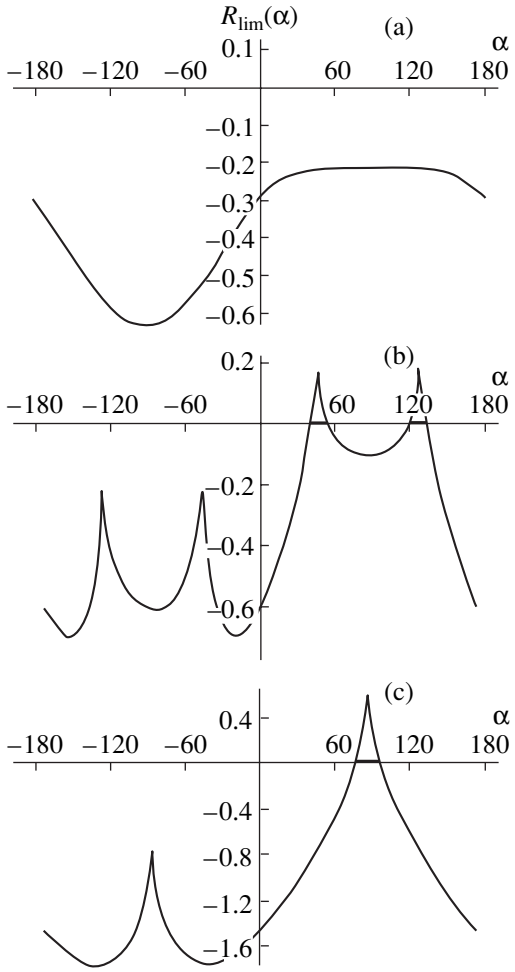

Fig. 2. Functions $R_{\lim} = R_{\lim}(\alpha)$ and intervals of the angle α where surface electromagnetic waves exist for $a = 0.9$ and $b = 0.3$. a' = (a) 0.5, (b) 0.7, and (c) 0.9.

All vectors **b** that specify allowed propagation directions lie in the interface plane on one side of the vector **c**:

$$[\mathbf{b} \times \mathbf{q}] \cdot \mathbf{c} > 0. \tag{16}$$

Below, we use relationship (15) to derive expressions for the extreme angles α_1 and α_2 of the intervals and establish conditions under which each of the cases listed above is realized.

Since the dispersion relation retains its form under the substitution $\alpha \longrightarrow \pi - \alpha$, it can be assumed that $\alpha \in [-\pi/2, \pi/2]$. Let the angle α be such that $a - b\cos\alpha < a'$. Then, $\hat{\nu}_{\lim} = \nu_{\lim} = \sqrt{a - b\cos\alpha}$ and, at $\nu = \hat{\nu}_{\lim}$, expression (9) for $\eta_{1,2}$ becomes

$$\eta_{1\lim} = \frac{1}{a}\sqrt{b\cos\alpha(2a - b\cos\alpha)}, \quad \eta_{2\lim} = 0$$

(the zero value of the coefficient η_2 means that one of the partial waves in the Faraday medium becomes a body wave). Eventually, condition (15) takes the form

$$\left(\tan\alpha - \frac{\sqrt{a'(a' - a + b\cos\alpha)}}{b\cos\alpha}\right)$$

$$\times \sqrt{b\cos\alpha(2a - b\cos\alpha)} + b\sin\alpha\sqrt{\frac{a'}{a' - a + b\cos\alpha}} \tag{17}$$

$$-(a - b\cos\alpha) - a' > 0.$$

If $a - b\cos\alpha > a'$, $\hat{\nu}_{\lim} = \nu'_{\lim} = \sqrt{a'}$ and the function $G_{\lim}(\alpha)$ increases without limit due to the singular term with $(a' - \nu^2)^{-1/2}$. The function is positive if the bracketed term in (13) is positive at $\nu = \sqrt{a'}$. Thus, we have

$$b\sin\alpha[a - a' + \sqrt{(a - a')^2 - b^2\cos^2\alpha}]$$

$$> \sqrt{(a - a')^2 - b^2\cos^2\alpha}$$

$$\times\left\{\left[a(a - a') - \frac{1}{2}b^2\cos^2\alpha + \frac{1}{2}b\cos\alpha\sqrt{4aa' + b^2\cos^2\alpha}\right]^{1/2}\right. \tag{18}$$

$$+\left[a(a - a') - \frac{1}{2}b^2\cos^2\alpha\right.$$

$$\left.\left.-\frac{1}{2}b\cos\alpha\sqrt{4aa' + b^2\cos^2\alpha}\right]^{1/2}\right\}.$$

Inequalities (17) and (18) define the intervals of α (a set of directions **b** at the interface between the Faraday and isotropic media) where surface electromagnetic waves can be excited. The extreme angles α_1 and α_2 of these intervals can be found from relationships (17) and (18) written as equalities. Then, we exclude the radical signs from (17) to obtain

$$c_0x^4 + c_1x^3 + c_2x^2 + c_3x + c_4 = 0, \tag{19}$$

where $x = \cos\alpha_1$ and

$$c_0 = 4a'^2b^4,$$

$$c_1 = 4a'b^3[3(a - a')^2 - b^2],$$

$$c_2 = b^2[(a - a')^2(a^2 - 18aa' + 9a'^2) + 2(a - a')(a + 3a')b^2 + b^4], \tag{20}$$

$$c_3 = 4b(a - a')[a'(a - a')^3 - (a - a')(a + 2a')b^2 - b^4],$$

$$c_4 = 4(a - a')^2[(a - a')a' + b^2]^2.$$

From (18), we find

$$d_0y^4 + d_2y^2 + d_4 = 0, \tag{21}$$

where $y = \cos\alpha_2$ and

$$d_0 = 4a'^2b^4,$$
$$d_2 = b^2[(a-a')^2(a^2-6aa'-3a'^2) - 2(a-a')(a-3a')b^2 + b^4], \quad (22)$$
$$d_4 = 4a'(a-a')^3[a(a-a')-b^2].$$

When $0 \le b < a$, $a' > 0$, and $a' < a$, the left of relationship (17) turns into zero if we take the greatest real root x of Eq. (19). If $a' < a$; the same takes place for the smallest real root. Relationship (18) turns into an equality if we take the following root of Eq. (21):

$$y = \frac{1}{2\sqrt{2}a'b}\left\{-(a-a')^2(a^2-6aa'-3a'^2) + 2(a-a')(a-3a')b^2 - b^4 - \sqrt{[(9a'-a)(a-a')+b^2][b^2-(a-a')^2]^3}\right\}^{1/2}. \quad (23)$$

The left boundary α_1 of the interval where surface electromagnetic waves exist is thus defined by a real root of Eq. (19) (the greatest one if $a' < a$ or the smallest one if $a' > a$). The right boundary α_2 can be calculated directly from formula (23).

Consider the variation of the angles α_1 and α_2 with a' at given a and b. If $a' < a - b$, conditions (17) and (18) do not hold whatever α is (case (i)). When $a' = a - b$, the roots of Eqs. (19) and (21) are $x = 1$ and $y = 1$ ($\alpha_1 = \alpha_2 = 0$). Further, the angles α_1 and α_2 increase monotonically with a' (Fig. 3). The angle α_2 becomes equal to $\pi/2$ when $y = 0$ and, hence, when the free term d_4 of Eq. (21) is zero. This condition is achieved at $a' = a - b^2/a$. Thus, when the condition

$$a - b < a' < a - \frac{b^2}{a}$$

is met, case (ii) is realized and there are two intervals (α_1, α_2) and $(\pi - \alpha_2, \pi - \alpha_1)$ of the angle α where surface waves exist. Finally, the angle α_1 becomes equal to $\pi/2$ when the free term c_4 of Eq. (19) vanishes; i.e., at $(a - a')a' + b^2 = 0$. Therefore, the condition

$$a - \frac{b^2}{a} < a' < \frac{1}{2}(a + \sqrt{a^2 + 4b^2}) \quad (24)$$

corresponds to case (iii), in which there is only one interval $(\alpha_1, \pi - \alpha_1)$ of surface wave existence, with α_1 being found from Eq. (19) as before. When $a' > (a + \sqrt{a^2 + 4b^2})/2$, surface waves cannot propagate (case (i)).

The variation of the angles α_1 and α_2 with b at fixed a and a' depends significantly on whether $a' < a$ or $a' > a$ (Fig. 4). However, surface waves can be excited in both cases when b exceeds a certain threshold. When $a' < a$, this threshold is

$$b^* = a - a'.$$

Case (ii) is realized when $b^* < b < \sqrt{a(a-a')}$; case (iii), when $b > \sqrt{a(a-a')}$. At $b = \sqrt{a(a-a')}$, coefficient d_4 of (22) turns to zero and $\alpha_2 = \pi/2$.

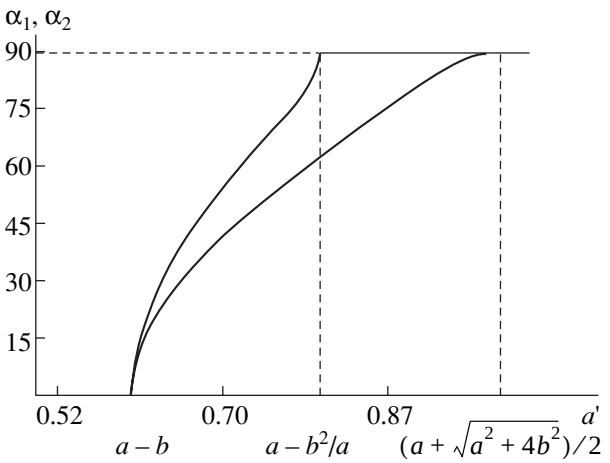

Fig. 3. Extreme angles α_1 and α_2 versus parameter a' at $a = 0.9$ and $b = 0.3$.

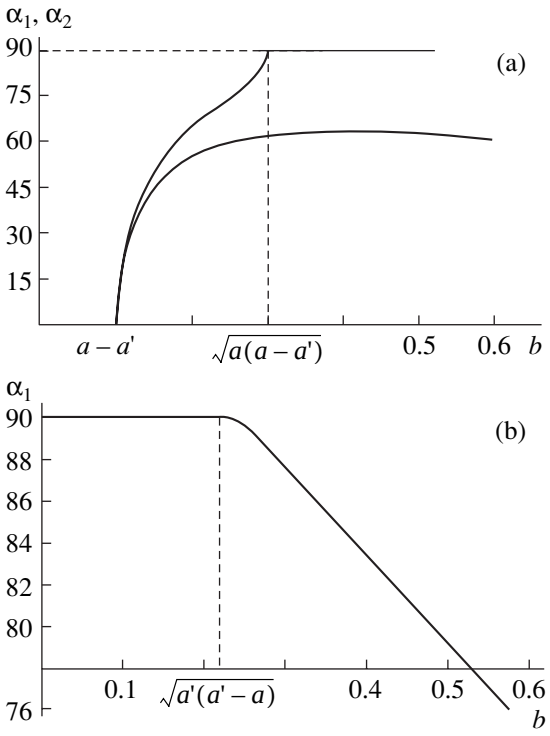

Fig. 4. Extreme angles α_1 and α_2 versus parameter b at $a = 0.9$ and $a' =$ (a) 0.8 and (b) 0.95.

When $a' > a$, the threshold value for b is determined from the condition $c_4 = 0$:

$$b^* = \sqrt{a'(a'-a)}.$$

When $b > b^*$, case (iii) is realized. With (3), the threshold of the external magnetic field intensity can easily be found:

$$\mathfrak{H}^* = \begin{cases} \dfrac{\pi n_0(\varepsilon' - n_0^2)}{V\lambda_0\varepsilon'}, & \varepsilon' > n_0^2 \\ \dfrac{\pi n_0^2\sqrt{n_0^2-\varepsilon'}}{V\lambda_0\varepsilon'}, & \varepsilon' < n_0^2. \end{cases} \tag{25}$$

At $a' = a$, the threshold for the external magnetic field is absent and case (iii) takes place at any b. From (19) and (20), we find that $\alpha_1 = \arccos(b/2a)$. Hence, the angular width of the range of directions **b** in which surface waves can propagate is $\Delta\alpha = 2\arccos(b/2a)$. When $b \ll a$, we have

$$\Delta\alpha = \frac{b}{a} = \frac{V\lambda_0\mathfrak{H}}{\pi n_0}.$$

For example, at a wavelength $\lambda_0 = 496$ nm, the refractive index and Verdet constant of crystalline strontium titanate $SrTiO_3$ are $n_0 = 2.48$ and $V = 0.31$ min/(Oe cm), respectively. If this crystal is adjacent to an isotropic inactive medium with the same refractive index, the width of the interval is $\Delta\alpha = 12''$ at the magnetic field intensity $\mathfrak{H} = 10^5$ Oe.

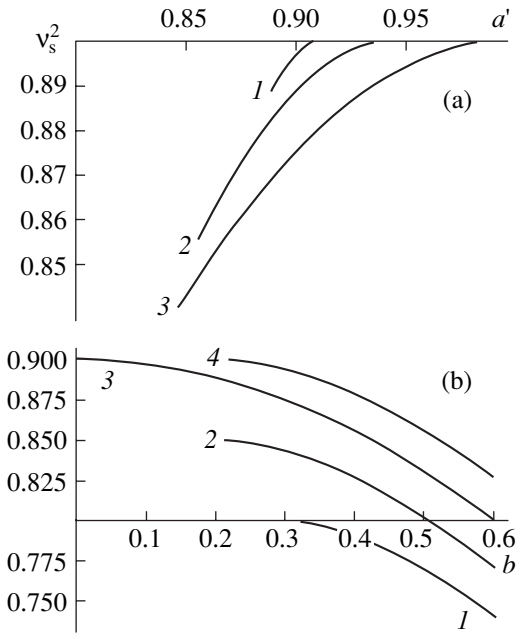

Fig. 5. Solutions to the dispersion relation for surface waves propagating perpendicularly to **H** at $a = 0.9$: (a) ν_s^2 versus a' at $b = 0.1$ (*1*), 0.2 (*2*), and 0.3 (*3*) and (b) ν_s^2 versus b at $a' = 0.8$ (*1*), 0.85 (*2*), 0.9 (*3*), and 0.95 (*4*).

In general, at arbitrary a, b, a', and α, the dispersion relation given by (12) and (13) can only be solved numerically. This equation is simplified when a surface wave propagates perpendicularly to an external magnetic field ($\alpha = \pi/2$, $\mathbf{c} = \mathbf{a}$). In this case,

$$\eta_{1,2} = \sqrt{\frac{a-\nu^2}{a}}, \quad \nu_{\lim}^2 = a, \quad \nu_{\lim}'^2 = a' \tag{26}$$

and (12) takes the form

$$[b - \sqrt{a(a-\nu^2)}]\sqrt{\frac{a'}{a'-\nu^2}} + [b - \sqrt{a'(a'-\nu^2)}]\sqrt{\frac{a}{a-\nu^2}} - a - a' = 0. \tag{27}$$

Note that the substitutions $a \rightleftharpoons a'$ do not change Eq. (27); i.e., the phase velocity of a surface wave propagating in a direction perpendicular to **c** at the interface between media with the inverse permittivity tensors $\varepsilon^{-1} = a' + ib\mathbf{c}^\times$ and $\varepsilon'^{-1} = a$ is the same as for media with the inverse permittivity tensors $\varepsilon^{-1} = a + ib\mathbf{c}^\times$ and $\varepsilon'^{-1} = a'$.

Equation (27) with positive a, b, and a' has the real root ν^2 when case (iii) is realized. Therefore, given a and b, a' must satisfy inequalities (24) (see also Fig. 3); given a and a', b must meet the inequalities

$$b > \sqrt{a(a-a')} \quad \text{when} \quad a' < a, \tag{28}$$

$$b > \sqrt{a'(a'-a)} \quad \text{when} \quad a' > a \tag{29}$$

(Fig. 4). Conditions (28) and (29) correspond to positive bracketed terms in (27).

Excluding the radicals from (27), we come to

$$(a-a')^2\nu^4 + 2(a+a')[b^2 - (a-a')^2]\nu^2 - [b^2 - (a-a')^2][(a-a')^2 - b^2] = 0. \tag{30}$$

The reduced frequency ν_s^2 squared equals the positive root of Eq. (30):

$$\nu_s^2 = \frac{1}{(a-a')^2}\{-(a+a')[b^2 - (a-a')^2] + 2b\sqrt{aa'[b^2 - (a-a')^2]}\}. \tag{31}$$

At $a = a'$, formula (30) yields

$$\nu_s^2 = a - \frac{b^2}{4a}.$$

In the limit $b = \sqrt{a(a-a')}$ (see (28)) (or, what is the same, the replacement of the first sign in inequality (24) by the equality sign), (31) yields $\nu_s^2 = \nu_{\lim}'^2 = a'$. The fact that the reduced frequency equals the limiting frequency means that the wave propagating in isotropic

medium II is a body wave (see (11)) with the Poynting vector parallel to the interface (a limiting electromagnetic wave [6]). In a similar fashion, passing to the equality in (29) (replacing the second sign in (24) by the equality sign), we arrive at $\nu_s^2 = \nu_{lim}^2 = a$. In this case, the body wave propagates in Faraday medium I (see (26)). Figure 5a plots the functions $\nu_s^2 = \nu_s^2(a')$ calculated from (31) for several values of the parameter b; Fig. 5b, the functions $\nu_s^2 = \nu_s^2(b)$ calculated for various a'. The leftmost points in the curves in Fig. 5a correspond to the limiting wave propagating in the isotropic medium; the rightmost points, in the Faraday medium. Accordingly, in Fig. 5b, the leftmost points in curves *1* and *2* refer to the limiting wave in the isotropic medium ($a' < a$), while the leftmost point in curve *4* refers to the limiting wave in the Faraday medium ($a' > a$).

CONCLUSIONS

Linearly polarized body electromagnetic waves can always propagate in Faraday media in both forward and backward directions. They experience a Faraday rotation twice when traveling through the medium back and forth. For surface waves, the interface between Faraday and isotropic optically inactive media serves as an isolator: these waves either cannot propagate in both directions **b** and –**b** (if condition (17) or (18) is not met) or can only propagate in the forward **b** direction defined by condition (16). Surface waves can be excited when the external magnetic field exceeds a threshold depending on the difference between the constitutive parameters a and a' (see also (25)). Thus, by varying the intensity or direction of the external magnetic field, one can effectively control the angular spectra of allowed propagation directions for surface waves.

APPENDIX: TENSORS Γ AND Γ'

In [5, 6], the Barnett–Lothe integral operator formalism for SAWs [12] was extended to surface electromagnetic excitations. The new approach involves a unified procedure for calculating the impedance tensors for surface electromagnetic waves in lossless anisotropic dielectric media and deriving dispersion relations with these tensors. For waves at the interface between anisotropic materials characterized by the permittivity tensors ε ($z = \mathbf{q}\cdot\mathbf{r} < 0$) and ε' ($z > 0$), the dispersion relation has the form of (7) with

$$\Gamma = -Q^- + iQ^-S, \quad \Gamma' = Q'^- + iQ'^-S'. \tag{A1}$$

The bar in (A1) means the pseudoinversion of planar tensors in the two-dimensional subspace of the interface (i.e., $QQ^- = Q^-Q = I$, where $I = -\mathbf{q}^{\times 2} = \mathbf{b}\otimes\mathbf{b} + \mathbf{a}\otimes\mathbf{a}$ is the projective operator). The tensors Q and S can be represented in the integral form:

$$Q = -\frac{1}{\pi}\int_0^\pi (\mathbf{e}_2\mathbf{e}_2)^- d\phi, \quad S = -\frac{1}{\pi}\int_0^\pi (\mathbf{e}_2\mathbf{e}_2)^-(\mathbf{e}_2\mathbf{e}_1)d\phi. \tag{A2}$$

The integrands in (A2) contain tensor bilinear forms of two vector arguments. For any vectors **u** and **v** (**uv** = 0), these forms are given as follows:

$$(\mathbf{u}\cdot\mathbf{v}) = -\frac{\mathbf{u}\tilde{\varepsilon}^{-1}\mathbf{v}\cdot(\mathbf{a}\otimes\mathbf{a}) + \nu^2 I\mathbf{u}^\times\varepsilon^{-1}\mathbf{v}^\times I}{\mathbf{a}\varepsilon^{-1}\mathbf{a} - \nu^2} + \nu^2\mathbf{b}\cdot\mathbf{u}\cdot\mathbf{b}\cdot\mathbf{v}I. \tag{A3}$$

The vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ are expressed through **b** and **q** as

$$\mathbf{e}_1 = \mathbf{b}\cos\phi + \mathbf{q}\sin\phi, \quad \mathbf{e}_2 = -\mathbf{b}\sin\phi + \mathbf{q}\cos\phi.$$

In formula (A3), $\tilde{\varepsilon}^{-1}$ is the transposed tensor adjugate to ε^{-1} [9, 10].

The tensors Q' and S' are also calculated from formulas (A2) and (A3) with the tensor ε in (A3) replaced by ε'.

To find Γ for a Faraday isotropic medium, we substitute expressions (2) and $\mathbf{c} = \mathbf{b}\cos\alpha + \mathbf{a}\sin\alpha$ into (A3). We here take into account that tensors of form (A3) are planar and represent them as expansions in dyadic projectors $\mathbf{b}\otimes\mathbf{b}$, $\mathbf{b}\otimes\mathbf{a}$, $\mathbf{a}\otimes\mathbf{b}$, and $\mathbf{a}\otimes\mathbf{a}$. Bearing in mind that $\mathbf{a}\varepsilon^{-1}\cdot\mathbf{a} = a$ and $\tilde{\varepsilon}^{-1} = a^2 - b^2\mathbf{c}\otimes\mathbf{c} + iab\mathbf{c}^\times$, we find from (A3) at $\mathbf{u} = \mathbf{v} = \mathbf{e}_2$ that

$$\begin{aligned}(\mathbf{e}_2\mathbf{e}_2) = (a-\nu^2)^{-1}\{&\nu^2(a-\nu^2\sin^2\phi)\mathbf{b}\otimes\mathbf{b}\\ &- i\nu^2 b\cos\alpha\sin\phi\cos\phi(\mathbf{b}\otimes\mathbf{a} - \mathbf{a}\otimes\mathbf{b})\\ &- [(a-\nu^2)(a-\nu^2\sin^2\phi) - b^2\cos^2\alpha\sin^2\phi]\mathbf{a}\otimes\mathbf{a}\}.\end{aligned} \tag{A4}$$

At $\mathbf{u} = \mathbf{e}_2$ and $\mathbf{v} = \mathbf{e}_1$, we have

$$\begin{aligned}(\mathbf{e}_2\mathbf{e}_1) = (a-\nu^2)^{-1}\{&\nu^4\sin\phi\cos\phi\mathbf{b}\otimes\mathbf{b}\\ &+ i\nu^2 b\cos\alpha\cos^2\phi\mathbf{b}\otimes\mathbf{a} + i\nu^2 b\cos\alpha\sin^2\phi\mathbf{a}\otimes\mathbf{b}\\ &- [\sin\phi\cos\phi(\nu^2(a-\nu^2) + b^2\cos^2\alpha)\\ &+ ib\sin\alpha(a-\nu^2)]\mathbf{a}\otimes\mathbf{a}\}.\end{aligned}$$

To calculate $(\mathbf{e}_2\mathbf{e}_2)^-$, we represent tensor $(\mathbf{e}_2\mathbf{e}_2)$ (A4) as a 2×2 matrix and invert it:

$$\begin{aligned}(\mathbf{e}_2\mathbf{e}_2)^- = &(a-\nu^2\sin^2\phi - b\cos\alpha\sin\phi)^{-1}\\ &\times(a-\nu^2\sin^2\phi + b\cos\alpha\sin\phi)^{-1}\\ \times\Big\{&\frac{1}{\nu^2}[(a-\nu^2)(a-\nu^2\sin^2\phi) - b^2\cos^2\alpha\sin^2\phi]\mathbf{b}\otimes\mathbf{b}\\ &- ib\cos\alpha\sin\phi\cos\phi(\mathbf{b}\otimes\mathbf{a} - \mathbf{a}\otimes\mathbf{b})\\ &- (a-\nu^2\sin^2\phi)\mathbf{a}\otimes\mathbf{a}\Big\}.\end{aligned} \tag{A5}$$

Next, we find the product of the tensors $(\mathbf{e}_2\mathbf{e}_2)^-$ and $(\mathbf{e}_2\mathbf{e}_1)$:

$$\begin{aligned}(\mathbf{e}_2\mathbf{e}_2)^-(\mathbf{e}_2\mathbf{e}_1) &= (a-\nu^2\sin^2\phi-b\cos\alpha\sin\phi)^{-1}\\ &\times(a-\nu^2\sin^2\phi+b\cos\alpha\sin\phi)^{-1}\\ &\times\{\nu^2(a-\nu^2\sin^2\phi)\sin\phi\cos\phi\,\mathbf{b}\otimes\mathbf{b}\\ &+ib\cos\alpha\cos\phi(a\cos\phi+ib\sin\alpha\sin\phi)\mathbf{b}\otimes\mathbf{a}\\ &-i\nu^2b\cos\alpha\sin^2\phi\,\mathbf{a}\otimes\mathbf{b}+[(\nu^2(a-\nu^2\sin^2\phi)\\ &+b^2\cos^2\alpha)\sin\phi\cos\phi+ib\sin\alpha(a-\nu^2\sin^2\phi)]\mathbf{a}\otimes\mathbf{a}\}.\end{aligned} \tag{A6}$$

Clearly, the tensors Q and S (see (A2)) are represented through the integrals

$$J_{(00;\,11;\,20;\,02;\,31)} = \frac{1}{\pi}\times\int_0^{\pi}\frac{(1;\ \sin\phi\cos\phi;\ \sin^2\phi;\ \cos^2\phi;\ \sin^3\phi\cos\phi)d\phi}{(a-\nu^2\sin^2\phi-b\cos\alpha\sin\phi)(a-\nu^2\sin^2\phi+b\cos\alpha\sin\phi)}, \tag{A7}$$

with $J_{02}=J_{00}-J_{20}$ and $J_{11}=J_{31}=0$. We assume that the denominator of the integrand is nonzero for all ϕ; i.e., the tensors Q, S, and Γ are defined if

$$0\le\nu^2<\nu_{\lim}^2\equiv a-b|\cos\alpha|. \tag{A8}$$

The quantity $\nu_{\lim}$ is called the limiting frequency [6]. Physically, condition (A8) means that each of the partial waves in a Faraday medium is localized near the interface, i.e., is not a body wave. By introducing the variable of integration $x=\cot\phi$, we obtain

$$J_{(00;\,20)} = \frac{1}{\pi a^2}\int_{-\infty}^{+\infty}\frac{(1+x^2;\,1)dx}{(x-i\eta_1)(x+i\eta_1)(x-i\eta_2)(x+i\eta_2)},$$

where $\eta_{1,2}$ given by formulas (9) have the meaning of the decay coefficients of the partial waves.

For J_{00} and J_{20}, we finally arrive at

$$J_{00} = \frac{1}{a^2(\eta_1+\eta_2)}\left(1+\frac{1}{\eta_1\eta_2}\right),\qquad J_{20}=\frac{1}{a^2\eta_1\eta_2(\eta_1+\eta_2)}. \tag{A9}$$

Next, substituting formulas (A5) and (A6) into (A2) in view of (A9) yields expressions for the tensors Q and S:

$$Q = \frac{1}{a(\eta_1+\eta_2)}\left[-\frac{1}{\nu^2}(a-\nu^2+a\eta_1\eta_2)\mathbf{b}\otimes\mathbf{b}+\left(1+\frac{a-\nu^2}{a\eta_1\eta_2}\right)\mathbf{a}\otimes\mathbf{a}\right], \tag{A10}$$

$$S = \frac{ib}{a(\eta_1+\eta_2)}\left[-\cos\alpha\,\mathbf{b}\otimes\mathbf{a}+\frac{\nu^2\cos\alpha}{a\eta_1\eta_2}\mathbf{a}\otimes\mathbf{b}-\sin\alpha\left(1+\frac{a-\nu^2}{a\eta_1\eta_2}\right)\mathbf{a}\otimes\mathbf{a}\right]. \tag{A11}$$

For the tensor Q^-, we have

$$Q^- = \frac{a(\eta_1+\eta_2)}{a-\nu^2+a\eta_1\eta_2}(-\nu^2\mathbf{b}\otimes\mathbf{b}+a\eta_1\eta_2\mathbf{a}\otimes\mathbf{a}). \tag{A12}$$

From formulas (A11), (A12), and (A1), one arrives at expression (8) for the tensor Γ for a Faraday optically active medium.

The tensor Γ' for the other bordering medium can be found by substituting a' for a in (9), (A11), and (A12) and setting $b=0$. Then,

$$Q'^- = -\nu^2\sqrt{\frac{a'}{a'-\nu^2}}\mathbf{b}\otimes\mathbf{b}+\sqrt{a'(a'-\nu^2)}\mathbf{a}\otimes\mathbf{a},\quad S'=0,$$

and the tensor Γ' coincides with Q'^-.

REFERENCES

1. *Surface Polaritons*, Ed. by V. M. Agranovich and D. L. Mills (North-Holland, Amsterdam, 1982; Nauka, Moscow, 1985).
2. M. I. D'yakonov, Zh. Éksp. Teor. Fiz. **94** (4), 119 (1988) [Sov. Phys. JETP **67**, 714 (1988)].
3. N. S. Averkiev and M. I. D'yakonov, Opt. Spektrosk. **68**, 1118 (1990) [Opt. Spectrosc. **68**, 653 (1990)].
4. A. N. Darinskiĭ, Kristallografiya **46**, 916 (2001) [Crystallogr. Rep. **46**, 842 (2001)].
5. A. N. Furs and L. M. Barkovsky, Microwave Opt. Technol. Lett. **14**, 301 (1997).
6. A. N. Furs and L. M. Barkovsky, J. Opt. A: Pure Appl. Opt. **1**, 109 (1999).
7. A. N. Furs and L. M. Barkovskiĭ, Kristallografiya **46**, 1102 (2001) [Crystallogr. Rep. **46**, 1018 (2001)].
8. Yu. I. Sirotin and M. P. Shaskol'skaya, *Fundamentals of Crystal Physics* (Nauka, Moscow, 1975).
9. F. I. Fedorov, *Optics of Anisotropic Media* (Izd. Akad. Nauk BSSR, Minsk, 1958).
10. F. I. Fedorov, *Theory of Gyrotropy* (Nauka i Tekhnika, Minsk, 1976).
11. L. M. Barkovsky, G. N. Borzdov, and A. V. Lavrinenko, J. Phys. A **20**, 1095 (1987).
12. J. Lothe and D. M. Barnett, J. Appl. Phys. **47**, 428 (1976).

Translated by A. Khzmalyan

Technical Physics, Vol. 48, No. 4, 2003, pp. 393–398. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 4, 2003, pp. 17–22.

THEORETICAL AND MATHEMATICAL PHYSICS

Determination of the Volumetric Heat Generation in a Furnace Chamber from the Radiation Flux Distribution

V. K. Shiff

Vavilov State Optical Institute All-Russia Research Center, Birzhevaya Liniya 12, St. Petersburg, 193171 Russia

e-mail: shiff@vs3886.spb.edu

Received September 11, 2002

Abstract—A thermal model of a furnace chamber involving both a three-dimensional radiation heat transfer equation and an energy equation describing the one-dimensional flow of a combustible mixture is proposed. Convective heat transfer at the walls and shields is taken into account by the approximate standardized method. The model allows one to calculate the temperature and heat flux distributions both in the volume and at the boundaries of the furnace chamber. The problem of finding the specific volumetric heat generation from the radiation fluxes measured at the furnace walls is considered with this model. *© 2003 MAIK "Nauka/Interperiodica".*

A mathematically rigorous description of complex heat transfer in furnace chambers of power installations is a challenge. The reason is that, first, heat-and-mass transfer and combustion are interrelated processes and, second, there is a great number of parameters determining the thermal state of a real furnace. Therefore, it seems to be topical to improve available engineering techniques so as to limit the number of parameters (the values of temperature and heat fluxes) measured in accessible points on a real boiler unit and, thereby, adjust the model (i.e., to finally determine unknown parameters). Heat transfer parameters of both surfaces (radiant emissivity, thickness, fouling factor, etc.) and thermophysical parameters of media involved in the heat transfer process (e.g., fuel burnup and densities of furnace gases) may be unknown parameters.

In view of the fact that energy transfer by radiation makes an overwhelming contribution (up to 90%) to heat transfer in a furnace, the three-dimensional radiation transfer equation [1, 2] is solved in this paper to take into account radiation transfer. In the framework of this model, both the turbulent flow and conductive heat transfer are neglected and the flow of a combustible mixture is assumed to be one-dimensional. The radiant emissivity of a furnace chamber and the total heat generation are taken into account by the standardized method. The total heat generation in a furnace is defined by the fuel rate (B_f) and also by the sum of the heat of fuel combustion (Q_H^f) and sensible heat of both the fuel and supplied air. The lowest fuel combustion heat is estimated by the Mendeleev approximate formula [3]. The specific volumetric heat generation upon fuel combustion is an important characteristic of the furnace process. It is related to heat-and-mass exchange and depends on the composition and properties of the fuel and fuel–oxidant (air) mixture, gasdynamic processes, and the completeness of combustion. This parameter greatly influences both the temperature field in the furnace volume and the distribution of the flux incident on the furnace walls. The straightforward calculation of the specific volumetric heat generation involves particular mathematical difficulties. In power engineering, empirical dependences of the fuel burnup on the furnace height, e.g., the so-called zone method [3], are used. Therefore, finding the volumetric heat generation from experimental data is of interest. In this paper, the thermal model proposed is applied both to the direct design of the furnace chamber of a DKVP-10 steam boiler and to the inverse problem of finding the specific volumetric heat generation in the furnace chamber.

MATHEMATICAL MODEL

The mathematical model of a furnace chamber involves the radiation heat transfer equation

$$(\Omega \cdot \nabla) I(\bar{r}, \Omega) + (k_{abs} + k_{dis}) I(\bar{r}\Omega) = k_{abs} I_P[T(\bar{r})] + \frac{k_{dis}}{4\pi} \int_{4\pi} P(\Omega, \Omega') I(\bar{r}, \Omega') d\Omega' \quad (1)$$

and the energy equation

$$\mathrm{div}(c_f \rho \nu T(\bar{r}) - \lambda \mathrm{grad} T(\bar{r})) = B_f Q_H^f S(\bar{r}) - \mathrm{div} Q_r. \quad (2)$$

Here,

$$\mathrm{div} Q_r = 4\pi k_{abs} I_P[T(\bar{r})] - k_{abs} G(\bar{r}) \quad (3)$$

is the radiation flux divergence; $P(\Omega, \Omega')$ is the scattering indicatrix; $S(r)$ is the specific volumetric heat generation due to fuel combustion;

1063-7842/03/4804-0393$24.00

$$I_P[T(\bar{r})] = \frac{n^2\sigma_0 T^4(\bar{r})}{\pi}; \quad (4)$$

$I_P[T(r)]$ is the Planck's function; k_{abs} and k_{dis} are the absorption and dissipation factors, respectively; σ_0 is the Stefan–Boltzmann constant; n is the refractive index of the medium; and

$$G(r) = \int_{4\pi} I(\bar{r}, \Omega)d\Omega \quad (5)$$

is the incident power.

Boundary conditions for the transfer equation are given by

$$I(\bar{r}, \Omega) = \varepsilon I_P[T(\bar{r})]$$
$$+ \frac{1-\varepsilon}{\pi} \int_{n\cdot\Omega'>0} (n\cdot\Omega')I(\bar{r}, \Omega')d\Omega', \quad n\cdot\Omega < 0, \quad (6)$$

where ε is the radiant emissivity of the boundary surface.

Boundary conditions for the energy equation. The continuity condition for the total heat flux (Q_w) at the boundary (interior surface of the furnace wall) makes it possible to write a set of two equations for two unknowns (the temperatures at the interior and exterior surfaces of the furnace wall T_w and T_{w0}):

$$Q_w = Q_r + Q_{con} = \frac{T_w - T_{w0}}{R_w}$$
$$= \alpha_0(T_{w0} - T_{air}) + \varepsilon_0\sigma_0(T_{w0}^4 - T_{air}^4). \quad (7)$$

One more equation for one unknown T_w,

$$Q_w = Q_r + Q_{con} = \frac{T_w - T_{sw}}{R_t}, \quad (8)$$

is written for a heat baffle (tube) at a given temperature T_{sw} of the steam-and-water mixture.

The total heat flux at the furnace interior surface is equal to the sum of the resultant radiation flux

$$Q_r = \varepsilon_w Q_{r,inc} - \frac{\varepsilon_w T_w^4}{4} \quad (9)$$

and the convective flux

$$Q_{con} = \alpha_{con}[T - T_w]. \quad (10)$$

The thermal resistance of a baffle tube, R_t, can be expressed via the heat-transfer factor of the steam-and-water mixture in the tubes (α_{sw}), the thermal resistance of the tube walls (σ_m/λ_m), and the thermal resistance of a surface deposit on the tubes (E_s):

$$R_t = \frac{1}{\alpha_{sw}} + E_s + \frac{\delta_m}{\lambda_m}. \quad (11)$$

In (7)–(11), T_{air} is the temperature of surrounding air; δ_m and λ_m are the thickness and thermal conductivity of the baffle tubes; R_w is the thermal resistance of the furnace wall (lining); α_0 is the coefficient of convective heat transfer from the furnace exterior surface to the environment; α_{con} is the coefficient of convective heat transfer from the furnace medium to the furnace walls and baffle tubes; and ε_w is the radiant emissivity.

SIMPLIFICATIONS AND APPROXIMATIONS

A real furnace is approximated by a parallelepiped. A gas inlet is modeled by setting the emissivity and reflection coefficient of the corresponding part of the surface equal to zero. We assume that a baffle entirely covers the furnace lateral surface. The absorption factor of a disperse furnace medium is estimated in the framework of the gray body approximation by the formulas of the standardized method [3]:

$$K_{abs} = -\frac{\ln(1-\varepsilon_\phi)}{S_{eff}}, \quad (12)$$

$$S_{eff} = 3.6\frac{V}{F}, \quad (13)$$

where ε_ϕ is the radiant emissivity, S_{eff} is the effective thickness of the radiating layer, V is the furnace volume, and F is the total surface area of the furnace walls.

As a fuel, we consider furnace oil. When burning, furnace oil produces fine soot (the mean particle radius $r_m < 0.02$ μm) with a negligible scattering coefficient of visible and infrared radiations. Therefore, in this paper, the integral term in transfer equation (1) is omitted.

A simple model is used to describe volumetric heat generation. A furnace is conventionally divided into two parts from top to bottom. The lower zone (maximum heat generation) extends from the bottom to the section that is roughly 1.5 m higher than the burner throat axis. The majority of the fuel burns in this zone. In the upper zone, the heat generation is assumed to decrease exponentially with height. Such a model of heat generation will be described by the six-parameter one-variable function

$$S(z) = \begin{cases} a_1, & \text{for} \quad z = z_1 \\ a_2, & \text{for} \quad z = z_2 \\ a_3, & \text{for} \quad z = z_3 \\ a_4, & \text{for} \quad z = z_4 \\ a_5, & \text{for} \quad z = z_5 \\ a_5\exp[-(z-z_5)a_6], & \text{for} \quad z_5 < z < zL, \end{cases} \quad (14)$$

where zL is the furnace height.

The heat transfer factors α_{sw}, α_0, and α_{con} are estimated by the empirical formulas [3, 4]

$$\alpha_0 = \begin{cases} -6.80 + 0.054T_{w0}, & T_{w0} \geq 473 \text{ K} \\ -20.56 + 0.83T_{w0}, & T_{w0} < 473 \text{ K}, \end{cases} \quad (15)$$

$$\alpha_{con} = \begin{cases} 0.15\xi\frac{\lambda}{D}\text{Pr}^{0.33}\text{Re}^{0.43}, & \text{Re} < 2000 \\ 0.023\xi\frac{\lambda}{D}\text{Pr}^{0.4}\text{Re}^{0.8}, & \text{Re} > 2000, \end{cases} \quad (16)$$

where Pr and Re are the Reynolds and Prandtl numbers, respectively, and ξ is the correction factor [1], which depends on the relation between the height of the furnace and its effective diameter D.

SOLUTION OF THE EQUATIONS

Transfer equations. In developing numerical methods for solving the transfer equation, much effort is made to preliminarily (fully or partially) integrate it with respect to frequency and angular variables. In this paper, the gray body approximation, which implies full integration over the radiation spectrum, is used; i.e., the radiative properties of the medium are assumed to be independent of the radiation wavelength. The discrete ordinate (DO) method, first applied to this situation by Chandrasekhar [2, 3], is used in this paper to simplify the angular dependence. The DO method can be divided into two steps. First, the angular discretization of the transfer equation is performed. At the second step, for each direction defined by zenith (Θ_L) and azimuthal (φ_m) angles on a given space grid, the finite-difference approximation of the transfer equation is constructed by the integro-interpolation method [2]:

$$I^{1,m}_{i,j,k} = \frac{|\mu_m|AI^{1,m}_{i\pm1/2,j,k} + |\xi_{1,m}|BI^{1,m}_{i,j\pm1/2,k} + |\eta_{1,m}|CI^{1,m}_{i,j,k\pm1/2} + \alpha k_{abs}I_P(T)\Delta V}{|\mu_m|A + |\xi_{1,m}|B + |\eta_{1,m}|C + \alpha k_{abs}(\Delta V)}. \quad (17)$$

Here, $\mu_L = \cos\Theta_L$, $\xi_{L,m} = \sin\Theta_L\cos\varphi_m$, and $\eta_{L,m} = \sin\Theta_L\sin\varphi_m$ are angular coordinates; α is the finite-difference weighting factor ($1/2 < \alpha \leq 1$); $A = \Delta y(j)\Delta z(k)$, $B = \Delta x(i)\Delta z(k)$, and $C = \Delta x(i)\Delta y(j)$ are the surface areas of the faces of control volumes; and $\Delta V = \Delta x(i)\Delta y(j)\Delta z(k)$. The plus and minus signs in the subscripts of formula (17) depend on whether the six faces of the domain of computation are counted in the positive (+) or negative (–) direction of the corresponding coordinate axis. The indices i, j, and k take the values $i = 1.Nx$, $j = 1.Ny$, and $k = 1.Nz$, where Nx, Ny, and Nz are the numbers of the nodes of the finite-difference grid along the axes OX, OY, and OZ, respectively.

Energy equation. In this equation, we neglect heat conduction (the second term on the left of Eq. (2)) and assume that the gas flow along the OZ ($v_x = v_y = 0$) axis is one-dimensional. Then, integrating energy equation (2) over the control volume with the indices i, j, and k in view of the continuity condition turns the differential energy equation into a nonlinear algebraic equation for the temperature at the center of the control volume:

$$4K_{abs}\sigma_0 T^4_{i,j,k} = \rho_f C_p T_{i,j,k-1/2} - \rho_f C_p T_{i,j,k+1/2} + B_f Q^f_H S_{i,j,k} + K_{abs}G_{i,j,k}, \quad (18)$$

where ρ_f is the fuel rate per unit volume:

$$\rho_f = \frac{B_f}{V}. \quad (19)$$

The first and second terms on the right of Eq. (17) are the increment of the gas enthalpy. The subscripts $k - 1/2$ and $k + 1/2$ correspond to opposite faces of the control volume. The gas temperatures at the lower faces of the control volumes with indices i, j, and 1/2 adjoining the furnace bottom are assumed to be equal to $T = 300$ K (initial gas temperature). Boundary control volumes with unknown temperatures of the furnace walls or heat baffle are complemented by the set of equations (7) for the walls or by equation (8) for the baffle.

RECONSTRUCTION OF SPECIFIC VOLUMETRIC HEAT GENERATION

The parameters of specific volumetric heat generation (i.e., n-dimensional vector $A = (a_1, a_2, \ldots, a_n)$) are used as unknowns and the densities of the total radiation flux incident onto the furnace walls, as measurands. The vector $A = (a_1, a_2, \ldots, a_n)$ is found by the minimization of the function

$$F(\bar{A}) = \sum_{k=1}^{Nk}\left(Q^{exp}_{r,inc}(k) - Q^{calc}_{r,inc}(k)\right)^2, \quad (20)$$

which is the sum of the squares of the differences between measured (exact) and calculated incident radiation fluxes, and Nk is the number of radiation detectors.

The minimization of the multidimensional function is performed by a simple method that does not require calculating the derivative of the function to be minimized, i.e., by the coordinate-wise random search method. For a chosen coordinate, we calculate a random trial step. The step is accepted if the function $F(A)$ decreases. Otherwise, we reverse the step direction for this coordinate and test the variation of the function being minimized. For each coordinate, the step size increases if a series of successive steps is successful and decreases otherwise. The minimization of the func-

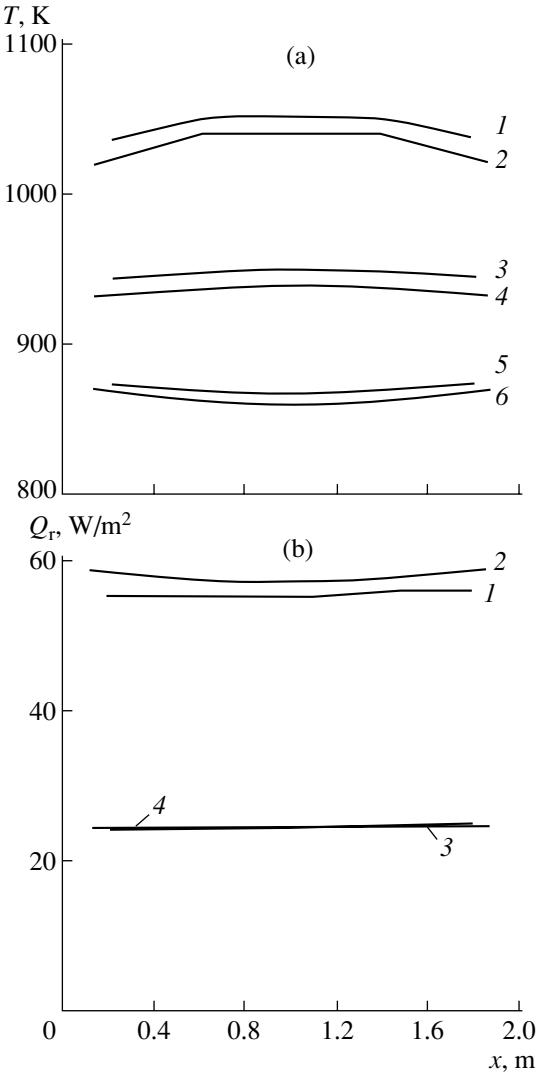

Fig. 1. (a) Temperature distribution in the plane $y = 1.0$ m along straight lines with applicates $z = (1, 2)$ 0.4, $(3, 4)$ 2.0, and $(5, 6)$ 3.6 m; $(1, 3, 5)$ zone method [2] and $(2, 4, 6)$ DO method. (b) Distribution of incident flux in the plane $y = 1.0$ m along the straight lines with applicates $z = (1, 2)$ 0 and $(3, 4)$ ZL; $(2, 4)$ DO method and $(1, 3)$ zone method [2].

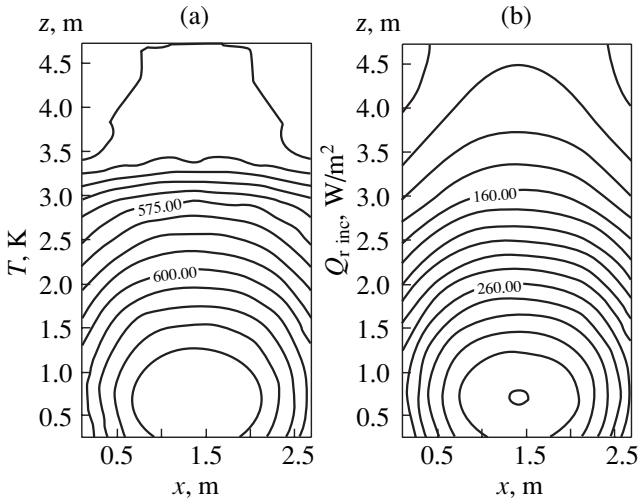

Fig. 2. (a) Isorads for radiation fluxes incident on the furnace wall ($y = YL$) and (b) isotherms on the furnace wall ($y = YL$).

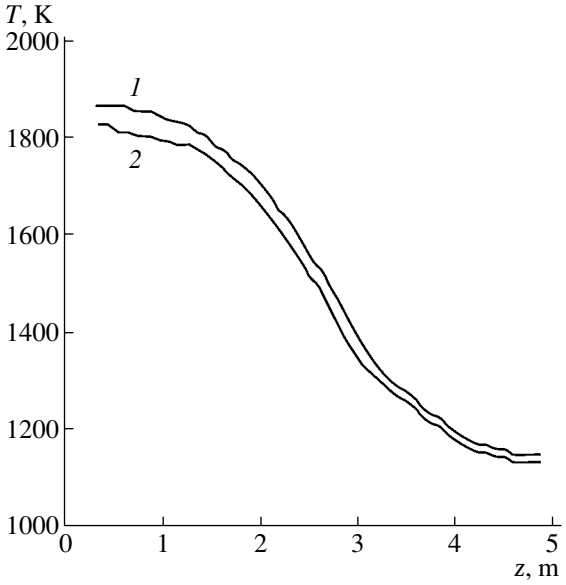

Fig. 3. Gas temperature distribution throughout the furnace height: (*1*) over the furnace surface ($x = 1.4$ m, $y = 3$ m) and (*2*) at the center of the furnace ($x = 1.4$ m, $y = 1.5$ m).

tion $F(A)$ (the search of unknown parameters) is finished when the variation of the function becomes smaller than a given value. The influence of random errors in measuring the radiation fluxes is modeled by introducing normally distributed noise into exact data:

$$\tilde{Q}^{\exp}_{\mathrm{r,\,inc}}(k) = Q^{\exp}_{\mathrm{r,\,inc}}(k) - \sigma_q(k)\zeta(k), \qquad (21)$$

$$\sigma_q(k) = \frac{Q^{\exp}_{\mathrm{r,\,inc}}(k)\gamma\%}{2.576}, \qquad (22)$$

where $\zeta(k)$ is a normally distributed random number with unit standard deviation and a zero mean value and $\sigma_q(k)$ is the root-mean-square deviation of the measured radiative heat flux at a kth detector for a relative measurement error of γ% at a 99% confidence level.

RESULTS

Comparative analysis of the transfer equation. Consider radiation transfer in an ideal furnace that has the form of a parallelepiped with sizes $XL = 2$ m, $YL = 2$ m, and $ZL = 4$ m. The temperatures and emissivities of the boundary surfaces are the following: $T = 1200$ K and $\varepsilon_{z0} = 0.85$ at $z = 0$; $T = 400$ K and $\varepsilon_{zH} = 0.70$ at $z = ZL$; and $T = 900$ K and $\varepsilon_{x0} = \varepsilon_{xL} = \varepsilon_{y0} = \varepsilon_{yL} = 0.70$ at $x = x0$, $x = xL$, $y = y0$, and $y = yL$. The specific volumetric heat generation and the absorption factor are assumed

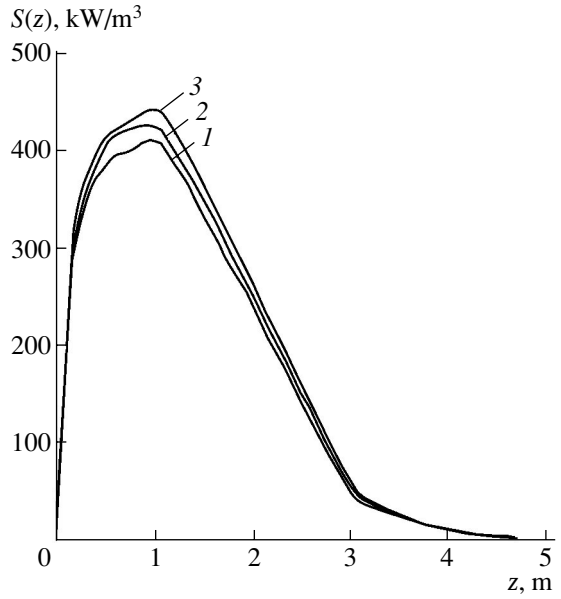

Fig. 4. Volumetric heat generation functions at $\gamma = 3\%$: (*1*, *3*) those constructed using only negative and only positive root-mean-square deviations of reconstructed parameters, respectively, and (*2*) the function constructed from the mean values of the parameters.

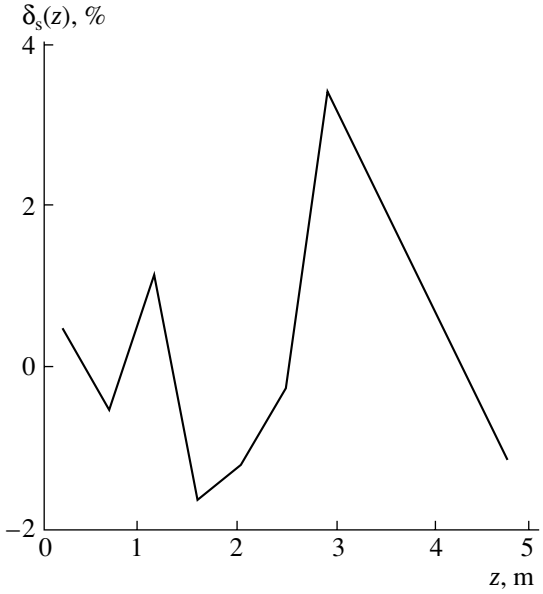

Fig. 5. Relative error of the specific volumetric heat generation as a function of height.

to be constant in the furnace volume and equal to S_v = 5.0 kW/m^3 and k_{abs} = 0.5 m^{-1}, respectively. The results found by the DO method for the temperature field and incident radiative fluxes (Figs. 1a, 1b) are in good agreement with the results of the zone method [1].

Thermal design of the furnace of a DKVR-10 steam boiler. The thermal design method is applied to a DKVR-10 industrial steam boiler. Its furnace is approximated by a parallelepiped with XL = 2.80 m, YL = 3.0 m, and ZL = 4.98 m. The gas inlet is located in the $Y = YL$ plane in the range $3.50 \le Z \le 4.98$ m. The radiant emissivities of the furnace interior surfaces are set equal to $\varepsilon_{z0} = 0.85$ and $\varepsilon_{zH} = \varepsilon_{x0} = \varepsilon_{xL} = \varepsilon_{y0} = \varepsilon_{yL} = 0.70$. The initial calculation data are listed in Table 1.

The parameters of the specific volumetric heat generation function are listed in Table 2 (second and third columns). The liquid fuel rate for the boiler is B_f = 0.193 kg/s. The compositions of the fuel and combustion products are listed in Tables 3 and 4, respectively. The computational results are presented in the form of temperature distribution and radiation fluxes both in the volume and on the walls of the furnace. Below the gas inlet on the furnace back wall ($Y = YL$) (Figs. 2a, 2b), the temperature of the interior surface and the incident heat flux decrease rapidly with increasing height. However, these parameters are almost uniform near the gas inlet. Such behavior of the thermal parameters of the furnace can be explained primarily by the heat generation model chosen, as well as by the presence of the gas inlet. The temperature along the furnace height (Fig. 3) behaves in a similar manner and coincides satisfactorily with the estimates made by the standardized method (Table 5).

Problem of reconstruction of volumetric heat generation for the DKVR-10. The results of the direct calculations of incident fluxes (see the previous section) are used as experimental values. The radiation detectors are placed at 14 points on the interior surface: six points at each of two adjacent lateral faces ($x = 0$ and $y = 0$) and one on the top and bottom faces. In measuring incident fluxes, the noise is modeled by a normal law (25 realizations). The measurement error is taken to be $\gamma = 3\%$. The mean values and the root-mean-square deviations of each ith parameter reconstructed are listed in Table 2 (columns 3, 4, and 5). In Fig. 4, the volumetric heat generation functions are shown for the mean values of the reconstructed parameters and their root-mean-square deviations. As is seen from Fig. 4, the range of the reconstructed function of volumetric heat generation can be considered satisfactory for the (rather high) flux measurement error $\gamma = 3\%$. The relative error of the reconstructed values of the volumetric heat gen-

Table 1. Initial calculation data

T_{sw}	464 K
α_{con}	(1.94–2.58) J/(m^2 s K) (T = 1000–2000) K
R_t	(0.299–3) m^2 s K/kJ
R_w	0.52 m^2 s K/kJ
E_s	0 m^2 s K/kJ
α_{sw}	1.00 kJ/(m^2 s K)

Table 2. Parameter vector for specific volumetric heat generation

N	Z_i, m	a_i	$\gamma = 3\%$		
			$\langle a_i \rangle$	$\sigma(a_i)$	δ_{air} %
1	0.226	333.2 kW/m^3	337.4 kW/m^3	12.30 kW/m^3	1.26
2	0.679	426.4 kW/m^3	414.1 kW/m^3	14.06 kW/m^3	2.88
3	1.13	412.6 kW/m^3	419.9 kW/m^3	16.42 kW/m^3	1.77
4	1.58	346.2 kW/m^3	340.4 kW/m^3	15.05 kW/m^3	1.68
5	2.94	66.5 kW/m^3	68.31 kW/m^3	58.24 kW/m^3	2.72
6	$2.94 < z < 4.80$	4.5 m^{-1}	4.605 m^{-1}	0.320 m^{-1}	2.33

Table 3. Elemental composition of the fuel

Carbon	C^c	85.9%
Hydrogen	H^c	11.5%
Nitrogen	N^c	0.26%
Oxygen	O^c	0.26%
Sulfur	S^c	2.05%
Ash content	A^c	0.1%
Humidity	W^p	3.0%

Table 4. Combustion product composition

Coefficient of excess of air	α_T	1.10	–
Theoretical volume of air	V_a	11.4	m^3/kg
Theoretical volume of chimney gases	V_g	12.3	m^3/kg
Theoretical volume of triatomic gases	V_{RO_2}	1.57	m^3/kg
Theoretical volume of steam	V_{H_2O}	1.46	m^3/kg
Theoretical volume of nitrogen	V_{N_2}	8.22	m^3/kg

Table 5. Characteristics of heat exchange in a furnace chamber (standardized method)

Theoretical combustion temperature in furnace	T_{com}	K	2253
Gas temperature outside furnace	T_{out}	K	1333
Heat of fuel combustion	Q_H^f	kJ/kg	0.339 + 5
Sensible heat of fuel	Q_{sens}	kJ/kg	0.244 + 3
Sensible heat of supplied air	Q_a	kJ/kg	0.304 + 3
Heat transferred by radiation	Q_r	kJ/kg	0.203 + 5

eration as a function of height $\delta_s(z)$ is shown in Fig. 5. The increase in the relative error may be related to two reasons: a considerable relative decrease in the volumetric heat generation at the furnace exit (Fig. 4) and the insufficient number of radiation detectors at the top of the furnace. It seems that the same reasons might explain the high value of the root-mean-square deviation for the fifth reconstructed parameter (Table 2), which is responsible for the exponential decay of the heat generation in the top region of the furnace.

CONCLUSIONS

(1) The proposed thermal model of a furnace, in contrast to the standardized method, allows one to estimate temperature fields and radiation fluxes, as well as to design small- and medium-size boiler units (for which the standardized method lacks empirical data). This technique makes it possible to estimate the most thermally stressed regions on heat-exchange surfaces.

(2) Within this thermal model, the possibility of reconstructing the volumetric heat generation profile at the relatively high errors involved in measurements of incident radiation flux is shown.

REFERENCES

1. W. A. Fiveland, Aerospace America, No. 9, 79 (1989).
2. V. K. Shiff, Opt. Zh., No. 6, 56 (2000).
3. N. V. Kuznetsov, V. V. Mitor, I. E. Mitor, and É. S. Karasina, *Thermal Design of Boilers: The Standardized Method* (Énergiya, Moscow, 1973).
4. J. Stanek, *Electric Melting of Glass* (Elsevier, Amsterdam, 1977; Moscow, 1979).

Translated by M. Fofanov

Technical Physics, Vol. 48, No. 4, 2003, pp. 399–403. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 4, 2003, pp. 23–27.

GASES AND LIQUIDS

On the Existence of Regular Structures in Liquid Human Blood Serum (Plasma) and Phase Transitions in the Course of Its Drying

T. A. Yakhno, O. A. Sedova, A. G. Sanin, and A. S. Pelyushenko

Institute of Applied Physics, Russian Academy of Sciences, Nizhni Novgorod, 603950 Russia

e-mail: tanya@awp.nnov.ru

Received September 3, 2002

Abstract—Experimental data show that regular three-dimensional liquid-crystal protein structures ranging in size from several hundredths to several tenths of a millimeter are present in the whole blood serum (plasma) of patients suffering from various diseases. When a drop of serum dries, some of the structures melt to produce a gel, whereas the rest of them undergo phase transition to form solid crystals. These crystals are shaped like immunoglobulin M molecules enlarged 1000-fold. In the serum (plasma) of ailing people, the amount of the gel formed upon drying increases, breaking the symmetry during the formation of nonequilibrium protein films. It is believed that low-intensity physical factors exert a therapeutic action by changing the phase state of protein in body fluids. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The phase composition of body fluids is vitally important for maintaining normal bodily functions [1]. The discovery of protein self-organization upon drying under nonequilibrium conditions [2, 3] has stimulated intense research on the structuring of biological fluids in naturally drying drops, which is aimed at using this phenomenon in diagnostic tests [4]. Blood plasma (serum) self-organization is an interesting and complex process involving closely related events that are the subject matter of biophysics, physical and colloid chemistry, chemical physics, immunology, and crystallography. In the previous work [5], we noted that the drops of blood serum taken from healthy people and patients suffering from viral hepatitis B and burn disease differ in the dynamic parameters of structuring upon drying. In the latter case, the time of drying is longer; the front of structuring moves at a lower rate, with the variance of this rate being greater; and the amount of gel produced is larger. When these serum samples were irradiated *in vitro* by low-intensity red or blue light at a dose of 100 J/m^2, the dynamic parameters of drying drops approached those of normal serum samples and the amount of gel decreased.

In this paper, we study phase transitions in the liquid blood serum (plasma) in greater detail, which allows us to consider their scenario from a new standpoint and hypothesize for a number of phenomena.

MATERIALS AND METHODS

The samples of blood plasma and serum were obtained from 30 clinically healthy people; 18 patients with viral hepatitis B and C in the acute stage (the material supplied by the Hepatological Center, Nizhni Novgorod); 30 patients with burn disease (supplied by the Federal Burn Treatment Center, Nizhni Novgorod Research Institute of Traumatology and Orthopedics); 40 women after normal or premature (second- and third-trimester) childbirth (supplied by the maternity and child-welfare services of Nizhni Novgorod); one patient with Waldenstrom's macroglobulinemia; and one patient with paraproteinemic hemoblastosis, whose blood contained a negligible amount of immunoglobulin M (IgM) (supplied by the Research Institute of Epidemiology and Microbiology, Nizhni Novgorod). In addition, the saliva of several patients with viral hepatitis was studied. The samples of blood plasma were analyzed before and after a freezing (–18°C)–thawing cycle in a household refrigerator.

The test fluids were applied on chemically clean glass slides either as small drops (six to eight 5-μl drops per slide) or in relatively large amounts (0.5 ml per slide). The slides were dried at room temperature for 24 h. Thus, the drying of the fluids occurred under different thermodynamic conditions depending on their volume and the form of the drops. Some of the samples (small drops) were dried under an MBS-10 microscope fitted with a television camera connected to a computer, so that phase transitions in the fluid could be recorded. After drying, the drops were studied in a Lyamam-IZ microscope under conventional illumination conditions with polarizing filters.

1063-7842/03/4804-0399$24.00 © 2003 MAIK "Nauka/Interperiodica"

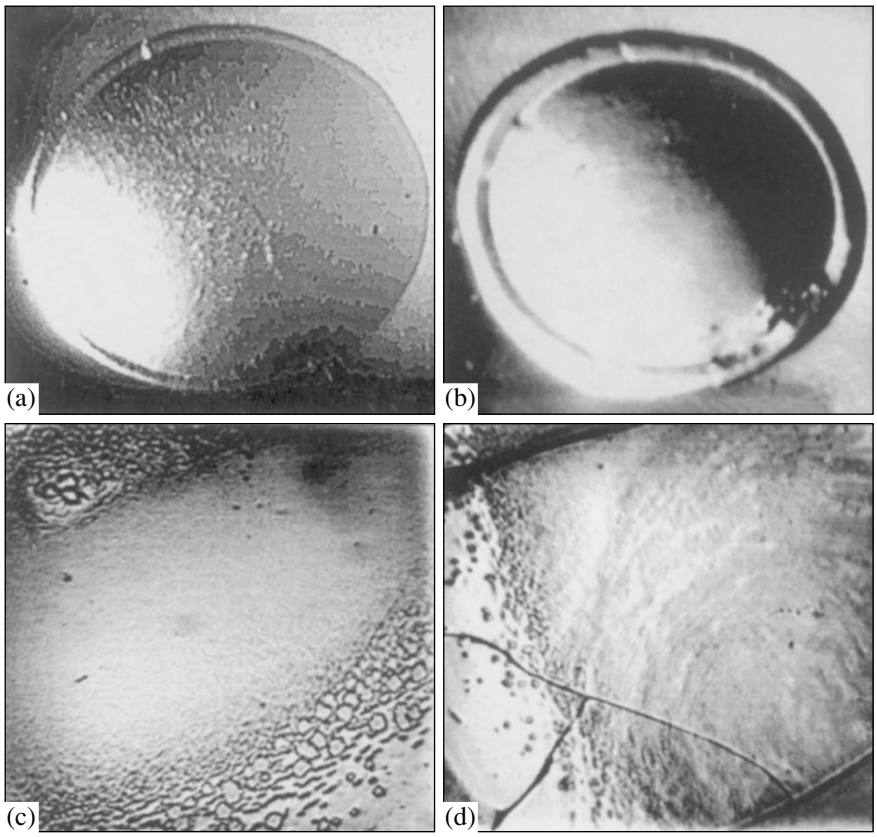

Fig. 1. Morphology of phase states in drying drops of blood serum taken from patients with different diseases: (a) exacerbation of chronic hepatitis B and C (×12); (b) viral hepatitis B, acute icteric form of medium severity (×12); (c) burn disease (×140); and (d) viral hepatitis B, acute icteric form (×140).

RESULTS AND DISCUSSION

Under the microscope, the liquid drops of serum taken from the patients with viral hepatitis or burn disease often looked rough due to the presence of small light-scattering structures (Fig. 1a). Such a phenomenon has never been observed in the sera of healthy donors. Sometimes, serum inhomogeneities were so great that they emerged on the surface, exhibiting a regular three-dimensional cluster structure (Fig. 1b). In the course of drying, some of these structures melted to produce the gel, whereas the remaining part underwent a phase transition to the solid (crystal) state (Figs. 1c, 1d). The formation of the abundant structureless gel breaks self-organization in drying drops and prevents their contents from being arranged into a regular centrosymmetric pattern, which is characteristic of the samples taken from healthy donors [4, 5]. In the plasma of women in the early postpartum period, solid and partially melted crystals usually formed a circle in the transition zone of the drop. In the cases of acute toxemia and premature childbirth, this zone was significantly wider and had larger crystals (Figs. 2a, 2b), which showed weak greenish fluorescence under polarized light (Figs. 2c, 2d). Upon slower drying (in a greater volume of the fluid), their structure persisted longer, with the shape of most of the crystals being similar to that of IgM (Fig. 3) [6] (with a 1000-fold difference in size). The same structures were also found in the saliva of some patients with viral hepatitis (at the edges of the drying drops). Phylogenetically, IgM is the most ancient class of immunoglobulins [7]. Its concentration increases primarily upon the first exposure to an antigen. An IgM molecule is a pentamer with a molecular weight of approximately 950 kDa. The IgM sedimentation constant is 7*S*, and its concentration in the blood of healthy people ranges from 0.5 to 1.9 mg/ml. A number of diseases are known to be accompanied by M hyperglobulinemia, which can be primary (e.g., in the case of Waldenstrom's disease, when only genetically modified IgM is produced) or secondary (in the cases of cancer, allergies, infectious and autoimmune diseases, hepatitis, cirrhosis of the liver, etc.) [8]. Publications available to us contain no description of regular macrostructures in the liquid phase of blood serum (plasma). In most cases, the liquid-crystal texture of biological fluids has been studied upon drying between two glass slides under conditions that are close to equilibrium [4, 9].

Both primary and secondary M hyperglobulinemias may be complicated by the deposition of amorphous

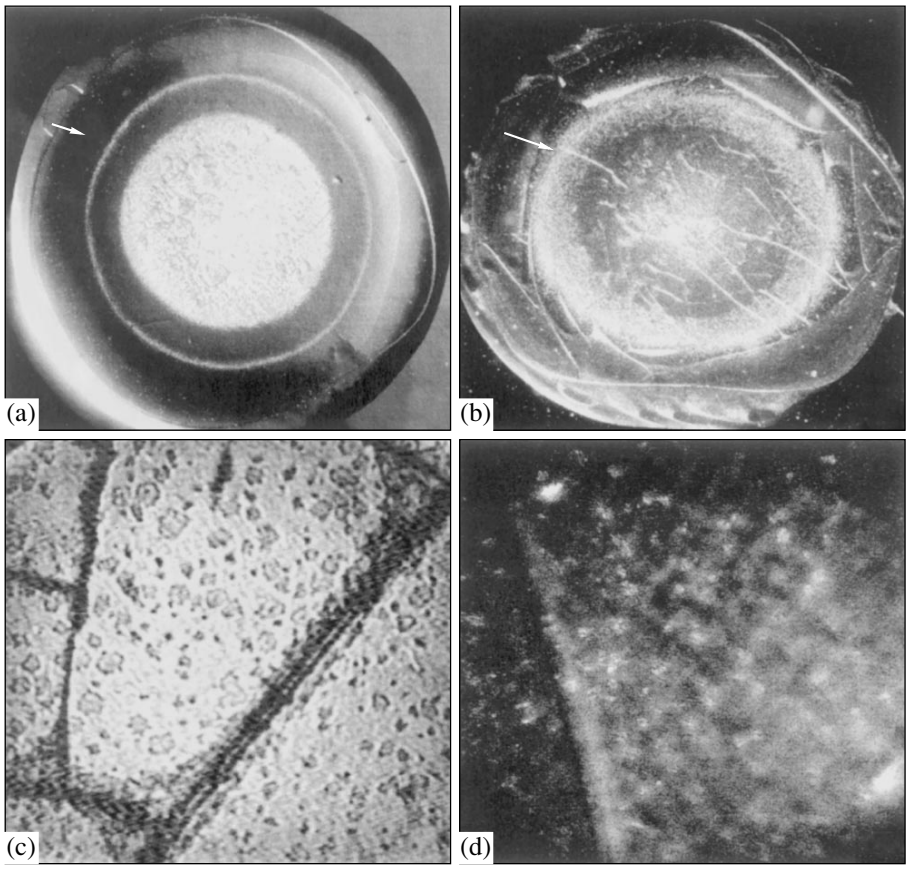

Fig. 2. Dried drops of blood plasma taken from women in the early postpartum period: (a) normal childbirth (after 40 weeks of gestation) (×17); (b) premature childbirth (after 34 weeks) (×17); (c) premature childbirth, protein crystals under normal illumination (×600); and (d) premature childbirth, protein crystals under polarized light (×600). The arrows indicate the location of protein crystals.

protein masses in blood vessels. These deposits are usually regarded as the swarms of circulating immune complexes [10, 11]. In our opinion, however, the crystallization of such complexes is unlikely. An antigen is bonded to specific sites in antibody molecules by means of ionic and hydrogen bonds, van der Waals forces, and hydrophobic interactions [12]. Bonding conditions are optimal only within the physiological ranges of pH, ionic strength, and salt concentration. When these conditions change, resulting immune complexes dissociate. Hence, it appears more probable that the liquid-crystal structures are formed by IgM molecules alone. Direct verification of this hypothesis (e.g., using fluorescein-labeled diagnostic immunoglobulins) is difficult and requires the development of new approaches. However, we obtained indirect evidence in favor of our hypothesis when analyzing the blood serum of the patient with Waldenstrom's macroglobulinemia (in this disease, functionally deficient IgM is the only immunoglobulin produced by the immune system). This serum in a tube segregated spontaneously into two transparent liquid fractions that differ in density (lower density at the top and higher density at the bottom). Upon drying, the crystals of interest were observed in the bottom fraction only, as well as in the thoroughly mixed whole serum (Fig. 3a). In the blood serum of the patient with paraproteinemic hemoblastosis, which contained only immunoglobulin IgG, no such crystals were found.

Thus, as the functioning of antibodies involves interactions at the molecular level, the apparent mesophase state of IgM in the liquid serum may evidence that IgM molecules are immunologically inactive. The same factor may be involved in pathogenesis of the endogenous intoxication syndrome [13–15].

Microcalorimetric data [16] indicated that the thermal denaturations of albumin and γ globulins, entering into the whole blood serum, occur independently: the enthalpy of γ-globulin denaturation was more than twice that of albumin denaturation. It was also found in [16] that the parameters of the thermal denaturation of proteins depend on the physiological state of the people examined [17]. These findings are consistent with the idea that γ globulins are in the mesophase state (accordingly a greater energy is needed to melt molecules of the liquid crystal) and also with our concept that several phase transitions take place in a multicomponent liquid

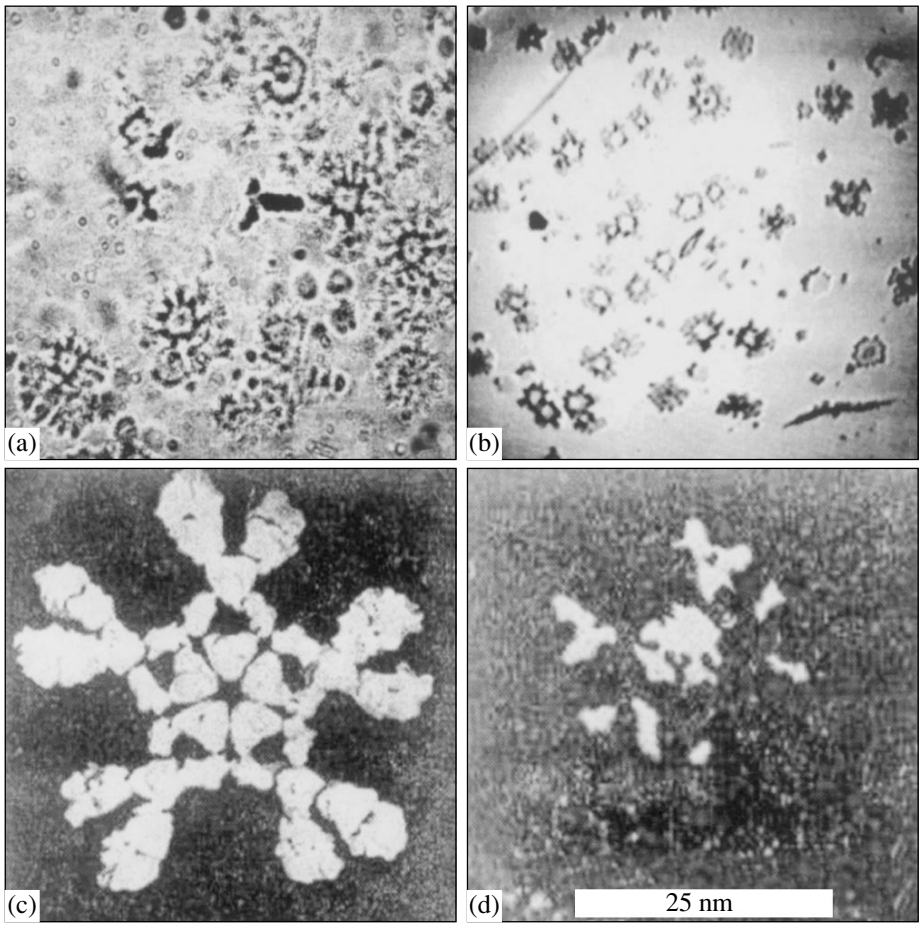

Fig. 3. Protein crystals in the dried 0.5-ml samples of blood serum from patients with (a) Waldenstrom's macroglobulinemia (×600) and (b) viral hepatitis B and chronic hepatitis C (×600). (c) Model of an IgM molecule [6] and (d) IgM molecule under an electron microscope [6].

medium subjected to a denaturing factor (high temperature, dehydration, etc.).

The freezing and thawing of the blood plasma (serum) samples taken from the patients, including women in the early postpartum period, resulted in an increase in the surface area of the drops applied on the glass slides (on average, by 24%). The same was observed when the samples were exposed *in vitro* to an eddy magnetic field [4], microwaves (in the millimeter range), and a He–Ne laser beam (our unpublished data). These facts indicate that, whatever agents act on this body fluid, its physical properties behave in a similar manner. Such behavior may be accounted for by changes in the liquid-crystal phase of protein: the macrostructures melt, antibodies recover functional activity, and the rate of their clearance (removal from circulation) increases. Recall that liquid crystals have found wide applications because of their high sensitivity to temperature and external electric fields [18]. The fact that physical therapeutic procedures influence protein liquid-crystal structures found in the blood plasma of patients may be a plausible explanation for the existence of a threshold irradiation level in He–Ne laser therapy and also for the absence of any effects when patients are irradiated after recovery [19].

Thus, the presence of regular three-dimensional liquid-crystal structures in the blood serum (plasma) of ailing people is probably explained by IgM overproduction under conditions of endogenous intoxication (i.e., in the presence of toxic metabolites in the blood, which may change the phase state of globular proteins). This entails disturbance in the cascade of protective and adaptive reactions. In such cases, measures aimed at detoxification, including physical therapy, may exert an effect via changing the phase composition of body fluids.

Research on phase transformations in protein during physiological and pathological processes holds considerable promise for both fundamental science and practical medicine. In the near future, new findings in this field will hopefully contribute to our knowledge of "mesomorphic pathogenesis" and provide a basis for novel approaches to the treatment of diseases accompanied by this phenomenon.

ACKNOWLEDGMENTS

The authors are grateful to Profs. G.Ya. Levin, O.V. Korochkina, and N.A. Egorova and also to M.M. Sheptun, T.A. Kochetkova, N.S. Antonova, V.G. Yakhno, and V.V. Nemov for valuable comments and discussions.

This work was partly supported by the Russian Foundation for Basic Research (project nos. 01-01-00388 and 02-04-49342).

REFERENCES

1. E. G. Rapis, Zh. Tekh. Fiz. **72** (4), 139 (2002) [Tech. Phys. **47**, 507 (2002)].
2. E. G. Rapis and G. Yu. Gasanova, Zh. Tekh. Fiz. **61** (4), 62 (1991) [Sov. Phys. Tech. Phys. **36**, 406 (1991)].
3. E. G. Rapis, Pis'ma Zh. Tekh. Fiz. **21** (5), 13 (1995) [Tech. Phys. Lett. **21**, 321 (1995)].
4. V. N. Shabalin and S. N. Shatokhina, *Morphology of Body Fluids* (Moscow, 2001).
5. T. A. Yakhno, V. G. Yakhno, G. Ya. Levin, *et al.*, in *Proceedings of the 4th International Conference on Mathematical Modeling, Stankin Moscow State Technol. Univ., Inst. of Math. Model., Russian Academy of Sciences, 2000*, Vol. 2, pp. 265–275.
6. *Structure and Function of Antibodies*, Ed. by L. Glynn and M. Steward (Wiley, Chichester, 1981; Mir, Moscow, 1983).
7. B. I. Kuznik, N. V. Vasil'ev, and N. N. Tsybikov, *Immunogenesis, Hemostasis, and Nonspecific Resistance of the Organism* (Meditsina, Moscow, 1989).
8. A. D. Amosov, *Hepatitis B* (Vektor Brest, Kol'tsovo, 1998).
9. R. I. Mints, S. A. Skopinov, S. V. Yakovleva, *et al.*, Pathological Physiology and Experimental Therapy, No. 6, 35 (1989).
10. L. D. Notarangelo, M. Duse, and A. G. Wgario, Immunodefic. Rev., No. 3, 101 (1992).
11. J. Levy, T. Espanol-Boren, C. Thomas, *et al.*, J. Pediatr. **131**, 47 (1997).
12. R. M. Khaitov, G. A. Ignat'ev, and I. G. Sidorovich, *Immunology* (Meditsina, Moscow, 2000).
13. N. P. Makarova and I. N. Konicheva, Anesthesiology and Intensive Care, No. 6, 4 (1995).
14. I. I. Dolgushin, L. Ya. Ébert, and R. I. Lifshits, *Trauma Immunology* (Sverdlovsk, 1989).
15. I. A. Eryukhin, B. V. Shashkov, V. F. Lebedev, *et al.*, *Endotoxicos at Severe Injury: Detoxification Therapy after Injury and Acute Surgical Diseases* (Lenizdat, Leningrad, 1989).
16. D. G. Khachidze and D. R. Monaselidze, Biophysics **45** (2), 325 (2000).
17. D. G. Khachidze and D. R. Monaselidze, Biophysics **45** (2), 320 (2000).
18. G. M. Zharkova and A. S. Sonin, *Liquid-Crystal Composites*, Ed. by V. P. Shibaev (Nauka, Novosibirsk, 1994).
19. T. Yakhno, in *Proceedings of International Workshop on Nonlinear Dynamics and Structures in Biology and Medicine: Optical and Laser Techologies, Saratov, 1996*, Vol. 3053, pp. 172–182.

Translated by N. Gorgolyuk

Technical Physics, Vol. 48, No. 4, 2003, pp. 404–414. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 4, 2003, pp. 28–37.

GASES AND LIQUIDS

Finite-Amplitude Waves on the Surface of a Viscous Deep Liquid

D. F. Belonozhko and A. I. Grigor'ev

Yaroslavl State University, Yaroslavl, 150000 Russia

e-mail: grig@uniyar.ac.ru

Received July 18, 2002; in final form, September 17, 2002

Abstract—For the first time a rigorous solution to the problem on time evolution of the periodic wave shape on the surface of a viscous infinitely deep liquid is found in the quadratic approximation with respect to the wave amplitude. It is found, in particular, that the damping rate of the quadratic component with respect to the wave amplitude is twice as high as the damping rate of the linear term. It is shown that inclusion of viscosity leads to asymmetry of the wave profile. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

In spite of continued study of finite-amplitude waves, until now all rigorous investigations were carried out within the model of an ideal liquid (see, for instance, [1–7] and the references therein). The most correct attempts at making allowance for the influence of viscosity on the nonlinear evolution of the free surface shape of a viscous liquid were performed in an approximation of low viscosity in the framework of the boundary layer theory [8–10]. Such an approach is valid only at large Reynolds numbers. Nevertheless, a correct analytical formulation of the problem of determining the shape of a wave propagating over the surface of an infinitely deep viscous liquid is quite real in a quadratic approximation with respect to the wave amplitude [11].

The problem of investigating the motion of a finite amplitude wave in a viscous liquid is relevant to numerous scientific, technical, and technological applications. For example, in [12–14] in a linear approximation in the wave amplitude, the existence of a surface instability in a viscous liquid against both elastic stresses and inactive surface-active substances (SAS) were predicted, as well as the existence of a vibrational liquid instability at a finite redistribution velocity of an electric charge over the liquid surface. A detail theoretical analysis of these effects is possible only in approximations with respect to the wave amplitude that are higher than first order. In connection with the above, the present problem was formulated.

MATHEMATICAL FORMULATION OF THE PROBLEM

Let $u = u(x, z, t)$ and $v = v(x, z, t)$ be horizontal and vertical velocity components, which, for simplicity, are assumed to be independent of the y coordinate; let $\mathbf{e}_x$ and $\mathbf{e}_z$ be unit vectors of the Ox and Oz axes. Then, wave profile $\xi = \xi(x, t)$ and velocity field $\mathbf{U} = u\mathbf{e}_x + v\mathbf{e}_z$ satisfy the following initial boundary value problem:

$$\partial_t \mathbf{U} + (\nabla \times \mathbf{U}) \times \mathbf{U} = -\nabla\left(\frac{1}{\rho}p + \frac{U^2}{2} + gz\right) + \nu\Delta\mathbf{U}; \quad (1)$$

$$\nabla\mathbf{U} = 0; \quad (2)$$

$$z = \xi: \partial_t\xi + u\partial_x\xi = v; \quad (3)$$

$$p - 2\rho\nu\mathbf{n}\cdot((\mathbf{n}\cdot\nabla)\mathbf{U}) - P_0 = \gamma\nabla\cdot\mathbf{n}; \quad (4)$$

$$\boldsymbol{\tau}\cdot((\mathbf{n}\cdot\nabla)\mathbf{U}) + \mathbf{n}\cdot((\boldsymbol{\tau}\cdot\nabla)\mathbf{U}) = 0; \quad (5)$$

$$z \longrightarrow -\infty: \mathbf{U} \longrightarrow 0; \quad (6)$$

$$t = 0: \xi = F(x); \quad (7)$$

$$z \le \xi: \mathbf{U} = \mathbf{U}^0 = \mathbf{U}^0(x, z) = u^0(x, z)\mathbf{e}_x + v^0(x, z)\mathbf{e}_z. \quad (8)$$

Here, t is time; p is the liquid pressure; ∂_t and ∂_x are the partial derivatives with respect to time and coordinate; $\boldsymbol{\tau}$ and $\mathbf{n}$ are the unit vectors tangent and normal to the liquid surface, respectively (explicit expressions for them and for divergence of the normal, div($\mathbf{n}$), are given in Appendix A); and Δ is Laplacian. The awkwardness of the problem (1)–(8) is directly related to the complexity of initial conditions (8). The particular form of the functions appearing in (8) is rather arbitrary. Therefore, in the present paper, initial conditions yielding the least awkward solution are chosen. In order to maintain orderliness of the reasoning, we shall refine the form of initial conditions in the course of solving the problem when it becomes clear how the initial conditions influence the awkwardness of the solution.

PRINCIPLE OF SOLVING THE PROBLEM

We use the perturbation method to find an analytical solution to problem (1)–(8). Usually in this method, equations are first written in dimensionless form. In the

present study, we do not specify a particular method of this reduction. Let an initial perturbation be periodic in x and form a wavy profile with a wavelength of $\lambda = 2\pi/k$ and an amplitude of a. Dimensionless parameter $\varepsilon = ka$ is independent of the method of reduction to dimensionless form, and its small value corresponds to the small amplitude approximation. In this study, all relationships are written in dimensional form. In reducing to dimensionless form, the parameter $a = \varepsilon k^{-1}$ and all variables proportional to it are transformed into dimensionless variables on the order of $O(\varepsilon)$. In dimensional expressions, such variables are related to the same order of magnitude.

At small ε, a solution of the problem is sought in the form

$$\begin{gathered}\mathbf{U} = \mathbf{U}_1 + \mathbf{U}_2 + O(\varepsilon^3); \quad \mathbf{U}_1 = O(\varepsilon); \quad \mathbf{U}_2 = O(\varepsilon^2);\\ p = p_0 + p_1 + p_2 + O(\varepsilon^3); \quad p_0 = O(1);\\ p_1 = O(\varepsilon); \quad p_2 = O(\varepsilon^2);\\ \xi = \xi_1 + \xi_2 + O(\varepsilon^3); \quad \xi_1 = O(\xi); \quad \xi_2 = O(\varepsilon^2).\end{gathered} \tag{9}$$

Let the initial perturbation profile of the free liquid surface be a sine wave in the first approximation with respect to ε:

$$F = F_1 + F_2 + O(\varepsilon^3); \quad F_1 = a\cos(kx); \quad F_2 = O(\varepsilon^2).$$

The forms of $F_2(x)$ and functions involved in an asymptotic representation of the initial distribution of the velocity field,

$$\mathbf{U}^0 = \mathbf{U}_1^0 + \mathbf{U}_2^0 + O(\varepsilon^3); \quad \mathbf{U}_1^0 = O(\varepsilon); \quad \mathbf{U}_2^0 = O(\varepsilon^2),$$

are chosen in order to find the least awkward solution.

Substituting (9) into (1) and (2) leads us to the zeroth-, first-, and second-order problems for these equations. In Appendix B, a procedure is described that allows one to decompose boundary conditions (3)–(5) into relationships for quantities of different orders of smallness in view of the expansions of these conditions in the vicinity of the unperturbed plane liquid surface.

In the zero-order approximation with respect to ε, the problem is reduced to determining the hydrostatic pressure:

$$p_0 = P_0 - \rho g z.$$

In what follows, we shall use special notation for linear differential operators,

$$\mathfrak{L} \equiv \begin{bmatrix} \partial_t - \nu(\partial_{xx} + \partial_{zz}) & 0 & (1/\rho)\partial_x \\ 0 & \partial_t - \nu(\partial_{xx} + \partial_{zz}) & (1/\rho)\partial_x \\ \partial_x & \partial_z & 0 \end{bmatrix};$$

$$\mathfrak{R} \equiv \begin{bmatrix} \partial_t \\ -\rho g + \gamma\partial_{xx} \\ 0 \end{bmatrix}; \quad \mathfrak{B} \equiv \begin{bmatrix} 0 & -1 & 0 \\ 0 & -2\rho\nu\partial_z & 1 \\ \partial_z & \partial_x & 0 \end{bmatrix}_{z=0}$$

and for column matrices,

$$\hat{0} \equiv \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}; \quad \hat{Y}_j \equiv \begin{bmatrix} u_j \\ v_j \\ p_j \end{bmatrix}.$$

Operator $\mathfrak{B}$ transforms objects of $\hat{Y}_j$ type in the following way: first, matrix operations are executed; then, all differentiation and arithmetic operations are performed; and after this, it is assumed that $z = 0$. The result of application of operator $\mathfrak{B}$ to a column vector consisting of three functions depending on three variables x, z, and t is a vector consisting of three functions independent of z.

THE PROBLEM OF THE FIRST ORDER OF SMALLNESS

For quantities of the first order of smallness in ε, the complete mathematical formulation of the problem has the form

$$\mathfrak{L}\hat{Y}_1 = \hat{0}; \tag{10}$$

$$\mathfrak{B}\hat{Y}_1 + \mathfrak{R}\xi_1 = \hat{0}; \tag{11}$$

$$z \longrightarrow -\infty: u_1 \longrightarrow 0; \quad v_1 \longrightarrow 0; \tag{12}$$

$$t = 0: \xi_1 = a\cos(kx); \quad z \le 0: \mathbf{U}_1 = \mathbf{U}_1^0. \tag{13}$$

Here, the initial condition has been formulated for the two first elements of $\hat{Y}_1$ but not for the entire symbolic vector of unknown values. In the expressions forming the initial conditions, there are no conditions for p_1 (the third element of this vector). Actually, values of p_1 at any instant of time are expressed through components of the $\mathbf{U}_1$ field in the liquid and on its free surface. Indeed, the linearized Navier–Stokes equation (the first equation of (10)) can be written in the form

$$\partial_t \mathbf{U}_1 = -\nabla\left(\frac{1}{\rho}p_1\right) + \nu\Delta\mathbf{U}_1.$$

If one applies divergence to both parts of this equation, taking into account that the liquid is incompressible, that is, $\mathrm{div}(\mathbf{U}_1) = 0$, and that the order of successive partial differentiation can be interchanged, then the Laplacian equation for p_1 is easily found:

$$\Delta p_1 = 0.$$

The linearized boundary condition for the normal stress and the condition of vanishing of the gradient of

a first-order additive to pressure at a large depth have the form

$$z = 0: p_1 = 2\rho\nu\partial_z u_1 + \partial_{xx}\xi_1;$$
$$z \longrightarrow -\infty: |\nabla p_1| \longrightarrow 0.$$

It is seen that, if field $\mathbf{U}_1$ is known on the liquid surface and in the liquid volume and an expression for ξ_1 is given, then p_1 satisfies the Dirichlet problem in an infinite domain (there exists a unique solution for this problem). The above considerations justify the absence of a condition for p_1 in the statement of problem (10)–(13).

Following [15], a complex solution of problem (10)–(13) is readily found and then, having separated the real part, the following solution of the first-order problem is obtained:

$$\xi_1 = a\cos(\Theta)\exp(T); \tag{14}$$

$$\begin{aligned} u_1 = a((&S_2\exp(kz) - 2\nu k(q_2\cos(q_2 z) + q_1\sin(q_2 z)) \\ &\times \exp(q_1 z))\cos(\Theta) + (D\exp(kz) \\ &- 2\nu k(q_1\cos(q_2 z) - q_2\sin(q_2 z)) \\ &\times \exp(q_1 z))\sin(\Theta))\exp(T); \end{aligned} \tag{15}$$

$$\begin{aligned} v_1 = a((&D_2\exp(kz) - 2\nu k^2\cos(q_2 z)\exp(q_1 z))\cos(\Theta) \\ &- (S_2\exp(kz) - 2\nu k^2\sin(q_2 z))\exp(q_1 z))\sin(\Theta))\exp(T); \end{aligned} \tag{16}$$

$$\begin{aligned} p_1 = a\rho k^{-1}((&-S_1 D + S_2^2)\cos(\Theta) \\ &+ 2S_2(S_1 + \nu k^2)\sin(\Theta))\exp(kz)\exp(T); \end{aligned} \tag{17}$$

$$\nu^2(k^2 + q^2)^2 - 4\nu^2 k^3 q + \left(kg + \frac{k^3\gamma}{\rho}\right) = 0; \tag{18}$$

$$q_1 = \mathrm{Re}q \geq 0; \quad q_2 = \mathrm{Im}(q) \geq 0; \tag{19}$$

$$S = \nu(q^2 - k^2); \quad S_1 = \mathrm{Re}S; \quad S_2 = \mathrm{Im}S; \tag{20}$$

$$\Theta = S_2 t - kx; \quad T = S_1 t; \quad D = S_1 + 2\nu k^2. \tag{21}$$

Here, q is calculated as a root of disperse equation (18) obeying conditions (19). The first of them necessarily follows from (12) and the second means that a traveling wave propagating along Ox is chosen as a solution. It is known (see, for example, [15, 16]) that only one root of Eq. (18) satisfies such conditions. This provides uniqueness of the procedure for calculation of the complex frequency S.

Using solution (14)–(17), we construct the substitution

$$\xi = a\cos(\Theta)\exp(T) + \xi_1^*; \quad \hat{Y}_1^* = \begin{bmatrix} u_1 \\ v_1 \\ p_1 \end{bmatrix} + \begin{bmatrix} u_1^* \\ v_1^* \\ p_1^* \end{bmatrix},$$

reducing (10)–(13) to a problem with the initially unperturbed surface,

$$\begin{aligned} &\mathfrak{Q}\hat{Y}_1^* = \hat{0}; \quad \mathfrak{B}_1\hat{Y}_1^* + \mathfrak{R}\xi_1^* = \hat{0}; \\ &z \longrightarrow -\infty: u_1 \longrightarrow 0; \quad v_1 \longrightarrow 0; \\ &t = 0: \xi_1^* = 0; \quad z \leq 0: \mathbf{U}_1^* = \mathbf{U}_1^0 - \mathbf{U}_1. \end{aligned} \tag{22}$$

The shape of the free liquid surface is represented in the form of the superposition of function ξ_1 coinciding with the initial surface perturbation at $t = 0$ and function $\xi_1^* = 0$ coinciding with the equilibrium liquid surface at the initial time instant ($\xi_1^* = 0$ at $t = 0$).

With the aim of obtaining the least awkward solution, it is reasonable to restrict our consideration to a condition of the absence of the initial velocity distribution in problem (22):

$$t = 0: \mathbf{U}_1^0 - \mathbf{U}_1 = 0. \tag{23}$$

From this it follows that problem (22) has a zero solution and relationships (14)–(21) are the solution of the first-order problem with initial condition (23), where components $\mathbf{U}_1^0$ are calculated using formulas (14) and (15) at $t = 0$.

THE SECOND-ORDER PROBLEM

Problem-solving outline. After decomposition of Eqs. (1) and (2) into relationships for quantities of different orders of smallness, one can obtain the following equation for the second-order quantities with the help of expression (9):

$$\begin{aligned} &\partial_t\mathbf{U}_2 + \nabla\left(\frac{1}{\rho}p_2\right) - \nu\Delta\mathbf{U}_2 \\ &= -\frac{1}{2}\nabla(U_1^2) - (\nabla\times\mathbf{U}_1)\times\mathbf{U}_1; \end{aligned} \tag{24}$$

$$\mathrm{div}\mathbf{U}_2 = 0. \tag{25}$$

Boundary conditions (72), (78), and (81) for them are obtained in Appendix B. All these relationships involve terms depending on a product of first-order magnitudes. These terms transform into particular expressions after the solution of first-order problem (14)–(21) is substituted into them. As a result, problem (24), (25), (72), (78), and (81) with the initial conditions as yet undetermined can be formulated in a new form:

$$\begin{aligned} \mathfrak{Q}\hat{Y}_2 = a^2\mathrm{Re}((&\hat{A}_1\exp(2q_1 z) + \hat{A}_2\exp(kz) \\ &+ \hat{A}_3\exp((k+q)z))\exp(2T) \\ &+ \hat{A}_4\exp((k+q)z)\exp(2(T + i\Theta))); \end{aligned} \tag{26}$$

$$\mathfrak{B}\hat{Y}_2 + \mathfrak{R}\xi_2 = a^2\mathrm{Re}(\hat{A}_5\exp(2T) + \hat{A}_6\exp(2(T+i\Theta))); \quad (27)$$

$$z \longrightarrow -\infty\colon u_2 \longrightarrow 0; \quad v_2 \longrightarrow 0; \quad (28)$$

$$t = 0\colon \xi_2 = F_2(x); \quad z \le 0\colon \mathbf{U}_2 = \mathbf{U}_2^0, \quad (29)$$

where $\hat{A}_1,\ldots,\hat{A}_6$ are three-component columns with complex coefficients independent of coordinates and time. Their expressions are given in Appendix C.

If one finds a particular solution $\hat{Y}^*, \xi^*$ of problem (26)–(29) obeying arbitrary initial conditions and uses the substitution

$$\hat{Y}_2 = \hat{Y}_* + \hat{Y}^*; \quad \hat{Y}_* = \begin{bmatrix} u_* \\ v_* \\ p_* \end{bmatrix}; \quad \hat{Y} = \begin{bmatrix} u^* \\ v^* \\ p^* \end{bmatrix}; \quad \xi_1 = \xi_* + \xi^* \quad (30)$$

this inhomogeneous problem is transformed into the homogeneous problem

$$\mathfrak{L}\hat{Y}_* = \hat{0}; \quad (31)$$

$$\mathfrak{B}\hat{Y}_* + \mathfrak{R}\xi_* = 0; \quad (32)$$

$$z \longrightarrow -\infty\colon u_* \longrightarrow 0; \quad v_* \longrightarrow 0; \quad (33)$$

$$t = 0\colon \xi_* = F_2(x) - \xi^*; \quad z \le 0\colon u_* = u_2^0 - u^*; \quad v_* = v_2^0 - v^*. \quad (34)$$

This problem can be solved in the same way as the first-order problem.

Then, it is required to find a particular solution $\hat{Y}^*$ and ξ^* of problem (26)–(28), and it is not necessary to write the initial conditions or worry about the completeness of the solution. In addition, due to the linearity of relationships (26)–(28), it is possible to omit the sign of Re and seek a particular solution $\hat{Y}^*$ and ξ^* in complex-valued form, considering the real part of the complex-valued solution as a physical solution.

Formulation of an auxiliary problem. Based on the structure of the right-hand side of (26), one can try to seek a particular solution of this equation in the form

$$\hat{Y} = a^2((\hat{C}_1\exp(2q_1z) + \hat{C}_2\exp(2kz) + \hat{C}_3\exp((k+q)z))\exp(2T) + \hat{C}_4\exp(k+q)z)\exp(2(T+i\Theta)). \quad (35)$$

Substitution of (35) into (26) leads us to an inhomogeneous set of linear algebraic equations for determining the coefficients forming columns $\hat{C}_1$–$\hat{C}_4$:

$$\hat{\Pi}_1C_1 = \hat{A}_1; \quad \hat{\Pi}_2C_2 = \hat{A}_2; \quad \hat{\Pi}_3C_3 = \hat{A}_3; \quad \hat{\Pi}_4C_4 = \hat{A}_4. \quad (36)$$

Matrices $\hat{\Pi}_1$, $\hat{\Pi}_2$, $\hat{\Pi}_3$, and $\hat{\Pi}_4$ are given in Appendix C.

Having found (35), it is easy to make the substitution of variables

$$\hat{Y}_2 = \hat{y}^* + \hat{y}; \quad \hat{y}^* = \begin{bmatrix} u_2^* \\ v_2^* \\ p_2^* \end{bmatrix}, \quad (37)$$

which transforms the problem into a simpler one:

$$\mathfrak{L}\hat{y}^* = 0; \quad (38)$$

$$\mathfrak{B}\hat{y}^* + \mathfrak{R}\xi_2 + \mathfrak{B}\hat{y} = a^2(\hat{A}_5\exp(2T) + \hat{A}_6\exp(2(T+i\Theta))); \quad (39)$$

$$z \longrightarrow -\infty\colon u_2^* \longrightarrow 0; \quad v_2^* \longrightarrow 0. \quad (40)$$

Now, we have homogeneous set of Eqs. (38) instead of inhomogeneous set (26). The simplification lies precisely in this. Similar to the first-order problem, the solution of problem (38) is expressed through scalar functions

$$\hat{y}^* = a^2\begin{bmatrix} \partial_x\varphi_2 - \partial_z\psi_2 \\ \partial_z\varphi_2 + \partial_x\psi_2 \\ f(t) - \rho\partial_t\varphi_2 \end{bmatrix}, \quad (41)$$

satisfying the equations

$$\Delta\varphi_2 = 0; \quad \partial_t\psi_2 - \nu\Delta\psi_2 = 0 \quad (42)$$

under the following restrictions:

$$z \longrightarrow -\infty\colon \partial_x\varphi_2 - \partial_z\psi_2 \longrightarrow 0; \quad \partial_z\varphi_2 + \partial_x\psi_2 \longrightarrow 0. \quad (43)$$

In considering the correctness of the transition to φ_2 and ψ_2, we use only properties of Eqs. (38), disregarding other relationships. This is why it is possible to use the scalarization procedure developed in solving the first-order problem. Such an approach is inapplicable in solving problem (26)–(28) because of the presence of the right-hand side in (26).

Insolvability of the auxiliary problem by the method of separation of variables. In order to satisfy relationships (39)–(40), it is necessary to look for φ_2 and ψ_2 in the form

$$\varphi_2 = \alpha(z)\exp(2(T+i\Theta)); \quad \psi_2 = \beta(z)\exp(2(T+i\Theta)).$$

Substitution of these expressions into (37) yields equations in $\alpha(z)$ and $\beta(z)$. Solving these equations, we readily find

$$\begin{aligned}\varphi_2(z) &= \lambda\exp(kz)\exp(2(T+i\Theta));\\ \psi_2(z) &= H\exp(rz)\exp(2(T+i\Theta));\end{aligned} \tag{44}$$

$$r = \sqrt{2(k^2+q^2)};\quad \mathrm{Re}(r) > 0. \tag{45}$$

The found functions φ_2 and ψ_2 satisfy Eqs. (42) at any values of complex-valued constants λ and H. Thus, after substituting (44) into (41), the resulting column vector

$$\begin{aligned}\hat{y}^* = a^2\begin{bmatrix}0\\0\\f(t)\end{bmatrix} &+ a^2 2k\begin{bmatrix}-i\\1\\2\rho S\end{bmatrix}\lambda\exp(kz)\exp(2(T+i\Theta))\\ &+ a^2\begin{bmatrix}-r\\-2ik\\0\end{bmatrix}H\exp(rz)\exp(2(T+i\Theta))\end{aligned} \tag{46}$$

is the solution of (38) regardless of the value of H and λ.

In order to find the expression in the left-hand side of (39) in the form of the right-hand side after substituting (46) into (39), one may seek a particular solution ξ^* for variable ξ_2 in the form

$$\xi^* = a^2\zeta\exp(2(T+i\Theta)). \tag{47}$$

After substituting (46) and (47) into (39), the left-hand side of (39) takes the form

$$\begin{aligned}&\hat{y}^* + \Re\xi^* + \mathfrak{B}\hat{y}\\ &= a^2\begin{bmatrix}0\\f(t)\\0\end{bmatrix} + a^2\begin{bmatrix}0\\C_1[3]+C_2[3]+C_3[3]\\\Lambda\end{bmatrix}\exp(2T)\\ &+ \hat{a}^2\hat{L}\left(\begin{bmatrix}H\\\lambda\\\zeta\end{bmatrix} - \hat{B}\right)\exp(2(T+i\Theta)),\end{aligned} \tag{48}$$

where

$$\Lambda = 2C_1[1]q_1 + C_3[1](k+q)$$

and the right-hand side is

$$\begin{aligned}&a^2(\hat{A}_5\exp(2T) + \hat{A}_6\exp(2(T+i\Theta)))\\ &= a^2\begin{bmatrix}0\\N_1\\M_0\end{bmatrix}\exp(2T) + a^2\begin{bmatrix}\Omega\\N\\M\end{bmatrix}\exp(2(T+i\Theta)).\end{aligned} \tag{49}$$

The form of square matrix $\hat{L}$ is given in Appendix C. In the same place, three-element vector $\hat{B}$ is written in explicit form with the coefficients depending on the elements of vector C_4. Hereafter, the mth element of vector C_n is denoted by $C_n[m]$. Expressions for Ω, M_0, M, N_1, and N are also given in Appendix C. If one focuses on the fact that $\Theta = S_2t - kx$, it becomes clear that equalities (48) and (49) at any values of x and t mean the equalities of corresponding real and imaginary parts at any Θ. As a result, we obtain equalities of Fourier series in sines and cosines of variable Θ. They are valid only if the corresponding coefficients of the Fourier expansions are equal to each other. Therefore, the problem is solvable if the equality $\Lambda = M_0$ is satisfied. It is readily verified that, if this equality is fulfilled, then one can equalize (48) and (49) by means of selecting constants H, λ, and ζ.

In the general case, $\Lambda \neq M_0$, and seeking a solution of (38)–(40) in the form (41) results in an insolvable problem.

Modified auxiliary problem and its solution by the method of separation of variables. A solution of the problem can be found if one considers the following auxiliary problem:

$$\mathfrak{L}\hat{y}_h = 0; \tag{50}$$

$$\begin{aligned}\mathfrak{B}\hat{y}_h + \Re\xi_h &= a^2(\hat{A}_5^*\exp(2T) + \hat{A}_6^*\exp(2(T+i\Theta)))\\ &- \mathfrak{B}\hat{y} + \begin{bmatrix}0\\0\\\Lambda - M_0\end{bmatrix};\end{aligned} \tag{51}$$

$$z \longrightarrow -\infty\colon u_h \longrightarrow 0;\quad v_h \longrightarrow 0. \tag{52}$$

Reasoning in the same way as in solving (38)–(40), we come to the step at which, in the previous case, condition $\Lambda \neq M_0$ appeared to transform the problem into an insolvable one. After substituting

$$\begin{aligned}\hat{y}_h &= a^2\begin{bmatrix}0\\0\\f(t)\end{bmatrix}\\ &+ a^2 2k\begin{bmatrix}-i\\1\\-2\rho S\end{bmatrix}\lambda\exp(kz)\exp(2(T+i\Theta))\\ &+ a^2\begin{bmatrix}-r\\-2ik\\0\end{bmatrix}H\exp(rz)\exp(2(T+i\Theta));\end{aligned} \tag{53}$$

$$\xi_h = a^2\zeta\exp(2(T+i\Theta)) \tag{54}$$

into (50)–(52), we obtain the set of equalities

$$\begin{aligned}\Lambda &= \Lambda;\\ f(t) &= (N_1 - C_1[3] - C_2[3] - C_3[3])\exp(2T);\end{aligned} \tag{55}$$

$$\hat{L}\begin{bmatrix} H \\ \lambda \\ \zeta \end{bmatrix} = \begin{bmatrix} C_4[2] + \Omega \\ 2C_4[2]\rho\nu(k+q) - C_4[3] + N \\ 2iC_4[2]k - C_4[1](k+q) + M \end{bmatrix}. \quad (56)$$

The first of them is an identity. The second uniquely determines $f(t)$. Equality (56) is an inhomogeneous set of linear algebraic equations in H, λ, and ζ. The form of matrix $\hat{\Lambda}$ and expressions for Ω, N, and M are written in Appendix C. Finally, H, λ, and ζ that are found should be substituted into (53) and (54).

The second auxiliary problem and its solution. The following change of the variables,

$$\hat{Y}_2 = \hat{Y}_\alpha + \hat{y} + \hat{y}_h; \quad \xi_2 = \xi_\alpha + \xi_h$$

transforms (26)–(28) into the following problem:

$$\mathfrak{L}\hat{Y}_\alpha = 0; \quad (57)$$

$$\mathfrak{B}\hat{Y}_\alpha + \mathfrak{R}\xi_\alpha = a^2\begin{bmatrix} 0 \\ 0 \\ M_0 - \Lambda \end{bmatrix}\exp(2T); \quad (58)$$

$$z \longrightarrow -\infty: u_\alpha \longrightarrow 0; \quad v_\alpha \longrightarrow 0. \quad (59)$$

If one rewrites these equations in the standard form,

$$\partial_t \mathbf{U}_\alpha + \nabla\left(\frac{1}{\rho}p_\alpha\right) - \nu\Delta\mathbf{U}_\alpha = 0; \quad \nabla\cdot\mathbf{U}_\alpha = 0;$$

$$z = 0: \partial_t\xi_\alpha - v_\alpha = 0; \quad p_\alpha - 2\rho\nu\partial_z v_\alpha + \partial_{xx}\xi_\alpha = 0;$$

$$\partial_z u_\alpha + \partial_x v_\alpha = a^2(M_0 - \Lambda)\exp(2T);$$

$$z \longrightarrow -\infty: u_\alpha \longrightarrow 0; \quad v_\alpha \longrightarrow 0,$$

then one can see that the problem has a particular solution with $\xi_\alpha = 0$, $p_\alpha = 0$, and $v_\alpha = 0$. Indeed, in this case, from the discontinuity equation it follows that $\partial_x u_\alpha \equiv 0$ and, consequently, u_α may depend only on z and t. As a result, the problem is transformed into a well-known problem of mathematical physics,

$$\partial_t u_\alpha - \nu\partial_{zz}u = 0; \quad (60)$$

$$z = 0: \partial_z u_\alpha = a^2(M_0 - \Lambda)\exp(2S_1 t) \quad (61)$$

with an additional condition,

$$z \longrightarrow -\infty: u_\alpha \longrightarrow 0. \quad (62)$$

This problem has a solution satisfying the initial condition

$$t = 0: u_\alpha = 0; \quad v_\alpha = 0. \quad (63)$$

According to [12], such a solution is

$$u_\alpha = a^2(M_0 - \Lambda)\sqrt{\frac{\nu}{\pi}} \times \int_0^t \exp\left(-\frac{z^2}{4\nu(t-\tau)}\right)\exp(2S_1\tau)d\tau. \quad (64)$$

It is easily seen that condition (62) is fulfilled in this case.

Results of solving the second-order problem. The reasoning presented allows us to formulate an algorithm for calculating the variables:

$$\xi^* = a^2(\zeta_1\cos(2\Theta) - \zeta_2\sin(2\Theta))\exp(2S_1 t); \quad \zeta_1 = \mathrm{Re}(\zeta), \quad \zeta_2 = \mathrm{Im}(\zeta); \quad (65)$$

$$\hat{Y}^* = \mathrm{Re}(\hat{y} + \hat{y}_h) + \begin{bmatrix} u_\alpha \\ 0 \\ 0 \end{bmatrix}. \quad (66)$$

The set of linear algebraic equations (36) allows one to determine vectors $\hat{C}_1$, $\hat{C}_2$, $\hat{C}_3$, and $\hat{C}_4$. Using them with the help of (35), one can construct $\hat{y}$. The elements of vector C_4 and values Ω, N, and M calculated with formulas from Appendix C are used for the construction of the right-hand side of (56). The inhomogeneous set of linear algebraic equations (56) is solved for H, λ, and ζ. Function $f(t)$ is calculated using formula (55), and vector $\hat{y}_h$ is defined by formula (53). Values ζ, $\hat{y}$, and $\hat{y}_h$ found in such a way, together with expression (64) for u_α, are substituted into (65) and (66). The separate steps of the algorithm represent a solution of the sets of linear equations and can be performed both analytically and numerically. Situations where no solution exists for any set are not considered here.

Since $\hat{y}$ is a particular solution of Eqs. (36), $\hat{y}_h$ and ξ^* are solutions of problem (50)–(52), and $\xi_\alpha = 0$ and $\hat{Y}_\alpha$ are solutions of problem (57)–(59), it is easy to verify that substitution (30), in which $\hat{Y}^*$ and ξ^* are constructed with the help of (65) and (66), transforms problem (26)–(29) into a simpler one (31)–(34). One can see that, under the initial conditions

$$t = 0: F_2(x) = \xi^*; \quad z = 0: u_2^0 = u^*; \quad v_2^0 = v^*, \quad (67)$$

the solution of problem (31)–(34) is zero, and (65) and (66) are solutions of problem (26)–(28) with initial conditions (67).

NONLINEAR WAVE PROFILE ON THE SURFACE OF A DEEP VISCOUS LIQUID

Analytical expression for the wave profile. Summation of (14) and (65) gives an expression for the

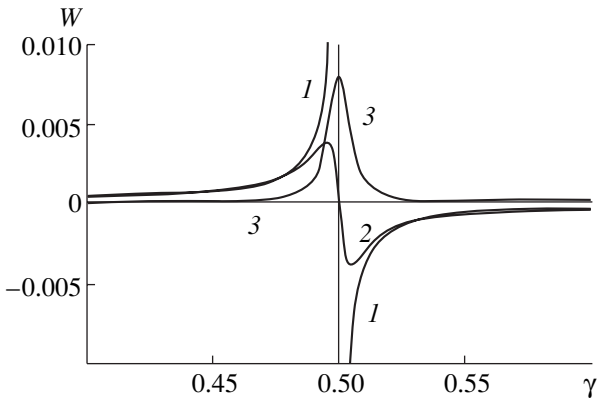

Fig. 1. Variations in the amplitudes of dimensionless quadratic terms $W(\gamma)$ entering into (68) and (69) describing the wave profile as functions of dimensionless surface tension. (*1*) For an ideal liquid $W(\gamma) = a^2\Lambda_0(\gamma)$; (*2*) and (*3*) for a viscous liquid $W(\gamma) = a^2\Lambda_1(\gamma, \nu)$ and $W(\gamma) = a^2\Lambda_2(\gamma, \nu)$, respectively, at $\nu = 10^{-3}$ cm^2/s.

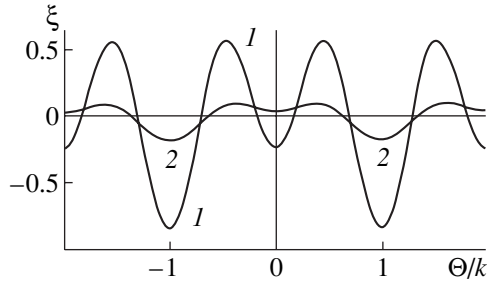

Fig. 2. Profiles of waves with a wavelength of $\lambda = 2$ cm on the water surface calculated using formula (61) at $\rho = 1$ g/cm^3, $\gamma = 72$ dyn/cm, $\nu = 0.01$ cm^2/s, and $g = 981$ cm/s^2 for various time instants t: (*1*) 0, (*2*) 5 s. The characteristic length in the reference frame attached to the wave is set off as an abscissa. The units of measure along the vertical and horizontal axes are centimeters.

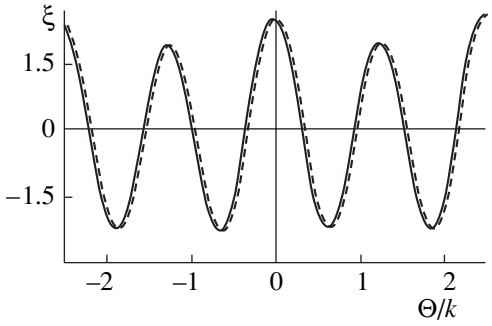

Fig. 3. Wave profiles on the water surface with the wavelength $\lambda = 2.5$ cm close to the resonance $\lambda = 2.4$ cm. The profiles are calculated using formulas (61) for $t = 0$ (solid line) and (62) (dashed line).

wave profile which is correct up to $O(\varepsilon^2)$:

$$\xi = a\cos(\Theta)\exp(S_1 t) + a^2(\zeta_1\cos(2\Theta) - \zeta_2\sin(2\Theta))\exp(2S_1 t). \quad (68)$$

Here, ζ_1 and ζ_2 are real and imaginary parts of variable ζ, which is calculated in the course of constructing (65) and (66), which is described directly below these expressions. Analytical expressions for ζ_1 and ζ_2 are very awkward and, therefore, uninformative. Values S_1, S_2, and Θ are calculated with the help of relationships (18)–(21).

Comparison of the solution found with known solutions. It is of interest to compare the solution found with the results by A.H. Nayfeh, who investigated a similar problem formulated for an ideal liquid [4]. In this case, according to [4], a direct expansion of the solution with respect to the amplitude of a deviation from the equilibrium shape in a quadratic approximation leads us to the following result:

$$\xi = a\cos(\Theta_0) + a^2\Lambda_0\cos(2\Theta_0);$$
$$\Lambda_0 = \frac{(\rho g k + \gamma k^3)}{2(\rho g - 2\gamma k^2)}; \quad \Theta_0 = kx - \omega_0 t. \quad (69)$$

When $\gamma = 0$, solution (69) is transformed into a Stokes wave [1, 2]. It is readily seen that expression (68) for the wave profile in a viscous liquid at $\nu \longrightarrow 0$ is reduced to the corresponding one for an ideal liquid (69).

Comparing solutions for viscous (68) and ideal (69) liquids shows that they differ to the greatest extent at values of dimensionless parameters corresponding to resonance interaction of the modes. For an inviscid liquid, from (69) it is seen that at $\gamma k^2 = 0.5\rho g$, a quadratic term with respect to the wave amplitude of the first approximation becomes infinitely large. Thus, the principal mode resonantly excited a wave with a half-wavelength.

In Fig. 1, in dimensionless variables chosen in such a way that $k = g = \rho = 1$, the amplitudes of the second terms in solutions (68) and (69) are plotted versus dimensionless surface tension γ for dimensionless values $\nu = 10^{-3}$ and $a = 0.01$. The set of dimensionless variables chosen allows us to compare viscous and inviscid models of intermode resonance for a wave on the liquid surface with a wavelength of 2.4 cm and amplitude of 0.3 mm. Figure 1 shows that in a region of dimensionless parameters corresponding to values of γ remote from the resonance conditions, solutions (68) and (69) coincide with one another. Near the resonance dimensionless value $\gamma = 0.5$, in the solution found in view of the viscosity, coefficient Λ_2 of $\sin(2\Theta)$ becomes nonzero. This shifts the wave phase found in the second order approximation relative to the phase of the principal wave. The amplitude of quadratic term $\sqrt{\Lambda_1^2 + \Lambda_2^2}$ remains less than the amplitude of the principal wave $a = 0.01$ even at the resonance value $\gamma = 0.5$. This means that solution (68) in the given case is suitable for all values of γ, whereas solution (69) based on an approximation corresponding to an inviscid liquid predicts a high resonance amplitude.

Examples of calculation of the wave profile. Calculations show that, on the water surface, profiles of waves remote from the resonance wavelength $\lambda \neq 2.4 \pm 0.1$ cm found using formula (69) and profiles calculated at the same values of the parameters at time instant $t = 0$ with formula (68) differ insignificantly from one another; however, the difference between the height of waves (68) and (69) grows quickly with time. The magnitude of the amplitude of wave (69) decreases due to the viscosity, and when this takes place, the amplitude of the first-order component decreases proportionally, $\exp(-S_1 t)$, where $S_1 < 0$, and the amplitude of the second-order term decreases as the square of that exponent. As a result, over the course of time, the decrease in the wave amplitude is accompanied by a change in the wave profile. In Fig. 2, there is an example of the profile of a nonlinear wave on the water surface with a wavelength of $\lambda = 2$ cm and $a = 0.3$ cm. Curve *1* is a profile calculated neglecting the viscosity by formula (69). It coincides up to the line width with a profile calculated for time instant $t = 0$ using formula (68) making allowance for the viscosity. According to (68), over a time interval of $t = 5$ s, the amplitude will decrease approximately by a factor of 2 (profile shown by curve *2*). In addition, the lower hollow of the curve will rise in that time above the zero level, changing the character of the profile.

Figure 3 shows that both expressions (68) and (69) are unsuitable for the description of a wave in the vicinity of a resonance with a wavelength of $\lambda = 2.5$ cm in the case under consideration. Indeed, at $a = 3$ mm, a term which should be of the second order of smallness makes a contribution that increases the amplitude almost up to 3 cm. This situation requires a special study, which will be performed in future research.

Influence of the viscosity on the wave profile symmetry. In an approximation of an ideal liquid (69), the wave profile is a sequence of large and small humps that are symmetrical relative to the vertical passing through the crests of the humps. It may be said that the front and rear (with respect to the wave propagation direction) slopes of the humps have the same steepness. Because of the viscosity, the symmetry is distorted. However, this manifests itself markedly only at sufficiently large values of viscosity. Figure 4 shows that for a liquid that has similar properties to water but with $\nu = 0.1$ cm^2/s (this value is ten times as much as that for water), the asymmetry of the humps becomes noticeable and at $\nu = 0.5$ cm^2/s, substantial. It can be seen in Fig. 4 that, for the wave with $\lambda = 2.6$ cm, the trailing edge of the large hump is steeper than the leading edge; for the small hump, the reverse is true. The viscosity of liquids with densities close to the density of water under normal conditions and surface tensions close to 50 dyn/cm may differ appreciably (from 0.01 cm^2/s for water to 7 cm^2/s for glycerine). Apparently, a solution with properties similar to that of Fig. 4 can be obtained by mixing liquids of different viscosity.

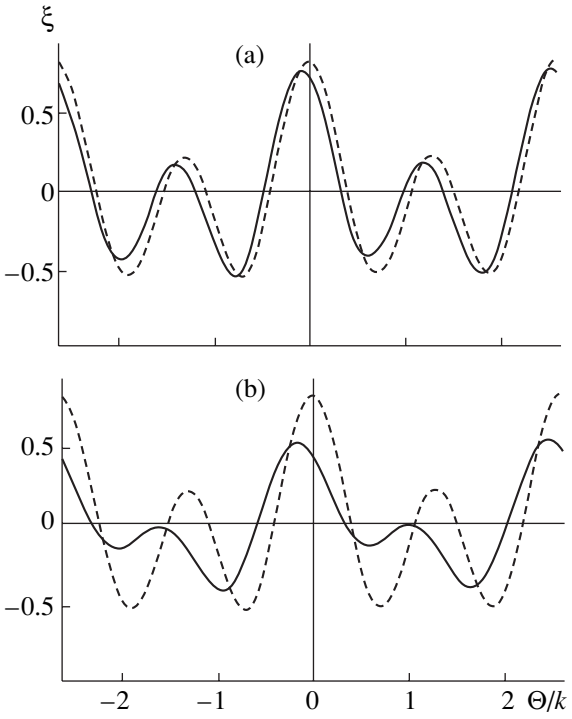

Fig. 4. Profiles of waves with a wavelength of $\lambda = 2.6$ cm on the surface of a liquid with $\rho = 1$ g/cm^3 and $\gamma = 60$ dyn/cm at $g = 981$ cm/s^2. The profiles are calculated using formulas (61) for $t = 0$ (solid line) and (62) (dashed line). ν, cm^2/s: (a) 0.1, (b) 0.5.

CONCLUSIONS

The asymptotic solution of the problem of wave propagation over the surface of an infinitely deep liquid with an arbitrary viscosity is correctly obtained in a quadratic approximation with respect to the wave amplitude. This solution allows us to generalize the idea of the Stokes wave (defined for an ideal liquid) to the case of a viscous liquid. Comparison of the solution found with that for an ideal liquid shows that even low viscosity plays a considerable role in the formation of the wave profile, and the role of viscosity in the time evolution of nonlinear waves is enhanced with growth of the order of smallness of the approximation. Allowing for the viscosity leads to a considerable change in the pattern of development of the resonance intermode interaction between waves and distorts the wave profile symmetry.

ACKNOWLEDGMENTS

This work was supported by the President of the Russian Federation (grant no. 00-15-9925).

APPENDIX A

Unit Vectors Normal and Tangential to the Free Liquid Surface: Divergence of the Normal Unit Vector

The equation of the plane free liquid surface perturbed by a wave motion has the form $z - \xi = 0$, where $\xi = \xi(x, z, t)$. Therefore, a unit vector normal to the liquid surface is calculated with the help of the relationship

$$\mathbf{n} = \frac{\nabla(z-\xi)}{|\nabla(z-\xi)|} = \frac{-\mathbf{e}_x\partial_x\xi + \mathbf{e}_z}{\sqrt{1+(\partial_x\xi)^2}}. \tag{70}$$

Unit vector $\boldsymbol{\tau}$ defined as

$$\boldsymbol{\tau} = \frac{\mathbf{e}_x + \mathbf{e}_z\partial_x\xi}{\sqrt{1+(\partial_x\xi)^2}} \tag{71}$$

satisfies condition $\boldsymbol{\tau}\cdot\mathbf{n} \equiv 0$, and, therefore, it is a unit vector tangential to the liquid surface perturbed by a wave motion.

The mean surface curvature, as is known, is equal to the divergence of an outward normal unit vector:

$$\nabla\cdot\mathbf{n} = \frac{-\partial_{xx}\xi}{\sqrt{1+(\partial_x\xi)^2}} + \frac{(\partial_x\xi)^2\partial_{xx}\xi}{(1+(\partial_x\xi)^2)^{\frac{3}{2}}}$$

$$= \frac{-\partial_{xx}\xi}{\sqrt{1+(\partial_x\xi)^2}}\left(-1 + \frac{(\partial_x\xi)^2}{1+(\partial_x\xi)^2}\right) \tag{72}$$

$$= -\frac{\partial_{xx}\xi}{(1+(\partial_x\xi)^2)^{\frac{3}{2}}} = -\partial_{xx}\xi + O(\xi^3).$$

APPENDIX B

Boundary Conditions for Magnitudes of Different Orders of Smallness

Kinematic boundary condition (3). Expanding (3) in a series in the vicinity of unperturbed liquid surface $\xi = 0$ and assuming hereafter that $u \sim O(\varepsilon)$, it is possible to obtain an asymptotic form of this condition up to an accuracy on the order of $O(\xi^3)$:

$$z = 0: \partial_t\xi + u\partial_x\xi = v + \xi\partial_z v. \tag{73}$$

Substituting into (73) the relationships $\xi = \xi_1 + \xi_2$, $u = u_1 + u_2$, and $v = v_1 + v_2$, where subscripts denote orders of smallness with respect to ε, one can decompose (73) into relationships for values of different orders of smallness:

$$z = 0: \partial_t\xi_1 - v_1 = 0; \tag{74}$$

$$\partial_t\xi_2 - v_2 = \xi_1\partial_z v_1 - u_1\partial_x\xi_1. \tag{75}$$

A condition for tangential tensions on the free liquid surface. If one uses formulas (70) and (71) as definitions of unit vectors $\mathbf{n}$ and $\boldsymbol{\tau}$, then, denoting

$$\chi = (1+(\partial_x\xi)^2)^{-1/2},$$

it is readily found that

$$(\mathbf{n}\cdot\nabla)\mathbf{U} = \mathbf{e}_x(\mathbf{n}\cdot\nabla)u + \mathbf{e}_z(\mathbf{n}\cdot\nabla)v$$

$$= \chi\mathbf{e}_x(-\partial_x\xi\partial_x u + \partial_z u) + \chi\mathbf{e}_z(-\partial_x\xi\partial_x v + \partial_z v);$$

$$(\boldsymbol{\tau}\cdot\nabla)\mathbf{U} = \mathbf{e}_x(\boldsymbol{\tau}\cdot\nabla)u + \mathbf{e}_z(\boldsymbol{\tau}\cdot\nabla)v$$

$$= \chi\mathbf{e}_x(\partial_x u + \partial_x\xi\partial_z u) + \chi\mathbf{e}_z(\partial_x v + \partial_x\xi\partial_z v):$$

$$\boldsymbol{\tau}\cdot((\mathbf{n}\cdot\nabla)\mathbf{U}) = \chi^2(-\partial_x\xi\partial_x u + \partial_z u - (\partial_x\xi)^2\partial_x v + \partial_x\xi\partial_z v);$$

$$\mathbf{n}\cdot((\boldsymbol{\tau}\cdot\nabla)\mathbf{U})$$

$$= \chi^2(-\partial_x\xi\partial_x u - (\partial_x\xi)^2\partial_z u + \partial_x v + \partial_x\xi\partial_z v).$$

With the help of the two last equalities, condition (5) can be rewritten in the form

$$z = \xi: (\partial_x v + \partial_z u)(1-(\partial_x\xi)^2) - 2(\partial_x u - \partial_z v)\partial_x\xi = 0. \tag{76}$$

Expanding (76) into a series in ξ in the vicinity of $z = 0$ leads us to approximation expression

$$z = 0: \partial_x v + \partial_z u + \xi\partial_z(\partial_x v + \partial_z u) = 2(\partial_x u - \partial_z v)\partial_x\xi$$

accurate up to magnitudes on the order of $O(\xi^3)$.

Substituting the necessary expressions from (10) into the latter expression, it is easy to find the following expression for magnitudes of the first order of smallness with respect to ε,

$$z = 0: \partial_x v_1 + \partial_z u_1 = 0, \tag{77}$$

and the expression for the second order of smallness,

$$\partial_x v_2 + \partial_z u_2 + \xi_1\partial_z(\partial_x v_1 + \partial_z u_1) = 2(\partial_x u_1 - \partial_z v_1)\partial_x\xi_1,$$

which is simplified in view of the discontinuity condition, written in the form $\partial_x u_1 = -\partial_z v_1$, as follows:

$$z = 0: \partial_x v_2 + \partial_z u_2 = -4\partial_z v_1\partial_x\xi_1 - \xi_1\partial_z(\partial_x v_1 + \partial_z u_1). \tag{78}$$

Condition (4) for the pressure at the perturbed liquid surface. If one uses formula (70) and definition $\mathbf{U} = \mathbf{e}_x u + \mathbf{e}_z v$, within an error on the order of $O(\xi^3)$, it is possible to obtain the expressions

$$z = \xi: \mathbf{n} = (-\partial_x\xi + O(\xi^3))\mathbf{e}_x + (1 + O(\xi^2))\mathbf{e}_z;$$

$$(\mathbf{n}\cdot\nabla)\mathbf{U} = (-\partial_x\xi\partial_x u + \partial_z u + O(\xi^3))\mathbf{e}_x + (-\partial_x\xi\partial_x v + \partial_z v + O(\xi^3))\mathbf{e}_z;$$

$$\mathbf{n}\cdot((\mathbf{n}\cdot\nabla)\mathbf{U}) = -\partial_x\xi\partial_z u - \partial_x\xi\partial_x v + \partial_z v + O(\xi^3).$$

An expansion of (4) in series in ξ in the vicinity of $\xi = 0$ up to terms on the order of $O(\xi^3)$ has the form

$$z = 0:\ p + \xi\partial_z p + \frac{1}{2}\xi^2\partial_{zz}p - 2\rho\nu\partial_z v - 2\rho\nu\xi\partial_{zz}v + 2\rho\nu\partial_x\xi(\partial_z u + \partial_x v) - P_0 = \gamma\,\mathrm{div}(\mathbf{n}). \quad (79)$$

If one takes into account that

$$p = p_0 + p_1 + p_2 + O(\xi^3);\quad p_1 = O(\varepsilon);\quad p_2 = O(\varepsilon^2);$$

$$\xi = \xi_1 + \xi_2 + O(\varepsilon^3);\quad \varepsilon_1 = O(\varepsilon);\quad \xi_2 = O(\varepsilon^2),$$

$$\nabla\cdot\mathbf{n} = -\partial_{xx}\xi + O(\varepsilon^3)$$

and that $p_0 = P_0 - \rho g z$, that is, $\partial_z p_0 = -\rho g$ and $\partial_{zz}p_0 = -\rho g$, then (79) is decomposed into the relationships for magnitudes of different orders of smallness

$$z = 0:\ p_0 = P_0;$$

$$-\rho g\xi_1 + p_1 - 2\rho\nu\partial_z v_1 + \gamma\partial_{xx}\xi_1 = 0;$$

$$-\rho g\xi_2 + p_2 - 2\rho\nu\partial_z v_2 + \gamma\partial_{xx}\xi_2 = 2\rho\nu\xi_1\partial_{zz}v_1 - 2\rho\nu\partial_x\xi_2(\partial_z u_1 + \partial_x v_1).$$

In view of (77), the last expression is simplified as follows:

$$-\rho g\xi_2 + p_2 - 2\rho\nu\partial_z v_2 + \gamma\partial_{xx}\xi_2 = 2\rho\nu\xi_1\partial_{zz}v_1 - \xi_1\partial_z p_1.$$

APPENDIX C

Auxiliary Vectors Arising in the Second-Order Problem

Values S, S_1, S_2, q, q_1, q_2, and D appearing in the relationships of this appendix are calculated with the help of expressions (18)–(21).

Vectors $\hat{A}_1, \ldots, \hat{A}_6$ entering into (26) and (27)

$$\hat{A}_1 = -4\nu^2k^3q_1\begin{bmatrix}q_2\\ k\\ 0\end{bmatrix};\quad \hat{A}_2 = \begin{bmatrix}0\\ -k(S_2^2 + D^2)\\ 0\end{bmatrix};$$

$$\hat{A}_3 = \nu k\begin{bmatrix}2(\bar{S} + 2\nu k^2)q_1q_2 + (S_2 + iD)(q_2^2 - q_1^2 + k^2)\\ -2ik(k+q)(S_2 + iD)\\ 0\end{bmatrix};$$

$$\hat{A}_4 = \begin{bmatrix}-i\nu k(q-k)^2(S + 2\nu k^2)\\ 0\\ 0\end{bmatrix};$$

$$\hat{A}_5 = \begin{bmatrix}0\\ N_1\\ M_0\end{bmatrix};\quad \hat{A}_6 = \begin{bmatrix}\Omega\\ N\\ M\end{bmatrix};\quad \begin{aligned}\Omega &= \Omega_1 + i\Omega_2;\\ N &= N_1 + iN_2;\\ M &= M_1 + iM_2,\end{aligned}$$

where $\bar{S}$ is a complex conjugate of S, and N_1, N_2, Ω_1, Ω_2, M_1, and M_2 are calculated by formulas

$$\Omega_1 = k(D - 2\nu kq_1);\quad \Omega_2 = k(S_2 - 2\nu kq_2);$$

$$N_1 = \rho\nu k^2(D + 2\nu(q_2 - q_1^2)) + \frac{1}{2}\rho(S_1D - S_2^2);$$

$$N_2 = \rho\nu k^2(S_2 - 4q_1q_2\nu) + \rho S_2(S_1 + \nu k^2);$$

$$M_0 = k(2k(S_2 - 2\nu kq_2) - kS_2 - q_2\nu(-k^2 + q_2^2 - 3q_1^2));$$

$$M_1 = -k(2k(S_2 - 2\nu kq_2) + kS_2 + q_2\nu(-k^2 + q_2^2 - 3q_1^2));$$

$$M_2 = k(2k(D - 2\nu kq_1) + kS_1 + \nu(2k^3 - k^2q_1 + 3q_1q_2^2 - q_1^3)).$$

Square matrices $\hat{\Pi}_1, \ldots, \hat{\Pi}_4$ entering into (36) are

$$\hat{\Pi}_1 = \begin{bmatrix}2(S_1 - 2\nu k^2) & 0 & 0\\ 0 & 0 & 2q_1\rho^{-1}\\ 0 & 2q_1 & 0\end{bmatrix};$$

$$\hat{\Pi}_2 = \begin{bmatrix}2(S_1 - 2\nu k^2) & 0 & 0\\ 0 & 0 & 2q_1\rho^{-1}\\ 0 & 2k & 0\end{bmatrix};$$

$$\hat{\Pi}_3 = \begin{bmatrix}2S_1 - \nu(k+q)^2 & 0 & 0\\ 0 & 0 & k+q\\ 0 & k+q & 0\end{bmatrix};$$

$$\hat{\Pi}_4 = \begin{bmatrix}2S + \nu(3k^2 - 2kq - q^2) & 0 & -2ik\rho^{-1}\\ 0 & 2S + \nu(3k^2 - 2kq - q^2) & (k+q)\rho^{-1}\\ -2ik & k+q & 0\end{bmatrix}.$$

Matrices $\hat{L}$, $\hat{B}$, and value Λ appearing in (48) are

$$\hat{L} = \begin{bmatrix} 2ik & -2k & 2S \\ 4i\rho\nu kr & -2\rho(S + 4\nu k^2) & -(\rho g + 4\gamma k^2) \\ -(r^2 + 4k^2) & -8ik^2 & 0 \end{bmatrix};$$

$$\hat{B} = \begin{bmatrix} C_4[2] \\ 2C_4[2]\rho\nu(k+q) - C_4[3] \\ 2iC_4[2]k - C_4[1](k+q) \end{bmatrix};$$

$$\Lambda = 2C_1[1]q_1 + C_3[1](k+q).$$

REFERENCES

1. J. Stoker, *Water Waves* (Wiley Interscience, 1957; Inostrannaya Literatura, Moscow, 1959).
2. G. B. Whitham, *Linear and Nonlinear Waves* (Wiley, New York, 1974; Mir, Moscow, 1977).
3. R. V. Goldshteĭn and V. A. Gorodtsov, *Mechanics of Continuous Media* (Nauka, Moscow, 2000), Chap. 1.
4. A. H. Nayfeh, J. Fluid Mech. **48**, 385 (1971).
5. I. Shugan and K. Voliak, J. Fluid Mech. **368**, 321 (1998).
6. K. Trulsen and I. Kliakhandler, Phys. Fluids **12**, 2432 (2000).
7. M. S. Long-Higgins, J. Fluid Mech. **423**, 275 (2000).
8. I. M. Mindlin, Izv. Ross. Akad. Nauk, Mekh. Zhidk. Gaza, No. 3, 135 (1994).
9. S. V. Nesterov, Izv. Ross. Akad. Nauk, Mekh. Zhidk. Gaza, No. 4, 116 (1995).
10. A. V. Fedorov and W. K. Melvil, J. Fluid Mech. **354**, 1 (1998).
11. S. O. Shiryaeva, D. F. Belonozhko, V. B. Svetovoĭ, *et al.*, Preprint No. 31, IMI RAN (Inst. of Microelectronics, Russian Academy of Sciences, Yaroslavl, 2001).
12. S. O. Shiryaeva, A. I. Grigor'ev, and V. A. Koromyslov, Pis'ma Zh. Tekh. Fiz. **22** (4), 89 (1996) [Tech. Phys. Lett. **22**, 173 (1996)].
13. S. O. Shiryaeva, D. F. Belonozhko, and A. I. Grigor'ev, Zh. Tekh. Fiz. **68** (2), 22 (1998) [Tech. Phys. **43**, 151 (1998)].
14. D. F. Belonozhko and A. I. Grigor'ev, Pis'ma Zh. Tekh. Fiz. **25** (22), 1 (1999) [Tech. Phys. Lett. **25**, 884 (1999)].
15. V. G. Levich, *Physicochemical Hydrodynamics* (Izd. Akad. Nauk SSSR, Moscow, 1952).
16. A. D. Polyanin, *Handbook of Linear Equations of Mathematical Physics* (Fizmatlit, Moscow, 2001).

Translated by N. Mende

Technical Physics, Vol. 48, No. 4, 2003, pp. 415–420. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 4, 2003, pp. 38–42.

GAS DISCHARGES, PLASMA

Study of the Gas-Filled Injector of a Plasma Opening Switch

V. V. Borovkov, K. I. Almazova, E. P. Volkov, V. G. Kornilov, C. Yu. Kornilov, V. D. Selemir, D. A. Tolshmyakov, A. A. Khizhnyakov, and V. I. Chelpanov

Russian Federal Nuclear Center All-Russia Research Institute of Experimental Physics, Sarov, Nizhni Novgorod oblast, 607190 Russia

e-mail: mailbox@ntc.vniief.ru

Received July 8, 2002

Abstract—A plasma injector with pulsed gas filling is investigated experimentally. Interferometric measurements of the formation dynamics of the plasma channel are carried out. Under optimal operating conditions, the injector is capable of producing a plasma channel 4 cm in diameter with an electron density of ~10^{17} cm^{-3}. The effect of the cathode diameter of the plasma opening switch on the conductivity of the plasma channel is studied. It is shown that the current flowing through the plasma channel of a single injector attains 400 kA.

INTRODUCTION

Over the last decade, the development of plasma opening switches (POSs) has been characterized by the advancement into the megaampere current range with a conduction phase duration as short as 1 μs [1, 2]. The advance into this range was accompanied by an increase in the POS plasma density from 10^{13}–10^{14} to 6×10^{15} cm^{-3}. The necessity of increasing the plasma density stems from the scaling obtained in the ACE and Decada devices [3, 4],

$$IT/rl \sim \sqrt{n}, \quad (1)$$

where I is the POS current that is reached during the conduction phase of duration T; n and l are the initial plasma density and the axial length of the plasma layer, respectively; and r is the POS cathode radius.

Scaling (1) determines the condition that is necessary for achieving the required quality of the current break, which is characterized by the ratio between the current rise and fall times.

Since 1999, work has been carried out at the Russian Federal Nuclear Center All-Russia Research Institute of Experimental Physics on creating pulsed microwave oscillators based on inductive energy storages with POSs, powered by magneto-cumulative generators (MCGs) [5].

The required value of the current in the load of an MCG (i.e., in the POS) can only be obtained at very low values of the load impedance: the inductance and resistance of the load must be no larger than several tens of nanohenry and milliohm, respectively. This limitation on the load unit of the MCG results in the limitation of the POS length and in the necessity of producing a high-density plasma layer.

At present, erosion plasma injectors are most widely used. They ensure the plasma density in the POS interelectrode gap at a level of $(4–6) \times 10^{15}$ cm^{-3} [6]. Interest in gas-plasma injectors stems from the possibility of varying the composition and density of the generated plasma, which, in turn, makes it possible to improve the POS parameters [7]. It is shown in [8] that a gas-plasma injector with a reversed Z-pinch allows the generation of plasma channels of different composition with electron densities from 6×10^{15} to 2×10^{16} cm^{-3}.

In this paper, we present the results of the investigation of a coaxial gas–plasma injector. It is shown that this type of plasma generator allows one to generate plasma jets with electron densities of up to several units of 10^{17} cm^{-3} and to increase the channel current produced by one injector to 400 kA. Experiments were carried out in two devices, one of which was specially designed for investigating the plasma injector. The POS operation in the current-rise mode was studied in the second device with the given plasma injector.

The question of the possibility of achieving a high-quality current break at such a high plasma density is still open, and it can only be answered experimentally.

EXPERIMENTAL SETUP AND PLASMA DIAGNOSTIC TECHNIQUE

Investigations of the plasma injector [9] were carried out in a vacuum chamber evacuated to $(5–8) \times 10^{-5}$ torr. The coaxial plasma injector (PI) with pulsed gas puffing was mounted on the side wall of the chamber. With the help of an electromagnetic valve, the gas was puffed into the inner electrode of the injector. The inner electrode was a 2-cm-diameter tube closed at one end. The diameter of the gas channel was 1.4 cm. A bar-

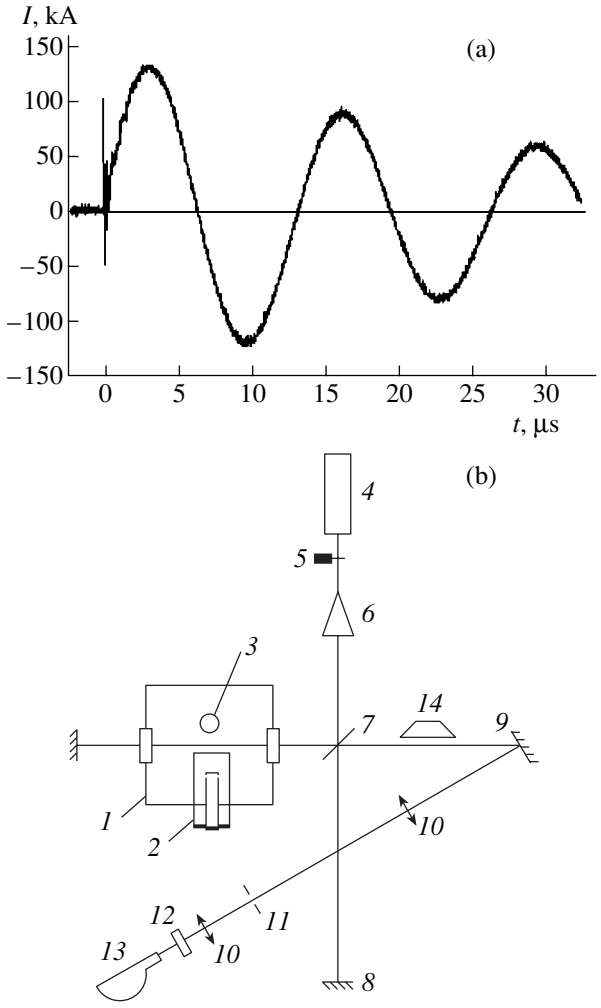

Fig. 1. (a) Waveform of the injector current and (b) the scheme of the optical plasma diagnostics: (*1*) vacuum chamber, (*2*) injector, (*3*) imitator of the POS cathode, (*4*) probing He–Ne laser, (*5*) electromagnetic gate, (*6*) beam-expanding telescope, (*7*, *8*) interferometer mirrors, (*9*) sweep mirror, (*10*) telescope system for imaging the injector onto the streak camera slit, (*11*) diaphragm, (*12*) KS-11 optical filter, (*13*) high-speed streak camera, and (*14*) Dove prism.

rel with an inner diameter of 4 cm served as an outer PI electrode. The length of the injector unit was 21 cm, and the distance between the gas puffing region (the end of the inner electrode) and the injector outlet was 5 cm. A negative voltage of 25 kV from a 12-μF capacitor was applied to the inner electrode of the injector. The maximum current amplitude was 125 kA, and the half-period was 6.5 μs. As an imitator of the POS cathode, we used a metal rod, which was located at a distance of 3 cm from the injector outlet.

Figure 1 shows the waveform of the PI current and the optical scheme of the plasma diagnostics. The plasma produced by the PI was studied with the help of a Michelson interferometer operating with the 0.63-μm radiation of a He–Ne laser. The interference pattern was recorded by a high-speed streak camera operating in the scanning-slit mode with temporal and spatial resolutions of 0.1 μs and 0.3 mm, respectively. To rotate the injector image on the slit of the streak camera, we used a Dove prism. The accuracy of determining the shift of the interference fringe was 1/20 of the fringe width.

In preliminary investigations, it was found that, when the gas valve was open but the PI discharge bank was not switched on, the gas effluent from the injector into the vacuum chamber did not lead to a shift of the interference fringes. In experiments with discharging the PI capacitor bank, we observed a decrease in the refractive index of the plasma outflowing from the injector. For this reason, we assumed that the main contribution to the change of the refractive index was made by free plasma electrons.

The integral electron density along the probing line (hereafter referred to as the line electron density) was calculated by the formula

$$n_e l = -1.76 \times 10^{17} \Delta k \text{ cm}^{-2}, \quad (2)$$

where n_e is the mean electron density over the optical path, l is the geometrical length of the plasma along the probing line, and Δk is the relative shift of the interference fringe.

The sensitivity of the technique for determining $n_e l$ was 9×10^{15} cm^{-2}. The numerical factor in formula (2) was determined taking into account the double pass of the probing radiation through the plasma and the polarizability of electrons at the probing radiation frequency.

RESULTS FROM THE OPTICAL DIAGNOSTICS OF THE INJECTOR PLASMA AND THEIR DISCUSSION

In the experiments we varied the time delay of the switching of the PI capacitor bank (T_g) relative to the switching of the gas valve for air puffing. It was assumed that, depending on the delay time, the gas density would change in the PI, which would lead to a change in the density of the generated plasma. The time between the switching of the gas valve and the start of the gas puffing into the discharge gap was ~350 μs.

Figure 2 shows typical interferograms illustrating the time evolution of the electron density distribution along the plasma channel for two values of T_g. The slit of the streak camera was oriented with the help of the Dove prism along the PI axis and perpendicular to the plane of the output nozzle of the PI, i.e. in the propagation direction of the plasma jet outflowing from the injector. Figure 3 shows the results of the interferogram processing.

Let us point out the characteristic features of the formation of the plasma channel at the PI outlet.

The maximum velocity of the front of the first plasma wave (~3×10^5 m/s) was observed for T_g = 400 μs. The plasma appeared in the interelectrode gap

of the POS practically without any delay after switching on the PI capacitor bank. Two characteristic features were observed in the time evolution of the plasma density in the probing region: (i) the line electron density in the plasma channel increased gradually to about 4×10^{17} cm^{-2} with a rise time of ~7 μs and then decreased to 1.5×10^{17} cm^{-2} and (ii) small-scale density fluctuations were observed in the plasma outflowing from the injector. About 12 μs after switching on the PI capacitor bank, the second plasma wave appeared at the injector outlet as a result of the oscillatory character of the discharge current of the injector capacitor bank. This wave propagated with a velocity of 3×10^4 m/s and had a sharp jump in the plasma density at its front (~5×10^{17} cm^{-2}).

When T_g was increased to 500 μs, the plasma appeared at the injector outlet with some delay, the propagation velocity of the leading front of the plasma wave decreased to 10^5 m/s, and the density jump appeared at the front of the first wave.

The above results allow us to formulate the general features of the formation of the plasma channel at the PI outlet.

The character of plasma density variations at T_g = 500 μs, namely, the generation of a shock wave at the front of the outflowing plasma, indicates that the mechanism for plasma channel generation in this regime is similar to the mechanism operating in coaxial electromagnetic shock tubes [10]. At large values of T_g, a portion of the gas leaves the PI and falls into the discharge volume of the POS. In fact, the velocity of the steady air stream outflowing into a vacuum is ~800 m/s. Consequently, in a time of ~100 μs (the time period between the start of gas puffing into the PI and the switching-on of the discharge bank), the size of the region occupied by the gas cloud can reach ~10 cm. During the breakdown in the PI discharge gap, a plasma layer is generated near the end of the gas-puffing tube, where the electric field strength is maximum. Under the action of the magnetic field pressure, the plasma layer accelerates toward the injector outlet and a shock wave is formed. Initially, the shock wave propagates inside the PI (for this reason, it appears at the injector outlet with a delay of 1.5 μs), and then it propagates through the gas cloud in the interelectrode gap of the POS. It follows from interferometric measurements that the air density at the PI outlet is no higher than 3.6×10^{16} cm^{-3} (which is the sensitivity limit of this technique for a neutral gas). A decrease in the plasma density behind the shock (Fig. 3a, curve *2*) is typical of shock tubes. The further smooth increase in the electron density is probably due to the subsequent ionization of the gas in the buffer area of the PI.

For T_g = 400 μs, before the discharge bank is switched on, the gas flowing out through the holes in the central electrode propagates toward the PI outlet and also into the PI buffer area over a distance of 3–4 cm; i.e., the gas cloud is concentrated inside the PI. The interferograms observed allow us to affirm that, when the discharge is initiated, the breakdown occurs in the outer shell of the gas cloud (on the side of the PI outlet). This is evidenced by the absence of a shock wave at the leading front of the plasma jet, which could be expected if the current layer were formed in another

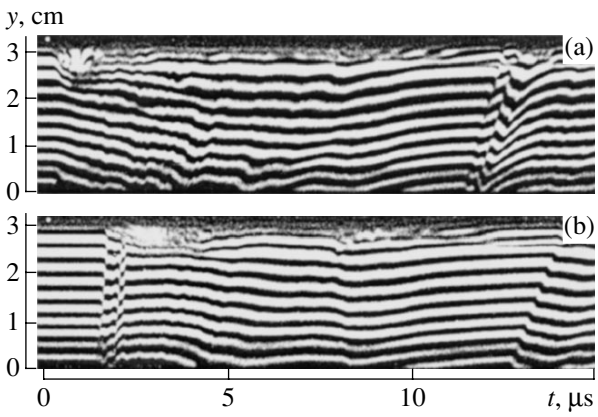

Fig. 2. Interferograms illustrating the time evolution of the electron density between the injector and the cathode for T_g = (a) 400 and (b) 500 μs; y = 0 is the coordinate of the injector exit nozzle, and t = 0 is the start time of the PI current.

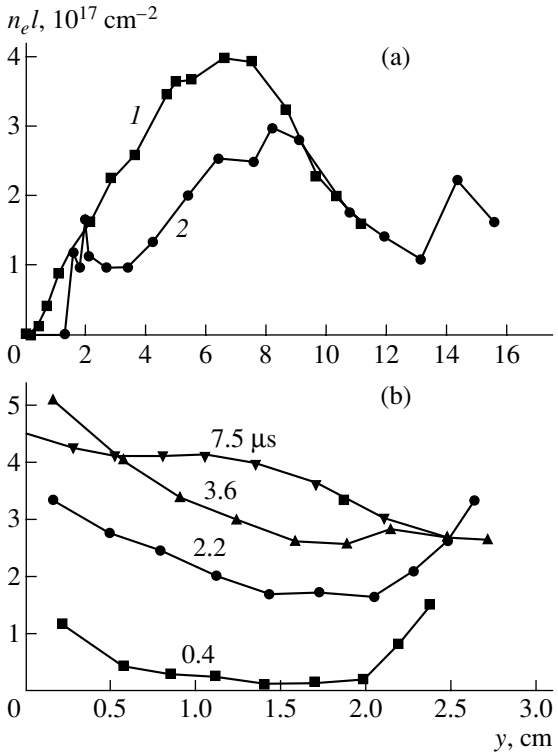

Fig. 3. (a) Line electron density as a function of time at a distance of 1.5 cm from the injector and (b) the electron density distribution along the plasma channel at different times for (a) T_g = (*1*) 400 and (*2*) 500 μs and (b) T_g = 400 μs.

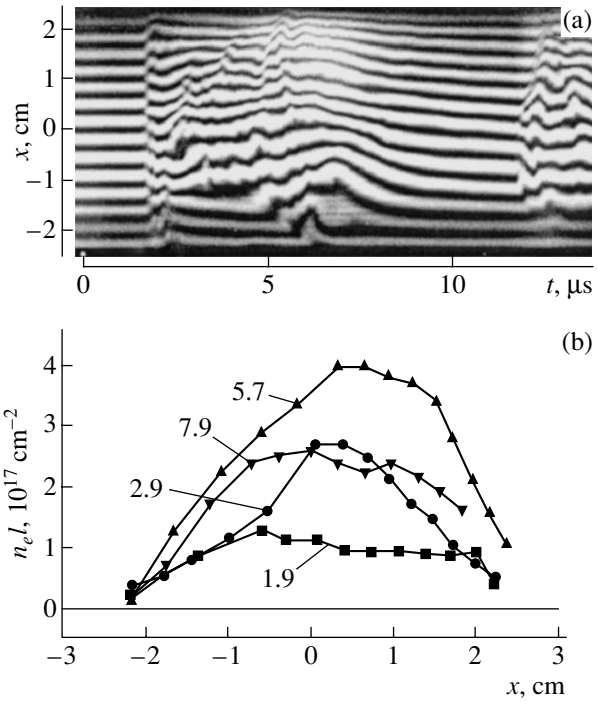

Fig. 4. (a) Interferogram illustrating the time evolution the plasma electron density across the plasma channel for T_g = 500 μs and (b) the results of its processing. Numerals by the curves show the times (in μs) corresponding to the given distributions of $n_e l$.

region of the gas cloud. The smooth increase in the electron density with time is related to the quasi-steady character of the plasma jet outflowing from the PI. We can assume that, at this stage, the current-carrying layer is displaced inside the injector. Small-scale fluctuations of the electron density are due to plasma bunches carried out from the PI by the relatively uniform plasma flow; the appearance of these bunches is probably caused by the onset of instabilities in the discharge region [11]. It follows from interferograms that the velocity of the density jumps (and, consequently, of the plasma jet as a whole) gradually decreases to 10^4 m/s. The decrease in the velocity of the plasma flow behind the wave front is also confirmed by the character of the interaction between the plasma and the POS cathode: the lower the plasma jet velocity, the smaller the amplitude of the wave reflected from the cathode (Fig. 2).

Figure 4 presents an interferogram and the results of its processing, which show the character of the time evolution of the electron density distribution over the transverse cross section of the plasma channel at the injector outlet. It can be seen from Fig. 4 that the leading front of the plasma jet is almost plane and its diameter is approximately equal to the diameter of the injector nozzle. The plasma density distribution across the plasma jet is nonaxisymmetric, which can be attributed to the misalignment of the electrode system of the PI.

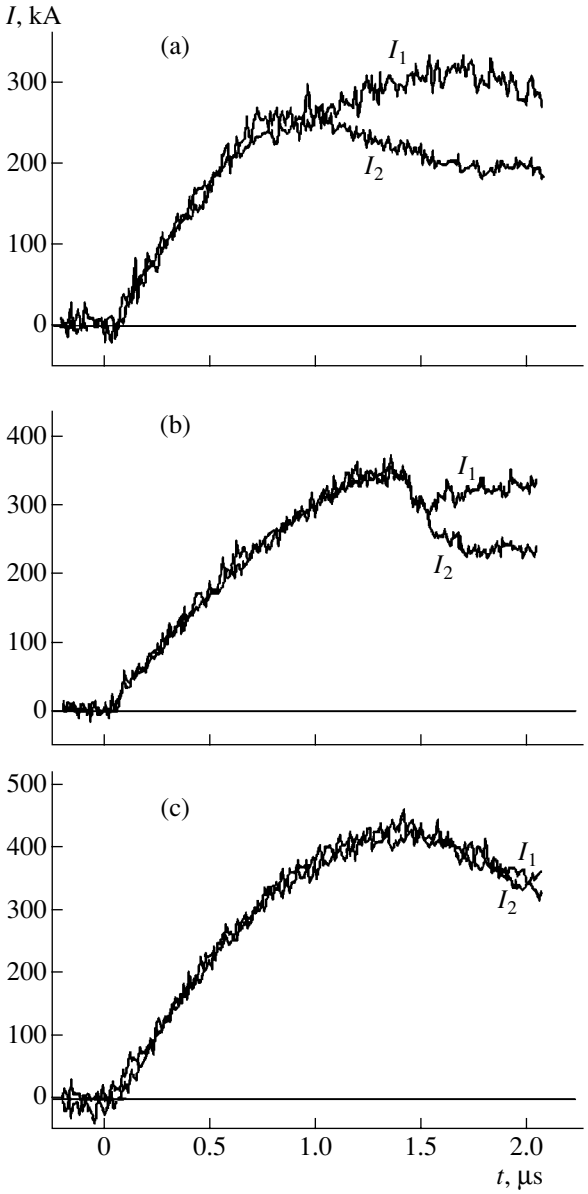

Fig. 5. Typical GIT–POS current pulses at different cathode diameters: (a) 2, (b) 8.5, and (c) 25 cm.

The above investigations allowed us to determine the optimal time delay for gas puffing (T_g = 400 μs) and the optimal time delay for switching on the discharge of the POS capacitive storage with respect to the beginning of the PI operation (5–7 μs). These time delays were then used in experiments in the GITaRa device.

INVESTIGATIONS OF THE POS IN THE GITaRa-1 DEVICE

The GITaRa-1 device contained a capacitive storage [a GIT-100 pulsed current generator (PCG)] connected to the POS through a bushing insulator. The GIT-100 parameters are the following: the capacitance is 8 μF,

the self-inductance is 35–40 nH, the charging voltage is 70–90 kV, and the resistance of the spark gaps is 15–20 mΩ. The plasma injector was mounted on the outer tube of the vacuum chamber (48 cm in diameter), which served as the POS anode. In the experiments we varied the configuration of the central electrode (cathode). Cathodes 2 or 8.5 cm in diameter were located at a distance of 3 cm from the injector nozzle, and a cathode 25 cm in diameter was aligned with the chamber axis. In the latter case, the distance between the nozzle of the plasma injector and the cathode surface was 8 cm. The total inductance of the circuit with the 2-cm-diameter cathode was 240 nH; in the two other cases, it was 160–180 nH. The current was measured by Rogowski coils, one of which was located nearby the bushing insulator and measured the total current I_1. The current I_2 through the plasma channel was measured with the help of a second loop located inside the vacuum chamber. The experiments were carried out at the optimal time delays for gas puffing and for switching on the discharge of the main capacitive storage, i.e., in the regimes in which the electron density in the plasma channel of the POS reached 10^{17} cm^{-3}. The time dependences of the current for different cathode diameters are shown in Fig. 5.

The plasma resistance in the conduction stage of the plasma channel is usually neglected. The experiments performed show that the plasma channel resistance, which depends on the size of the region over which the cathode is overlapped, can limit the current flowing through the channel. Besides, the maximum current depends on the strength of the magnetic field, which in the case of azimuthal nonuniformity is determined not only by the cathode diameter, but also by the size of the plasma cathode. Indeed, as can be seen in Fig. 5, the maximum value of the current was obtained with the 25-cm-diameter cathode and the 8-cm interelectrode gap. In this case, the size of the plasma region overlapping the cathode surface increased to ~6 cm due to the divergence of the plasma jet. The coincidence of the time dependences of I_1 and I_2 shows that the entire current flows through the plasma channel. In the case presented in Fig. 5b, the difference between the currents I_1 and I_2 can be attributed to both an increase in the plasma channel resistance in the break stage and the subsequent breakdown of the bushing insulator of the POS chamber.

The table gives the values of the maximum current I_m flowing through the plasma channel in the conduction stage and the plasma channel resistance r_{pl} for different cathode diameters d_k. The circuit resistance R averaged over the quarter of the current oscillation period was determined by choosing the parameters of the function describing the current measured in the circuit: $I(t) = (U_0/\rho)\exp(-R/2L)\sin(t/\sqrt{LC})$, where U_0 is the charge voltage and ρ and L are the wave resistance and inductance of the circuit, respectively.

POS characteristics at different cathode diameters

d_k, cm	I_m, kA	R_{pl}, mΩ
2	250	135
8.5	350	50
25	400	35

Hence, by varying the cathode diameter and the distance from the injector, it is possible to optimize the plasma channel resistance and maximize the peak current value. It can be expected that, for an azimuthally uniform plasma bridge with an electron density of ~10^{17} cm^{-3}, the current can be increased substantially.

CONCLUSION

Our investigations have shown that, depending on the time delay of the switching of the PI capacitor bank, two different regimes of the plasma channel formation are realized. At small values of T_g, a quasi-steady regime of the plasma outflow from the PI was observed; in this case, the electron density at the injector outlet increased smoothly. When T_g was increased to 100 μs, the plasma channel generation in the POS interelectrode gap was accompanied by the formation of a shock wave. The maximum electron density (10^{17} cm^{-3}) in the channel was observed for the former regime. It has been shown that the resistance of the POS plasma channel depends on the cathode configuration and that the current flowing through the channel can attain 400 kA.

In the MCG–POS experiments (which will be reported in a separate paper), six injectors were symmetrically placed over the azimuth of the outer POS tube. This made it possible to increase the current through the plasma channels to 2 MA.

ACKNOWLEDGMENTS

This work was supported in part by the Sandia National Laboratories (USA), contract no. 17665.

REFERENCES

1. K. D. Ware, P. G. Filios, R. L. Gullicckson, *et al.*, IEEE Trans. Plasma Sci. **25**, 160 (1997).
2. S. P. Bugaev, A. M. Volkov, A. A. Kim, *et al.*, Izv. Vyssh. Uchebn. Zaved. Fiz., No. 12, 38 (1997).
3. N. Rix, P. Coleman, J. R. Thompson, *et al.*, IEEE Trans. Plasma Sci. **25**, 169 (1997).

4. D. Price, K. Childers, D. Kortbwi, *et al.*, in *Proceedings of the 12th IEEE International Pulsed Power Conference, Monterey, CA, 1999*, Vol. 2, p. 1095.
5. V. D. Selemir, V. A. Demidov, A. V. Ivanovskiĭ, *et al.*, Fiz. Plazmy **25** (12), 1085 (1999) [Plasma Phys. Rep. **25**, 1000 (1999)].
6. B. V. Weber, D. D. Hinshelwood, and R. J. Commisso, IEEE Trans. Plasma Sci. **25**, 189 (1997).
7. P. S. Anan'in, V. B. Karpov, Ya. E. Krasik, *et al.*, Zh. Tekh. Fiz. **61** (8), 84 (1991) [Sov. Phys. Tech. Phys. **36**, 894 (1991)].
8. J. J. Moschella, R. C. Hazelton, C. Vidoli, *et al.*, IEEE Trans. Plasma Sci. **28**, 2247 (2000).
9. S. Yu. Kornilov, V. V. Borovkov, V. E. Chelpanov, *et al.*, in *Proceedings of the 1st International Congress on Radiation Physics and Modification Materials, Tomsk, 2000*, Vol. 2, p. 332.
10. Ya. B. Zel'dovich and Yu. P. Raizer, *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena* (Nauka, Moscow, 1966; Academic, New York, 1967).
11. R. L. Baker, D. Q. Hwang, R. W. Evans, *et al.*, Appl. Phys. Lett. **79**, 1237 (2001).

Translated by E. Satunina

Technical Physics, Vol. 48, No. 4, 2003, pp. 421–426. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 4, 2003, pp. 43–48.

SOLIDS

Effect of an Ultralow-Frequency Electric Field on the Explosive Instability Threshold of Ice

E. G. Fateev

Institute of Applied Mechanics, Ural Division, Russian Academy of Sciences, Izhevsk, 426000 Russia

e-mail: fateev@udman.ru

Received May 30, 2002; in final form, August 21, 2002

Abstract—It is found that a weak (10^3–10^5 times lower than breakdown fields) ultralow-frequency ($1 < \omega <$ 1000 Hz) electric field has a strong effect on the explosive instability threshold of ice uniaxially compressed by high pressures in the temperature range 210–240 K. The explanation for the high electromechanical sensitivity of ice is based on the concept that ice undergoing structure modifications due to highly nonuniform compression is a heterogeneous system with cooperative phenomena in space-bounded sets of dipoles. The dipoles form around new-phase nuclei, defects, or air microbubbles and occupy domains with a typical size of 10^{-2}–10^{-5} mm. When exposed to ultralow-frequency electric fields, such systems may exhibit resonant bursts of polarization, causing the ice stability to drastically drop because of dipole compression or microbreakdowns. *© 2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Recently, the high mechanical sensitivity of crystal hydrate insulators under ultralow-frequency (ULF) electric fields has been discovered [1–6]. It is remarkable that this effect has been found in experiments where the threshold of explosive (Bridgman) instability against ULF electric field was studied. The Bridgman effect is observed in many solid insulators when they undergo slow uniaxial compression with rates $dP/dt \leq$ 0.1 GPa/s under high pressures (up to 20 GPa) in open-ended anvils [7–14].

The Bridgman explosive effect is usually attended by the formation of shock waves and the fast (0.5–2.0 km/s) emission of finely dispersed breakdown products outside the compression facility (Fig. 1). Such explosive instabilities occur when the elastic energy of a highly compressed body is converted into mechanical work with ultrafast volume relief of the system once certain critical pressure and temperature values have been attained [11, 13]. Under compression, instabilities due to phase transitions and polymorphic transformations in solid insulators may somewhat decrease the explosion threshold [13]. Note also that the explosion generates a high-energy pulse of electromagnetic radiation over a wide range of wavelengths (up to X-rays [10]) and causes the emission of electrons [11]. The threshold P_c (or the mean critical pressure inside a body that initiates the Bridgman effect) decreases with increasing temperature and compression rate [13], correlates with the thermodynamic parameters of the material [14], and is size dependent [12].

When highly compressible crystal hydrates are subjected to a very weak ULF electric field (10^3–10^4 times lower than breakdown fields), the explosive instability threshold P_c drops by a factor of 1.5–2.0 at certain frequencies. Recent theoretical models of this phenomenon [15–17] rely on the fundamental idea that similar highly sensitive effects must be observed in many insulators that are in a heterogeneous state or are experiencing phase transitions. Such phenomena may also be present in materials with proton conduction. The strong variation of the plasticity of ice in weak constant electric fields is an example [18]. Ice is a most interesting and easily accessible material for checking the above idea, since it may be in 12 phase states [19–21] and has two amorphous modifications (see, e.g., [22–26]). In experiments with ice, one can easily cover the ranges of pressures and temperatures where one or another phase transition can be observed.

Interest in the possible high-sensitivity mechanical behavior of ice in ULF fields is associated with the discovery of multikilometer ice shells on several satellites

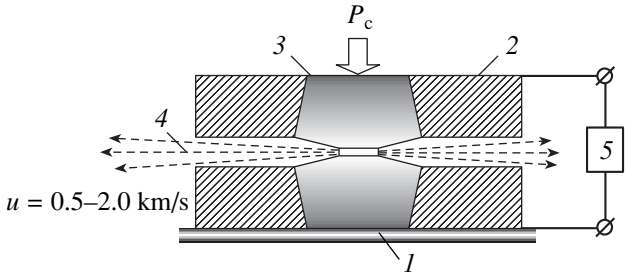

Fig. 1. Compression of ice by two Bridgman anvils with superhard inserts, which are necessary for creating an explosion with fast emission of a part of the material outside the compression device, under the action of ULF electric pulses. (*1*) Ice, (*2*) Bridgman anvil, (*3*) superhard insert, (*4*) emitted material, and (*5*) low-frequency pulse generator.

1063-7842/03/4804-0421$24.00 © 2003 MAIK "Nauka/Interperiodica"

of large planets in the solar system [27–30]. The ice thickness on Jupiter's satellites Europe and Ganimed was estimated as 100–150 km with temperatures in the range 130–273 K and pressures from 0 to 0.25 GPa [31]. This means that the ice crust on the satellites may contain vast regions where ice is in a metastable state [32]. The stability of ice in this state might be considerably affected by the ULF variations of the Jupiter's giant magnetic field and induced ULF electric fields. In the ice crust of Europe, the skin depth for magnetic waves of amplitude 220 nH is at least 110 km thick [33].

In this work, we for the first time report experimental data demonstrating a high electromechanical sensitivity of ice to weak ULF fields.

EXPERIMENT

The high electromechanical sensitivity of ice was demonstrated in experiments where a ULF electric field influenced the threshold of explosive instability, which has been recently discovered over wide pressure and temperature ranges [32]. We experimented with Bridgman anvils that had VK-8 superhard alloy inserts in the form of a truncated cone. The ground and polished operating area had a diameter $d = 10$ mm. The experiments were performed by the same method as in [32] at temperatures between 210 and 240 K. The rate of uniaxial compression was fixed: $dP/dt \approx 0.02$ GPa/s. The temperature of the sample was measured by a copper/Copel thermocouple with the junction placed in the neighborhood of the sample edge. At the initial (preparatory) stage, the anvils were directly cooled by liquid nitrogen to 230–240 K. Then, the operating area of one anvil was covered by a given amount of distilled water, which was converted to a thin ice disk of thickness $h \approx 0.4$ mm by slightly compressing the cooling water film with an insulating plane. At the second stage, the ice disk was cooled to any given temperature in the range 273–100 K. The rate of ice cooling was $dT/dt \approx -20$ K/s.

To obtain statistically significant values of P_c, many runs at a given temperature T have to be done. The desired temperature was achieved as follows. The anvils with the sample were cooled to a temperature somewhat below the desired value, and the entire system was placed under a press. Once the desired temperature had been established by natural heating with a low rate ($dT/dt \approx 0.1$ K/s), the system was compressed and ULF rectangular electrical pulses were simultaneously applied to the anvils (according to the technique described in [1–6]) up to the onset of explosive instability. The amplitude of the pulses was varied between 0.05 to 65 V, and their duration was 10^{-5} s throughout the experiment. They were applied from the beginning of compression to the onset of Bridgman instability with a frequency of 8.0–10^5 Hz. Note that the experiments were carried out in the temperature range 210–240 K. At $T > 220$ K, the temperature dependence of the ice stability threshold is nearly linear; for $210 < T < 220$ K, it exhibits a minimum, which is related to the phase transition I*h*–II in polycrystalline ice [32].

Figure 2 shows ULF spectra of the ice instability threshold for 65-V pulses at temperatures $T_1 = 210$ K, $T_2 = 220$ K, $T_3 = 230$ K, and $T_4 = 240$ K. Each data point was obtained from seven explosion runs. At certain ultralow frequencies (at the early stage of excitation by the method described in [3]), permittivity bursts were observed. For $T_2 = 220$ K, the qualitative ULF spectrum $\varepsilon(\omega)$ was constructed (Fig. 3). In addition, for each of the temperatures listed, we found the dependences of the instability threshold on the pulse amplitude (0.05–65 V) at 10 Hz (Fig. 4). To gain a better insight into the electrical properties of ice at these temperatures, the relative resistance of the ice disk was plotted against the uniaxial compression P in the absence of explosion (Fig. 5).

RESULTS AND DISCUSSION

As follows from Figs. 2a–2d, the ULF spectra of the explosive instability threshold $P_c(\omega)$ for ice have a nontrivial shape for all four temperatures. In the frequency range $8 < \omega < 10^5$ Hz and for the pulse amplitude $U = 65$ V, all four spectra exhibit several dips. Earlier, similar dips were found for several model and natural crystal hydrates [1–6]. For $\omega > 10^3$ Hz, the drop of the threshold $P_c(\omega)$ is most likely to be associated with dielectric loss bursts, as is the case for crystal hydrates under the same conditions. However, the nontrivial shape of the ULF instability spectra for ice at $\omega < 10^3$ Hz can apparently be explained by taking into consideration that, when nonuniformly compressed, ice (as well as crystal hydrates) exhibits giant resonance-like permittivity bursts at ULFs (Fig. 3). The strong effect of permittivity bursts on the instability threshold P_c is probably related to enhanced dipole compressions or microbreakdowns under ULF fields in a heterogeneous system into which a heavily compressed solid material turns [16, 17]. At $T_2 = 220$ K, the frequencies at which the permittivity bursts are observed are the same as those at which the ULF instability threshold spectrum exhibits deep dips. In essence, these dips for ice exposed to a ULF electric field indicate the high mechanical sensitivity of ice; in other words, it shows a high mechanical response to a weak electrical action. Such a strong response could be expected under breakdown fields. For ice, the breakdown field is on the order of $E \sim 10^6$ V/cm [34], while the strong response is observed in variable fields of strength $E = 10^1$–10^3 V/cm. From Fig. 4, it is seen that with $T_1 = 210$ K, $T_2 = 220$ K, and $T_3 = 230$ K, the high electromechanical response of ice at 10 Hz is retained up to $E \sim 10^2$ V/cm. With $T = 240$ K, the curve $P_c(V)$ has an unexpected minimum in still lower fields, $E = 1$–10 V/cm, at the same frequency.

Note that the "oscillatory" spectra $\varepsilon(\omega)$ obtained for highly compressible ice in this work differ considerably from the permittivity spectra taken from ice crystals by conventional techniques [35–37]. A normal permittivity spectrum for ice has the form of Debye dispersion

$$\varepsilon(\omega) = \varepsilon_\infty + \frac{\varepsilon_m - \varepsilon_\infty}{1 + (\omega\tau)^2},$$

where ε_m is the maximal value at ULFs, ε_∞ is the high-frequency value in the dispersion law $\varepsilon(\omega)$, and τ is the relaxation time of bound charges.

However, the spectrum $\varepsilon(\omega)$ (Fig. 3) for highly compressible ice has a set of peaks and dips. Such a radical difference between the spectra $\varepsilon(\omega)$ obtained in this work and those known previously can be explained only by invoking the fundamentally new concept of relaxation processes taking place in dense disperse systems at the beginning of their low-frequency excitation [16, 17]. At the initial stage of excitation, resonance responses gradually (for a time $\Delta t > 0.2$–1.0 s, which depends on many parameters of the system) transforming into dispersion-like spectra can be observed even in the one-dimensional approximation of a complicated model system with interacting dipoles.

The parallels observed between ULF spectra of explosive instability for ice and crystal hydrates suggest the temporal ($\Delta t < 0.2$–1.0 s) circulation of bound charges in compressed ice exposed to a ULF field. Such a circulation may appear [17] if a heterogeneous structure with disperse phase grains of typical size $\sim 10^{-2}$–10^{-5} mm that are surrounded by liquid or quasi-liquid sheaths and contain mobile anions and cations (most likely, protons or OH^- and H_3O^+ ions) forms in compressed ice. Diffuse phase transitions like I*h*–II, which may result in such a heterogeneous structure, are a real possibility at temperatures of 210 and 220 K and mean pressures of 0.06–0.10 GPa [32]. The same is indicated by an inflection in the dependences of the relative resistance of the ice disks on the mean anvil pressure at 210 and 220 K (Fig. 5). At 230 and 240 K, the inflection is nearly or completely absent, which means the absence of phase transitions for the given combinations of *P* and *T*. With such parameters, however, quasi-liquid layers may exist between crystal grains in the I*c* phase even at $T = 140$–210 K, as follows from IR spectroscopy data [38]. It seems likely that these intergranular layers may serve as effective paths for charge circulation (around new-phase grains, defects, or microbubbles).

It should also be noted that, when cooling, gases dissolved in distilled water may form spherical microvoids with unusual thermodynamic properties. For example, their inner surfaces may turn out to be quasi-liquid ion-conducting sheaths. In water, gas bubbles can form fractal clusters of size 10–30 μm (see, e.g., [39]). It appears that the quasi-liquid properties of the inner sheaths in the bubbles may exist over the wide temperature range 180–273 K [39].

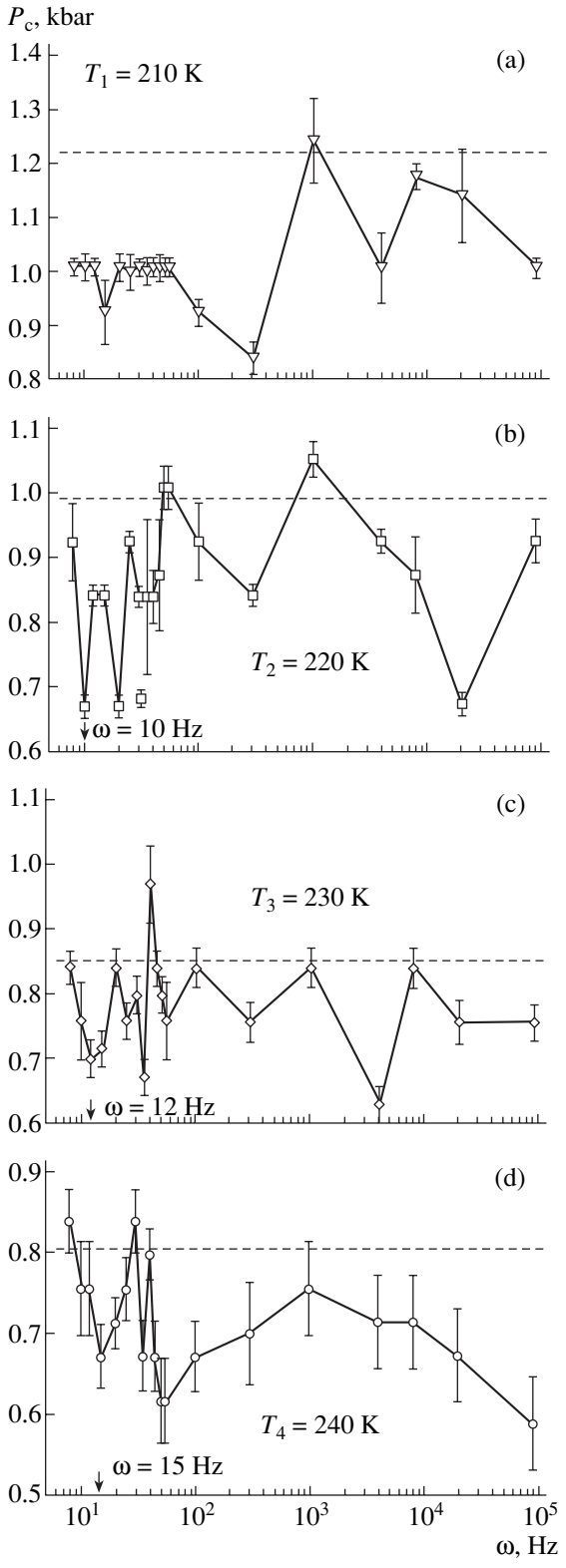

Fig. 2. ULF spectra of explosive instability threshold P_c for the case when ice between Bridgman anvils is exposed to ULF rectangular pulses of amplitude $U = 56$ V at (a) 210, (b) 220, (c) 230, and (d) 240 K. The dashed lines show the P_c level without electric fields. The arrows in panels (b–d) demonstrate the slight shift of the first minima in the ULF spectra $P_c(\omega)$ as the temperature rises from 220 to 240 K.

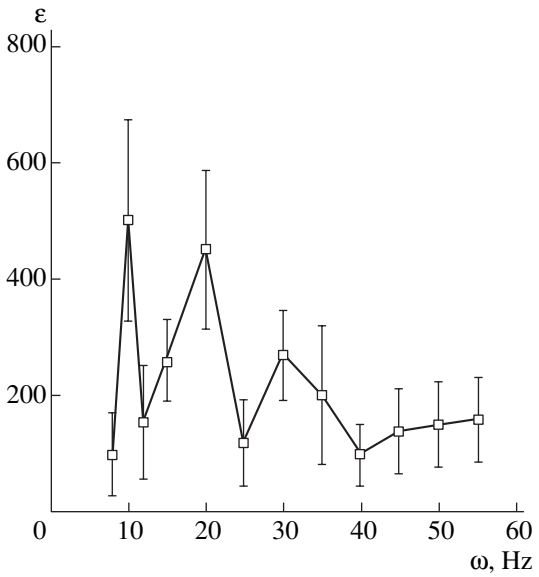

Fig. 3. ULF permittivity spectrum ε(ω) at the instant of maximal burst under strong compression of ice exposed to 65-V pulses at 220 K.

Numerous defects that are generated by phase transitions in ice [44] can also act as effective spatially confined traps of charges circulating in ULF fields [40–43]. The same is equally true for gas microbubbles and new-phase grains.

Yet the oscillatory behavior of the permittivity is hard to explain by the presence of spatial traps with charges circulating around. Indeed, during compression, ice is continuously in the ULF electric field but the permittivity bursts appear only after the compression has reached certain critical preexplosive values. We are thus led to the assumption that the traps rapidly form and rapidly disappear at such pressures, which may take place at the instant of phase transitions in ice.

Let us estimate the difference in the relaxation times $\tau = a^2/2D$ of charges in the liquid films around new-phase grains and in traps around defects (here, a is the characteristic size of a grain or trap and D is the diffusion coefficient of charges in the liquid or quasi-liquid sheath or in a trap). The parameters typical of a disperse system under normal pressure and temperature are $a = 10^{-4}$–10^{-6} m and $D < 10^{-12}$ m^2/s; accordingly, $\tau \le 10^{-10}$–10^{-4} s and the characteristic dispersion frequencies lie in the range $\omega = 1/\tau = 1$–10^4 Hz [16]. For a system with traps, $a \le 10^{-6}$–10^{-8} m, $D \sim 10^{-10}$ m^2/s, and, accordingly, $\tau = 10^{-2}$–10^{-6} s and $\omega = 10^2$–10^6 Hz [41]. Thus, at least two characteristic frequency of dielectric dispersion must exist in ice.

At lower temperatures, the characteristic frequencies must decrease and the resonance-like dips in the ULF spectrum $P_c(\omega)$ of instability threshold correspondingly shift [5]. Such a shift is really observed for several initial dips in the ULF spectra with $\omega = 10$–15 Hz and temperatures of 220, 230, and 240 K (Figs. 2b–2d). A slight shift can also be identified at 10^2–10^4 Hz for the spectra $P_c(\omega)$ with 210, 220, and 230 K (Figs. 2a–2c). However, the dips in this part of the spectrum are most likely to be due to resonances associated with characteristic relaxation times for smaller disperse particles or defect-induced traps.

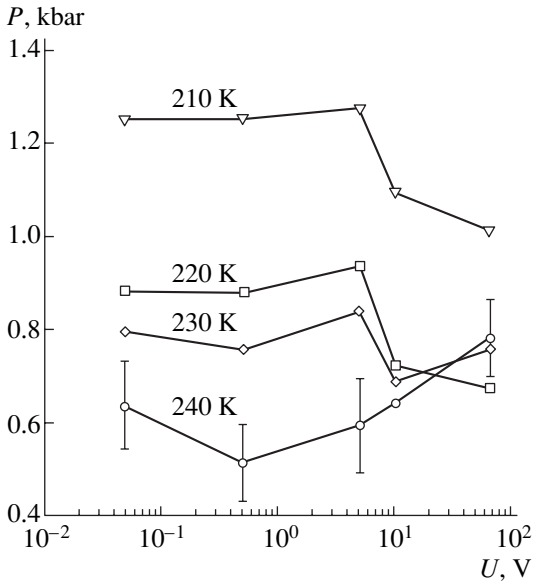

Fig. 4. Explosive effect threshold P_c vs. pulse amplitude at 10 Hz for four temperatures. For each of the temperatures, the P_c level without the field is the same as shown in Fig. 2.

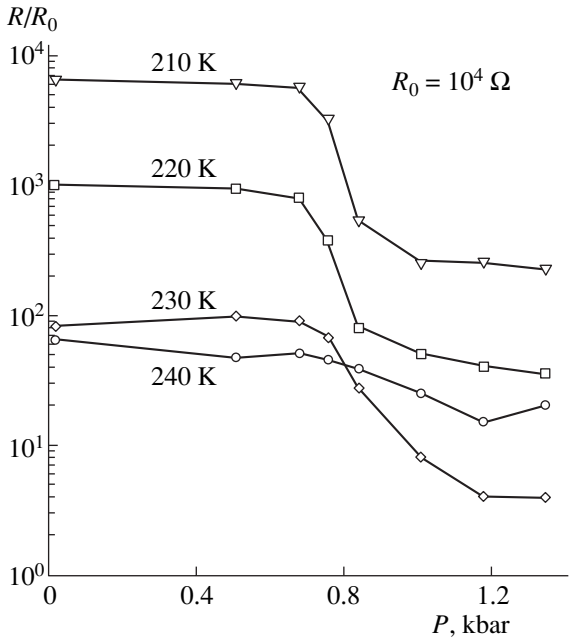

Fig. 5. Relative resistance of the ≈0.4-mm-thick ice disk vs. uniaxial compression for four temperatures in the absence of the Bridgman effect.

In experiments with "warmer" ice, additional minima in the spectra $P_c(\omega)$ arise (Figs. 2b–2d). They can be explained in terms of the theoretical model of system ultrasensitivity, where a system, in the first approximation, is represented as a chain of extended dipole oscillators with widely varying moments [17].

Of interest is the dependence of the explosive effect threshold on the pulse amplitude, $P_c(V)$, at 240 K, which has a minimum at $U \approx 0.5$ V (Fig. 4). One may suppose that such behavior is related to the potential of H_2O dissociation into hydrogen and oxygen: $U_c \approx 1.23$ V [42]. It can be assumed that carrier clouds generated in ice at $U > U_c$ effectively shield dipole interactions in chains of spatially closed oscillators, thereby suppressing local polarization bursts [16]. The degree of shielding may increase with the ULF pulse amplitude, since more ions like OH^- and H_3O^+ are activated in ice under these conditions [41–43]. At the same time, the dependence $P_c(V)$ with a minimum at $E \approx 12$ V/cm is observed only near a certain threshold temperature $T \approx 244$ K. Above this temperature, the behavior of ice is plastic rather than explosive [32]. At this temperature, a distinct phase transition is not observed either (Fig. 5). It is likely that in this temperature range, ice has (at an appropriate pressure) a number of liquid-like spacers with a thickness [40] much larger than under other conditions, which may increase the effective mobility of ions.

At temperatures of 210, 220, and 230 K, most of the charge carriers that must be injected from the anvils (electrodes) possibly remain at the anvil boundaries because of the absence of through quasi-liquid spacers. If so, dipole interactions at these temperatures are shielded by the field-induced ion clouds to a lesser extent. Nevertheless, the phase transition duration at 210, 220, and 230 K seems to be insufficient to initiate giant oscillations in dipole systems, which could have a still greater effect on the mechanical stability of ice. According to estimates [16], the durations of such phase transitions do not exceed 0.1–0.2 s. However, at 240 K, the existence time of structural elements with charge circulation is probably much longer, since the nature of these structures is somewhat different. As a result, the probability of generating local giant polarization bursts in dipole systems at this temperature grows and these bursts may considerably affect the ice stability even in ultralow ULF fields. This means that, as the ULF pulse amplitude rises at 240 K, two competing trends arise: on the one hand, the effect of the pulses on the systems of oscillators increases and, on the other hand, dipole interactions weaken because activated ion clouds shield them more effectively; hence, the appearance of a minimum in the curve $P_c(V)$.

CONCLUSION

Thus, it was shown that ice may exhibit a high mechanical sensitivity to weak ULF electric fields over a wide range of low temperatures. This effect is unrelated to the dielectric loss of ice at ULFs, as indicated by a correlation between the ULF spectra of mechanical instability excitation (the Bridgman effect) and those of the permittivity. The parallels in the behavior of ice and crystal hydrates in weak ULF electric fields lead us to conclude that ultrahigh-sensitivity effects are observed when ice undergoes phase transitions or is in the metastable state. These results can be explained in terms of nonlinear cooperative phenomena in sets of nonmolecular spatial dipoles with a characteristic size of 10^{-2}–10^{-5} mm and strongly varying moments.

ACKNOWLEDGMENTS

The author is indebted to Yu.I. Prokhorov for the valuable discussions.

REFERENCES

1. E. G. Fateev, Pis'ma Zh. Tekh. Fiz. **19** (10), 48 (1993) [Tech. Phys. Lett. **19**, 313 (1993)].
2. E. G. Fateev, Pis'ma Zh. Tekh. Fiz. **20** (20), 83 (1994) [Tech. Phys. Lett. **20**, 847 (1994)].
3. E. G. Fateev, Zh. Tekh. Fiz. **66** (6), 93 (1996) [Tech. Phys. **41**, 571 (1996)].
4. E. G. Fateev, Dokl. Akad. Nauk **354**, 252 (1997).
5. E. G. Fateev, Pis'ma Zh. Éksp. Teor. Fiz. **65**, 876 (1997) [JETP Lett. **65**, 919 (1997)].
6. E. G. Fateev, Zh. Tekh. Fiz. **71** (6), 37 (2001) [Tech. Phys. **46**, 682 (2001)].
7. P. W. Bridgman, Phys. Rev. **48**, 825 (1935).
8. P. W. Bridgman, Proc. Am. Acad. Arts Sci. **71** (9), 387 (1937).
9. P. W. Bridgman, *Studies in Large Plastic Flow and Fracture with Special Emphasis on the Effects of Hydrostatic Pressure* (McGraw-Hill, New York, 1952).
10. T. Ya. Gorazdovskiĭ, Pis'ma Zh. Éksp. Teor. Fiz. **5** (3), 78 (1967) [JETP Lett. **5**, 64 (1967)].
11. M. A. Yaroslavskiĭ, *Rheological Explosion* (Nauka, Moscow, 1982).
12. N. S. Enikolopyan, A. A. Mkhitaryan, A. S. Karagezyan, and A. A. Khzardzhyan, Dokl. Akad. Nauk SSSR **292**, 887 (1987).
13. E. G. Fateev and V. P. Khan, Pis'ma Zh. Tekh. Fiz. **17** (20), 51 (1991) [Sov. Tech. Phys. Lett. **17**, 736 (1991)].
14. V. P. Khan and E. G. Fateev, Pis'ma Zh. Tekh. Fiz. **16** (8), 81 (1990) [Sov. Tech. Phys. Lett. **16**, 317 (1990)].
15. E. G. Fateev, Pis'ma Zh. Tekh. Fiz. **26**, 103 (2000) [Tech. Phys. Lett. **26**, 640 (2000)].
16. E. G. Fateev, Zh. Tekh. Fiz. **71** (1), 92 (2001) [Tech. Phys. **46**, 89 (2001)].
17. E. G. Fateev, Phys. Rev. E **65** (2), 021403 (2002).
18. V. F. Petrenko and E. M. Schulson, Philos. Mag. A **67**, 173 (1993).
19. P. W. Bridgman, J. Chem. Phys. **3**, 597 (1935).
20. P. V. Hobbs, *Ice Physics* (Clarendon, Oxford, 1974).
21. C. Lobban, J. L. Finney, and W. F. Kuhs, Nature **391**, 268 (1998).

22. O. Mishima, L. D. Calvert, and E. Whalley, Nature **310**, 393 (1984).
23. O. Mishima, L. D. Calvert, and E. Whalley, Nature **314**, 76 (1985).
24. O. Mishima, J. Chem. Phys. **100**, 5910 (1994).
25. E. G. Ponyatovskiĭ, V. V. Sinitsyn, and T. A. Pozdnyakova, Pis'ma Zh. Éksp. Teor. Fiz. **60**, 352 (1994) [JETP Lett. **60**, 360 (1994)].
26. O. V. Stal'gorova, E. L. Gromnitskaya, and V. V. Brazhkin, Pis'ma Zh. Éksp. Teor. Fiz. **62**, 334 (1995) [JETP Lett. **62**, 356 (1995)].
27. D. Morrison and D. P. Cruikshank, Space Sci. Rev. **15**, 641 (1974).
28. J. D. Anderson, E. L. Lau, W. L. Sjorgen, *et al.*, Science **276**, 1236 (1997).
29. M. H. Carr, M. J. Belton, C. R. Chapman, *et al.*, Nature **391**, 363 (1998).
30. P. Helfenstein and E. M. Parmentier, Icarus **61**, 173 (1985).
31. R. Sullivan, R. Greeley, K. Homan, *et al.*, Nature **391**, 371 (1998).
32. E. G. Fateev, Pis'ma Zh. Éksp. Teor. Fiz. **73**, 482 (2001) [JETP Lett. **73**, 432 (2001)].
33. F. Neubauer, Nature **395**, 749 (1998).
34. A. A. Vorob'ev and G. A. Vorob'ev, *Electric Breakdown and Destruction of Solid Dielectrics* (Vysshaya Shkola, Moscow, 1966).
35. R. P. Auty and R. H. Cole, J. Chem. Phys. **20**, 1309 (1952).
36. R. Ruepp and M. D Kab, *Physics of Ice* (Plenum, New York, 1969).
37. M. P. Tonkonogov, Usp. Fiz. Nauk **168** (1), 29 (1998) [Phys. Usp. **41** (1), 25 (1998)].
38. P. Jenniskens, S. F. Banham, D. F. Blake, *et al.*, J. Chem. Phys. **107**, 1232 (1997).
39. N. F. Bunkin and A. V. Lobeev, Pis'ma Zh. Éksp. Teor. Fiz. **58**, 91 (1993) [JETP Lett. **58**, 94 (1993)].
40. M. A. Anisimov and R. U. Tankaev, Zh. Éksp. Teor. Fiz. **81**, 215 (1981) [Sov. Phys. JETP **54**, 110 (1981)].
41. V. F. Petrenko and I. A. Ryzhkin, Zh. Éksp. Teor. Fiz. **87**, 558 (1984) [Sov. Phys. JETP **60**, 320 (1984)].
42. V. F. Petrenko and N. Maeno, J. Phys. (France) **48**, 115 (1987).
43. V. F. Petrenko and R. W. Whitworth, *Physics of Ice* (University Press, Oxford, 1999).
44. O. V. Stal'gorova, E. L. Gromnitskaya, and V. V. Brazhkin, Zh. Éksp. Teor. Fiz. **112**, 200 (1997) [JETP **85**, 109 (1997)].

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 4, 2003, pp. 427–430. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 4, 2003, pp. 49–52.

SOLIDS

Magnetoimpedance Effect in FeCoMoSiB Amorphous Sheets

A. A. Anashko, A. V. Semirov, and A. A. Gavrilyuk

Irkutsk State Pedagogical University, Irkutsk, 664653 Russia

e-mail: semirov@isttu.irk.ru

Received June 11, 2002; in final form, October 7, 2002

Abstract—The magnetoimpedance effect in sheets made of $Fe_4Co_{67}Mo_{1.5}Si_{16.5}B_{11}$ amorphous metallic alloy is studied in relation to the mutual orientations of the sheet axis, permanent magnetic field, and variable electrical current. Also, the effective permeability is studied as a function of the mutual orientations of the sheet axis, permanent magnetic field, and rf magnetic field. Under certain orientations of the rf magnetic field and rf electrical current relative to the sheet axis, experimental dependences of the magnetoimpedance effect and effective permeability on the permanent magnetic field are found to correlate qualitatively. The experimental data are explained in terms of domain reconfiguration in the alloys. © *2003 MAIK "Nauka/Interperiodica".*

Today, the giant magnetoimpedance effect is one of the most topical areas of research in the physics of magnetic phenomena [1–4]. In spite of many publications devoted to this effect, a number of issues necessary for a complete understanding of its nature still remain unclear.

This work generalizes the complex study of this effect. Our goal was to trace how the mutual orientations of the axis of rolling of a fast-quenched amorphous metallic alloy (sheet axis), permanent external magnetic field H, and rf current I passing through the alloy influence the H dependence of the sample impedance Z. We also studied the effect of the orientations of the rf magnetic field h in a measuring coil and of the field H relative to the sheet axis on the H dependence of the effective permeability μ_{eff}.

The samples used were 25-μm-thick disks of diameter 9.5 mm cut from sheets of $Fe_4Co_{67}Mo_{1.5}Si_{16.5}B_{11}$ amorphous metallic alloy. The saturation magnetostriction and saturation induction of the alloy were, respectively, $\lambda_s = -3 \times 10^{-7}$ and $B_s = 0.55$ T.

The amount of the magnetoimpedance effect was measured with a high-value resistor series-connected to the sample. The amount of the effect $\Delta Z/Z_0$ was determined as [4]

$$\Delta Z/Z_0 = (Z_H - Z_0)/Z_0 = (U_H - U_0)/U_0, \qquad (1)$$

where Z_0 is the sample impedance at $H = 0$, Z_H is that in a magnetic field H, U_H is the voltage across the sample in the field H, and U_0 is the voltage at $H = 0$.

The magnetoimpedance effect was studied in magnetic fields up to 9600 A/m and current frequencies 0.1–10 MHz.

When the sheet axis is aligned with the current direction and magnetic field H, the impedance Z remains virtually the same as H grows in the interval 0–1200 A/m. With a further increase in H, the impedance decreases monotonically and the dependence $Z(H)$ tends toward saturation (Fig. 1a). The magnetoimpedance effect is negative: in a field of 9600 A/m and current frequency of 1 MHz, it equals 50%.

With the sheet axis parallel to the rf current and normal to the direction of the applied permanent magnetic field H, the impedance decreases monotonically (Fig. 1b) with increasing H. In this case, the decrease is faster than when H is parallel to the sheet axis and passing current and $\Delta Z/Z_0$ tends toward saturation at lower magnetic fields. The negative value of the magnetoimpedance effect is 54% for an rf current frequency of 1 MHz and magnetic field 9600 A/m.

When the sheet axis is normal to both the magnetic field H and rf current, the impedance Z first grows with H, peaks at $H = 1680$ A/m, and then declines monotonically, tending toward saturation (Fig. 1c). The positive effect has a maximal value of 60% at an rf current frequency of 1.5 MHz; the negative effect, 25% for $H = 9600$ A/m and a current frequency of 1.5 MHz.

Finally, when the sheet axis is parallel to H and normal to the rf current, the impedance Z remains practically the same for $H = 0$–1200 A/m. As H grows further, the impedance decreases monotonically and the dependence $Z(H)$ tends toward saturation (Fig. 1d). The effect is negative: 28% for $H = 9600$ A/m and a current frequency of 1.5 MHz.

In the saturation field, the impedance is minimal irrespective of the H and current orientations relative to the sheet axis.

According to [5], the magnetoimpedance effect is related to the skin depth, which in magnetic materials is given by

$$\delta = (\rho/\pi f \mu_0 \mu_{eff})^{1/2}, \qquad (2)$$

where ρ is the bulk resistivity of the sample, f is the variable current frequency, μ_{eff} is the effective permeability, and μ_0 is the permeability of free space.

1063-7842/03/4804-0427$24.00 © 2003 MAIK "Nauka/Interperiodica"

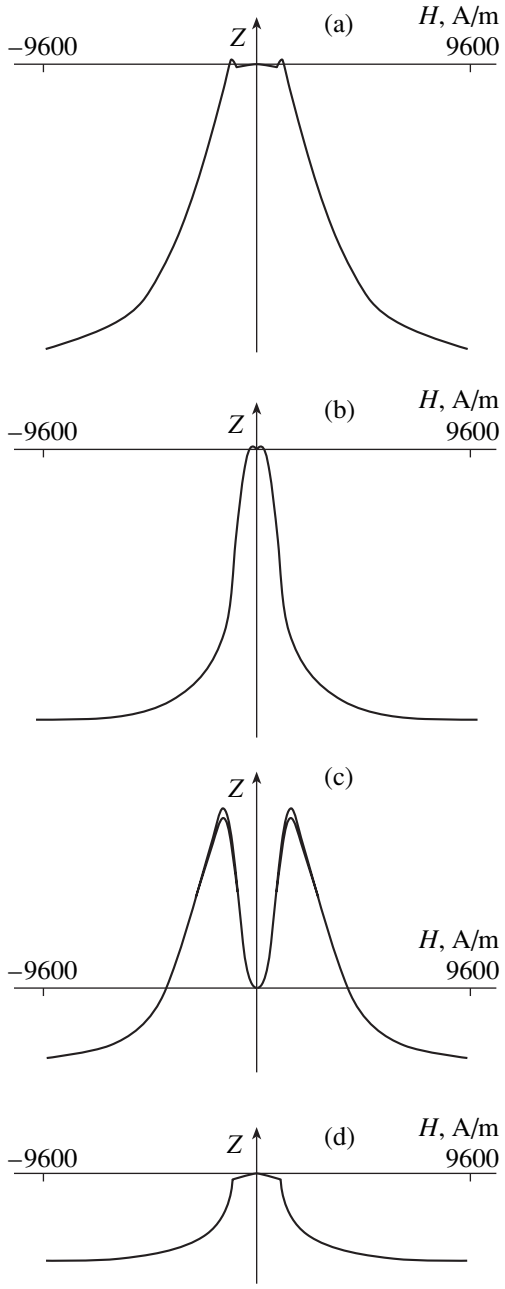

Fig. 1. Sample impedance vs. applied permanent magnetic field. The sheet axis is (a) parallel to both the rf current and external permanent magnetic field, (b) parallel to the rf current and normal to the magnetic field, (c) normal to both the rf current and the magnetic field, and (d) normal to the rf current and parallel to the external magnetic field.

The sample impedance is

$$Z \sim (\rho f \mu_{eff})^{1/2}. \quad (3)$$

In amorphous magnetically soft materials, an applied magnetic field may change μ_{eff} [5] by two or three orders of magnitude. Accordingly, the skin depth varies significantly and the effect may be very high. It should be noted that μ_{eff} and, hence, Z are complicated functions of H. Therefore, to adequately explain the experimental dependences $Z(H)$, it is necessary to study the dependence $\mu_{eff}(H)$ at various mutual orientations of the sheet axis, H, and a probing rf magnetic field h.

The value of μ_{eff} was measured with a series-connected resistor and measuring coil. The sample was placed into the coil with an rf current generating a probing rf magnetic field h. The voltage picked up from the coil is proportional to the effective permeability of the sample. The effective permeability was measured at frequencies of the rf magnetic field from 0.1 to 10 MHz and H = 0–9600 A/m with various orientations of the sheet axis, permanent magnetic field H, and rf probing field h.

Figure 2 shows the H dependences of the effective permeability at different frequencies of the probing field h. In Fig. 2a, the sheet axis is parallel to H and normal to h; in Fig. 2b, the sheet axis is normal to both H and h; in Fig. 2c, the sheet axis is normal to H and parallel to h; and in Fig. 2d, the sheet axis is aligned with both H and h. It is seen that the run of the curves is similar to that of the dependences $Z(H)$ provided that the rf current passing through the sample during the impedance measurements is perpendicular to the probing rf field during the permeability measurements. In measuring the magnetoimpedance effect, the magnetization of the sample interacts with the rf magnetic field, which is generated by the rf current and is normal to it.

Qualitatively, our experimental results can be explained as follows. In fast-quenched cobalt-based amorphous alloys, the surface axis of easy magnetization was shown [6] to coincide with the axis of rolling of the sheet. Therefore, it can be assumed that strip domains prevail on the sheet surface.

Experimental study [7] of the permeability in relation to magnetization reversal processes showed that the permeability is insensitive to the displacement of 180° domain walls. In the case when magnetization reversal is accomplished by rotation, μ_{eff} first grows with H, reaches a maximum, and then monotonically decreases, tending toward saturation under high fields. Generally, μ_{eff} can be represented as

$$\mu_{eff} = \mu_d + \mu_r,$$

where μ_d is the component due to domain wall displacement and μ_r is that due to magnetization rotation.

The components μ_d and μ_r depend on the frequency of the variable magnetic field in a different manner [7]. The former component decreases more sharply than the latter when the variable magnetic field frequency grows.

With the sheet axis normal to h and parallel to H, the effective permeability μ_{eff} is defined by μ_r. Magnetization reversal in this case is accomplished by domain wall displacement under the action of the permanent magnetic field H. At the initial portion of the curve

$\mu_{eff}(H)$, the effective permeability is nearly constant. As H grows further, μ_{eff} declines and tends to unity in the saturation field (Fig. 2a). The effect is negative and reaches a maximum in the saturation field:

$$\Delta Z/Z_0 = (\mu_s^{1/2} - \mu_r^{1/2})/\mu_r^{1/2}, \quad (4)$$

where μ_s is the effective permeability of the sample in the saturation state.

When the sheet axis is normal to h and H (Fig. 2b), the component μ_r also makes a major contribution to the effective permeability. As the permanent magnetic field rises, the magnetization rotates and the magnetization component that is aligned with H grows; as a result, μ_r decreases and μ_d increases. If the condition $\mu_r > \mu_d$ is met at a certain frequency of the rf current, μ_{eff} drops more sharply than in the previous case, saturating at lower fields, and the initial portion with constant μ_{eff} is absent. Under the saturation field, μ_{eff} tends to unity; therefore, the maximal negative effect is determined by relationship (4) irrespective of the H direction.

With the sheet axis running parallel to h and normally to H, μ_d makes a major contribution to μ_{eff}. As H increases, the magnetization of the sample rotates. In this case, the contribution from μ_d to μ_{eff} diminishes and that from μ_r grows. If the condition $\mu_r > \mu_d$ is met at a certain frequency of the rf current, μ_{eff} increases with H. When the external magnetic field becomes sufficient to rotate the magnetization about H, i.e., becomes equal to the field of anisotropy H_a, μ_{eff} is completely specified by μ_r and attains a maximum (Fig. 2c). Under such a magnetic field, the sample impedance and positive magnetoimpedance effect reach maxima. The magnetoimpedance effect depends on the variation of μ_{eff} with H and is given by

$$\Delta Z/Z_0 = (\mu_r^{1/2} - \mu_d^{1/2})/\mu_d^{1/2}. \quad (5)$$

As H grows further, μ_{eff} decreases and tends to unity under the saturation field. In this case, the maximum negative magnetoimpedance effect is observed under the saturation field and depends on μ_{eff}, which varies with H from μ_d at $H \approx 0$ to μ_r at $H \approx H_a$:

$$\Delta Z/Z_0 = (\mu_s^{1/2} - \mu_d^{1/2})/\mu_d^{1/2}. \quad (6)$$

With the sheet axis running parallel to h and H (Fig. 2d), μ_d makes a major contribution to μ_{eff}. As H increases, 180° domain walls are displaced. In the initial portion of the curve $\mu_{eff}(H)$, the effective permeability remains virtually invariable in this case. Then, μ_{eff} drops with increasing H and tends to unity in the saturation magnetic field. In this case, the positive magnetoimpedance effect is absent and the maximal negative effect depends on μ_{eff}, which varies from μ_d to $\mu_s = 1$ according to (6).

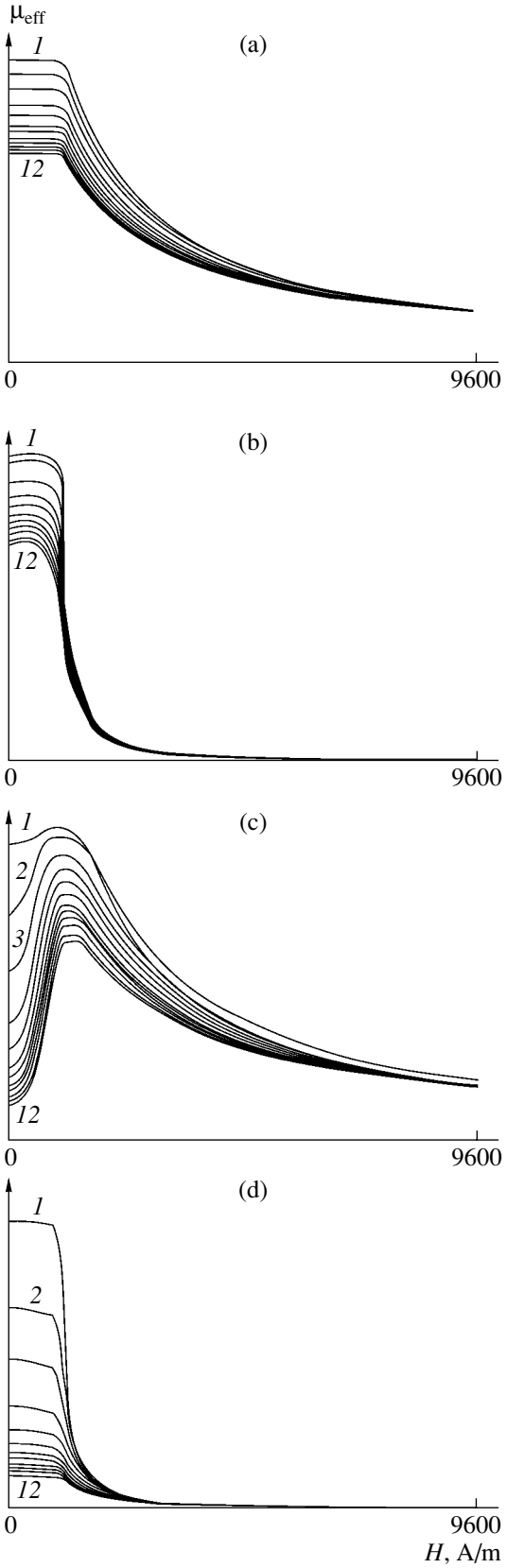

Fig. 2. Family of the effective permeability vs. permanent magnetic field curves. The frequencies of the probing magnetic field are (*1*) 0.1, (*2*) 0.5, (*3*) 1, (*4*) 2, (*5*) 3, (*6*) 4, (*7*) 5, (*8*) 6, (*9*) 7, (*10*) 8, (*11*) 9, and (*12*) 10 MHz.

From the results obtained, the following conclusions can be drawn.

(1) In cobalt-based amorphous metallic alloys produced by fast quenching from the melt, the dependence of the magnetoimpedance effect on the permanent magnetic field is governed by the mutual orientations of the sheet axis, permanent magnetic field, and rf current passing through the alloy.

(2) The dependences of the magnetoimpedance effect and effective permeability on the applied permanent magnetic field correlate when the probing magnetic field and rf current are appropriately oriented relative to the sheet axis.

(3) The differences in the impedance vs. permanent magnetic field curves under different mutual orientations of the sheet axis, permanent magnetic field, and rf current are associated with different magnetization reversal mechanisms in the samples studied.

REFERENCES

1. L. V. Panina, K. Mohri, K. Bushida, *et al.*, J. Appl. Phys. **76**, 6198 (1994).
2. R. S. Beach and A. R. Berkowitz, J. Appl. Phys. **76**, 6209 (1994).
3. L. Panina and K. Mohri, J. Magn. Magn. Mater. **157/158**, 137 (1996).
4. O. L. Sokol-Kutylovskiĭ, Fiz. Met. Metalloved. **84** (3), 54 (1997).
5. A. S. Antonov, S. N. Gadetskiĭ, A. B. Granovskiĭ, *et al.*, Fiz. Met. Metalloved. **83** (6), 60 (1997).
6. K. Suzuki, H. Fujimori, and K. Hasimoto, *Amorphous Metals* (Metallurgiya, Moscow, 1987), translated from Japanese.
7. N. S. Chislyakov and B. P. Tushkov, *Apparatus and Methods for Thin Magnetic Film Studies* (Krasnoyarsk, 1968), pp. 291–296.

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 4, 2003, pp. 431–436. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 4, 2003, pp. 53–58.

SOLID-STATE ELECTRONICS

Current–Voltage Characteristics of a Spin Half-Metallic Transistor

A. K. Zvezdin, A. S. Mishchenko, and A. V. Khval'kovskiĭ

Institute of General Physics, Russian Academy of Sciences, Moscow, 119991 Russia

e-mail: khvalkov@ran.gpi.ru

Received June 24, 2002

Abstract—A new design of spin transistor based on half-metallic ferromagnets (referred to as a spin half-metallic transistor) is suggested, and its current–voltage characteristics are studied theoretically. Like a bipolar transistor, the new device can amplify current. At the same time, the properties of a spin half-metallic transistor depend considerably on the mutual orientation of the magnetizations of its three contacts. We also propose a device based on an $F^{\uparrow}$–$F^{\downarrow}$ junction. This device consists of two single-domain half-metallic parts with opposite magnetizations. There is a range of voltages where the current–voltage characteristics of an $F^{\uparrow}$–$F^{\downarrow}$ junction and a semiconductor diode are similar. The behavior of an $F^{\uparrow}$–$F^{\downarrow}$ junction under different conditions is studied.

INTRODUCTION

A new direction, spin electronics (spintronics), has recently evolved in applied physics. To process information, spintronic devices, unlike conventional microelectronic devices, use not only the charge of an electron but also its spin [1]. Spintronics is based on the so-called spin-dependent transport of electrons, a physical phenomenon discovered late in the 20th century, the essence of which is the dependence of electron transport properties on the spin degree of freedom. Examples of spin-dependent transport are giant magnetoresistance [2–5], the magnetoresistance of tunnel junctions and nanocontacts [6–9], the spin accumulation effect, and the surface resistance of heterojunctions [10, 11]. Currently available spintronic devices include sensing heads for high-density hard disks and magnetic RAMs. Spintronic devices based on the spin injection effect seem to be even more promising [12, 13], since they allow one to process and read information written in spin degrees of freedom. Spin transistors belong to such devices.

Two types of spin transistors are being intensely discussed in the literature: a spin field-effect transistor (spin FET) [14] and an $F^{\uparrow}$–N–$F^{\downarrow}$ transistor [15], where F and N mean a ferromagnetic and normal metal, respectively. If a spin field-effect transistor is very close to a bipolar transistor in principle of operation, the apparent analogy between an $F^{\uparrow}$–N–$F^{\downarrow}$ and a bipolar transistor fails from the viewpoint of device functioning. In the former, a base layer (normal metal) does not serve to amplify the current between the emitter and collector. Instead, it is often intended for measuring the potential difference at the base–collector heterojunction. This potential difference arises when the spin is injected from the emitter and depends on the spin orientation in the collector junction. Such a device is not an amplifier in the strict sense of the word.

This paper considers the design of an $F^{\uparrow}$–$F^{\downarrow}$–$F^{\uparrow}$ spin transistor, which has similar current amplification properties as an n–p–n (or p–n–p) bipolar transistor but its characteristics depend considerably on the mutual orientation of electron spins in the emitter, base, and collector.

Our transistor is based on half-metallic ferromagnets [16–18]. These materials are metals, but their band structure has the following feature: the densities of states are different for spins of opposite orientation, and one of them has a discontinuity (gap). Such materials behave like a metal for electrons of one (major) polarization and like a semiconductor for opposite (minor) polarization.

In order to stress how it differs in band structure from a semiconductor transistor, the transistor discussed will be referred to as a spin half-metallic (SHM) transistor. To describe electron transport in ferromagnets, the two-current approach (currents of electrons with spins up and down) was proposed [10, 19, 20].

Such materials as Fe_3O_4, Cr_2O, $LaPrMnO_3$, Heusler alloys (PtMnSb and PtMnSn [21, 22]), etc., are examples of half-metals. In these materials, the gap 2Δ is about several tenths of an electronvolt. At room temperature, the minority electron conductivity can be several orders of magnitude lower than the majority electron

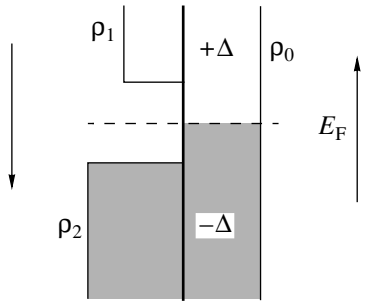

Fig. 1. Schematic band structure of a ferromagnetic half-metal.

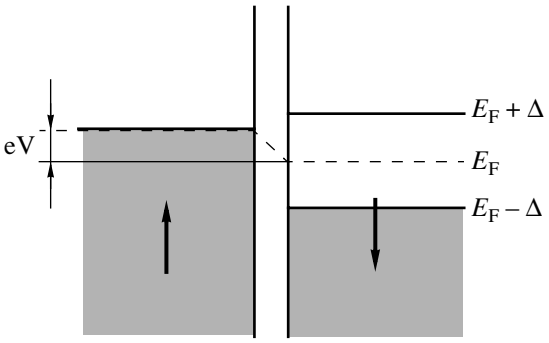

Fig. 2. Band structure of the $F^{\uparrow}$–$F^{\downarrow}$ junction for major polarization in the left contact. The arrows indicate the domain magnetization direction.

conductivity. The model band structure of these materials is shown in Fig. 1.

$F^{\uparrow}$–$F^{\downarrow}$ JUNCTION

Let us consider a sample comprising two domains of opposite orientation. To prevent the formation of the domain wall, a thin insulating layer of thickness from one to several atomic layers should be inserted between the domains. For electrons of major polarization at the left junction, the band structure is shown in Fig. 2. For the opposite polarization, the structure is similar.

It is assumed that the tunneling resistance of the insulating layer is lower than the resistance due to the effect of spin accumulation [10, 22, 23].

It should be noted that the resulting current in the insulating layer decreases by a factor on the order of $\exp(-\sqrt{2m(E-U)}\,L/\hbar)$, where U is the potential barrier height, L is the barrier length, E is the electron kinetic energy, and m is the electron effective mass. For example, for $(E - U) \approx 4$ eV and $L \approx 5$ Å, the current decay is about 6×10^{-3}. Although in the case considered the boundary between the domains is an insulating layer rather than a domain wall, we assume below that the polarization of electrons when they pass through this layer remains unchanged. The theoretical dependence of the current through a spin diode made of ferromagnetic semiconductors on the domain wall thickness and temperature [24] shows that electrons retain their spin polarization for sharp domain walls (i.e., with a thickness of less than 20 nm).

Consider the *I–V* characteristic of such a structure. The tunnel electron current related to one of the polarizations can be calculated by the formula (see, for example, [26])

$$I_p = \frac{2\pi e|M|^2}{\hbar} \times \int_{-\infty}^{\infty} \{f(E-eV) - f(E)\}\rho_R(E-eV)\rho_L(E)dE, \quad (1)$$

where $f(x)$ is the Fermi function; ρ_L and ρ_R are the density of states for electrons of the conducting polarization in the left and right ferromagnets, respectively; V is the potential difference between contacts to the $F^{\uparrow}$–$F^{\downarrow}$ junction; M is a tunneling matrix element (assumed to be constant); and e is the electron charge.

In our case, the functions ρ_L and ρ_R for half-metallic ferromagnets are given by (Fig. 1)

$$\rho_L = \rho_0 = \text{const}, \qquad \rho_R = \begin{cases} \rho_1; & E > E_F + \Delta \\ 0; & E_F - \Delta < E < E_F + \Delta \\ \rho_2; & E < E_F - \Delta, \end{cases} \quad (2)$$

where E_F is the Fermi energy.

Integral (1) can be taken analytically:

$$I_p = \frac{2\pi e|M|^2}{\hbar}\rho_0\rho_1 \times \left\{ eV + kT\frac{\rho_2}{\rho_1}\ln\left(\frac{1+\exp((eV-\Delta)/kT)}{1+\exp(-\Delta/kT)}\right) - kT\ln\left(\frac{1+\exp((eV+\Delta)/kT)}{1+\exp(\Delta/kT)}\right)\right\}. \quad (3)$$

For the special case $\rho_1 = \rho_2 = \rho_0$ and $\Delta = 0$, Eq. (3) yields the tunnel current for a junction where the magnetizations of edge contacts are codirected (an $F^{\uparrow}$–$F^{\downarrow}$ junction). This current can be given in the form[1] (cf. [24–26])

$$I_{\uparrow\uparrow} = GV, \quad (4)$$

where G is the conductivity of such a structure. In our

[1] This is the current of majority carriers only. The current of minority carriers is neglected, since it is much less than the former at voltages $|V| \ll 2\Delta$ [10, 22].

case,

$$G = \frac{2\pi e^2 |M|^2}{\hbar}\rho_0^2.$$

The total current $I_{\uparrow\downarrow}$ of an $F^{\uparrow}$–$F^{\downarrow}$ junction is equal to the sum of the currents of both polarizations. For low voltages ($|eV| < kT$), it can be expressed as

$$I_{\uparrow\downarrow} = \left(\frac{\rho_1 + \rho_2}{\rho_0} G e^{-\Delta/kT}\right) V = \left(\frac{\rho_1 + \rho_2}{\rho_0} e^{-\Delta/kT}\right) I_{\uparrow\uparrow}; \quad (5)$$

for $V \gg \Delta$,

$$I_{\uparrow\downarrow} = \frac{\rho_1 + \rho_2}{\rho_0} G(V - \Delta). \quad (6)$$

Thus, at moderate temperatures, the current of an $F^{\uparrow}$–$F^{\downarrow}$ junction is roughly zero for voltages $|eV| < \Delta$; otherwise, it can be represented by straight line (6) (Fig. 3).

An $F^{\uparrow}$–$F^{\downarrow}$ junction may be called a spin diode, since the nonlinear part of the I–V curve near the point $V = \Delta$ behaves like the characteristic of a conventional diode. Obviously, this seeming similarity breaks for $V < -\Delta$, where the I–V curve of the junction has a nonzero value.

NONCOLLINEAR POLARIZATIONS

Let us discuss the case when the magnetizations of the electrodes lie at an angle ϑ to each other (Fig. 4). Consider an electron that has come from the left domain to the right. Since, according to our assumption, the polarization of the electron has not changed, its spin in the right domain will make the same angle ϑ with the magnetization axis. The quantum-mechanical state of this electron $|\vartheta_l^{\uparrow}\rangle$ [27] is given by

$$|\vartheta_l^{\uparrow}\rangle = \sin\frac{\vartheta}{2}|\psi_r^{\downarrow}\rangle + \cos\frac{\vartheta}{2}|\psi_r^{\uparrow}\rangle. \quad (7)$$

According to the Fermi golden rule, the transition probability per unit time is proportional to the transition matrix element squared and the states $|\psi_r^{\downarrow}\rangle$ and $|\psi_r^{\uparrow}\rangle$ are orthogonal to each other; therefore, the resulting current is

$$I = \cos^2\frac{\vartheta}{2} I_{\uparrow\downarrow} + \sin^2\frac{\vartheta}{2} I_{\uparrow\uparrow}. \quad (8)$$

Upon simple transformations, Eqs. (3) and (8) yield a general expression for the spin diode current in the case of arbitrarily magnetized domains:

$$I = \frac{G}{e}\left\{\left(eV + kT\ln\left(\frac{1 + \exp((eV - \Delta)/kT)}{1 + \exp(-\Delta/kT)}\right)\right.\right.$$

$$\left.\left. - kT\ln\left(\frac{1 + \exp((eV + \Delta)/kT)}{1 + \exp(\Delta/kT)}\right)\right)\right.$$

$$\left. \times \frac{\rho_1 + \rho_2}{\rho_0}\cos^2\left(\frac{\vartheta}{2}\right) + eV\sin^2\left(\frac{\vartheta}{2}\right)\right\}. \quad (9)$$

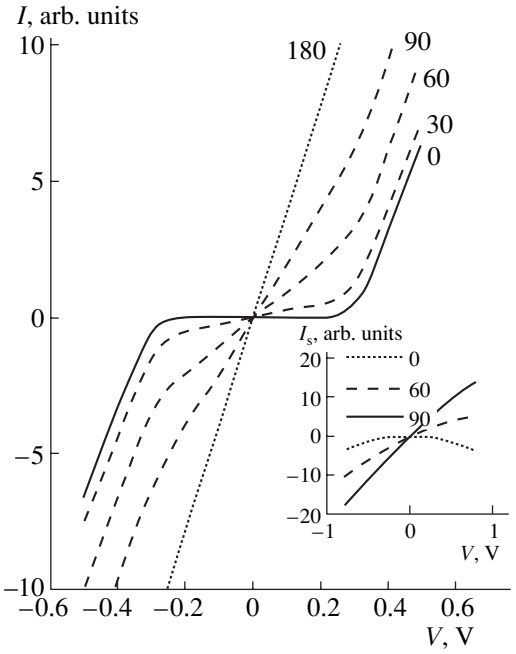

Fig. 3. I–V curve of the $F^{\uparrow}$–$F^{\downarrow}$ junction for various angles ϑ between the domain magnetization. The case $\vartheta = 0$ corresponds to the oppositely magnetized domains; $\vartheta = 180°$, to the domains magnetized in one direction ($F^{\uparrow}$–$F^{\downarrow}$ junction). The inset shows the behavior of the spin current for several ϑ. The figures by the curves are ϑ in degrees.

The results of calculation by this formula are shown in Fig. 3 for different values of ϑ and $\rho_0/\rho_2 = 3$, $\rho_1/\rho_2 = 1.5$, $kT = 25$ meV, and $\Delta = 0.3$ eV.

The behavior of the spin current for an $F^{\uparrow}$–$F^{\downarrow}$ junction is shown in the inset to Fig. 3. Here, the spin current is equal to the difference between the currents of the two polarizations. It is shown for several values of ϑ and the same values of the problem parameters as above. The expression for the angular dependence of the spin current is

$$I_s = \frac{G}{e}\left\{\left(eV - kT\ln\left(\frac{1 + \exp((eV - \Delta)/kT)}{1 + \exp(-\Delta/kT)}\right)\right.\right.$$

$$\left.\left. - kT\ln\left(\frac{1 + \exp((eV + \Delta)/kT)}{1 + \exp(\Delta/kT)}\right)\right)\right.$$

$$\left. \times \frac{\rho_1 - \rho_2}{\rho_0}\cos^2\left(\frac{\vartheta}{2}\right) + eV\sin^2\left(\frac{\vartheta}{2}\right)\right\}. \quad (10)$$

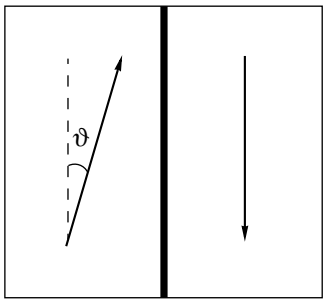

Fig. 4. Spin diode where the magnetization of one domain makes an angle ϑ with the magnetization of the other.

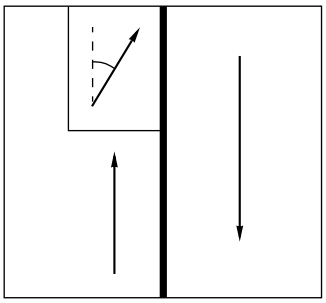

Fig. 5. Spin diode with an additional obliquely magnetized domain inside one of the contacts.

Let us also consider the case when one more domain is present in one of the electrodes that is magnetized at an angle to the magnetizations of both contacts (Fig. 5). In this case, the current is given by

$$I = \frac{\left((S - S_{\rightarrow}) + S_{\rightarrow}\cos^2\frac{\vartheta}{2}\right)}{S} I_{\uparrow\downarrow} + \frac{S_{\rightarrow}}{S}\sin^2\frac{\vartheta}{2} I_{\uparrow\uparrow} \quad (11)$$
$$= \left(1 - \delta S \sin^2\frac{\vartheta}{2}\right) I_{\uparrow\downarrow} + \delta S \cos^2\frac{\vartheta}{2} I_{\uparrow\uparrow},$$

where S is the total contact surface area, $S_{\rightarrow}$ is the surface area of the domain with inclined magnetization, and $\delta S = S_{\rightarrow}/S$ is the relative surface area of this domain. In this case, the I–V characteristic is similar to that for oppositely directed polarizations.

SHM TRANSISTOR

We now turn to a three-contact device based on half-metallic ferromagnets. The band structure of a spin transistor for electrons polarized so that base carriers have minor polarization is shown in Fig. 6. Such a polarization will be called conducting polarization. The contribution from the current of the opposite polarization can be neglected for $kT \ll \Delta$, $V_{EB} > 0$, and $V_{BC} < \Delta$.[2]

We consider a common-base transistor. For this configuration, the dependence of the potential energy on the coordinate x along the transistor for electrons of conducting polarization is modulated as shown in Fig. 7.

The diffusion equation for nonequilibrium carriers in the base (namely, electrons injected from the emitter) has the form

$$D\frac{\partial^2 n}{\partial x^2} = \frac{n - n_0}{\tau}, \quad (12)$$

[2] This is because, for this polarization, the carriers in the base are majority carriers (in the emitter and collector, they are minority carriers). Therefore, the emitter current cannot influence the collector current in this range of temperatures and voltages.

where D is the electron diffusion coefficient, τ is the average spin lifetime of an electron in the base, $(n - n_0)$ is the concentration of electrons injected from the emitter into the base, and n_0 is the equilibrium electron concentration (all the parameters are considered for minority carriers).

The boundary conditions for the electron concentrations at the base boundaries are specified by the Boltzmann factor:

$$n(+0) = n_0 e^{-\frac{\Delta_1 - eV_{EB}}{kT}},$$
$$n(W - 0) = n_0 e^{-\frac{\Delta_2 + eV_{BC}}{kT}}. \quad (13)$$

Here, W is the base width; $n(+0)$ and $n(W - 0)$ are the conduction electron concentrations at the emitter–base and base–collector interfaces; Δ_1 and Δ_2 are the potential barriers for electrons tending to penetrate from the emitter to the base and from the collector to the base in the absence of the voltage; and V_{EB} and V_{BC} are the emitter-to-base and base-to-collector voltages, respectively. For generality, we consider the case $\Delta_1 \neq \Delta_2$, which corresponds to variously magnetized emitter and collector.

Given the distribution of the minority electron concentration in the base, $n(x)$, we can calculate the emitter, I_E, collector, I_C, and base, I_B, currents by the formulas [28]

$$I_E = eD\left.\frac{\partial n}{\partial x}\right|_{x=0},$$
$$I_C = eD\left.\frac{\partial n}{\partial x}\right|_{x=W}, \quad (14)$$
$$I_B = I_E - I_C.$$

Solving Eq. (12) with boundary conditions (13) and substituting the solution into Eq. (14), we arrive at the

following expressions for the transistor currents:

$$I_E = \frac{eDSn_0}{L}\frac{1}{\sinh\frac{W}{L}}\left\{\alpha\coth\frac{W}{L} - \beta\right\}, \quad (15)$$

$$I_C = \frac{eDSn_0}{L}\frac{1}{\sinh\frac{W}{L}}\left\{\alpha - \beta\coth\frac{W}{L}\right\}, \quad (16)$$

$$I_B = \frac{eDSn_0}{L}\frac{\cosh\frac{W}{L} - 1}{\sinh\frac{W}{L}}\{\alpha + \beta - 2\}. \quad (17)$$

Here,

$$\alpha = \exp\left(-\frac{\Delta_1 - eV_{EB}}{kT}\right), \quad (18)$$

$$\beta = \exp\left(-\frac{\Delta_2 + eV_{BC}}{kT}\right), \quad (19)$$

S is the cross-sectional area of the transistor and $L = \sqrt{D\tau}$ is the electron mean free path in the base.

DISCUSSION

As follows from the *I–V* characteristic (Fig. 8), calculated for the potential energy profile in Fig. 7, the spin transistor behaves similarly to its semiconductor analogue. This means that this device can amplify current. However, the base size in semiconductor transistors is known to be limited by the thickness of the *p–n* junction and cannot be narrower than a tenth of a micrometer. In an SHM transistor, this restriction is removed: the junction area may be shrunk to several interatomic spacings.

One more advantage of an SHM transistor is that it can operate as a switching device and, correspondingly, can be used as an element of a magnetic RAM. If the central domain is reversely magnetized, all the domains will be magnetized in the same direction and the current will increase many times. Magnetization reversal in only one domain can be accomplished by properly selecting the coercive forces of the emitter, base, and collector. For example, if the coercive force of the base is less than that of the emitter and collector, there appears a range of magnetic fields where the base magnetization is switched but the emitter and collector magnetizations are not.

Although this work focuses on an SHM transistor made of half-metallic ferromagnets, the basic idea of an $F^{\uparrow}$–$F^{\downarrow}$–$F^{\uparrow}$ spin transistor as a current (or voltage) amplifier can be generalized to normal ferromagnetic metals [28, 29]. Of fundamental importance here is the presence of two electron currents (with spins up and down) and spin-dependent transport properties in a ferromagnet [20].

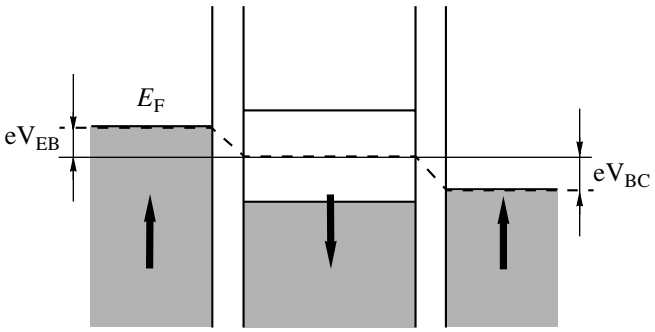

Fig. 6. Band structure of a spin transistor for electrons of conducting polarization.

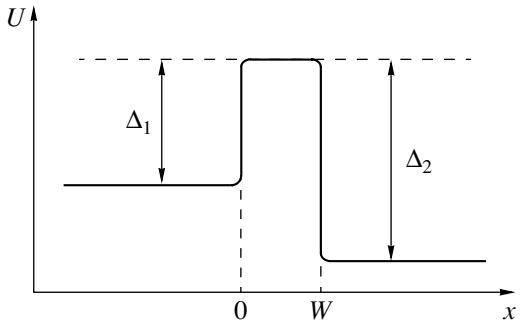

Fig. 7. Potential energy profile for electrons of conducting polarization.

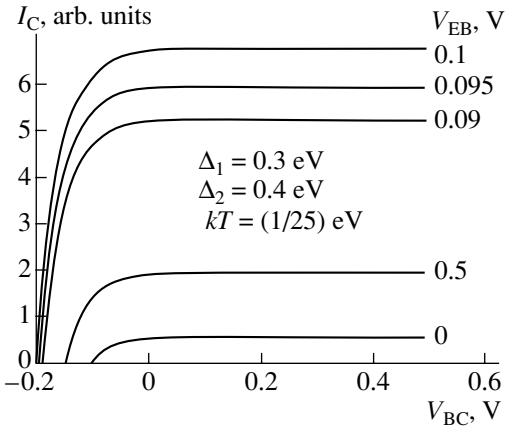

Fig. 8. Family of the *I–V* characteristics for a spin transistor with various emitter-to-base voltages V_{EB}.

CONCLUSION

We considered an $F^{\uparrow}$–$F^{\downarrow}$–$F^{\uparrow}$ device (SHM transistor) based on half-metallic ferromagnets. Unlike conventional spin transistors, an SHM transistor is capable of amplifying current like its semiconductor analogue. In addition, an SHM transistor can operate as a switch

and, therefore, can be used as an element of magnetic RAMs. In this paper, we also considered an $F^{\uparrow}$–$F^{\downarrow}$ spin junction. This device has a nonlinear *I–V* characteristic, which resembles that of a semiconductor diode in several voltage ranges. The *I–V* characteristic of the $F^{\uparrow}$–$F^{\downarrow}$ spin junction for different operating regimes, as well as the characteristic of the spin SHM transistor, was constructed.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project nos. 02-02-17389 and 99-02-17830) and the federal program "Physics of Solid-State Nanostructures."

REFERENCES

1. G. A. Prinz, Science **282**, 1660 (1988); I. Zutic, cond-mat/0112368 (2001).
2. M. N. Bainbich *et al.*, Phys. Rev. Lett. **61**, 2472 (1987); G. Bransch *et al.*, Phys. Rev. B **39**, 4829 (1989).
3. M. A. M. Gujs and G. E. W. Bauer, Adv. Phys. **46**, 285 (1997); V. V. Dobrovitskiĭ and A. K. Zvezdin, Usp. Fiz. Nauk **166**, 439 (1996) [Phys. Usp. **39**, 407 (1996)].
4. A. K. Zvezdin and S. N. Utochkin, Pis'ma Zh. Éksp. Teor. Fiz. **57**, 433 (1993) [JETP Lett. **57**, 439 (1993)].
5. R. Meservey and P. M. Tedrow, Phys. Rep. **238**, 173 (1994).
6. N. Garcia, M. Munoz, and Y.-W. Zhao, Phys. Rev. Lett. **82**, 2923 (1999).
7. G. Tatara, Y.-W. Zhao, M. Munoz, and N. Garcia, Phys. Rev. Lett. **83**, 2030 (1999); M. Munoz *et al.*, Appl. Phys. Lett. **79**, 2946 (2001); N. Garcia *et al.*, Appl. Phys. Lett. **79**, 4550 (2001).
8. A. K. Zvezdin and A. F. Popkov, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 304 (2000) [JETP Lett. **71**, 209 (2000)]; H. Imamura, N. Kobayashi, S. Takahashi, *et al.*, Phys. Rev. Lett. **84**, 1003 (2000).
9. L. R. Tagirov, B. P. Vodopyanov, and K. B. Efetov, Phys. Rev. B **63**, 104428 (2001); L. L. Savchenko, A. K. Zvezdin, A. F. Popkov, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **43**, 1449 (2000).
10. P. C. Van Son, H. Van Kempen, and P. Wyder, Phys. Rev. Lett. **58**, 2271 (1987).
11. M. V. Tsoi *et al.*, Phys. Rev. Lett. **80**, 4281 (1998).
12. A. G. Aronov, JETP Lett. **24**, 32 (1976).
13. M. Johnson and R. H. Silsbee, Phys. Rev. Lett. **55**, 1790 (1985); M. Johnson and R. H. Silsbee, Phys. Rev. B **35**, 4959 (1987); Y. Ji *et al.*, Phys. Rev. Lett. **86**, 5585 (2001).
14. S. Datta and B. Das, Appl. Phys. Lett. **56**, 665 (1990).
15. M. Johnson, Science **260**, 320 (1993).
16. H. Y. Hwang, S.-W. Cheong, N. P. Ong, *et al.*, Phys. Rev. Lett. **77**, 2041 (1996).
17. J. M. D. Coey *et al.*, Phys. Rev. Lett. **87**, 026601 (2001); J. M. D. Coey, Phys. Rev. B **64**, 020407-1 (2001); J. M. D. Coey, Phys. Rev. Lett. **80**, 3815 (1998).
18. J. M. De Teresa *et al.*, Phys. Rev. Lett. **82**, 4288 (1999).
19. A. Fert and I. A. Campbell, J. Phys. (Paris), Colloq. **32**, C1 (1971).
20. T. Valet and A. Fert, Phys. Rev. B **48**, 7099 (1993).
21. Yu. A. Uspenskiĭ, E. T. Kulatov, and S. V. Khalilov, Zh. Éksp. Teor. Fiz. **107**, 1708 (1995) [JETP **80**, 952 (1995)].
22. A. K. Zvezdin and V. A. Kotov, *Modern Magnetooptics and Magnetooptical Materials* (IOP, Bristol, 1997), Sec. 5.2.
23. E. I. Rashba, Phys. Rev. B **62**, R16267 (2000).
24. G. Vignale and M. Flatté, cond-mat/0202002 (2002).
25. E. I. Wolf, *Principles of Electron Tunneling Spectroscopy* (Oxford University Press, 1985; Naukova Dumka, Kiev, 1990).
26. M. Tinkkham, *Introduction to Superconductivity* (McGraw-Hill, New York, 1975; Atomizdat, Moscow, 1980).
27. I. V. Savel'ev, *Principles of Theoretical Physics* (Nauka, Moscow, 1996), Vol. 2.
28. S. M. Sze, *Physics of Semiconductor Devices* (Wiley, New York, 1981; Mir, Moscow, 1984).
29. R. Vlutters, O. M. J. van't Erve, R. Jansen, *et al.*, Phys. Rev. B **65**, 024416 (2002).

Translated by M. Astrov

Technical Physics, Vol. 48, No. 4, 2003, pp. 437–440. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 4, 2003, pp. 59–62.

SOLID-STATE ELECTRONICS

Structure and Composition of Gallium Nitride Films Produced by Processing Gallium Arsenide Single Crystals in Atomic Nitrogen

G. A. Sukach*, V. V. Kidalov, M. B. Kotlyarevsky**, and E. P. Potapenko***

** Institute of Semiconductor Physics, National Academy of Sciences of Ukraine, pr. Nauki 45, Kiev, 03028 Ukraine*
e-mail: sukach@isp.kiev.ua
*** Berdyansk State Pedagogical Institute, Berdyansk, 71118 Ukraine*
Received September 9, 2002

Abstract—Thin GaN films are grown on (001) single-crystal GaAs substrates processed in an atmosphere of active nitrogen radicals. Auger electron spectroscopy is applied to take the depth profiles of the basic chemical elements that enter into the composition of the epitaxial GaN films and single-crystal GaAs substrates. It is found that the surface composition of the GaN films is characterized by considerable nonstoichiometry (the excess nitrogen achieves ≈9%), which is caused by the presence of atomic nitrogen in the discharge chamber. With a high-resolution X-ray diffraction method, the structural perfection of the epitaxial layers is investigated. It is shown that low-temperature annealing (at temperatures below 700°C) is responsible for the formation of cubic GaN films on the (001) surface of cubic GaAs, whereas higher temperature annealing results in the growth of the hexagonal films. *© 2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Group-III nitrides (InN, GaN, and AlN) and related ternary compounds are wide-gap materials with unique properties, which are promising for electronic and optoelectronic device technologies (see, for example, [1, 2]). The probability of defect formation in the nearly perfect lattice of nitrides is less than in the lattices of arsenides and phosphides, since metal–nitrogen bonds are stronger than bonds between metal and As (or P) atoms. Accordingly, the performance of nitride-based devices (service life; output power; operating frequency; thermal, mechanical, radiation, and chemical stabilities; speed; and supply voltage) is superior to arsenide- and phosphide-based [2] devices.

Among the many fundamental problems to be tackled in order to improve the quality and reliability of opto- and microelectronic devices based on Group-III nitrides, one may note the lack of good substrates from the standpoint of their crystallographic (inadequate lattice parameter and temperature coefficient of expansion), morphological, topological, structural, and electrophysical properties and the associated problem of the buffer layer. Therefore, today's technologies for structure modification, specifically, the synthesis of novel materials by processing single-crystal semiconductors in an atmosphere of active gas radicals, attract much attention because of their potential application [3].

In the commercial-scale production of semiconductor (primarily GaN-based) heterostructures, the most frequently used substrates are made of Al_2O_3 (basal (lattice mismatch ≈13.9%) and other planes are used for growth), SiC, Si (≈4.5%), AlN (≈2.5%), $LiGaO_2$ (≈1%), and many other materials [1].

The best method of producing high-quality thin GaN films is homoepitaxy on substrates made of GaN single crystals. Another type of substrate is a structure consisting of a thin GaN layer grown on a so-called quasi-substrate (Si, GaAs, or other materials). The latter approach may significantly simplify the process of epitaxy and make it cheaper. Basically, the problem of the buffer layer can also be solved in this way.

In this work, we study thin GaN films that are grown on GaAs substrates exposed to active nitrogen radicals and investigate their structural characteristics. It should be noted that the equipment for our method is 100 times cheaper than that for MO hydride and molecular-beam epitaxies, which are widely used in the production of nitrides.

EXPERIMENTAL

GaN films were grown on (001)GaAs single-crystal substrates. The substrates were first ground and then finished by thin diamond pastes. After degreasing and removing the imperfect surface layer by etching in the standard etchant $H_2SO_4 : H_2O_2 : H_2O = 3 : 1 : 1$ for 2 min, the samples were rinsed in distilled water and placed into a high-frequency discharge chamber. GaN layers were produced by annealing the GaAs substrates in active (i.e., capable to incorporate into a growing semiconductor film) nitrogen radicals generated in a high-power rf plasma discharge. This is an attractive

1063-7842/03/4804-0437$24.00

technique for growing complex semiconductors, since it is simple and cheap (for details, see [4]).

The difficulty with using nitrogen for the epitaxial growth of semiconductor compounds is associated with the high binding energy in an N_2 molecule (9.76 eV at temperatures of 300–750°C), which makes it impossible to activate molecular nitrogen. Today, as an activation source, one usually uses different types of high-frequency plasma sources. We use the discharge of a high-frequency generator with an operating frequency of 40 MHz and power of up to 2 kW [5]. Under these conditions, the formation of gallium nitride is basically different from other methods of epitaxy. The major difference is the presence of the ionic component in the flow of particles toward the surface. In addition, the neutral particles themselves have an elevated energy and are in an excited state. These three factors make it possible to obtain GaN films virtually at room temperature.

To exclude surface damage due to nitrogen atoms, the high-frequency plasma was passed through a strong permanent magnetic field to remove its ionic component. Thus, the subsequent annealing of GaAs was performed in an atmosphere where neutral excited nitrogen atoms (active radicals) dominate. The concentration of active nitrogen radicals was measured with the help of an LT-2 manometer tube placed into the reaction chamber. The tube was sealed-in so that its catalytic filament was near the surface of the substrate being processed. Under the same conditions, the concentration of active nitrogen radicals in our method was 4 to 5 orders of magnitude higher than that in the nonactivated atmosphere. The yield of active nitrogen radicals under our experimental conditions attained 15% of the total amount of N_2 molecules. For ammonia, the yield was even higher. The concentration of the radicals was maximum when the pressure in the reaction chamber was varied from 10^{-3} to 10^{-1} torr.

RESULTS AND DISCUSSION

The elemental surface composition of epitaxial films obtained by our method was analyzed by Auger electron spectroscopy (AES). Figure 1 shows the AES spectra taken after irradiating the sample surface by the flux of primary 3-keV electrons. These spectra contain information on the elemental composition of only a thin (0.3- to 0.5-nm-thick) surface layer (the maximum AES depth is 3 nm [6]). A comparison between the energies of the spectral lines of Auger electrons emitted by the atoms to be identified and the known characteristic peaks of atomic spectra allowed us to clarify the chemical nature of the atoms. The energy positions and shapes of the spectral lines indicate that the main elements in the thin surface layer of the films grown are gallium (the energy of the dominating peak is about 1070 eV) and nitrogen (the energy of the peak is about 380 eV). Thus, one can conclude that the substrate surface is covered by a thin GaN film. One of the factors that favors this process is that Ga–N binding energy (7 eV) is higher than Ga–As binding energy (4 eV) [7].

To investigate the chemical composition across the thickness of the GaN/GaAs heterostructures, we used layer-by-layer ion sputtering of the surface film and found the concentration profiles of the main chemical elements entering into the composition of the epitaxial GaN film and single-crystal GaAs substrate. Figure 2 shows the intensities of Auger electron peaks in the Ga–N–As system as functions of the sputtering (etching) time or, in other words, the concentration profiles of the elements in the epitaxial film–substrate structure. As is seen from the experimental concentration profiles for this structure, the basic components in the film are Ga and N, and the decrease in the N concentration toward the substrate is fairly smooth. Inside the substrate, only Ga and As are present, as it should be. Consequently, the processing of GaAs by nitrogen radicals favors the formation of a stable thin GaN layer on its surface. This can be explained in the following way. First, Ga atoms migrate from the bulk of the GaAs substrate to its surface and react with active nitrogen radicals of the gas phase, thereby forming the GaN compound. Second, As atoms evaporate intensely from the GaAs near-surface layer and from the growing GaN film into the gas phase. The sharp decay of the As concentration profile near the surface suggests that the transition GaAsN layer is depleted by As atoms and its thickness is insignificant. In addition, it is seen from the profilograms

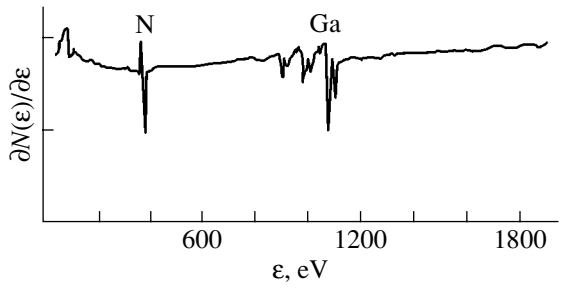

Fig. 1. AES spectra from the surface of the epitaxial GaN film.

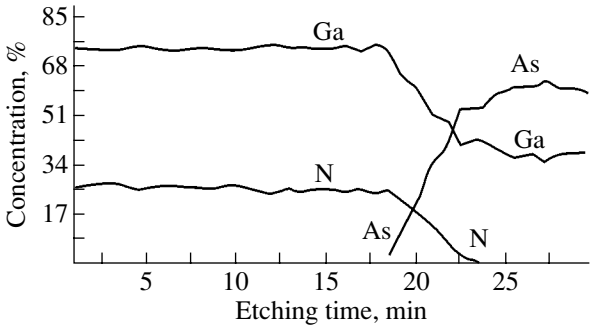

Fig. 2. Element concentration in the GaN/GaAs heterostructure as a function of the etching time.

that Ga atoms migrate to the surface faster than As atoms. Hence, Ga atoms play a limiting role in the formation of GaN films.

One more factor that encourages the growth of epitaxial GaN films is the penetration and diffusion of active nitrogen radicals into the GaAs substrate (which is of importance for thin film growth) followed by their capture by resulting As vacancies. This also contributes to the formation of the GaN compound. In general, nitrogen atoms can occupy not only sites in the As sublattice but also interstices because of their small radius. In the former case, the lattice parameter decreases; in the latter, it remains virtually unchanged.

It is necessary to note that the GaN film exhibits a considerable surface nonstoichiometry: the excess nitrogen amounts to ≈9%. This is caused primarily by a large amount of nitrogen atoms present in the chamber, which prevent the evaporation of nitrogen from the film. This indicates that the first of the above-mentioned factors (quasi-epitaxial growth of thin GaN films) prevails.

The nitrogen concentration and structural perfection of the epitaxial layers were investigated by high-resolution X-ray diffraction. The lattice parameters (interplanar spacings) of the GaN films grown and GaAs substrates were determined from the angular position and intensity of diffracted X rays, which were found from X-ray diffraction patterns (rocking curves). Figure 3 shows X-ray diffraction patterns for three GaN samples grown on the (001)GaAs surface at different temperatures. In addition to the peaks located near $2\Theta \approx 32°$, which correspond to symmetric reflections from the (002) plane of cubic single-crystal GaAs, one can also see the broadened asymmetric peaks near $2\Theta \approx 35°$ and ≈40°. These peaks correspond to the hexagonal and cubic GaN structures; i.e., they are the reflections from the (0002) and (002) planes of the hexagonal and cubic GaN modifications, respectively. For the cubic structure, the lattice constant found from the positions of the peaks was close to 4.515 ± 0.008 Å, which is in good agreement with many published results (see, for example, [8]).

The half-width of the peaks (not exceeding 1°) indicates that the crystal quality of our GaN layers is satisfactory, while not as good as that of nearly perfect MBE-grown films (the half-width does not exceed 25″ [9]). In Fig. 3, only the GaN and GaAs peaks are observed instead of those for the GaAsN solid solution. The identification of the interplanar spacings corresponding to the diffraction peaks showed that the thin epitaxial GaN films may have both the cubic and the hexagonal structures, depending on processing (growth) conditions. With an increase in the temperature, the diffraction peaks shift toward smaller angles. Generally, among all the process parameters (substrate temperature, vapor pressure, discharge current, growth rate, etc.), the substrate temperature is the most critical. The low-temperature process (the annealing temperature below 700°C) is responsible for the formation of cubic thin GaN films on the (001) surface of cubic GaAs [10]. In our opinion, this is because the As vapor pressure at low temperatures is lower than at high temperatures. As the temperature in the reaction chamber grows, the reflection corresponding to hexagonal GaN layers arises and begins to dominate. This process involves mixing the cubic and hexagonal phases in equal amounts. In this situation, it is easy to control the structure of the growing films.

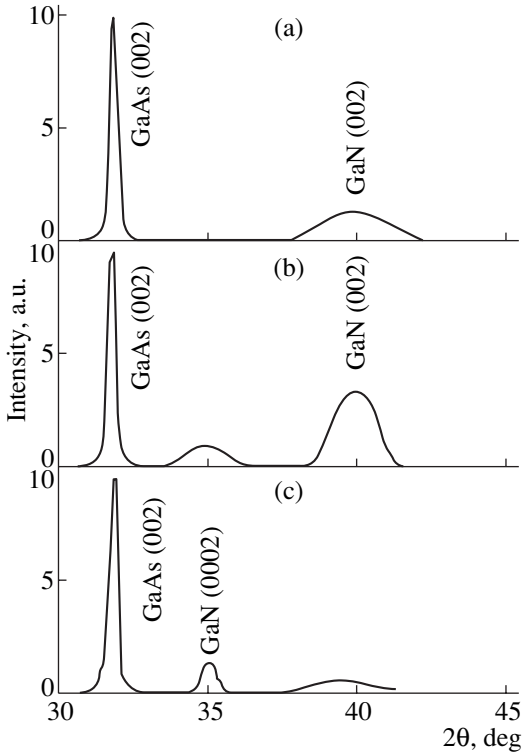

Fig. 3. X-ray diffraction patterns for GaN/GaAs heterostructures epitaxially grown at different temperatures: (a) 600, (b) 700, and (c) 750°C.

Thus, the annealing of GaAs single crystals in the atmosphere of active nitrogen radicals leads to the growth of thin single-crystal GaN layers. The minimum film thickness roughly estimated from the etching rate of the GaN compound was no less than 200–300 Å. The actual thickness of the epitaxial films did not exceed 200–300 nm, which results in the partial relaxation of mechanical stresses. The concentration profiles of the basic chemical elements entering into the epitaxial GaN films and GaAs single-crystal substrate were obtained. It was found that, depending on the temperature of GaAs substrate annealing in the atmosphere of active nitrogen radicals, thin epitaxial GaN layers may have both the hexagonal and the cubic structure.

ACKNOWLEDGMENTS

This work was financially supported by the DFFD of Ukraine (project no. 04.07/256).

REFERENCES

1. S. Strite and H. Morkoc, J. Vac. Sci. Technol. B **10**, 1237 (1992).
2. B. Monemar and G. Pozina, Prog. Quantum Electron. **24**, 239 (2000).
3. L. A. DeLouise, J. Vac. Sci. Technol. A **11**, 609 (1993).
4. A. N. Georgobiani, M. B. Kotljarevsky, U. A. Aminov, *et al.*, Nucl. Instrum. Methods Phys. Res. A **388**, 431 (1997).
5. G. A. Sukach, V. V. Kidalov, A. I. Vlasenko, *et al.*, Optoélektron. Poluprovodn. Tekh., No. 37, 65 (2002).
6. G. V. Abdullaev and T. D. Dzhafarov, *Atomic Diffusion in Semiconductor Structures* (Atomizdat, Moscow, 1980).
7. J. W. Orton, D. E. Lacklison, N. Baba-ali, *et al.*, J. Electron. Mater. **24**, 263 (1994).
8. H. Morkoc, S. Strite, G. B. Gao, *et al.*, J. Appl. Phys. **76**, 1363 (1994).
9. Z. Q. He, X. M. Ding, X. Y. Hou, *et al.*, Appl. Phys. Lett. **64**, 315 (1994).
10. M. Mizuta, S. Fujieda, Y. Matsumoto, *et al.*, Jpn. J. Appl. Phys. **25**, L945 (1986).

Translated by Yu. Vishnyakov

Technical Physics, Vol. 48, No. 4, 2003, pp. 441–448. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 4, 2003, pp. 63–70.

SOLID-STATE ELECTRONICS

Phase Interactions in $TiN_x(TiB_x)$–n-Si–n^+-Si Contacts and Their Thermal Degradation Due to Rapid Thermal Annealing

N. S. Boltovets*, V. N. Ivanov*, R. V. Konakova, V. V. Milenin**, and D. I. Voĭtsikhovskiĭ****

** Orion Research Institute State Enterprise, ul. É. Pot'e 8a, Kiev, 03057 Ukraine*

*** Institute of Semiconductor Physics, National Academy of Sciences of Ukraine, pr. Nauki 45, Kiev, 03028 Ukraine*

e-mail: konakova@isp.kiev.ua

Received September 9, 2002

Abstract—The temperature stability of $TiN_x(TiB_x)$–n-Si–n^+-Si, Au–$TiN_x(TiB_x)$–n-Si–n^+-Si, and Au–Ti(Mo)–$TiN_x(TiB_x)$–n-Si–n^+-Si Schottky-barrier contacts subjected to rapid thermal annealing in hydrogen at temperatures T = 400, 600, and 800°C is studied. It is shown that structural and morphological transformations and the related degradation of electrophysical characteristics in interstitial alloys (titanium nitrides and borides) start at 600°C. Reasons for the degradation of the barrier properties of titanium borides and nitrides are discussed. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

In recent years, increasingly stringent requirements for the stability of metal–semiconductor contacts in solid-state electronic devices have stimulated the search for novel metallization materials that provide the stability of contact parameters under severe environmental conditions. Among the promising contact materials are nitrides, carbides, and borides of transition metals (from Groups III–VI). These compounds offer metallic properties and also are chemically inert and temperature resistant. Titanium nitride ranks first in this list, as indicated by the number of publications.

Titanium nitride films are already being used as ohmic and barrier contacts in electronics [1]. It seems, however, that their usage as barriers preventing diffusion in multilayer thin-film metallization systems of semiconductor devices will be even more promising. Data for contacts using boride-based interstitial alloys are scarce, although available information indicates that interdiffusion, including reactive diffusion, at interfaces is suppressed when they are employed in metallization systems [2, 3].

The application of interstitial alloys is limited because detailed knowledge of the mechanisms behind the degradation and failure of contacts using these alloys is lacking. Of great importance in this respect is to understand the nature and properties of the transition layer (especially its diffusion-preventing capability).

In this work, we study the temperature stability of Ti-based interstitial alloy layers applied on Si alone and in combination with other metal layers.

EXPERIMENTAL

The samples used were standard silicon n–n^+ structures prepared by vapor-phase epitaxy. The thickness of the n layer and the donor concentration in it were, respectively, 1–2 μm and $(8–9) \times 10^{16}$ cm^{-3}. For the n^+ substrate, these values were, respectively, 300 μm and $(2–3) \times 10^{19}$ cm^{-3}. TiB_x layers were applied on the chemically cleaned surface of the epitaxial layer by the magnetron sputtering of titanium boride. TiN_x layers were applied both by the magnetron sputtering of compact stoichiometric TiN targets and by thermoreactive ion synthesis. The contact layers of titanium, molybdenum, and gold were applied by electron-beam evaporation at a pressure of $\approx 6.6 \times 10^{-4}$ Pa. The thicknesses of these layers did not exceed 100 nm. Rapid thermal annealing (RTA) of the contacts was carried out at 400, 600, and 800°C in the hydrogen atmosphere for 60 s.

The mechanisms of contact formation and the temperature stability of the contacts were studied by SEM, AES, XPS, and microprobe analysis, as well as by taking static I–V characteristics. The structure and morphology of the contacts, as well as the component distributions in them, were investigated using 10×10-mm metallized test systems formed on n-Si–n^+-Si substrates. The I–V characteristics were taken from Schottky-barrier diodes with a diameter of 100 μm that were made by photolithography in the form of mesas.

RESULTS AND DISCUSSION

(i) TiN_x-based contacts. Consider first a simple TiN_x/Si system. Factors influencing the barrier properties of such contacts were revealed by studying the

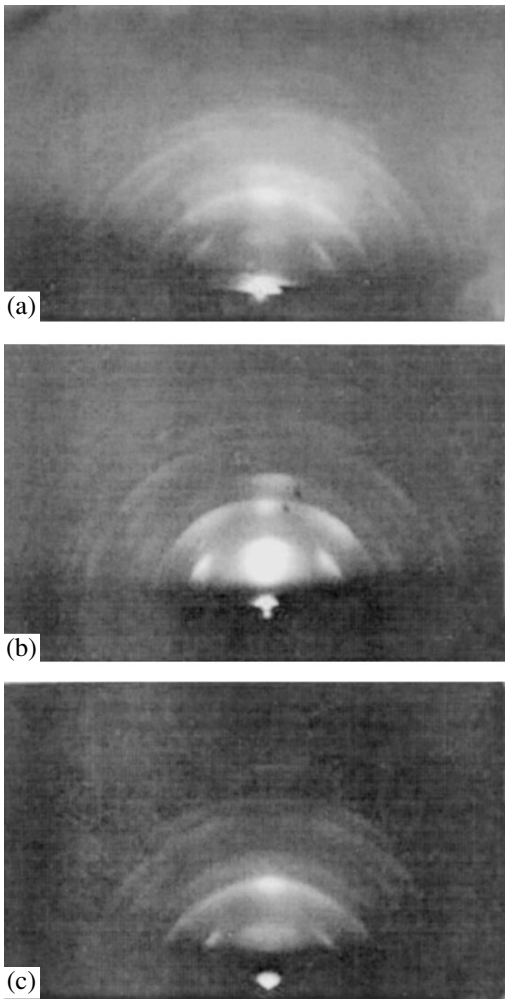

Fig. 1. Electron diffraction patterns from TiN_x films applied on the Si substrate by thermoreactive ion synthesis. (a) As-prepared and after RTA at T = 600 (b) and 800°C (c).

structure, chemical composition, and morphology of the TiN_x films before and after temperature actions.

Figure 1 shows electron diffraction patterns from the TiN_x films obtained by thermoreactive ion synthesis. The diffuse diffraction lines indicate their quasi-amorphous state. This result was supported by X-ray diffraction studies. From X-ray diffraction data for the TiN_x films, we calculated the fractions of the amorphous and polycrystalline phases in the films, grain size L, lattice spacing a, and macrostresses σ (see Table 1).

As follows from Table 1, as the RTA temperature increases, the quasi-amorphous TiN_x layer recrystallizes into polycrystalline with a grain size of ≈40 nm. Simultaneously, residual compression stresses in the TiN_x film are reduced, the reduction being the stronger, the thinner the film.

Along with this change in the mechanical stresses in the heterosystem, the recrystallization of the TiN_x films should also have an effect on the parameters of the transition layer, i.e., on the barrier properties of the TiN_x film. This supposition was supported by AES depth profiles in the film (Fig. 2). These results show that the extent of the interface (transition layer) noticeably changes at 600°C (i.e., under RTA conditions), when intense transformations of the structure and morphology of the TiN_x film are observed. The intriguing fact is that the expansion of the transition layer does not follow the simple exchange mechanism of interface formation. Since titanium nitrides are chemically inert [6] and the oxygen present in the film seems to be in the bound state, producing titanium oxides and titanium oxynitride, the properties of the TiN_x/Si interface should be governed by competitive Ti–Si and N–Si reactions and depend on the number of free or dissociated Si, Ti, and N atoms and on the permeability of the TiN_x film. Note that the latter reaction is a high-temperature process because of the high energy of N diffusion activation in Si (≈3.7 eV) [7].

Thus, the interface of the contact is nonuniform. As the RTA temperature grows, the structural–phase nonuniformity of the transition layer may increase because of both the recrystallization and diffusion mobility, which causes mixing in the TiN_x/Si system and chemical reactions between the components of the contact pair. However, in view of the low permeability of the TiN_x film and an insufficient amount of free Si atoms, one may expect that the interface in the heteropair will remain sharp up to high annealing temperatures, just as follows from Fig. 2.

It should be noted that the above scenario of chemical and structural transformations of the interface ignores the formation of a thin oxide layer on the surface of the semiconductor. The presence of SiO_2 may appreciably suppress chemical processes in the transition layer and, thereby, improve the temperature stability of the contact.

Actually, contacts comprise several films of different metals in order to satisfy many requirements for a contact structure. When combining with one another and also with TiN_x through mutual boundaries, these metals may significantly affect the TiN_x barrier properties and, accordingly, the electrophysical parameters of the contacts.

The basic multilayer structures used in this work were Au/TiN_x/Si, Au/Ti/TiN_x/Si, and Au/Mo/TiN_x/Si. It turned out that the degradation mechanism for all three structures is the same. The only difference is that the Mo layer serves as an additional diffusion barrier, because the mutual solubility of Mo and Au is low [8].

Let us consider the thermal degradation of multilayer contact structures in detail with Au/Ti/TiN_x/Si.

Figure 3 shows the morphology of the layered contact subjected to RTA. Table 2 demonstrates the variation of the atomic structure at different metallization sites (sites 1–4) according to the microprobe analysis data.

Table 1. Lattice spacing, microstresses, and size of coherently scattering blocks in 1-μm-thick TiN layers

Parameter	TiN layer		
	as-prepared	600°C	800°C
Percentage of polycrystalline phase, %	30	80	95
Si lattice spacing a, Å	5.4276(0)	5.4293(0)	5.4292(2)
TiN lattice spacing a, Å	4.2524(7)	4.2420(4)	4.2362(2)
Macrostresses in TiN layers σ, GPa	3.4	1.3	0.1
TiN block size L, nm	15.0	20.0	40.0

Note: Macrostresses in TiN films were calculated by the formula [4] $\sigma \approx -\frac{E}{\mu}\frac{(d_1 - d_0)}{d_0}$, where $E \approx 256$ GPa is the Young's modulus of TiN [5], $\mu \approx 0.3$ is the Poisson's ratio, d_1 is the interplanar spacing for a set of planes that make the greatest contribution to reflection under the normal incidence of an X-ray beam on the sample, and d_0 is the interplanar spacing for the same set of planes without stresses. The value of L was estimated by the formula [4] $L = 0.94\lambda/\rho\cos\Theta$, where λ is the X-ray wavelength, Θ is the Bragg angle, $\rho = \sqrt{B^2 - b^2}$, B is the total half-width of the line from the sample, and b is the instrumental half-width of the reference.

As follows from the results obtained, structural and chemical transformations in the contact metallization do not occur at RTA temperatures below 400°C. The structural–phase reconstruction of the system is observed at temperatures between 400 and 600°C. It is accompanied by changes in the metal atom distribution and also by the appearance of pores and cracks in the TiN_x layer. The layered structure of the contact turns into a crab structure, which decorates defects in the TiN_x underlayer. Pores and cracks are filled with the products of the Ti–Au reaction. It was shown [9, 10] that this reaction, which starts even at 350°C, may pro-

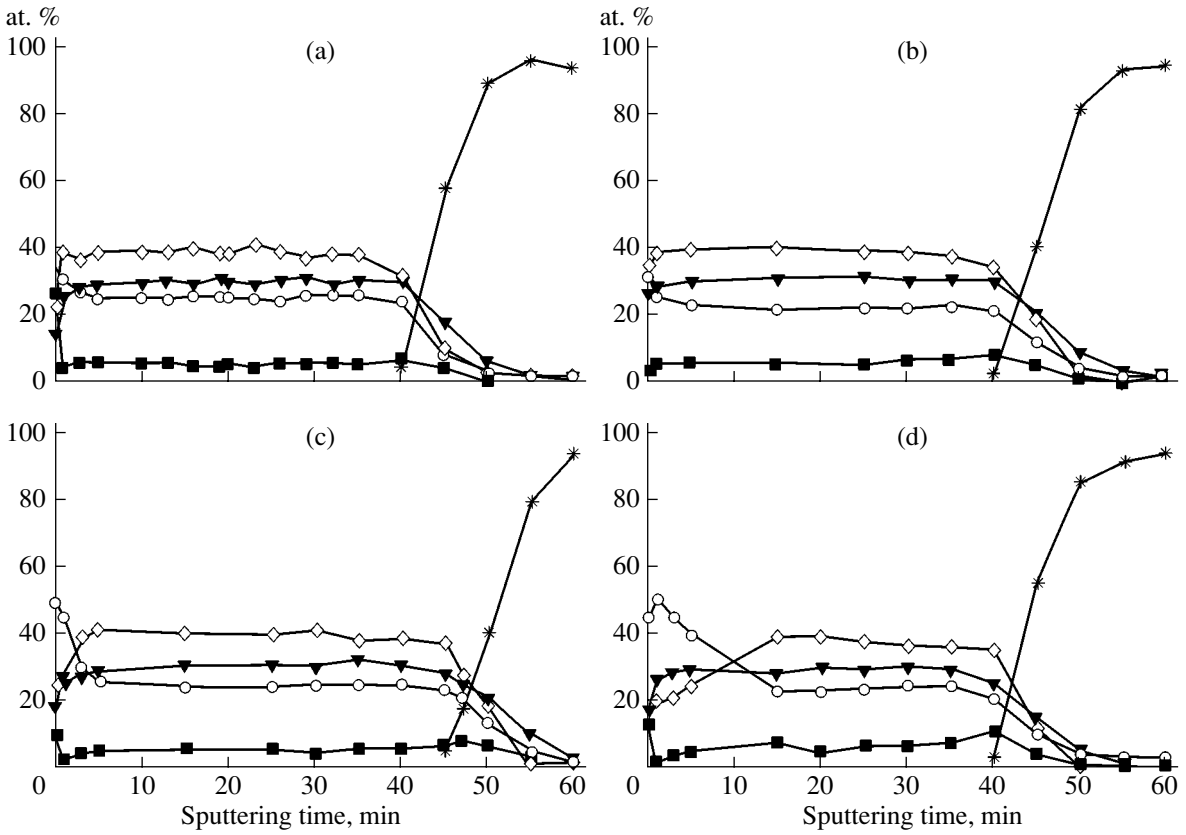

Fig. 2. Depth profiles of the components in the TiN_x/Si contact. (a) As-prepared contact and (b–d) after RTA at T = 400, 600, and 800°C, respectively, for 60 s in hydrogen. (–*–) Si, (–▼–) Ti, (–◇–) N, (–■–) C, and (–○–) O.

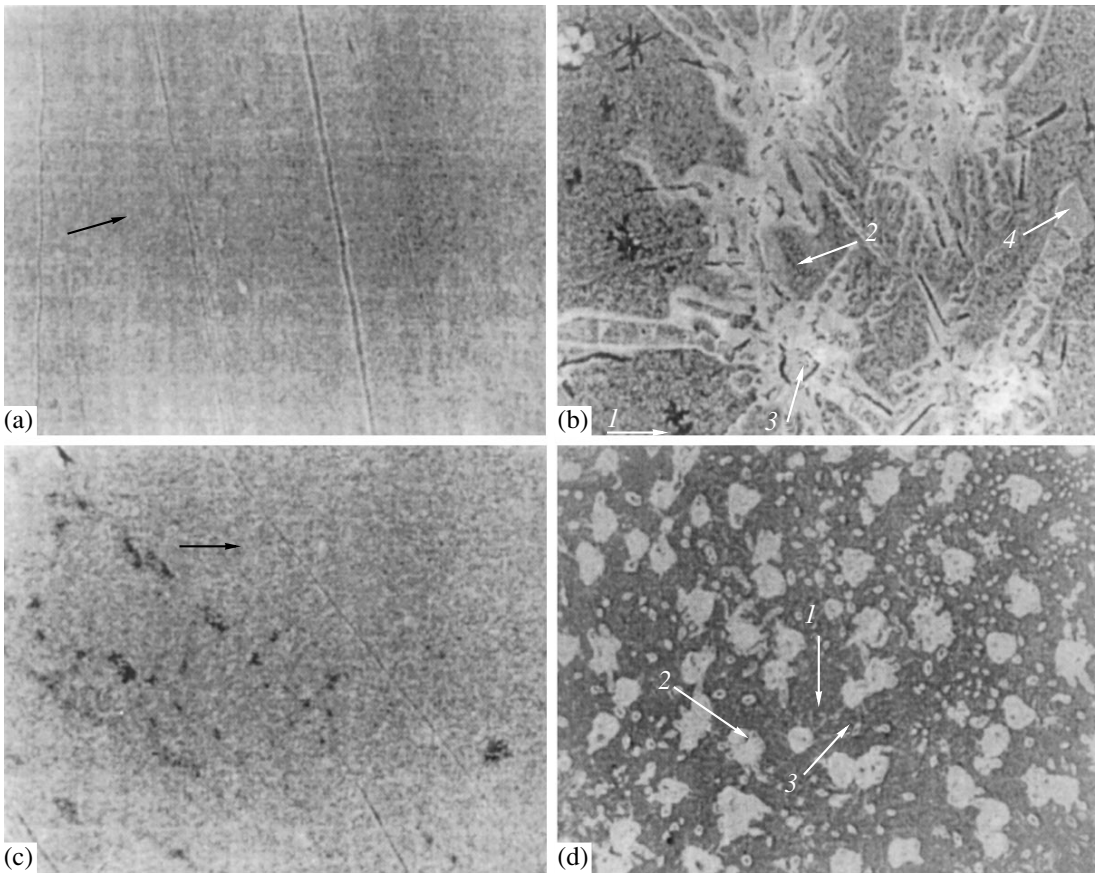

Fig. 3. Surface morphology of the Au/Ti/TiN_x/Si contact (a) before and after RTA at (b) 400, (c) 600, and (d) 800°C for 60 s in hydrogen.

duce intermetallic compounds, which are then converted to the stable phase $AuTi_3$.

At the early stage, the failure of the antidiffusion layer appears to be associated with micropores present in the TiN_x film, which give rise to microcracking. The presence of micropores in films based on interstitial alloys was noted in [11]. The high brittleness (small ductility margin) of metallic nitrides and their extremely low capability to plastically deform upon annealing are also factors favoring cracking. As a result, elastic stresses have no time to occupy the entire volume of the material and are localized at sites where the concentration of structure defects is increased. It is at these sites where the layer starts to fail with the subsequent formation and interaction of microcracks.

Thus, the number of pores and microcracks originating in the initial TiN_x film is responsible for the degradation of the contact structure under study. The number of these defects depends on the TiN_x fabrication conditions, TiN_x film thickness, and mechanical stresses [12]. Microcracks serve as an additional path for transfer through the interface and, hence, alter the electrical properties of the contact. The further change in the electrophysical properties of the structure is due to the complete failure of the TiN_x barrier film. Diffusing toward the substrate along cracks in the antidiffusion layer, Au creates sharp protrusions of eutectic Si-based alloys, which change the electrophysical properties of the contact.

Such a scenario of thermal degradation of the contacts was completely supported by depth profiling data for the Au/Ti/TiN_x/Si system (Fig. 4). Although Ti reacts with Au at an RTA temperature of 400°C, the barrier properties of the TiN_x layer are retained.

As follows from the data in Figs. 3 and 4, RTA at T = 600 (800°C) breaks the layered structure of the contact and considerably smears the interface. The latter effect is the most pronounced at an RTA temperature of 800°C. In this case, AuTi alloy, TiN_x-based alloy, Au, and Si/SiAuTi eutectic/Si composition become the dominant components of the contact. As a result, the "electrical" boundary of the contact shifts deeper into the semiconductor and its structural–chemical homogeneity breaks.

(ii) TiB_x-based contacts. The temperature stability of TiB_x barrier layers will first be considered using the most-studied Si/TiB_2/Mo/Au system.

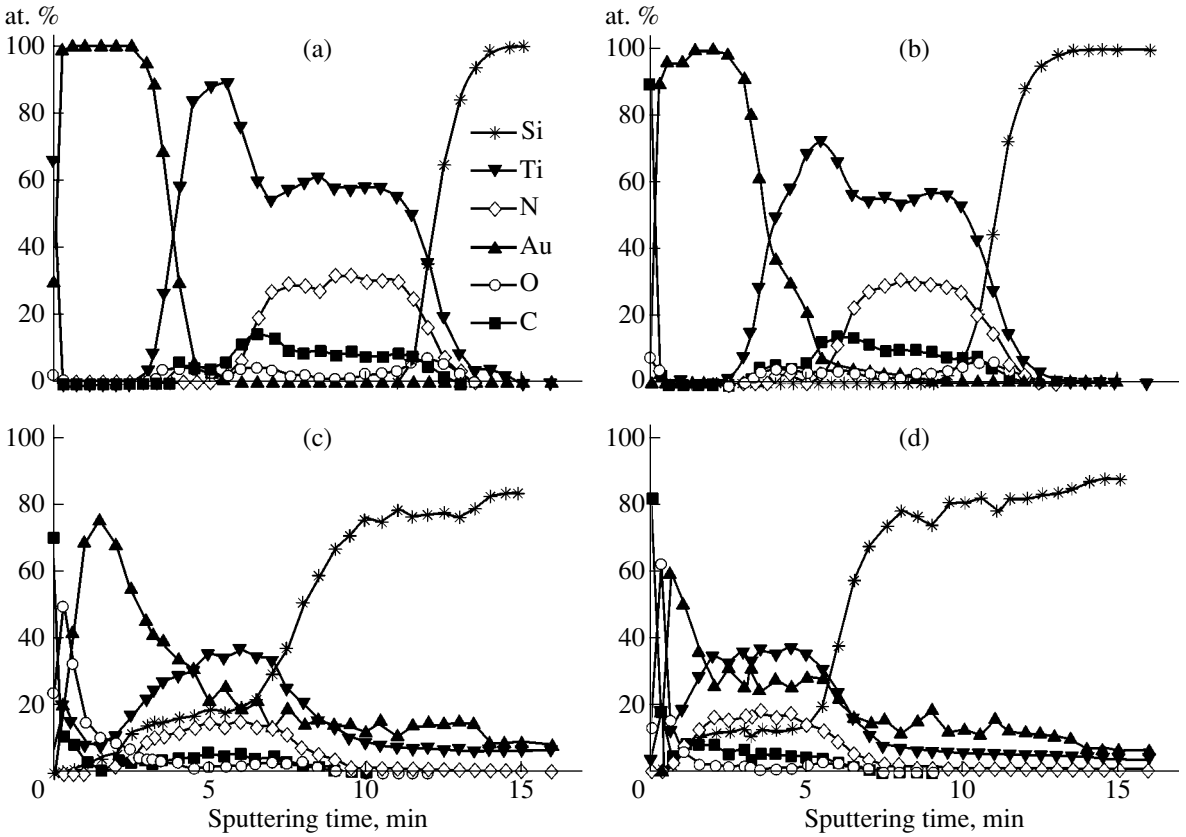

Fig. 4. The same as in Fig. 2 for the Au/Ti/TiN_x/Si contact.

Figure 5 shows the depth profiles for the components of the Si/TiB_2/Mo/Au contact structure before and after RTA, as well as the binding energies for core electrons of the contact components. These data allow one to judge the annealing-induced modifications of the structure and phase composition, as well as interfacial interactions, at different annealing temperatures. At $T \leq$ 600°C, the layered structure of the contact remains unchanged and the contact components are redistributed insignificantly, although the phase composition somewhat changes.

Since the position of the core level peak in basic atoms depends on their interaction with atoms of other elements in the matrix, it can be assumed that the antidiffusion barrier includes elements for which the binding energy of Ti 2*p* electrons is 454.8 eV and that of B 1*s* electrons, 192.8 and 181.1 eV. Therefore, the phase composition of the film can be identified as a mixture of TiB_x compounds with a small amount of oxyborides and/or boron suboxide [13]. This conclusion agrees with the results obtained in [14]. One can also argue that the antidiffusion layer–Si interface is free of TiSi silicides with consideration for the binding energies of

Table 2. Effect of RTA on the planar redistribution of atoms in the Au/Ti/TiN_x/Si contacts

Annealing temperature, °C	Element percentage, %											
	Au				Ti				Si			
	site number											
	1	2	3	4	1	2	3	4	1	2	3	4
Without annealing	70.6				5.0				24.4			
400	71.0				5.1				23.9			
600	9.0	16.8	80.5	11.0	5.3	4.3	8.3	37.2	85.7	78.9	11.2	43.8
800	10.8	81.0	26.9	–	4.6	4.7	33.1	–	84.6	14.3	40.7	–

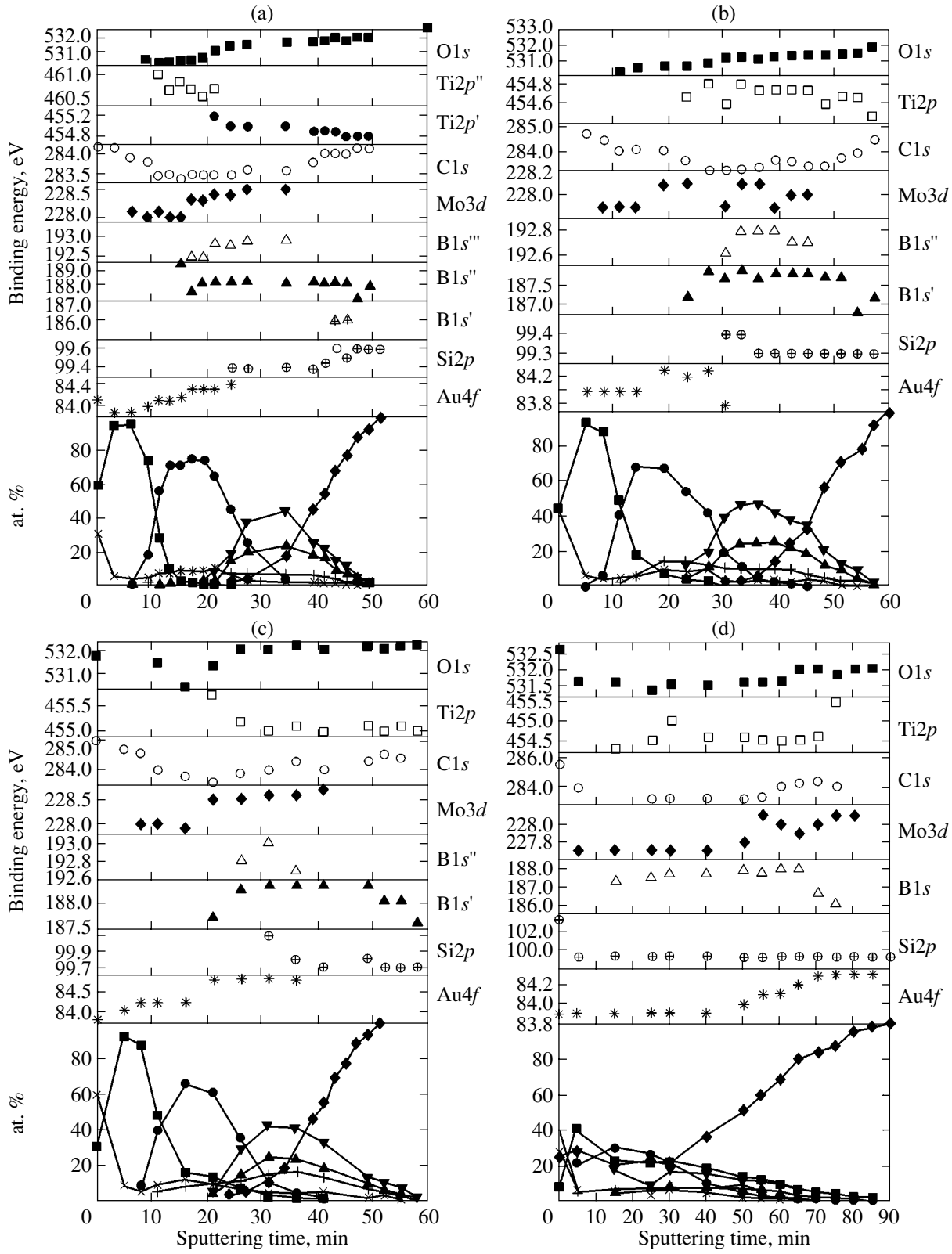

Fig. 5. Binding energy of the core electrons and the AES profiles of the components in the Si/TiB_2/Mo/Au contact. (a) As-prepared and after RTA at (b) 400, (c) 600, and (d) 800°C in hydrogen for 60 s. (■) Au, (●) Mo, (▲) Ti, (▼) B, (◆) Si, (+) O, and (∗) C.

Table 3. Parameters of TiN- and TiB-based Schottky barriers on the *n*-Si–n^+-Si structure before and after RTA in a hydrogen atmosphere for 60 s

Contact structure		Annealing conditions					
		before annealing		400°C		600°C	
		φ_b, V	*n*	φ_b, V	*n*	φ_b, V	*n*
TiN_x/*n*-Si/n^+-Si	*T*	0.59	1.36	0.58	1.11	0.55	1.26
Au/TiN_x/*n*-Si/n^+-Si	*M*	0.59	1.30	0.59	1.20	0.55	1.38
Au/Ti/TiN_x/*n*-Si/n^+-Si	*T*	0.59	1.3	0.59	1.18	0.50	1.65
TiB_x/*n*-Si/n^+-Si	*M*	0.55	1.11	0.57	1.13	0.59	1.24
Au/TiB_x/*n*-Si/n^+-Si	*M*	0.55	1.2	0.56	1.2	0.56	1.28
Au/Mo/TiB_x/*n*-Si/n^+-Si	*M*	0.55	1.3	0.60	1.32	0.56	1.44

Note: *M*, layers obtained by magnetron sputtering; *T*, layers obtained by thermoreactive ion synthesis.

2*p* electrons of Ti (454.8 eV) and 2*p* electrons of Si (99.6 eV). The binding energies of 3*d* electrons of Mo and 4*f* electrons of Au imply that the Mo–Au interface is chemically inactive [13].

Annealing at moderate temperatures removes boron anhydride. This possibly causes micropores to form and stimulates minor reactions in the Ti–Si system. Thus, prerequisites for the local chemical restructuring of the contact are provided.

Annealing at 800°C completely breaks the layered structure of the contact. In this case, irregularities on the contact surface cannot be described by the normal distribution [15]. This points to the essential role of activation processes at the interface. Under these conditions, the microrelief of the interface is defined by chemical reactions between the contact components and semiconductor. The presence of different phases and the associated roughness of the interface result in the degradation of the electrophysical properties.

Thus, the thermal threshold of degradation of the contact depends on the temperature stability of the TiB_x layer. This conclusion was supported by studies of the TiB_x/Si, Au/TiB_x/Si, and Au/Mo/TiB_x/Ti/Si systems.

In fact, TiB_x layers have structural elements that deform and strengthen the basic metallic lattice [11]. These deformations cannot be released by diffusion. They can be removed by high-temperature processing, which loosens metal–metalloid bonds. With this in mind, we can put forward several reasons for the local breakdown of the TiB_x film.

The first one is intrinsic mechanical stresses in TiB_x films. It is known that TiB_x layers obtained by magnetron sputtering possess high compression stresses [16]. The layers under study are quasi-amorphous; hence, one might expect that both the elastic properties of these materials and the mechanisms of their structural relaxation differ considerably from those typical of crystalline layers. To date, a correlation between the microstructure and stressed state of the film has been confirmed only on a qualitative basis [11]. One can, however, argue that in quasi-amorphous layers, the temperature threshold of microcracking due to structure relaxation upon annealing will be different from that in crystalline layers. In addition, the temperature threshold in our case is affected by the contact layers with other thermal expansion coefficients.

The second reason is that the reactions proceeding at local sites of the barrier layers favor the formation of pores in TiB_x. Porosity leads to an increase in the thermal expansion coefficient with temperature. In this case, the strain distribution in the structure becomes even more nonuniform, which may cause microcracking, especially in view of the low ductility margin of TiB_x.

Third, under certain conditions (an oxygen atmosphere and a sufficiently high temperature), the TiB_x layer may decompose according to the reaction

$$4TiB_x + (2m + 3x)O_2 \longrightarrow 4TiO_m + 2xB_2O_3.$$

In this case, the oxidation rate and oxide structure are controlled by two factors: boron anhydride evaporation and borate formation. The former factor is dominant at the relatively low temperatures used in this work. The other factor becomes essential at higher temperatures; hence, it can be disregarded.

The results reported above were confirmed by investigating the electrophysical characteristics of $TiN_x(TiB_x)$/*n*-Si/n^+-Si surface-barrier structures subjected to RTA. RTA-induced changes in the parameters of the Schottky barriers (the barrier height φ_b and the ideality factor *n*) are listed in Table 3. It is seen that the values of φ_b and *n* for the diode structures before and after RTA at 400°C differ insignificantly. Conversely, at $T = 600$°C, the barrier properties of the contact degrade.

Thus, by varying the conditions for titanium nitride and titanium boride formation, as well as the metallization composition, one can prepare contact structures that are stable against RTA at temperatures no higher than 600°C.

ACKNOWLEDGMENTS

This work was partially supported by the program INCO-COPERNICUS (project no. 977131).

REFERENCES

1. A. A. Seĭdman, Élektron. Tekh., Ser. 2: Poluprovodn. Prib., No. 6, 13366 (1988).
2. J. E. Sundgren, Thin Solid Films **22** (1), 21 (1985).
3. G. Lemperiere and J. M. Poitevin, Thin Solid Films **111**, 339 (1984).
4. A. A. Rusakov, *X-Ray Radiography of Metals* (Atomizdat, Moscow, 1977).
5. G. V. Samsonov and K. I. Portnoĭ, *Refractory Compound Alloys* (Oborongiz, Moscow, 1961).
6. R. B. Kotel'nikov, S. N. Bashlykov, Z. G. Galiakbarov, *et al.*, *Super-Refractory Elements and Compounds* (Metallurgiya, Moscow, 1968).
7. S. Murarka, *Silicides for VLSI Applications* (Academic, New York, 1983; Mir, Moscow, 1986).
8. M. Harris, E. Lugujjo, and S. U. Gampisano, J. Vac. Sci. Technol. **12**, 524 (1975).
9. O. M. Barabash and Yu. N. Koval, *Structure and Properties of Metals and Alloys* (Naukova Dumka, Kiev, 1986).
10. *Thin Films: Interdiffusion and Reactions*, Ed. by J. M. Poate, K. Tu, and J. Meier (Wiley, New York, 1978; Mir, Moscow, 1982).
11. R. N. Andrievskiĭ, Usp. Khim. **66** (1), 57 (1997).
12. J. Suni, M. Maenpaa, M. A. Nicolet, *et al.*, J. Electrochem. Soc. **130**, 1215 (1983).
13. V. I. Nefedov, *Handbook of X-ray Electron Spectroscopy of Chemical Compounds* (Khimiya, Moscow, 1984).
14. A. S. Seal, T. Barr, N. Sobczak, *et al.*, J. Vac. Sci. Technol. A **A15**, 505 (1997).
15. N. S. Boltovets, N. M. Goncharuk, V. A. Krivutsa, *et al.*, Semicond. Phys. Quantum Electron. Optoelectron. **3**, 352 (2000).
16. J. Chem and J. A. Barnard, Mater. Sci. Eng. **191**, 233 (1995).

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 4, 2003, pp. 449–454. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 4, 2003, pp. 71–76.

OPTICS, QUANTUM ELECTRONICS

Quasi-Diffraction Effects during Irradiation of Moving Surfaces: Part II

D. N. Doinikov*, K. I. Zor'ko*, M. F. Kudoyarov*, A. V. Matyukov*, S. A. Mukhin*, and M. Ya. Patrova**

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

*** NPF TREM Closed Joint-Stock Society, St. Petersburg, 194021 Russia*

e-mail: mkud@cycla.ioffe.rssi.ru

Received May 27, 2002

Abstract—In the second part of this work, the general principles of the quasi-diffraction approach developed in the first part are applied to analyzing the irradiation of moving surfaces by pulsed fluxes of fast charged particles. Special emphasis is given to the fact that the operating aperture is scanned by a narrow focused beam during the formation of these fluxes.[1] © *2003 MAIK "Nauka/Interperiodica".*

1. SPACE–TIME BEATS ARISING UPON PERIODICALLY SCANNING THE OPERATING APERTURE BY A PULSED BEAM

Charged particle fluxes differ from electromagnetic and neutron fluxes in that they can be controlled by electrostatic and magnetic fields. Specifically, electrostatic or magnetic scanning is used to provide a uniform irradiation density over a coordinate [1, 2]. If an accelerator generating an initial electron or ion beam operates in the pulsed mode, space–time beats arise, which result from the presence of two independent frequencies: the scanning frequency and the pulse-repetition frequency. Let us consider the formation of the beats more closely.

Let irradiating pulses be rectangular and have a duration T_p (pattern *1* in Fig. 1). Also, let the pulsed beam scan the irradiating aperture by a linear periodic law with a period T_{sc} such that $T_{sc} \le T_p < T_0$, where T_0 is the pulse-repetition period. This condition is met in the patterns on the left of Fig. 1. Consider the case where the period T_0 is an almost exact multiple of T_{sc}, i.e., there exists a natural number $n > 1$ such that

$$\left|\frac{nT_{sc} - T_0}{T_{sc}}\right| \ll 1. \qquad (1)$$

With the arrival of each pulse, the scan function $x(t)$ slightly shifts relative to the beginning of the pulse. In pattern *2* (Fig. 1), the initial phases of the processes are selected so that the irradiation density $\rho(x)$ in the upper part of the aperture is twice as high as that in the lower part. We assume that the phase shift ΔT in this case is zero. In pattern *3*, $\Delta T \neq 0$. It is easy to check that here the density distribution over the aperture has the form of a three-step function depicted to the right of pattern *3*. With ΔT increasing further, first the irradiation density distribution becomes uniform in the interval $-x_m \le x \le x_m$ (pattern *4* in Fig. 1) and then the irradiation density in the lower part starts dominating. The cycle of $\rho(x)$ variation ends up at $\Delta T = T_p$. Mathematically, the function $\rho(x, \Delta T)$ for the cases shown in the patterns can be represented as

$$\rho(x, \Delta T) = \begin{cases} 4 & \dfrac{4x_m}{T_{sc}}\Delta T < x < x_m \\ 3 & -\dfrac{4x_m}{T_{sc}}\Delta T \le x \le \dfrac{4x_m}{T_{sc}}\Delta T \\ 2 & -x_m < x < -\dfrac{4x_m}{T_{sc}}\Delta T. \end{cases} \qquad (2)$$

From the right-hand side of Fig. 1, it is seen that if the dependence $\rho(x, \Delta T)$ for each ΔT is approximated by a linear function passing through the point ($x = 0$, $\rho = \rho_0$), where ρ_0 is the aperture-averaged incident flux intensity, the approximating function "precesses" about this point with a period, which will be called the beat period T_b. It is obvious that

$$T_b = T_0\frac{T_{sc}}{\Delta T}. \qquad (3)$$

With T_{sc} and T_p constant, the values of T_0 and ΔT are not independent. They are related to each other as

$$\Delta T = |T_0 - nT_{sc}|. \qquad (4)$$

[1] Initially, we designed this article as consisting of two parts, with experimental data being reported in the second part. Later, however, we considered that it would be more appropriate to describe the experimental data processing techniques in greater detail. They will be discussed in the third part.

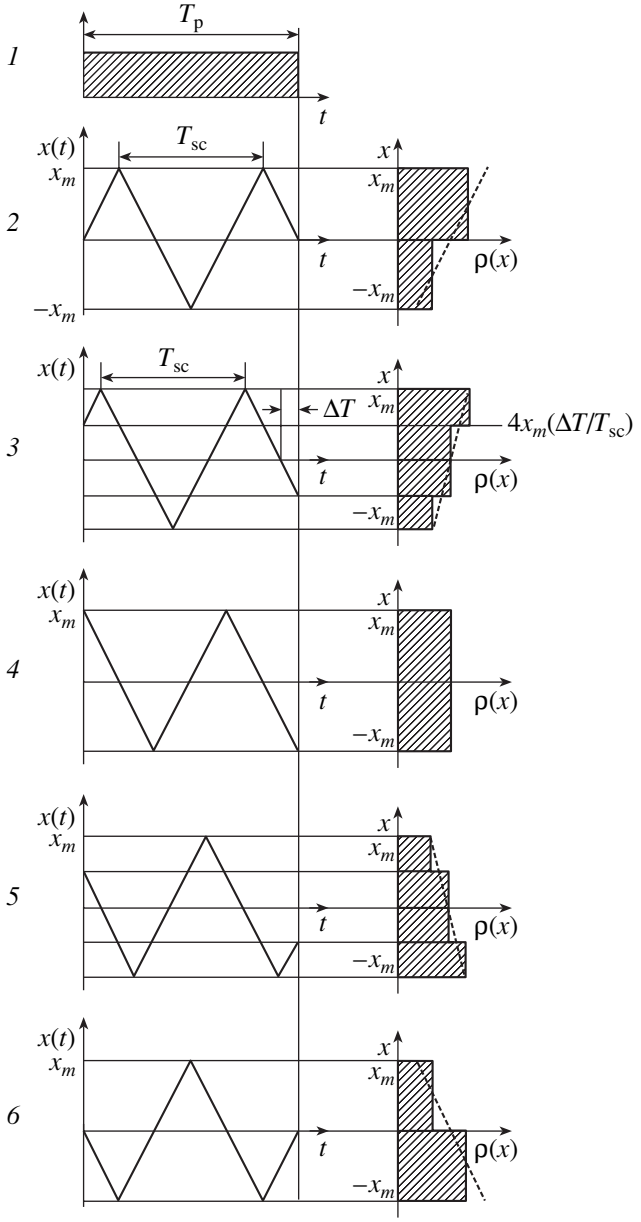

Fig. 1. Time waveforms and the corresponding distributions of the irradiation density $\rho(x)$ over the pulse duration T_p for different phases of the scanning function $x(t)$.

From (3) and (4), we get

$$T_b = \frac{T_0 T_{sc}}{|T_0 - nT_{sc}|} = \frac{1}{f_0 f_{sc}\left|\frac{1}{f_0} - \frac{n}{f_{sc}}\right|} = \frac{1}{|f_{sc} - nf_0|} = \frac{1}{f_b}, \quad (5)$$

where $f_{sc} = 1/T_{sc}$, $f_0 = 1/T_0$, and $f_b = 1/T_b$ are the associated frequencies.

From (5), it follows that $f_b = |f_{sc} - nf_0|$. If f_b and f_0 are constant, the scan frequencies to which a given beat frequency corresponds are given by

$$f_{sc} = nf_0 \pm f_b. \quad (6)$$

If $f_{sc} \gg f_b$, the same beat frequency will be repeated very often as the scanning frequency varies. Let, for example, $f_{sc} \approx 100f_0$. Then, if f_{sc} varies within a 10% interval, say, from f_{sc} to $1.1f_{sc}$, beats with a given low frequency f_b will take place at 20 points within this narrow frequency interval (at integer $n = 100$–109 for both signs). The pulse duration does not enter into the expressions given above; however, one can easily show that T_p has an effect on the beat amplitude and not on the beat frequency.

Thus, if an irradiating ion beam with a pulsed temporal structure scans the operating aperture along one coordinate, the pulse-duration-averaged flux density distribution along this coordinate has a variable component of frequency $f_b = |f_{sc} - nf_0|$.

In the linear approximation, this variable component is given by a space–time expression for a standing wave with a wavelength far exceeding the aperture width and with the node at the aperture center. Adding up the variable and constant components of the irradiation density, one can write an expression for the density of the irradiating flux thus formed as a function of coordinate x and time t:

$$I(t, x) = I_0 + \frac{I_m}{x_m} x \cos(\omega_b t), \quad (7)$$

where I_0 is the constant component of the intensity, $\omega_b = 2\pi f_b$ is the angular beat frequency, and I_m is the variable component amplitude at the aperture edge $x = x_m$.

In (7), the linearly periodic time dependence is changed to sinusoidal. Experience shows that such a replacement is valid: the associated expression describes a real process and at the same time the calculation is greatly simplified.

2. PULSED-BEAM MODULATION OF THE MOVING SURFACE IRRADIATION DENSITY USING LONGITUDINAL SCANNING: THE CALCULATION OF THE MODULATION COEFFICIENT

In the first part of this work [3], it was shown that the irradiation of a moving surface by a sinusoidally modulated flux through a slot screen can be considered as the Fraunhofer diffraction of a wave field in a specific quasi-diffraction space. Here, this analogy will help us to develop an appropriate mathematical apparatus. Namely, first the calculation of the density of flux (7) irradiating a moving surface through a single-slot screen will be reduced to the analysis of the diffraction pattern in a quasi-diffraction space. Then, we will pass from the coordinates of this space to the physical quantities of interest.

To do this, consider a two-dimensional Euclidean space XY with a standard Euclidean metric. The physical meaning of the coordinates of this space will be

refined later; for the moment, we will consider them as coordinates of a usual physical space. Let a slot screen $B'B''$ be placed at a distance R from the plane of observation $A'A''$ (Fig. 2). A wave flux of conventionally electromagnetic nature falls on the screen from above. Let this flux generate a field varying linearly across the slot (i.e., in the x direction). In complex form, the amplitude of this field is given by

$$I_a(x, t) = \left(\frac{I_m\Delta x}{x_m + \Delta x} - \frac{I_m x}{x_m + \Delta x}\right)\exp(-j\omega t) = \left(I_m\eta\frac{I_m x}{x_m}\right)\frac{\exp(-j\omega t)}{1+\eta}, \quad (8)$$

where Δx is the shift of the zero amplitude point relative to the slot center, $\eta = \Delta x/x_m$ is the relative value of this shift, and I_m is the field amplitude at the slot edge that is opposite to the shift Δx.

Figure 2 shows the field distribution at the time instant $t = 0$, when $\mathrm{Re}[I_a(x, t)]$ reaches one of its maximal absolute values. Field (8) can be viewed as a fragment of a standing wave that is distributed along x and has the node at $x = \Delta x$. Its wavelength $\lambda \gg x_m$. According to the Huygens–Fresnel principle, the representation of a field in the slot plane uniquely specifies its distribution behind the screen. If the analysis of the diffraction pattern is restricted to the Fraunhofer approximation, the angles φ and α should be assumed to be small. In this case, the analysis of the diffraction pattern on the plane $A'A''$ is reduced to the analysis of angular diffraction at infinity.

Let us introduce the wavevector $k = \omega/c$ in the direction of field propagation, where c is the field propagation velocity in the quasi-diffraction space. Then, the desired dependence of the diffraction field on the angle α has the form [4, 5]

$$\bar{I}_a(\alpha, \eta) = \int_{-x_m}^{x_m} I_a(x, t)\exp(jk\alpha x)dx$$

$$= \frac{I_m\eta}{(1+\eta)}\exp(-j\omega t)\int_{-x_m}^{x_m}\exp(jk\alpha x)dx$$

$$-\frac{I_m}{(1+\eta)x_m}\int_{-x_m}^{x_m} x\exp(jk\alpha x)dx$$

$$= \frac{I_m\eta}{1+\eta}\exp(-j\omega t)\int_{-x_m}^{x_m}\cos(k\alpha x)dx$$

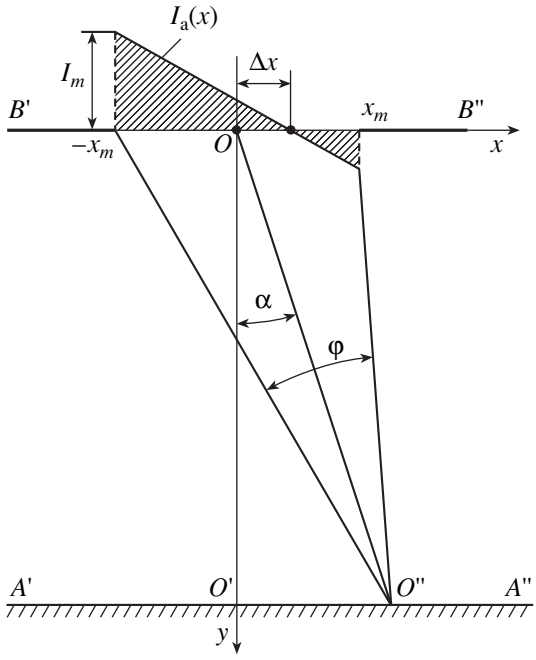

Fig. 2. On the analysis of the quasi-diffraction produced by a single-slot screen with a linear amplitude distribution across the slot (the angles α and φ are assumed to be small).

$$-j\frac{I_m}{(1+\eta)x_m}\exp(-j\omega t)\int_{-x_m}^{x_m} x\sin(k\alpha x)dx$$

$$= \frac{2I_m x_m\eta}{1+\eta}\frac{\sin(k\alpha x_m)}{k\alpha x_m}\exp(-j\omega t)$$

$$-j\frac{2I_m x_m}{(1+\eta)x_m}\left[\frac{\sin(k\alpha x_m)}{k\alpha x_m} - \cos(k\alpha x_m)\right]\exp(-j\omega t).$$

With $k\alpha x_m = \xi$, we have

$$\bar{I}_a(\xi, \eta) = \frac{2I_m x_m\eta}{1+\eta}\sqrt{\frac{\pi}{2\xi}}J_{1/2}(\xi)\exp(-j\omega t) - j\frac{2I_m x_m}{1+\eta}\sqrt{\frac{\pi}{2\xi}}J_{3/2}(\xi)\exp(-j\omega t), \quad (9)$$

where $J_{1/2}(\xi)$ and $J_{3/2}(\xi)$ are the Bessel function of the first kind with a half-integral index.

The magnitude of (9) is

$$\left|\bar{I}_a(\xi, \eta)\right| = \frac{I_m l}{1+\eta}\sqrt{\frac{\pi}{2\xi}[\eta^2 J_{1/2}^2(\xi) + J_{3/2}^2(\xi)]}, \quad (10)$$

where $l = 2x_m$ is the slot width.

According to the general statement put forward in [3], we can revise formula (10) and assign the variables entering into this formula a new physical meaning. Namely, we will consider the irradiation of a surface moving with a velocity v through an aperture of width l. It is assumed in this case that the irradiating flux

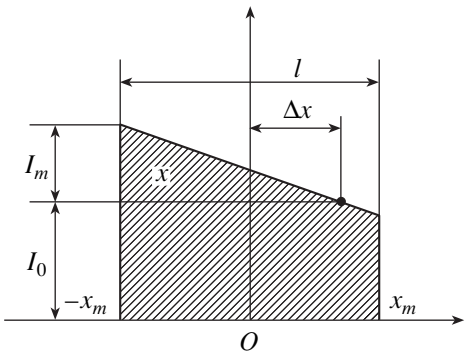

Fig. 3. Distribution of the mean flux intensity across the slot (aperture) of width l at a fixed time instant. The scanning axis is displaced by a distance Δx relative to the aperture center.

obeys the laws of geometrical, rather than wave, optics. Then, the distribution of its intensity across the aperture is given by

$$I(x, t) = I_0 + \mathrm{Re}[I_a(x, t)]$$
$$= I_0 + \left(\frac{I_m\eta}{1+\eta} - \frac{I_m x}{(1+\eta)x_m}\right)\cos(\omega t), \quad (11)$$

where I_0 is the constant component of the intensity and $I_a(x, t)$ is given by (8).

Distribution (11) for $t = 0$ is shown in Fig. 3. From physical considerations, it is clear that $I_0 \geq I_m$. Then, according to [3], we can argue that behind the limiting aperture the coefficient of irradiation density modulation along the propagation direction is given by

$$A_p(\xi, \eta) = C\left|\bar{I}_a(\xi, \eta)\right|$$
$$= C\frac{I_m l}{1+\eta}\sqrt{\frac{\pi}{2\xi}[\eta^2 J_{1/2}^2(\xi) + J_{3/2}^2(\xi)]}, \quad (12)$$

where C is a constant factor.

Consider the limit

$$\lim_{\eta\to\infty} A_\rho(\xi, \eta) = CI_m l\sqrt{\frac{\pi}{2\xi}J_{1/2}^2}. \quad (13)$$

As follows from (11), to this limit corresponds the x-independent irradiation intensity, which harmonically varies with time. This case was discussed in [3], where the following expression for the coefficient of irradiation density modulation was derived:

$$A_{\rho 1} = \frac{I_m}{I_0}\left|\frac{\sin\xi}{\xi}\right|. \quad (14)$$

Here, $\xi = \omega_b l/2v$, v is the surface motion velocity, ω_b is the angular frequency of modulation (the beat frequency in our case), I_m is the maximal intensity modulation amplitude (in our case, it is the amplitude of the variable intensity component at the slot edge opposite to the shift Δx, Fig. 3), and I_0 is the constant component of the irradiation intensity.

Comparing (13) and (14), we conclude that $C = 1/(I_0 l)$. Since I_m and I_0 do not depend on η in both the limiting and general cases, the final form of expression (12) is

$$A_\rho(\xi, \eta) = \frac{I_m}{I_0(1+\eta)}\sqrt{\frac{\pi}{2\xi}[\eta^2 J_{1/2}(\xi) + J_{3/2}^2(\xi)]}. \quad (15)$$

Turning back to the conclusions reached in Section 1 of this work, we can argue that expression (11) generalizes expression (7) for the case $\eta \neq 0$, when the scanning axis does not pass through the aperture center and may even run outside the aperture. If the scanning axis passes through the aperture center, $\eta = 0$ and, according to (15), the coefficient of irradiation density modulation is calculated by the formula

$$A_\rho(\xi, 0) = \frac{I_m}{I_0}\sqrt{\frac{\pi}{2\xi}J_{3/2}^2(\xi)}. \quad (16)$$

For the parameter values such that $A_\rho = 0$, the variable component of the density is absent; hence, the surface is uniformly irradiated.

Figure 4 shows relationships (15) for three values of η. When the scanning axis passes through the aperture center (curve *2*), the density modulation coefficient is zero; in other words, the irradiation density remains uniform if the beat frequency is sufficiently low. In practice, however, to keep this frequency at a low level is a challenge, as was noted above. Even a minor variation of f_{sc} or f_p may drastically increase the beat frequency f_b, causing the irradiation density to be periodically nonuniform. For the same reason, it is difficult to keep the beat frequency constant at any point corresponding to $A_\rho = 0$. The situation can hardly be remedied by increasing the scanning frequency, since the low-frequency beat spectrum for high f_{sc} is also extremely wide. The introduction of controllable aperiodicity into the scanning process seems to be the most efficient way to avoid irradiation nonuniformity. In this case, beats cannot concentrate at a certain point of the quasi-diffraction pattern and the unfavorable effect appears to spread over it.

Of interest is the case $\eta = 1$. Unlike the cases $\eta = 0$ and $\eta = \infty$, here the ξ dependence of A_ρ is fairly smooth and monotonic. Such an irradiation regime is useful if it is necessary to obtain a constant modulation coefficient when the beat frequency or the velocity of the surface irradiated significantly vary.

To conclude this section, we give (without proof) an expression for the density of moving surface irradiation through a space-periodic grating as a function of ξ. This expression was derived on the assumption that the symmetry axis of scanning runs through the grating center

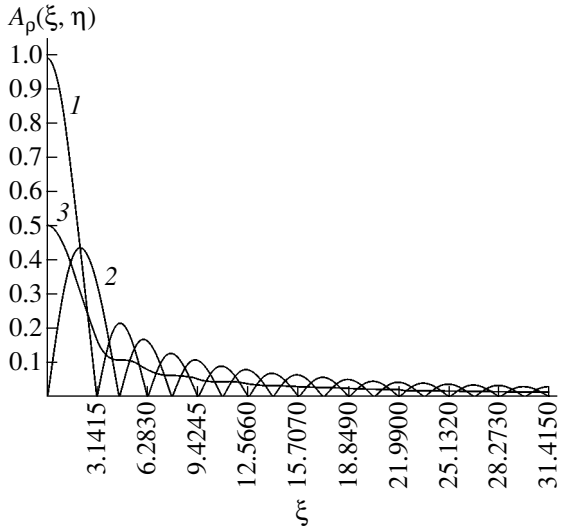

Fig. 4. Irradiation density modulation coefficient A_ρ vs. ξ for the relative shift $\eta = \infty$ (*1*), 0 (*2*), and 1 (*3*).

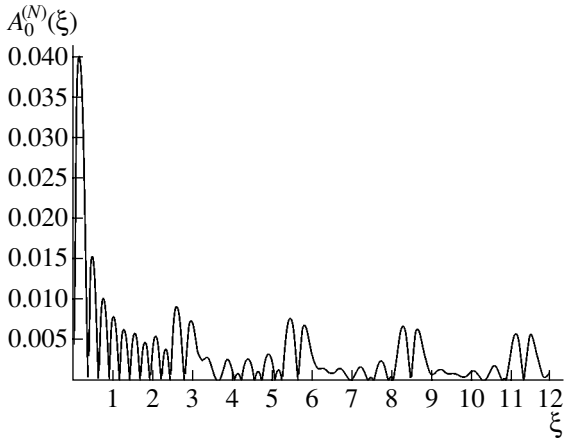

Fig. 5. Dependence $A_\rho^{(N)}(\xi)$ for $N = 11$ and $(d-l)/d = 1$. The scanning axis passes through the center of the middle slot of the grating.

and that the number N of slots is odd. Clearly, the latter assumption becomes insignificant for large N.

$$A_\rho^{(N)}(\xi) = \frac{I_m}{I_0 N l[d(N-1)+l]} \times \sum_{i=1}^{N} (-1)^{i+1} l_i^2 \sqrt{\frac{\pi}{2\xi_i}} J_{3/2}(\xi_i), \quad (17)$$

where

$$l_i = (i-1)d + \Psi(i)(d-l) = \Psi(i+1)l;$$

$$\xi_i = (\xi/l)\lfloor (i-1)d + \Psi(i)d + l(-1)^{i+1} \rfloor;$$

$$\Psi(i) = \frac{1+(-1)^i}{2};$$

$\xi = l\omega_b/2v$, l is the slot width, d is the grating spacing, I_m is the irradiation intensity at the grating edges, and I_0 is the constant component of the irradiation intensity.

From physical considerations, it is evident that $A_\rho^{(N)}$ cannot exceed unity. Hence, $I_0 \geq I_m$, as in the previous cases.

Figure 5 plots function (17) for $N = 11$ and $(d - l)/d = 0.1$. It is seen that pulsed irradiation combined with scanning causes the characteristic split of quasi-diffraction maxima, including the zero one.

3. INTERPRETATION OF QUASI-DIFFRACTION SPACE COORDINATES

In [3], we noted that it is rather difficult to indicate grounds for the unambiguous identification of the coordinates of the *XY* space and assign the metric of this space the physical meaning of basic transformation invariant. However, without pretending to the uniqueness and generality of an approach, this problem can be completely solved. Indeed, let us assume that the coordinate system in the *XY* space is fixed as is shown in Fig. 2. With the boundary conditions imposed by the slot taken into account, diffraction can be completely described by the wave equation

$$\frac{1}{c^2}\frac{\partial^2 I(x, y, t)}{\partial t^2} = \Delta I(x, y, t), \quad (18)$$

where I is the wave function of the field, c is the velocity of field propagation in the *XY* space, and Δ is two-dimensional Laplacian.

With α/c substituted for $1/v$ in Eq. (14), where α is a small diffraction angle, the dependence $A_{\rho 1}(\alpha)$ coincides with the absolute value of the solution to (18) in the plane of observation *A'A"* in the Fraunhofer approximation [6, 7]. From this substitution, it follows that $\alpha = c/v$. Since the Fraunhofer approximation assumes that $\alpha \ll 1$, we have $c \ll v$; that is, the velocity of wave propagation in the *XY* space must be much less than the velocity of the irradiated surface in real space.

Let an arbitrary point on the surface travel a distance R for a real time τ. Then,

$$\alpha = \frac{c}{v} = \frac{c}{(R/\tau)} = \frac{c\tau}{R}. \quad (19)$$

We assume that the *XY* space has the metric of a usual Euclidean space where the *X* and *Y* axes are mutually orthogonal. If a plane wave is incident on a slot screen in the *Y* direction and diffraction shifts are observed in the *X* direction, we have $\alpha \approx \tan\alpha = x/y$ for small angles. Hence, in view of (19), we can assign the coordinates x and y the physical meaning

$$x = c\tau, \quad y = R. \quad (20)$$

In other words, in a quasi-diffraction space, the coordinate x is the normalized real time τ (c is the normalizing factor) and the coordinate y is the length of the surface area irradiated within this time. If the frequency of modulation of the irradiating flux remains constant, the transition to another point of a quasi-diffraction pattern means a change in the angle of observation and, hence, in the velocity of the moving surface.

ACKNOWLEDGMENTS

The authors thank. Prof. G.V. Ostrovskaya for criticism and valuable comments.

REFERENCES

1. M. F. Kudoyarov, A. V. Matyukov, and S. A. Mukhin, in Abstracts of Papers, Membrane-98 (Moscow, 1988), p. 123.
2. G. M. Gusinski, *et al.*, Prib. Tekh. Éksp., No. 1, 25 (1973).
3. K. I. Zor'ko, M. V. Kudoyarov, A. V. Matyukov, *et al.*, Zh. Tekh. Fiz. **72** (10), 87 (2002) [Tech. Phys. **47**, 1291 (2002)].
4. Ya. P. Terletskiĭ and Yu. P. Rybakov, *Electrodynamics* (Vysshaya Shkola, Moscow, 1990).
5. F. T. S. Yu, *Introduction to Diffraction, Information Processing and Holography* (MIT, Cambridge, Mass., 1973; Sov. Radio, Moscow, 1979).
6. I. M. Nagibina, *Interference and Diffraction of Light*, (Leningrad, 1974).
7. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Nauka, Moscow, 1973, 6th ed.; Pergamon, Oxford, 1975, 4th ed.).

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 4, 2003, pp. 455–459. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 4, 2003, pp. 77–81.

OPTICS, QUANTUM ELECTRONICS

Optical Characteristics of Laser-Produced Antimony Plasma

A. K. Shuaibov, M. P. Chuchman, and L. L. Shimon

Uzhhorod National University, vul. Pidgirna 46, Uzhhorod, 88000 Ukraine

Received July 15, 2002

Abstract—The processes occurring in expanding laser-produced antimony plasma are investigated by the emission spectroscopy method. The plasma expansion velocity, the recombination time of SbII, and the electron temperature and density are determined from the dynamics of SbI line emission. Based on the results obtained, the processes occurring during the formation and expansion of laser-produced antimony plasma are qualitatively analyzed. *© 2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Optical diagnostics of laser plasmas and the study of their characteristics are of importance for optimizing laser technology processes [1]. The great interest in this research stems from the absence of adequate comprehensive laser plasma models, especially for certain semiconductors, metals, and compounds that are widely used in microelectronics (for film depositing and micropolishing) and photochemistry, as well as in developing short-wavelength radiation sources [2, 3].

Pure antimony and antimony-containing crystals (such as $CuSbS(Se)_2$) are widely used in microelectronics [4]. Since the emission spectra of the laser plasmas produced from multicomponent compounds are rather complicated [5], the study of the laser plasmas produced from individual crystal components at similar pumping conditions is of significant interest.

EXPERIMENTAL SETUP

The laser plasma was produced by irradiating a massive block of very-high-purity antimony with an LTIPCh-5 Q-switched neodymium laser. The laser wavelength was 1.06 μm, the pulse duration was 20 ns, and the pulse repetition rate was 12 Hz. With the help of a deflecting prism and a lens with a focal distance of F = 50 cm, the laser beam was focused onto the target in a spot 0.3–0.5 mm in diameter; the focal-spot intensity on the target surface was $(3–5) \times 10^8$ W/cm^2. The plasma expanded into air at a residual pressure of 3–7 Pa.

The laser plasma emission in the spectral range 200–600 nm was analyzed using an MDR-2 monochromator equipped with a 1200-line/mm diffraction grating, a FÉU-106 photomultiplier, and a KSP-4 recorder. To take into account the relative spectral sensitivity of the monochromator and photomultiplier, the recording system was calibrated with hydrogen and tungsten lamps.

The pulsed radiation was recorded at the distances r = 1 and 7 mm from the sample surface. Pulses with a duration longer than 1 μs were recorded with a pulsed Foton photomultiplier and a S1-99 oscilloscope, and the shorter pulses were recorded using an ÉLU-14FS photomultiplier and a 6LOR-04 oscilloscope.

The target was positioned at an angle of 60° with respect to the laser beam, and the emitted radiation was received at a right angle to the beam. The emission spectra were interpreted using the data from [6].

Based on the measured emission intensities of the SbI spectral lines and their spatiotemporal dynamics, one can calculate some plasma parameters, assuming that the plasma is in equilibrium and the spectral transitions are homogeneous.

The average propagation velocity of the laser plume was determined by monitoring the positions of the emission maxima at different distances from the target. Under the assumption that the upper level is populated via recombination, the recombination time can be determined from the time dependence of the logarithm of the emission intensity I normalized to its maximum value I_m. In this case, the slope of this straight line, $\tan\alpha$, gives the time τ_r of the ion recombination from the nearest upper ionization state [1]

$$\tan\alpha = \tau_r = \Delta t/\Delta\ln(I/I_m). \quad (1)$$

The electron temperature T_e was determined by the Ornstein method using the Boltzmann statistics and from the ratio of the emission intensities of the SbI spectral lines [7]

$$T_e = \Delta(E_2 - E_1)/\Delta\ln(I_1/I_2), \quad (2)$$

$$T_e = (E_2 - E_1)/k\ln(I_1A_2g_2\lambda_1/I_2A_1g_1\lambda_2), \quad (3)$$

where E and g are the energy and statistical weight of the upper level, respectively; I and λ are the spectral

1063-7842/03/4804-0455$24.00 © 2003 MAIK "Nauka/Interperiodica"

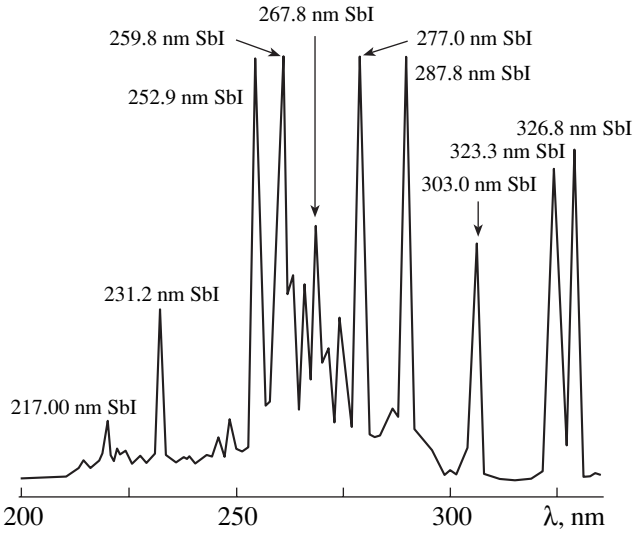

Fig. 1. Fragment of the emission spectrum from a laser-produced antimony plasma.

line intensity and wavelength, respectively; and A is the transition probability.

From the relation between τ_r and the recombination coefficient [1], one can estimate the electron density n_e:

$$n_e = \tau_r T_e^{-9/2}/8.75 \times 10^{-27} z^3, \quad (4)$$

where z is the ion charge.

DISCUSSION OF THE RESULTS

To take into account all the processes in laser plasma, one needs to know the parameters of the laser beam, the properties of the target material, the surface conditions, and the mechanism for the interaction of radiation with matter. The interaction of the laser beam with a solid target begins with the efficient electron photoemission and the photosublimation of the target material, which are followed by the thermodestruction of the target and the expansion of the plasma produced. Then, the plasma particles are created due to the processes occurring in the plasma itself [8, 9]. As a result, a significant electric potential arises on and near the target surface [10], which facilitates the formation of plasma particles and stimulates plasma expansion.

An analysis of the spectra and waveforms of the plasma emission provides enough information about the plasma formation and propagation. Based on the calculated electron temperature and density, it is possible to reveal the most typical processes in the laser plasma.

The emission spectrum of the laser-produced antimony plasma contains intense spectral lines of Sb atoms in the range 217.0–388.8 nm and the spectral lines of single-charged Sb ions in the range 347.4–461.3 nm. The upper energy states of these transitions are E_{up} = 5–7 eV for SbI and E_{up} = 11–12 eV for SbII. A fragment of the emission spectrum of the laser-produced antimony plasma is shown in Fig. 1. The emission intensity distribution over the spectrum indicates the presence of a recombination bottleneck at an energy level of E = 7.51 eV and significantly disagrees with the data on the electron excitation cross sections from [11].

Typical waveforms of the emission intensities (Figs. 2, 3) have two maxima, except for the case of ion emission near the target surface. For transitions from the low-lying levels of SbI and SbII, the increase in the distance from the target results in the onset of another maximum. At the given laser power and wavelength, the processes of target photodestruction and thermodestruction are feasible, and the combined action of these processes governs the emission time behavior near the target. The energy of the heated surface is insufficient to form the second maximum of ions, and they are produced exclusively via multiphoton ionization. This process is facilitated due to the significant width of energy states in a solid. A comparison of the lifetime of the SbI excited state (≈5 ns) with the characteristic emission time allows us to conclude that there is an extra channel for the particle formation in the course of recombination. The recombination and the effect of the potentials of both the plasma itself and the target surface stretch the plasma and its emission in space and time, which is seen in the waveforms of the spectral lines recorded at a distance of 7 mm from the target (Fig. 3).

The continuum was recorded at $r \le 3$ mm, and its characteristic waveform also had two maxima. It can be seen in the spectrum that the emission pedestal is sig-

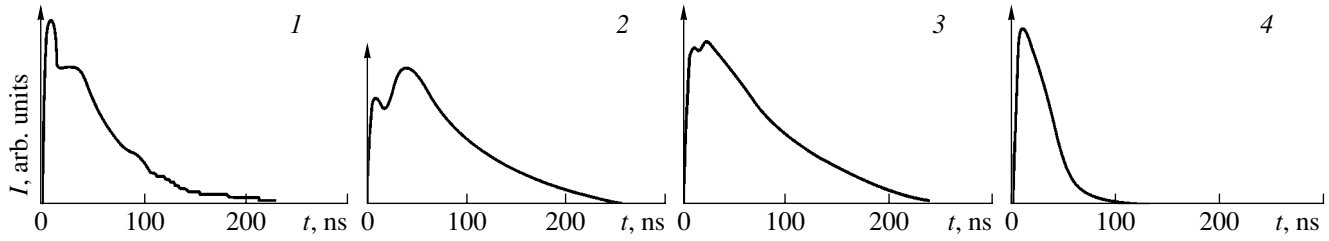

Fig. 2. Waveforms of the line emission intensity at a distance of 1 mm from the target: (*1*) SbI 277.0-nm, (*2*) SbI 287.8-nm, (*3*) SbI 326.8-nm, and (*4*) SbII 461.3-nm lines.

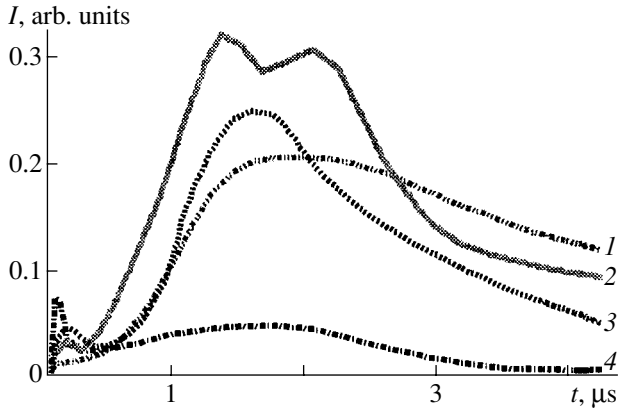

Fig. 3. Waveforms of the line emission intensity at a distance of 7 mm from the target: (*1*) SbI 277.0-nm, (*2*) SbI 287.8-nm, (*3*) SbI 326.8-nm, and (*4*) SbII 461.3-nm lines.

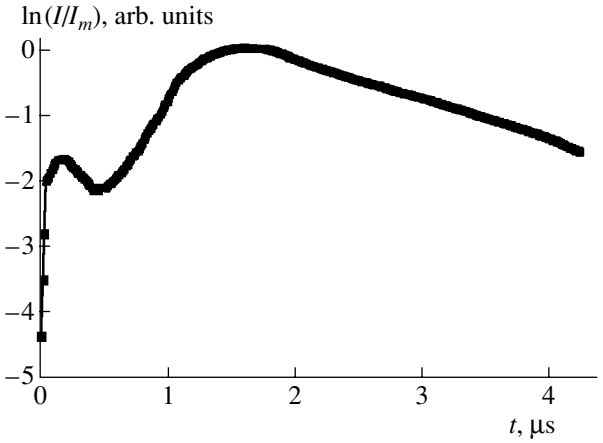

Fig. 4. Logarithm of the normalized emission intensity of the SbI 326.8-nm line at a distance of 7 mm from the target vs. time.

nificantly higher in the region of the supposed locations of the antimony dimer bands, λ = 210–230 nm.

The positions of the maxima in the emission intensity waveforms indicate the most efficient creation of ions with E_{up} = 11.2 eV and atoms with E_{up} = 5.7 eV, which are followed by the emission intensity maxima from the atoms with high E_{up} values. At the end of the laser pulse, there arises the emission maximum corresponding to the transitions from the level E_{up} = 5.4 eV. Although the second maximum for all the other spectral lines arises in the reverse order, it nevertheless remains up to the pulse end. At a distance of 7 mm from the target, the first maxima somewhat change their order of appearance in time: E_{up} = 11.2, 5.8, 5.4, and 5.7 eV. The second maximum in the waveforms appears in the order of decreasing E_{up} energies. The additional maxima of the emission intensity from Sb atoms and ions coincide in time.

Taking into account that the molecular evaporation of antimony is feasible [6] and that almost all the incident photon energy can be converted into the energy of the atomic electrons [12], the presence of additional maxima in the waveforms (Figs. 2, 3) and the fact that the plasma particles are most efficiently produced when the E_{up} value is a multiple of the laser photon energy can be explained as a result of the dissociation and dissociative recombination of Sb_2^+ and Sb_4^+ ions.

The average recombination time of SbII ions calculated from the radiation decay rate in the waveforms (Figs. 2, 3) transformed into curves like that presented in Fig. 4 amounts to 6.2 ns at r = 1 mm and 2.7 μs at r = 7 mm. Here, the main difficulty is to exclude other channels for the production of SbI; these channels should either be related to recombination or be insignificant, which is true during plasma cooling in the late stage of plasma expansion.

It is also easy to calculate the average propagation velocity of the plasma and the length of laser plume from the waveforms of the SbI spectral line intensities. Thus, the average propagation velocity from r = 1 to 7 mm is 5 km/s for the first maximum and 2.3 km/s for the second maximum. It follows from this that the length of the laser plume is 1.1 mm at a distance of 1 mm from the target and 36 mm at a distance of 7 mm. Additional information about the plasma propagation can be obtained taking into account that $\ln(I/I_m) = \ln(N/N_m)$ (where N is the density of atoms with the energy E_{up}) and that the emission intensity and the plasma density vary in time by an exponential law. In Fig. 4, straight-line segments can be seen, which indicate the propagation of plasma layers with equal velocities and densities.

Hence, we can conclude that plasma layers with different velocities are mixed, which leads to the redistribution of energy in the plasma; i.e., there is an extra channel for the production of particles.

In Fig. 5, the plasma electron temperatures calculated by the most intense SbI spectral lines indicate that, during plasma expansion, T_e varies in time within the range 0–6.5 eV, the average value being 0.63 eV. These data agree with the data from [13–15], where the presence of spikes and a decrease in T_e from the front to the tail were observed in laser-produced metal plasmas.

It can be seen that, in the early stage, the electrons are heated more intensely when interacting with the SbI atoms in the 5.826-eV state than with those in the 5.696-eV state; however, as time elapses, the situation becomes reversed. This means that the photo- and thermoexcitation processes are dominant only in the initial stage and that their role decreases with time, which is accompanied by the redistribution of the SbI level populations as the plasma cools. An important circumstance is the high average electron density $n_e = 8.27 \times 10^{17}$ cm^{-3}, which is two orders of magnitude higher than

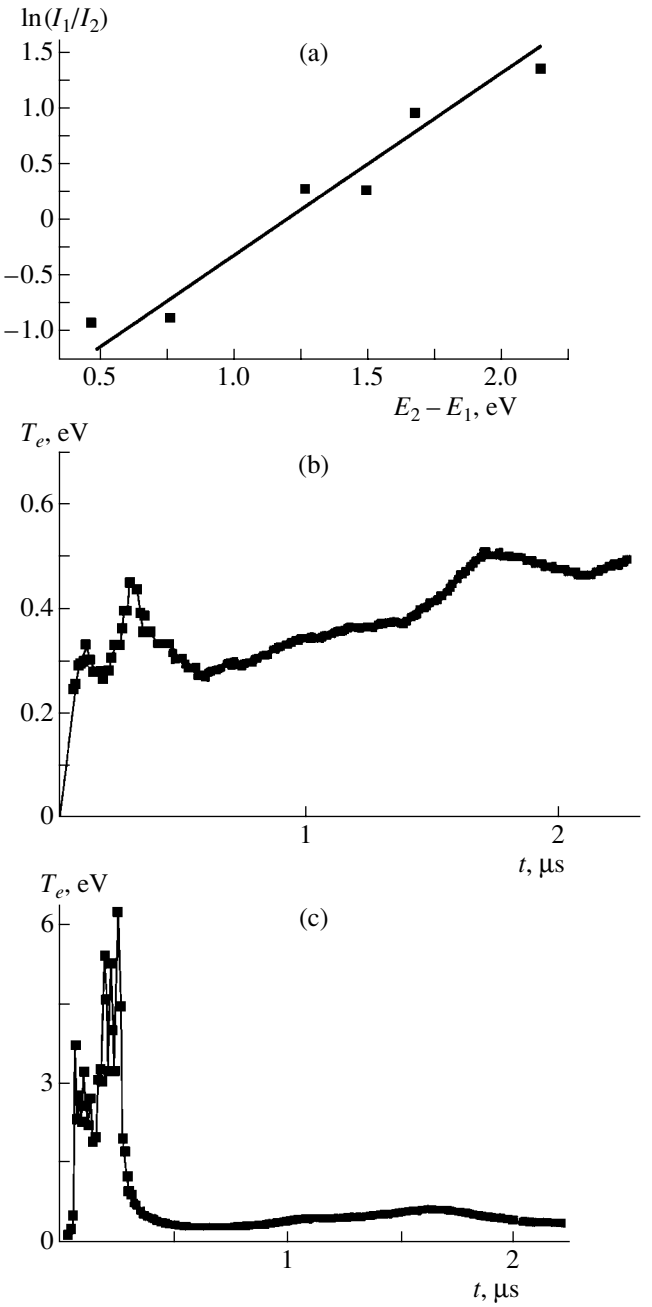

Fig. 5. (a) Logarithm of the ratio between the averaged intensities of the SbI spectral lines vs. difference between their upper level energies and (b, c) the time evolution of the electron temperature calculated by the ratio of the intensities of the (b) λ_1 = 287.8 nm and λ_2 = 277.0 nm lines and (c) λ_1 = 287.8 nm and λ_2 = 326.8 nm lines of SbI at a distance of 7 mm from the target.

that for metal plasmas under similar conditions [16]. The error in all the measurements did not exceed 30%.

A more detailed examination of the T_e dynamics reveals a characteristic set of repetitive steplike maxima. A comparison of their positions with the maxima of the emission intensity shows that the highest plasma temperature is attained in the region of the combined action of the photo- and thermoexcitation. The reproduction of the structure of these maxima indicates the self-stimulation of the heating and the excitation due to reabsorption.

Thus, for the above regime of plasma formation in the region where the combined action of different excitation factors take place, both the dissociation of complex ions and the formation of Sb^+ should proceed efficiently, which is confirmed by the additional increase in the intensity between the two main maxima in the ion emission waveform (Fig. 3).

CONCLUSION

The above qualitative analysis of the plasma emission allows one to model the behavior of an expanding erosion laser plasma, which is important for various laser applications.

In our case, the laser plasma is characterized by the following features: the presence of two excitation stages due to the multiphoton and thermal mechanisms for target destruction; the self-stimulation of plasma processes due to reabsorption; the overlapping of the plasma layers with different kinetic energies; the formation and destruction of heavy complex particles; and a distinctive mechanism for the energy redistribution in plasma due to the high electron density, which enables efficient recombination.

ACKNOWLEDGMENTS

We are grateful to A.I. Dashchenko for his help in carrying out the experiments.

REFERENCES

1. L. T. Sukhov, *Laser Spectral Analysis* (Nauka, Novosibirsk, 1990).
2. K. V. Yumashev, A. M. Molyarevich, P. V. Prokoshin, *et al.*, Kvantovaya Élektron. (Moscow) **25**, 73 (1998).
3. B. K. Kotlyarchuk, D. I. Popovich, and V. Ya. Pentko, Zh. Tekh. Fiz. **57**, 1824 (1987) [Sov. Phys. Tech. Phys. **32**, 1091 (1987)].
4. D. Besson, L. Bardotti, A. Hoareau, *et al.*, Mater. Sci. Eng. B **56**, 51 (1999).
5. A. K. Shuaibov, L. L. Shimon, and M. P. Chuchman, Zh. Tekh. Fiz. **71** (5), 85 (2001) [Tech. Phys. **46**, 590 (2001)].
6. Yu. M. Smirnov, Khim. Vys. Énerg. **26**, 7 (1992).
7. A. N. Zeĭdel', N. I. Kaliteevskiĭ, L. V. Lipis, and M. P. Chaĭka, *Emission Spectral Analysis of Atomic Materials* (Fizmatgiz, Leningrad, 1960).
8. S. I. Anisimov, Ya. A. Imis, G. S. Romanov, and Yu. V. Khodyko, *Effect of High-Power Radiation on Metals* (Nauka, Moscow, 1970).

9. R. V. Aryutyunyan, V. Yu. Baranov, V. A. Bol'shov, D. D. Malyuta, and A. Yu. Sebrant, *Effect of Laser Radiation on Metals* (Nauka, Moscow, 1989).
10. Yu. A. Bykovskiĭ, I. Yu. Konyukhov, and V. D. Peklenkov, Kvantovaya Élektron. (Moscow) **31**, 45 (2001).
11. A. K. Shuaibov, L. L. Shimon, A. I. Dashchenko, *et al.*, Ukr. Fiz. Zh. (Russ. Ed.) **46**, 1144 (2001).
12. U. Fano and J. W. Cooper, Rev. Mod. Phys. **40**, 441 (1968) [Russ. transl. Nauka, Moscow, 1972].
13. O. A. Novodvorskiĭ, E. O. Filippova, O. D. Khramova, *et al.*, Kvantovaya Élektron. (Moscow) **31**, 159 (2001).
14. V. S. Burakov, A. F. Bokhonov, P. A. Naumenkov, *et al.*, Zh. Prikl. Spektrosk. **65**, 426 (1998).
15. R. W. Dreyfus, J. Appl. Phys. **69**, 1721 (1991).
16. Yu. M. Smirnov, Usp. Fiz. Nauk **164**, 665 (1994) [Phys. Usp. **37**, 621 (1994)].

Translated by N. Ustinovskiĭ

Technical Physics, Vol. 48, No. 4, 2003, pp. 460–468. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 4, 2003, pp. 82–89.

OPTICS, QUANTUM ELECTRONICS

Asymmetric Fabry–Perot Resonator Containing an Arbitrary Inhomogeneous Layer

A. H. Gevorgyan*, A. Zh. Khachatryan, and N. M. Ispiryan****

** Yerevan State University, ul. Manukyana 1, Yerevan, 375025 Armenia*

*** Armenian State Engineering University, Yerevan, 375046 Armenia*

e-mail: agevorgyan@ysu.am

Received September 17, 2002

Abstract—A new method for determining the transmission and reflection coefficients for an arbitrarily polarized electromagnetic wave incident obliquely on an inhomogeneous insulating layer inside an asymmetric Fabry–Perot resonator is proposed. Algebraic relationships between these coefficients for a layer bounded by different homogeneous semi-infinite media and for the same layer in a vacuum are derived. Three examples corresponding to real situations are analyzed, and the results of corresponding numerical calculations are discussed. *© 2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

For many years, the propagation of electromagnetic waves in random disordered media has been a central problem in the wave theory [1–6]. It is well known that even an approximate analysis of this problem for 2D and 3D systems involves mathematical difficulties. At the same time, interest in various properties of one-dimensional random and arbitrary layered structures has increased [7–13]. The reasons are the technological progress in creating artificial systems with the desired structure and composition and their expanded application.

Recently, various optical devices based on planar and thin-film layered structures have found wide applications [9, 14–16]. For example, one-dimensional metal–dielectric photonic crystals are frequently used as high-quality mirrors [15, 17–19]. Of interest also are various combinations of photonic crystals with layered structures. In particular, they can be used in real-time optical delay lines [20]; high-performance substrates for antennas, optical diodes, and limiters [21–23]; and optical sensors and modulators [24, 25].

In this work, we consider the amplitude coefficients of reflection and transmission for an arbitrarily polarized plane electromagnetic wave incident obliquely on a one-dimensional inhomogeneous insulating layer bounded by two different semi-infinite layers (asymmetric Fabry–Perot resonator with a heterogeneous layer inside, Fig. 1). Such structures have many applications, such as laser cavities; various acousto-, electro-, and magnetooptical devices; diffraction filters; mirrors; etc.

STATEMENT OF THE PROBLEM

Consider an isotropic insulating layer that is inhomogeneous in the z direction and homogeneous in the x and y directions. Let this layer be sandwiched in two different homogeneous semi-infinite dielectric media (Fig. 1); that is,

$$\varepsilon(z) = \begin{cases} \varepsilon_1, & z < 0 \\ \varepsilon_0, & 0 \le z < a - \Delta/2 \\ \varepsilon(z), & a - \Delta/2 \le z \le a + \Delta/2 \\ \varepsilon_0, & a + \Delta/2 < z \le d \\ \varepsilon_2, & z > d, \end{cases} \qquad (1)$$

where $\varepsilon(z)$ is an arbitrary function and, in the general case, ε_1 differs from ε_2.

Assume that the wave vector $\mathbf{k}$ lies in the (x, z) plane and that the electric and magnetic fields represent the real parts of complex vectors $\mathbf{E}\exp\{-i\omega t\}$ and

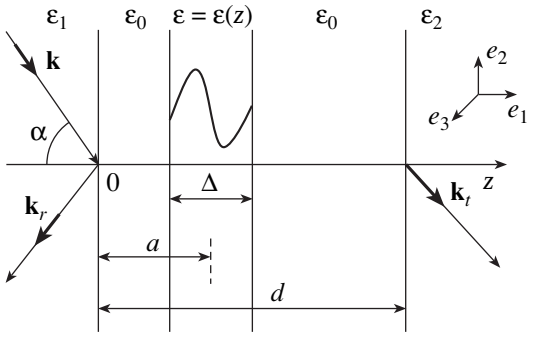

Fig. 1. Propagation of light through a 1D layer between two different isotropic semi-infinite media.

1063-7842/03/4804-0460$24.00

$\mathbf{H}\exp\{-i\omega t\}$. Then, the spatial dependences of $\mathbf{E}$ and $\mathbf{H}$ are written as

$$\mathbf{E}(\mathbf{r}) = \begin{cases} \mathbf{E}_0\exp\{i\mathbf{k}\cdot\mathbf{r}\} + \mathbf{E}_r\exp\{i\mathbf{k}_r\cdot\mathbf{r}\}, & z<0 \\ \mathbf{E}_t\exp\{i\mathbf{k}_t\cdot\mathbf{r}\}, & z>d, \end{cases} \quad (2)$$

where $\mathbf{E}_0$, $\mathbf{E}_r$, and $\mathbf{E}_t$ are the respective amplitudes of the incident, reflected, and transmitted waves.

The vectors $\mathbf{k}$ and $\mathbf{k}_r$ have equal absolute values ($k = \omega\sqrt{\varepsilon_1}/c$) and belong to the xz plane. Note that the magnetic field exhibits the same asymptotic behavior as that given by expression (2). The wave vectors of the incident and reflected waves are given by

$$\mathbf{k} = k_1\cos\alpha\mathbf{e}_1 + k_1\sin\alpha\mathbf{e}_2$$

and

$$\mathbf{k}_r = -k_1\cos\alpha\mathbf{e} + k_1\sin\alpha\mathbf{e}_2, \quad (3)$$

respectively, where α is the angle of incidence and $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$ are unit base vectors for the z, x, and y axes, respectively.

Since the energy flux density of an electromagnetic wave remains constant, the wave vector of the transmitted wave is independent of the refractive index of the layer. It depends only on the angle of incidence and refractive indices of the first and second semi-infinite media:

$$\mathbf{k}_t = k_2\cos\beta\mathbf{e}_1 + k_2\sin\beta\mathbf{e}_2, \quad (4)$$

where β is the angle of refraction ($\sqrt{\varepsilon_1}\sin\alpha = \sqrt{\varepsilon_2}\sin\beta$) and $k_2 = \omega\sqrt{\varepsilon_2}/c$.

One can represent an arbitrarily polarized plane wave as the superposition of s- and p-polarized waves:

$$\mathbf{E}_{i,r,t} = E^s_{i,r,t}\mathbf{n}_s + E^p_{i,r,t}\mathbf{n}_p. \quad (5)$$

Here, the subscripts i, r, and t correspond to the incident, reflected, and transmitted waves, respectively, and $\mathbf{n}_s$ and $\mathbf{n}_p$ are the unit vectors of s and p polarizations. Note that the polarization of the s and p waves is invariant under scattering.

For p polarization, the electric field vector is in the plane parallel to the plane of the layer. For s polarization, this vector lies in the plane of incidence. Let us represent $E^s(x, z)$ and $H^p(x, z)$ as $E^s(x, z) = E^s(z)U(x)$ and $H^p(x, z) = H^p(z)V(x)$. Based on the Maxwell equation and asymptotic equation (2), we can then derive the following wave equations [1, 2]:

$$\frac{d^2E^s(z)}{dz^2} + \frac{\omega^2}{c^2}(\varepsilon(z) - \varepsilon_1\sin^2\alpha)E^s(z) = 0, \quad (6)$$

$$\frac{d}{dz}\left(\frac{1}{\varepsilon(z)}\frac{dH^p(z)}{dz}\right) + \frac{\omega^2}{c^2}\left(1 - \frac{\varepsilon_1\sin^2\alpha}{\varepsilon(z)}\right)H^p(z) = 0. \quad (7)$$

It follows from the Snell law that the term $\varepsilon_1\sin^2\alpha$ in expressions (6) and (7) can be replaced by the term $\varepsilon_2\sin^2\beta$. This fact reflects the independence of the angle of refraction β on the parameters (width and refractive index) of the layer. This angle depends only on the angle of incidence α and the permittivities ε_1 and ε_2 of the semi-infinite media.

Using boundary conditions (2), we introduce the complex amplitude reflection and transmission coefficients for the s and p waves in the form

$$T^s_{1,2} = \frac{E^s_t}{E^s_0}, \quad R^s_{1,2} = \frac{E^s_r}{E^s_0}$$

and

$$T^p_{1,2} = \frac{H^p_t}{H^p_0}, \quad R^p_{1,2} = \frac{H^p_r}{H^p_0}. \quad (8)$$

It can be deduced from wave equations (6) and (7) that the following quantities related to the electric and magnetic components of the s and p waves are independent of z:

$$E^s = \frac{d(E^s)^*}{dz} - (E^s)^*\frac{dE^s}{dz} = \text{const}, \quad (9)$$

$$\frac{1}{\varepsilon(z)}\left[H^p\frac{d(H^p)^*}{dz} - (H^p)^*\frac{dH^p}{dz}\right] = \text{const}. \quad (10)$$

These conditions represent the law of conservation of the energy flux density for the s- and p-polarized waves, respectively. Using the asymptotics of the field in the semi-infinite media (expression (2)) and also conditions (9) and (10), we obtain

$$k_1\cos\alpha(|E^s_0|^2 - |E^s_r|^2) = k_2\cos\beta|E^s_t|^2, \quad (11)$$

$$\frac{k_1\cos\alpha}{\varepsilon_1}(|H^p_0|^2 - |H^p_r|^2) = \frac{k_2\cos\beta}{\varepsilon_2}|H^p_t|^2. \quad (12)$$

In view of (8), expressions (11) and (12) are represented as

$$\begin{aligned} 1 + |R^s_{1,2}|^2 &= \frac{k_2\cos\beta}{k_1\cos\alpha}|T^s_{1,2}|, \\ 1 + |R^p_{1,2}|^2 &= \frac{k_1\cos\beta}{k_2\cos\alpha}|T^p_{1,2}|^2. \end{aligned} \quad (13)$$

Our goal is to find amplitude coefficients (8) for given parameters of the system (ε_1, ε_2, $\varepsilon(z)$, d, Δ, and a), angle of incidence, and wavelength.

SCATTERING BY AN INSULATING LAYER BOUNDED BY TWO HOMOGENEOUS SEMI-INFINITE MEDIA

Consider the scattering of an electromagnetic wave by an insulating layer with an arbitrary permittivity $\varepsilon(z)$

sandwiched between two different semi-infinite media with permittivities ε_1 and ε_2. For the domains $z < 0$ and $z > d$, the general solution to Eqs. (6) and (7) is written as

$$\begin{aligned} &A^{s,p}\exp\{ik_{1z}z\} + B^{s,p}\exp\{-ik_{1z}z\}, \quad z < 0, \\ &C^{s,p}\exp\{ik_{2z}z\} + D^{s,p}\exp\{-ik_{2z}z\}, \quad z > d, \end{aligned} \tag{14}$$

where $k_{1z} = k_1\cos\alpha$ and $k_{2z} = k_2\cos\beta$.

The transfer matrix method [26–29] yields a linear relation between the coefficients entering into expressions (14):

$$\begin{pmatrix} C^{s,p} \\ D^{s,p} \end{pmatrix} = \frac{k_{1,z}^{s,p}}{k_{2,z}^{s,p}} \begin{pmatrix} (1/T_{1,2}^{s,p})^* & (-R_{1,2}^{s,p}/T_{1,2}^{s,p})^* \\ -R_{1,2}^{s,p}/T_{1,2}^{s,p} & 1/T_{1,2}^{s,p} \end{pmatrix} \begin{pmatrix} A^{s,p} \\ B^{s,p} \end{pmatrix}. \tag{15}$$

Here, $T_{1,2}^{s,p}$ and $R_{1,2}^{s,p}$ are the amplitude transmission and reflection coefficients, respectively, given by relationships (8) for the *s* and *p* waves.

Note that in the case of the *s*-polarized wave, $k_{1,z}^s = k_{1,z}$ and $k_{2,z}^s = k_{2,z}$. For the *p*-polarized wave, we have $k_{1,z}^p = k_{1,z}/\varepsilon_1$ and $k_{2,z}^p = k_{2,z}/\varepsilon_2$.

Let us introduce the matrices

$$Q_{1,0}^{s,p} = \frac{k_{1z}}{k_{0z}} \begin{pmatrix} (1/t_{1,0}^{s,p})^* & (-r_{1,0}^{s,p}/t_{1,0}^{s,p})^* \\ -r_{1,0}^{s,p}/t_{1,0}^{s,p} & 1/t_{1,0}^{s,p} \end{pmatrix}$$

and

$$Q_{0,2}^{s,p} = \frac{k_{0z}}{k_{2z}} \begin{pmatrix} (1/t_{0,2}^{s,p})^* & (-r_{0,2}^{s,p}/t_{0,2}^{s,p})^* \\ -r_{0,2}^{s,p}/t_{0,2}^{s,p} & 1/t_{0,2}^{s,p} \end{pmatrix}, \tag{16}$$

where $k_{0x} = \omega/c\cos\gamma$ and $\sin\gamma = \sqrt{\varepsilon_1}\sin\alpha = \sqrt{\varepsilon_2}\sin\beta$.

In expressions (16), $t_{1,0}$ and $r_{1,0}$ ($t_{0,2}$ and $r_{0,2}$) are the amplitude transmission and reflection coefficients, respectively, for the first (second) semi-infinite medium bordering a vacuum on the right (left). The amplitudes $t_{1,0}^{s,p}$, $r_{1,0}^{s,p}$, $t_{0,2}^{s,p}$, and $r_{0,2}^{s,p}$ can be represented as

$$\begin{aligned} \frac{1}{t_{0,2}^s} &= \frac{k_{0z}+k_{2z}}{2k_{0z}}\exp\{i(k_{2z}-k_{0z})d\}, \\ \frac{r_{0,2}^s}{t_{0,2}^s} &= \frac{k_{0z}-k_{2z}}{2k_{0z}}\exp\{i(k_{2z}+k_{0z})d\}, \end{aligned} \tag{17}$$

$$\begin{aligned} \frac{1}{t_{0,2}^p} &= \frac{\varepsilon_2 k_{0z}+k_{2z}}{2\varepsilon_2 k_{0z}}\exp\{i(k_{2z}-k_{0z})d\}, \\ \frac{r_{0,2}^p}{t_{0,2}^p} &= \frac{\varepsilon_2 k_{0z}-k_{2z}}{2\varepsilon_2 k_{0z}}\exp\{i(k_{2z}+k_{0z})d\}. \end{aligned} \tag{18}$$

Expressions for $t_{1,0}^s$ and $r_{1,0}^s$ can be obtained from relationships (17) by substituting $d = 0$, k_{0z} for k_{1z}, and k_{2z} for k_{0z}. Similarly, expressions for $t_{1,0}^p$ and $r_{1,0}^0$ can be obtained from relationships (18) by substituting $d = 0$, k_{1z} for $\varepsilon_2 k_{0z}$, and $\varepsilon_1 k_{0z}$ for k_{2z}.

The transfer matrix for the entire system can be represented as the product of matrices (16) and the transfer matrix for a layer bordering a vacuum on both sides:

$$U_{1,2}^{s,p} = Q_{0,2}^{s,p} U^{s,p} Q_{1,0}^{s,p}. \tag{19}$$

Here, $U^{s,p}$ is the transfer matrix for the layer bordering a vacuum on both sides:

$$U^{s,p} = \begin{pmatrix} (1/T^{s,p})^* & (-R^{s,p}/T^{s,p})^* \\ -R^{s,p}/T^{s,p} & 1/T^{s,p} \end{pmatrix}, \tag{20}$$

where $T^{s,p}$ and $R^{s,p}$ are the amplitude transmission and reflection coefficients of the *s* and *p* waves with the transfer matrix for this layer.

Using expressions (17)–(20), we can derive algebraic relationships between the amplitude coefficients $T_{1,2}^{s,p}$, $R_{1,2}^{s,p}$, $T^{s,p}$, and $R^{s,p}$. For the *s* wave, these relationships are given by

$$\begin{aligned} \frac{1}{T_{1,2}^s} &= \frac{\exp\{ik_{2z}d\}}{4k_{1z}k_{0z}}[(k_{2z}-k_{0z})(k_{0z}-k_{1z})a^s \\ &\quad + (k_{2z}+k_{0z})(k_{0z}+k_{1z})(a^s)^* \\ &\quad + (k_{0z}+k_{2z})(k_{1z}-k_{0z})b^s \\ &\quad + (k_{0z}-k_{2z})(k_{0z}+k_{1z})(b^s)^*], \end{aligned} \tag{21}$$

$$\begin{aligned} \frac{R_{1,2}^s}{T_{1,2}^s} &= \frac{\exp\{ik_{2z}d\}}{4k_{1z}k_{0z}}[(k_{0z}-k_{2z})(k_{0z}+k_{1z})a^3 \\ &\quad + (k_{1z}-k_{0z})(k_{0z}+k_{2z})(a^s)^* \\ &\quad + (k_{0z}+k_{2z})(k_{0z}+k_{1z})b^s \\ &\quad + (k_{2z}-k_{0z})(k_{0z}-k_1)(b^s)^*], \end{aligned} \tag{22}$$

where

$$a^s = \exp\{ik_{0z}d\}/(T^s)^*, \quad b^s = R^s\exp\{-ik_{0z}d\}/T^s. \tag{23}$$

For the *p* wave, these are

$$\begin{aligned} \frac{1}{T_{1,2}^p} &= \frac{\exp\{ik_{2z}d\}}{4\varepsilon_2 k_{1z}k_{0z}}[(k_{2z}-\varepsilon_2 k_{0z})(\varepsilon_1 k_{0z}-k_{1z})a^p \\ &\quad + (k_{2z}+\varepsilon_2 k_{0z})(\varepsilon_1 k_{0z}+k_{1z})(a^p)^* \\ &\quad + (\varepsilon_2 k_{0z}+k_{2z})(k_{1z}-\varepsilon_1 k_{0z})b^p \\ &\quad + (\varepsilon_2 k_{0z}-k_{2z})(\varepsilon_1 k_{0z}+k_{1z})(b^p)^*], \end{aligned} \tag{24}$$

$$\frac{R^p_{1,2}}{T^p_{1,2}} = \frac{\exp\{ik_{2z}d\}}{4\varepsilon_2 k_{1z}k_{0z}}[(\varepsilon_2 k_{0z} - k_{2z})(\varepsilon_1 k_{0z} + k_{1z})a^p + (k_{1z} - \varepsilon_1 k_{0z})(\varepsilon_2 k_{0z} + k_{2z})(a^p)^* + (\varepsilon_2 k_{0z} + k_{2z})(\varepsilon_1 k_{0z} + k_{1z})b^p + (k_{2z} - \varepsilon_2 k_{0z})(\varepsilon_1 k_{0z} - k_{1z})(b^p)^*], \quad (25)$$

where

$$a^p = \exp\{ik_{0z}d\}/(T^p)^*, \quad b^p = R^p\exp\{-ik_{0z}d\}/T^p. \quad (26)$$

It follows from expressions (21)–(26) that the determination of the amplitude coefficients $T^{s,p}_{1,2}$ and $R^{s,p}_{1,2}$ for a layer bounded by two different homogeneous semi-infinite media is reduced to the determination of the amplitude coefficients $T^{s,p}$ and $R^{s,p}$ for the same layer bordering a vacuum on both sides. Note that in the case of normal incidence ($R^s = -R^p$ and $T^s = T^p$), the amplitude coefficients $T^s_{1,2}$ and $R^s_{1,2}$ for the electric component of the *s* wave and the amplitude coefficients $T^p_{1,2}$ and $R^p_{1,2}$ for the magnetic component of the *p* wave are related as $R^s_{1,2} = -R^p_{1,2}$ and $\sqrt{\varepsilon_2}T^s_{1,2} = \sqrt{\varepsilon_1}T^p_{1,2}$.

THE PROBLEM OF SCATTERING AS THE CAUCHY PROBLEM FOR A WAVE EQUATION

It has been demonstrated in the previous section that there exist relationships (expressions (21–(26)) between the amplitude coefficients of an electromagnetic wave incident on an insulating layer bounded by two homogeneous semi-infinite media and the amplitude coefficients of this wave incident on the same layer bordering a vacuum on both sides (Fig. 1). Below, we will demonstrate that the problem of determining the coefficients $T^{s,p}_{1,2}$ and $R^{s,p}_{1,2}$ can be formulated as the Cauchy problem for wave equations (6) and (7).

In accordance with the approach developed in [30, 31], the amplitude coefficients $T^{s,p}$ and $R^{s,p}$ for a layer that has a continuous refractive index $\varepsilon(z)$ and borders a vacuum on both sides can be expressed in terms of real functions $H^{s,p}_{1,2}(z)$ and $N^{s,p}_{1,2}(z)$ at the point $z = d$:

$$\frac{1}{T^{s,p}} = \frac{1}{2}\exp\{ik_{0x}d\} \times [(H^{s,p}_1(d) + N^{s,p}_2(d)) - i(N^{s,p}_1(d) - H^{s,p}_2(d))], \quad (27)$$

$$\frac{R^{s,p}}{T^{s,p}} = -\frac{1}{2}\exp\{ik_{0x}d\} \times [(H^{s,p}_1(d) - N^{s,p}_2(d)) - i(N^{s,p}_1(d) + H^{s,p}_2(d))]. \quad (28)$$

The functions $H^{s,p}_{1,2}(z)$ and $N^{s,p}_{1,2}(z)$ are solutions to the following system of differential equations:

$$\frac{dN^s_{1,2}}{dz} = \frac{\omega^2}{c^2}\left[\frac{\varepsilon_1\sin^2\alpha - \varepsilon(z)}{k_{0z}}\right]H^s_{1,2}, \quad \frac{dH^s_{1,2}}{dz} = -k_{0z}N^s_{1,2} \quad (29)$$

and

$$\frac{dN^p_{1,2}}{dz} = \frac{\omega^2}{c^2}\frac{1}{k_{0z}}\left[1 - \frac{\varepsilon_1\sin^2\alpha}{\varepsilon(z)}\right]H^p_{1,2}, \quad \frac{dH^p_{1,2}}{dz} = -\varepsilon(z)k_{0z}N^p_{1,2} \quad (30)$$

with the initial conditions

$$H^{s,p}_1 = 1, \quad H^{s,p}_2 = 0 \quad \text{and} \quad N^{s,p}_1 = 0, \quad N^{s,p}_2 = 1. \quad (31)$$

It is seen from Eqs. (29) and (30) that the functions $H^{s,p}_{1,2}(z)$ and $N^{s,p}_{1,2}(z)$ also obey wave equations (6) and (7). Note that in Eqs. (29) and (30), $k_{0z} = \frac{\omega}{c}\cos\gamma$, where γ is the angle at which the wave propagating from a vacuum is incident on the layer, and that to determine $T^{s,p}$ and $R^{s,p}$, it is necessary to take into account the relationships $\sin\gamma = \sqrt{\varepsilon_1}\sin\alpha = \sqrt{\varepsilon_2}\sin\beta$.

Using expressions (21)–(26), (29), and (30), one can show that the amplitude coefficients $T^{s,p}_{1,2}$ and $R^{s,p}_{1,2}$ for a layer bounded by different homogeneous semi-infinite media are expressed in terms of the functions $H^{s,p}_{1,2}(z)$ and $N^{s,p}_{1,2}(z)$ at the point $z = d$ as follows:

$$\frac{1}{T^s_{1,2}} = \frac{1}{2}\exp\{ik_{2z}d\}\left[\left(\frac{k_{2z}}{k_{1,z}}H^s_1(d) + N^s_2(d)\right) - i\left(\frac{k_{0z}}{k_{1z}}N^s_1(d) - \frac{k_{2z}}{k_{0z}}H^s_2(d)\right)\right], \quad (32)$$

$$\frac{R^s_{1,2}}{T^s_{1,2}} = -\frac{1}{2}\exp\{ik_{2z}d\}\left[\left(\frac{k_{2z}}{k_{1,z}}H^s_1(d) - N^s_2(d)\right) - i\left(\frac{k_{0z}}{k_{iz}}N^s_1(d) + \frac{k_{2z}}{k_{0z}}H^s_2(d)\right)\right], \quad (33)$$

$$\frac{1}{T^p_{1,2}} = \frac{1}{2}\exp\{ik_{2z}d\}\left[\left(\frac{\varepsilon_1 k_{2z}}{\varepsilon_2 k_{1,2}}H^p_1(d) + N^p_2(d)\right) - i\left(\frac{\varepsilon_1 k_{0z}}{k_{1z}}N^p_1(d) - \frac{k_{2z}}{\varepsilon_2 k_{0z}}H^p_2(d)\right)\right], \quad (34)$$

$$\frac{R^p_{1,2}}{T^p_{1,2}} = -\frac{1}{2}\exp\{ik_{2z}d\}\left[\left(\frac{\varepsilon_1 k_{2z}}{\varepsilon_2 k_{1,z}}H^p_1(d) - N^p_2(d)\right) - i\left(\frac{\varepsilon_1 k_{0z}}{k_{1z}}N^p_1(d) + \frac{k_{2z}}{\varepsilon_2 k_{0z}}H^p_2(d)\right)\right]. \quad (35)$$

It follows from expressions (29)–(31) and (32)–(35) that the problem of determining the amplitude coefficients $T^{s,p}_{1,2} = T^{s,p}_{1,2}$ ($z = d$) and $R^{s,p}_{1,2} = R^{s,p}_{1,2}$ ($z = d$) for the *s* and *p* waves is reduced to the Cauchy problem for Eqs. (6) and (7), respectively.

It is expedient to apply (29)–(35) to the simple case of a homogeneous layer ($\varepsilon(z) = \varepsilon = \text{const}$) bounded by two semi-infinite media. In this case, from Eqs. (29)–(31), we have

$$H^s_1 = \cos k_z z, \quad N^s_1 = \frac{k_z}{k_{0z}}\sin k_z z, \quad H^s_2 = -\frac{k_{0z}}{k_z}\sin k_z z, \quad N^s_2 = \cos k_z z, \quad (36)$$

$$H^p_1 = \cos k_z z, \quad N^s_1 = \frac{k_z}{\varepsilon k_{0z}}\sin k_z z, \quad H^s_2 = -\frac{\varepsilon k_{0z}}{k_z}\sin k_z z, \quad N^s_2 = \cos k_z z, \quad (37)$$

where $k_z = \omega/c\sqrt{\varepsilon}\cos\varphi$ and $\varepsilon\sin\varphi = \sin\gamma$.

Substituting expressions (36) into Eqs. (32) and (33) and expressions (37) into Eqs. (34) and (35), we arrive at

$$\frac{1}{T^s_{1,2}} = \exp\{ik_{2z}d\}\left[\frac{1}{2}\left(1 + \frac{k_{2z}}{k_{1z}}\right)\cos\{k_z d\} - i\frac{k_{2z}k_{1z} + k_z^2}{2k_{1z}k_z}\sin\{k_z d\}\right], \quad (38)$$

$$\frac{R^s_{1,2}}{T^s_{1,2}} = \exp\{ik_{2z}d\}\left[\frac{1}{2}\left(1 - \frac{k_{2z}}{k_{1z}}\right)\cos\{k_z d\} + i\frac{k_z^2 - k_{1z}k_{2z}}{2k_{1z}k_z}\sin\{k_z d\}\right], \quad (39)$$

$$\frac{1}{T^p_{1,2}} = \exp\{ik_{2z}d\}\left[\frac{1}{2}\left(1 + \frac{\varepsilon_1 k_{2z}}{\varepsilon_2 k_{1z}}\right)\cos\{k_z d\} - i\frac{\varepsilon^2 k_{2z}k_{1z} + \varepsilon_2\varepsilon_1 k_z^2}{2\varepsilon\varepsilon_2 k_{1z}k_z}\sin\{k_z d\}\right], \quad (40)$$

$$\frac{R^p_{1,2}}{T^p_{1,2}} = \exp\{ik_{2z}d\}\left[\frac{1}{2}\left(1 - \frac{\varepsilon_1 k_{2z}}{\varepsilon_2 k_{1z}}\right)\cos\{k_z d\} + i\frac{\varepsilon_2\varepsilon_1 k_{0z}^2 - \varepsilon^2 k_{1z}k_{2z}}{2\varepsilon\varepsilon_2 k_{1z}k_z}\sin\{k_z d\}\right] \quad (41)$$

with

$$\sqrt{\varepsilon_1}\sin\alpha = \sqrt{\varepsilon}\sin\varphi = \sqrt{\varepsilon_2}\sin\beta,$$

where α is the angle at which the wave propagating from the first semi-infinite medium is incident on the layer, φ is the angle of refraction in the layer, and β is the angle of refraction in the second semi-infinite medium.

Expressions (38)–(41) enable one to determine the transmission and reflection coefficients for the *s*- and *p*-polarized waves in the case of a homogeneous layer bounded by two different semi-infinite media.

Assume for definiteness that k_{2z} entering into expressions (38)–(41) is a real quantity. In other words, we suppose that $\varepsilon_2 > \varepsilon_1$ and that, if $\varepsilon_2 < \varepsilon_1$, the angle of incidence α is less than the critical angle α' ($\sin\alpha' = \sqrt{\varepsilon_2/\varepsilon_1}$) of total internal reflection from the interface between the first and second semi-infinite media. Expressions (38)–(41) allow us to determine the condition for the total transmission for the *s* and *p* waves. Note that if k_z is an imaginary quantity ($\varepsilon_1 < \varepsilon$ or $\alpha > \alpha''$ and $\sin\alpha'' = \sqrt{\varepsilon/\varepsilon_1}$), the reflected and transmitted waves exist for both polarizations.

If k_z is a real quantity, the condition $R^s_{1,2} = 0$ means that the parameters of the problem satisfy simultaneously two equations

$$k_{1z} = k_{2z} \quad \text{and} \quad \sin k_z d = 0. \quad (42)$$

It is seen that the total transmission of the *s* wave through the layer is possible only if the semi-infinite media are identical ($\varepsilon_1 = \varepsilon_2$) and $k_z d = \pi n$, where $n = 1, 2, \ldots$.

In the case of *p* polarization and $R^p = 0$, two equations can be derived from expression (41):

$$\varepsilon_2 k_{1z} = \varepsilon_1 k_{2z}, \quad \sin k_z d = 0. \quad (43)$$

Note that the first equation is equivalent to the Brewster condition for the total transmission of a *p*-polarized wave at the interface between the first and the second semi-infinite media ($\tan\alpha = \sqrt{\varepsilon_2/\varepsilon_1}$) [1]. It follows from Eqs. (43) that the condition $R^p = 0$ is satisfied only

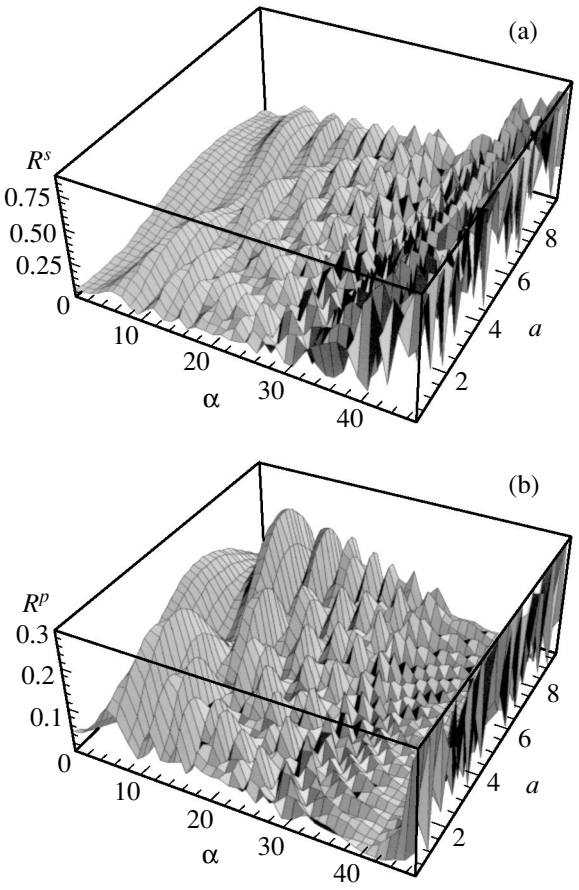

Fig. 2. Reflection coefficient for the (a) *s* and (b) *p* waves versus angle of incidence α and parameter *a* for λ = 0.5 μm, $\varepsilon_0 = 1$, $\varepsilon_1 = 1.69$, $\varepsilon_2 = 3.24$, $\varepsilon = 2.25 + i0.1$, $d = 10$ μm, and $\Delta = 0.5$ μm.

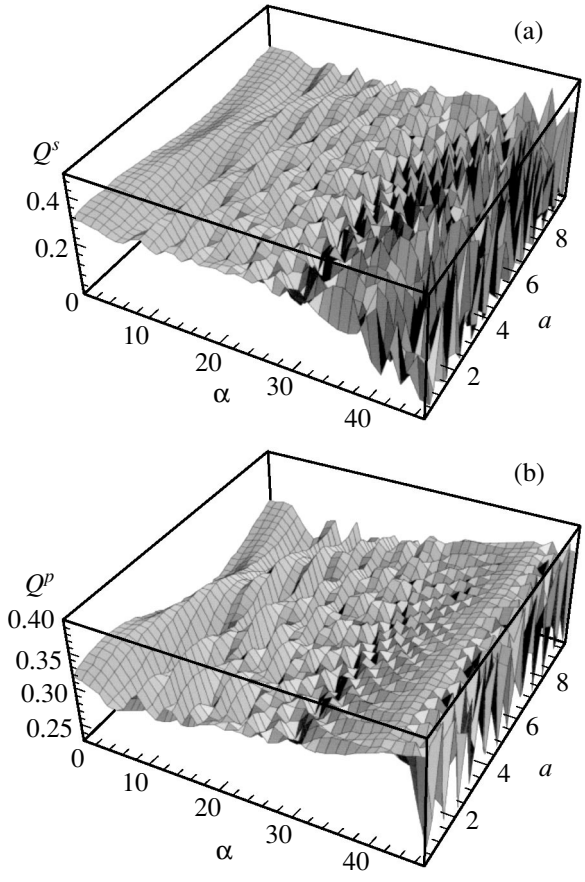

Fig. 3. Absorption coefficient for the (a) *s* and (b) *p* waves versus angle of incidence α and parameter *a*. The remaining parameters are the same as in Fig. 2.

if the angle of incidence equals $\alpha = \arctan\sqrt{\varepsilon_2/\varepsilon_1}$ and $k_z d = \pi n$, where $n = 1, 2, \ldots$.

NUMERICAL CALCULATIONS AND CONCLUSIONS

To illustrate the results obtained, we will consider three examples. The first one corresponds to a homogeneous layer with a finite width Δ placed inside an asymmetric Fabry–Perot resonator of width *d*. Let *a* be the distance between the center of the homogeneous layer and the left arm of the interferometer (Fig. 1). Note that in accordance with the statement of the problem, the parameters *d*, Δ, and *a* must satisfy the conditions $\Delta \le d$ and $\Delta/2 \le a \le d - \Delta/2$. Assume that the permittivity ε_0 of the free space inside the interferometer is constant. Then, expression (27) and (28) for the amplitude coefficients take the form

$$\frac{1}{T^s} = \exp\{ik_{0z}\Delta\}\left[\cos\{k_z\Delta\} - i\frac{k_{0z}^2 + k_z^2}{2k_{0z}k_z}\sin\{k_z\Delta\}\right], \quad (44)$$

$$\frac{R^s}{T^s} = i\exp\{i2k_{0z}a\}\frac{k_z^2 - k_{0z}^2}{2k_{0z}k_z}\sin\{k_z\Delta\}, \quad (45)$$

$$\frac{1}{T^p} = \exp\{ik_{0z}\Delta\} \times\left[\cos\{k_z\Delta\} - i\frac{(\varepsilon/\varepsilon_0)k_{0z}^2 + (\varepsilon_0/\varepsilon)k_z^2}{2k_{0z}k_z}\sin\{k_z\Delta\}\right], \quad (46)$$

$$\frac{R^p}{T^p} = i\exp\{i2k_{0z}a\}\frac{(\varepsilon_0/\varepsilon)k_z^2 - (\varepsilon/\varepsilon_0)k_{0z}^2}{2k_{0z}k_z}\sin\{k_z\Delta\}. \quad (47)$$

Figure 2 plots the reflection coefficients $R^s = |R_{12}^s|^2$ and $R^p = |R_{12}^p|^2$ for the *s*- and *p*-polarized waves, respectively, versus the angle of incidence α and parameter *a*. The refractive index of the isotropic layer is written as

$$n = \sqrt{\varepsilon} = n' + n'', \quad (48)$$

where *n*" is the absorption coefficient.

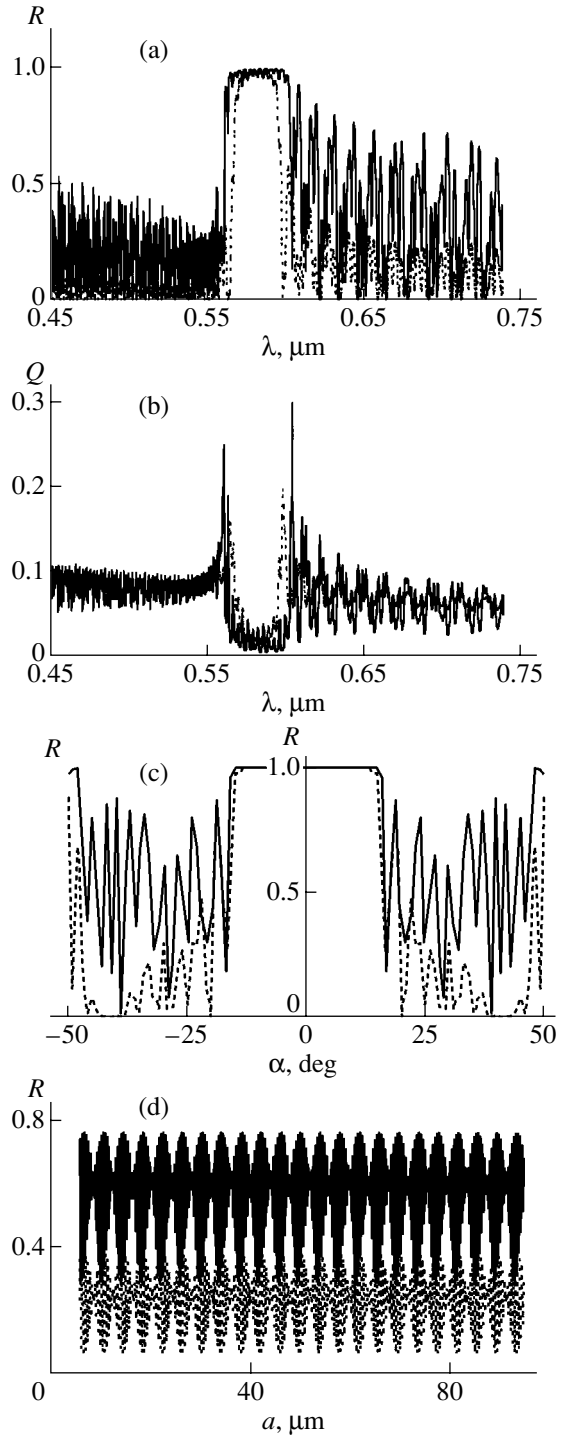

Fig. 4. Reflection coefficient R versus (a) wavelength, (c) angle of incidence, and (d) parameter a. (b) The absorption coefficient Q versus wavelength for $\sigma = 0.4$ μm, $c = 0.5$, $d = 100$ μm, and $\Delta = 25\sigma$. $\varepsilon = 2.25 + i0.001$ (a, b) and 2.25 (c, d), $\alpha = 30°$ (a, b, d), $a = 40$ μm (a–c), and $\lambda = 0.63925$ (c) and 0.64 μm (d). The remaining parameters are the same as in Fig. 2.

Figure 3 plots the parameters $Q^{s,p} = 1 - (R^{s,p} + T^{s,p})$, which characterize the electromagnetic wave energy absorbed in the medium, versus the angle of incidence α and parameter a. It is seen that the quantities $Q^{s,p}$ oscillate and the amplitudes of these oscillations are different for the s and p waves. The oscillations are modulated. The values of $Q^{s,p}$ vary over wide ranges from anomalously strong to anomalously weak absorption [33]. These variations can be used, in particular, for designing systems with controlled reflection, transmission, and absorption.

Consider the second case where a layer between semi-infinite media is periodically inhomogeneous with $\varepsilon(z) = \varepsilon(1 + c\sin^2 bz)$, where $b = 2\pi/\sigma$ and σ and c are the period and amplitude of modulation, respectively. Figure 4a shows the wavelength dependence of the reflection coefficient R; Fig. 4b demonstrates the absorption Q as a function of the wavelength; and Figs. 4c and 4d show the dependences of the reflection coefficient R on the angle of incidence α and parameter a, respectively. The solid and dashed lines correspond, respectively to the s-and p-polarized waves. It is seen from Figs. 4a–4c that there exist finite ranges of the wavelength and angle of incidence where the reflection coefficient equals unity (the so-called forbidden band, where $|R|^2 \approx 1$). This is a manifestation of the diffraction of the electromagnetic wave by the periodic structure of the medium similarly to the Bragg reflection of X rays from crystal planes. The dependence of the reflection coefficient on the system's parameters for the s wave differs from that for the p wave. Namely, in contrast to the s wave, the p wave does not reflect in the case of incidence at the Brewster angle. Our calculations show that this difference is observed at any width of the layer. In addition, the width and position of the Bragg reflection range for the s wave differ from those for the p wave. Oscillations accompanying the decrease in the reflection coefficient outside the Bragg reflection range are noteworthy. The oscillations are modulated and are related both to the diffraction of light in the finite volume and to multiple reflections from the dielectric interfaces. If the absolute value of the angle of incidence exceeds 50°, we observe total internal reflection. It is seen from Fig. 4b that the wavelength dependence of Q exhibits a minimum in the forbidden band, where the absorption is suppressed (Bormann effect). Outside the forbidden band, Q oscillates: anomalously strong absorption changes to anomalously weak absorption and vice versa. In the case of nonuniform periodic absorption, the mechanism of this phenomenon is the same as behind the Bormann effect. For uniform absorption (in this case, Q also oscillates outside the Bragg reflection range), this phenomenon is related to the variation in the group velocity of light, which is a newly discovered mechanism of anomalous absorption [32]. Figure 4d and the results of numerical analysis show the modulated oscillation of the a dependence of the reflection coefficient R. If the wavelength of incident light falls into the range of diffraction reflection, the oscillation amplitude is small (diffraction reflection suppresses the reflection coefficient oscillations).

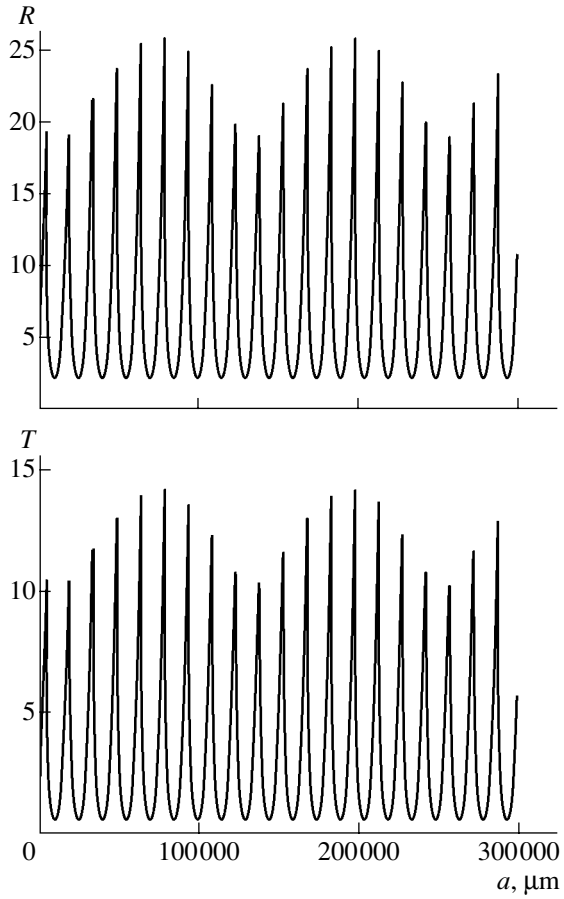

Fig. 5. Reflection, *R*, and transmission, *T*, coefficients versus parameter a for $\Delta = 3000$ μm and $d = 300\,000$ μm. The remaining parameters are the same as in Fig. 2.

Consider another practically important example. Let the layer bounded by two semi-infinite media be an amplifying medium (such a system is equivalent to a resonator with an active element inside). In this case, the parameter n'' entering into expression (48) is negative. In the approximation of linear optics, the relationships obtained in this work describe both amplification and lasing. Figure 5 shows the dependences of the reflection, *R*, and transmission, *T*, coefficients on the parameter a. We consider normal incidence. The calculations were made for the wavelength of incident ruby laser radiation ($\lambda = 0.6943$ μm) and the parameters of the medium $n' = 1.763$ and $n'' = 10^{-4}$ (the parameters of ruby at the given wavelength). Narrow peaks in Fig. 5 correspond to the values of the parameter a satisfying the phase conditions for lasing [33]. The oscillations observed are modulated. The highest peak corresponds to the value of a at which both phase and amplitude conditions for lasing are satisfied. Thus, the lasing conditions can be controlled by varying the parameter a.

Note that the problem of transmission through a resonator whose active element has a constant gain $\chi = 4\pi n/\lambda$ is not adequate for the real process. Actually, the gain decreases with increasing intensity of radiation propagating in a lasing medium, which is related to the creation of population inversion. If the energy accumulated in the active element of a laser is very high, the frequency of stimulated transitions exceeds the pumping frequency. The difference in the populations between the ground and excited states decreases sharply, the gain drops, and, hence, the lasing intensity saturates [33]. The interaction of the radiation with the amplifying medium becomes nonstationary and nonlinear. Therefore, the linear approximation fails.

ACKNOWLEDGMENTS

We are grateful to D.M. Sedrakyan for helpful remarks and discussion.

REFERENCES

1. M. Born and E. Wolf, *Principles of Optics* (Pergamon, Oxford, 1969; Nauka, Moscow, 1970).
2. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media* (Nauka, Moscow, 1982; Pergamon, New York, 1984).
3. I. M. Lifshits, S. A. Gredeskul, and L. A. Pastur, *Introduction to the Theory of Disordered Systems* (Nauka, Moscow, 1982).
4. N. A. Khizhnyak, *Integral Equations of Macroscopic Electrodynamics* (Naukova Dumka, Kiev, 1988).
5. V. I. Klyatskin, *Embedding Methods in the Theory of Wave Propagation* (Nauka, Moscow, 1986).
6. L. M. Brekhovskikh and O. A. Godin, *Acoustics of Layered Media* (Nauka, Moscow, 1989; Springer-Verlag, New York, 1990).
7. V. Freilkher, M. Pustilnik, and I. Yurkevich, Phys. Rev. B **56**, 5974 (1997).
8. J. M. Elson and P. Tran, J. Opt. Soc. Am. A **12**, 1765 (1995).
9. S. Strehlke, D. Satri, *et al.*, Thin Solid Films **297**, 291 (1997).
10. A. Karlsson and R. Stewart, J. Opt. Soc. Am. A **12**, 1513 (1995).
11. Y. Sah and G. S. Ranganath, Opt. Commun. **114**, 18 (1995).
12. N. P. K. Cotter, T. W. Preist, and J. R. Sambles, J. Opt. Soc. Am. A **12**, 1097 (1995).
13. M. Schubert, Phys. Rev. B **53**, 4265 (1996).
14. A. Yariv and P. Yeh, *Optical Waves in Crystals: Propagation and Control of Laser Radiation* (Wiley, New York, 1984; Mir, Moscow, 1987).
15. J. D. Joannopoulos, R. D. Meade, and J. N. Winn, *Photonic Crystals* (Princeton Univ. Press, Princeton, 1995).
16. M. D. Tocci, M. J. Bloemer, M. M. Scolara, *et al.*, Appl. Phys. Lett. **66**, 2324 (1995).
17. A. J. Ward, J. B. Pendry, and W. J. Stewart, J. Phys.: Condens. Matter **7**, 2217 (1995).
18. M. J. Bloemer and M. Scalora, Opt. Spektrosk. **87**, 595 (1999) [Opt. Spectrosc. **87** (4), 545 (1999)].
19. E. F. Schubert, N. E. J. Hunt, A. M. Vredenberg, *et al.*, Appl. Phys. Lett. **63**, 2603 (1993).

20. M. Scalora, R. J. Plynn, S. B. Reinhardt, *et al.*, Phys. Rev. E **54**, 1078R (1996).
21. E. R. Brown and O. B. McMahon, Appl. Phys. Lett. **68**, 1300 (1996).
22. M. Scalora, J. P. Dowling, C. M. Bowden, *et al.*, Phys. Rev. Lett. **73**, 1368 (1994).
23. M. Tocci, M. Scalora, M. J. Bloemer, *et al.*, Phys. Rev. A **53**, 2799 (1996).
24. M. Tocci, M. J. Bloemer, M. Scalora, *et al.*, Appl. Phys. Lett. A **66**, 2324 (1995).
25. N. J. Tao, S. Boussaad, W. L. Huang, *et al.*, Rev. Sci. Instrum. **70**, 4656 (1999).
26. E. Fontana, R. H. Pantell, and S. Strober, Appl. Opt. **29**, 4694 (1990).
27. P. Erdos and R. C. Herndon, Adv. Phys. **31**, 65 (1982).
28. M. Ya. Azbel, Phys. Rev. B **28**, 4106 (1983).
29. A. Zh. Khachatrian, D. M. Sedrakian, G. M. Andreasyan, and Yu. N. Ayrapetyan, Journal of Contemp. Phys. (Armenian Acad. Sci.) **36**, 118 (2001).
30. D. M. Sedrakian, A. H. Gevorgyan, and A. Zh. Khachatrian, Opt. Comm. **192**, 135 (2001).
31. D. M. Sedrakian and A. Zh. Khachatrian, Ann. Phys. (Leipzig) **11**, 3 (2002).
32. A. H. Gevorgyan, Mol. Cryst. Liq. Cryst. **378**, 129 (2002).
33. O. Zvelto, *Principles of Lasers* (Plenum, New York, 1989, 3rd ed.; Mir, Moscow, 1990).

Translated by A. Chikishev

Technical Physics, Vol. 48, No. 4, 2003, pp. 469–478. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 4, 2003, pp. 90–99.

OPTICS, QUANTUM ELECTRONICS

Effect of Photoexcitation on the Electrical Performance of ZnS : Mn Thin-Film Electroluminescent Structures

N. T. Gurin, A. V. Shlyapin, O. Yu. Sabitov, and D. V. Ryabov

Ul'yanovsk State University, Ul'yanovsk, 432700 Russia

e-mail: soy@sv.uven.ru

Received July 29, 2002

Abstract—It is shown that the kinetics of the charge and current passing through a thin-film electroluminescent emitter, as well as the *I–V* characteristics of the emitter, greatly diverge under blue, red, and IR pulsed illumination with photon energies of ≈2.6, ≈1.9, and ≈1.3 eV, respectively, and a photon flux density of 4×10^{14}–3×10^{15} mm^{-2} s^{-1}. Results obtained indicate that, during the operation of the emitter, deep centers associated presumably with V_{Zn}^{2-} zinc vacancies and V_S^+ and V_S^{2+} sulfur vacancies exchange charge. These centers lie above the valence band by ≈1.1, ≤1.9, and ≤1.3 eV, respectively. Their concentrations are estimated as (3–4) $\times 10^{16}$ cm^{-3} for V_{Zn}^{2-} and V_S^+ and $\approx 1.5 \times 10^{16}$ cm^{-3} for V_S^{2+}. It is demonstrated that positive and negative space charges forming in the near-anode and near-cathode regions of the phosphor layer specify the electric performance of the emitters. © *2003 MAIK "Nauka/Interperiodica".*

Studies of the photoelectric properties of ZnS : Mn-based thin-film electroluminescent emitters (TFELEs), as well as the behavior of their blue EL band, indicate that deep centers associated with zinc vacancies V_{Zn} that lie 2.7–1.8 eV below the conduction band bottom play a key role in EL initiation [1–4]. At the same time, a number of TFELE properties cannot be explained without considering the effect of deep centers due to sulfur vacancies V_S [4–8].

The aim of this work is to study the effect of excitation by light from different spectral ranges on the electrical performance of TFELEs in the active operating regime to clarify the role of deep centers in EL initiation and refine their energy position in the ZnS : Mn forbidden gap.

The object of our experiments was metal–insulator–semiconductor–insulator–metal (MISIM) TFELEs, where M is a lower 0.2-μm-thick transparent SnO_2 electrode applied on a glass substrate and an upper 0.15 μm-thick opaque TF Al electrode of diameter 1.5 mm; S, a Mn-doped (0.5 wt %) 0.48-μm-thick ZnS : Mn phosphor layer; and I, a 0.17-μm-thick Y_2O_3-stabilized (13 wt %) ZrO_2 insulating layer. The phosphor layer was thermally evaporated *in vacuo* in a quasi-static volume at a substrate temperature of 250°C with subsequent annealing at 250°C for 1 h, the opaque electrode was also applied with thermal evaporation, and the insulating layers were applied by electron-beam evaporation.

The time variation of the current through the TFELE $I_e(t)$ was recorded upon the excitation of the device by an alternating-sign triangular voltage $V(t)$ from a G6-34 voltage generator equipped with a driver amplifier and G5-89 external trigger generator. The maximal amplitude of pulses was V_m = 160 V at a voltage nonlinearity factor of no more than 2%. We used the continuous mode of excitation with frequencies of 20 and 50 Hz and the pulsed mode. In the latter case, a train of pulses with a duration equal to two periods of the triangular voltage and a repetition rate f = 4, 20, and 50 Hz was used with the positive and negative voltage half-waves applied to the upper electrode in the first half-period (variants (+Al) and (–Al), respectively). The pulse-repetition interval of the train was T = 0.2, 2, and 100 s. The current I_e was measured with a 0.1- to 10 kΩ-resistor series-connected to the TFELE. The voltage drop across the resistor was no higher than 0.5% of V_m. The instantaneous values of the TFELE brightness were measured with an FÉU-84-3 photoelectric multiplier. The dependences $V(t)$ and $I_e(t)$ were recorded with an S9-16 two-channel storage oscilloscope linked to a computer via an interface. For either channel, 2048 points with a given period of discretization and 256 levels of amplitude quantization were measured and stored. Mathematical and graphical processing was accomplished with the Maple V Release4 Version 4.00b and GRAPHER Version 1.06 2-D Graphing System application packages. The time variations of the mean field in the phosphor $F_p(t)$, as well as that of the current, $I_p(t)$, and charge, $Q_p(t)$, passing through the phosphor layer in the active mode of operation, were calculated by the method described in [6, 7] for the insulator capacitance C_i = 730 pF and phosphor capacitance C_p = 275 pF. These values were determined with an E7-14

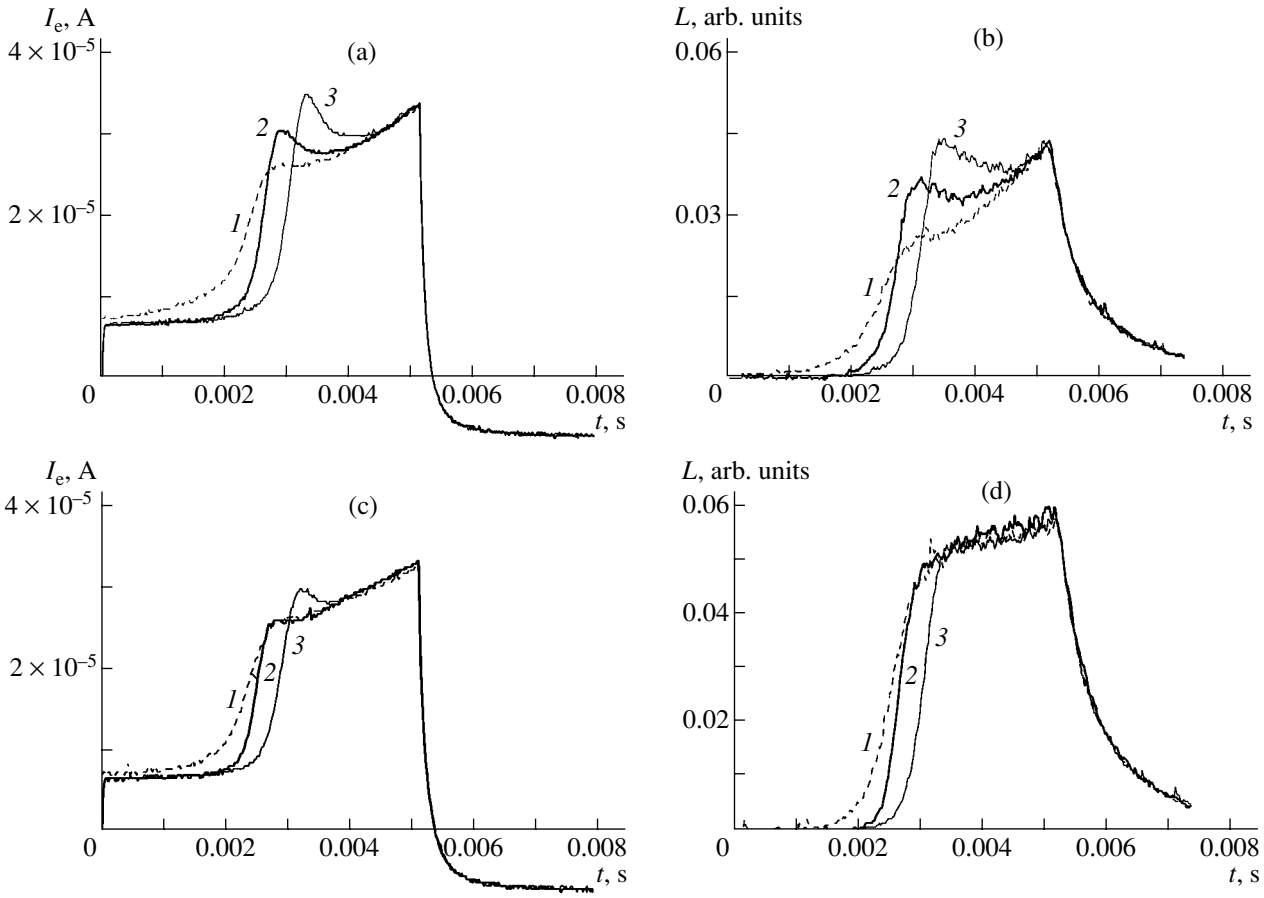

Fig. 1. (a, c) $I_e(t)$ and (b, d) $L(t)$ at f = 50 Hz. (*1*) Continuous excitation regime, (*2*) excitation by a single pulse with T = 2 s, and (*3*) blue illumination pulse applied in the interval between exciting voltage pulses with T = 2 s is 500 ms long. (a, b) Variant (–Al) and (c, d) variant (+Al).

immitance meter for known geometrical parameters of the device. The device was illuminated (excited) on the substrate side under continuous and pulsed conditions. In the blue range, it was illuminated with an E1L51-3B light-emitting diode with the following radiation parameters: the maximal amplitude wavelength λ_m = 475 nm, the half-height width of the radiation spectrum $\Delta\lambda_{0.5} \approx 35$ nm, luminous intensity ≈1 cd, the radiation power ≈5 mW, and the photon flux density $\Phi \approx 1.6 \times 10^{15}$ mm^{-2} s^{-1}. In the red range, the device was illuminated with a semiconductor laser with λ_m = 656 nm, $\Delta\lambda_{0.5}$ = 15 nm, $P \approx 1$ mW, and $\Phi \approx 4 \times 10^{14}$ mm^{-2} s^{-1}. In the IR range, the device was illuminated by two AL107A diodes with λ_m = 950 nm, $\Delta\lambda_{0.5}$ = 25 nm, total power $P \approx 12$ mW, and total flux density $\Phi \approx 3 \times 10^{15}$ mm^{-2} s^{-1}.

Pulsed photoexcitation was accomplished in two regimes: (1) a light pulse is applied during the action of the train of triangular voltage pulses (the duration of the pulses equals the doubled period of the exciting voltage) and (2) a pulse is applied between the trains immediately after the last pulse with the radiation pulse width being equal to the pause between the trains.

The basic results of our investigation are as follows.

The effect of the illumination in the pause between the trains on the shapes of the current pulse $I_e(t)$ and brightness pulse $L(t)$ is similar to that of the pause without the illumination (Fig. 1). This is because the charge state of the deep centers in the phosphor layer changes equally during the pause both with and without [6, 7] the illumination.

The curves $I_e(t)$, $L(t)$, $I_p(t)$, and $I_p(F_p)$ for the variants (–Al) and (+Al) are asymmetric (Figs. 1, 2). As in [6, 7], this is explained by the nonuniform distribution of structural defects and Mn^{2+} ions across the phosphor layer, with the Mn^{2+} ion concentration growing toward the upper (Al) electrode [6, 7].

As compared with the no-illumination conditions in the pause, additional effects due to the pulsed illumination between the trains are the following. (i) For blue illumination, the current in the initial portion of the curve $I_p(t)$ grows significantly up to the point *r,* which is a demarcation line between the "fast" and "slow" rise of the curves $I_e(t)$. After this point, the rate of rise for the curves decreases [6, 7]. Simultaneously, for the variant

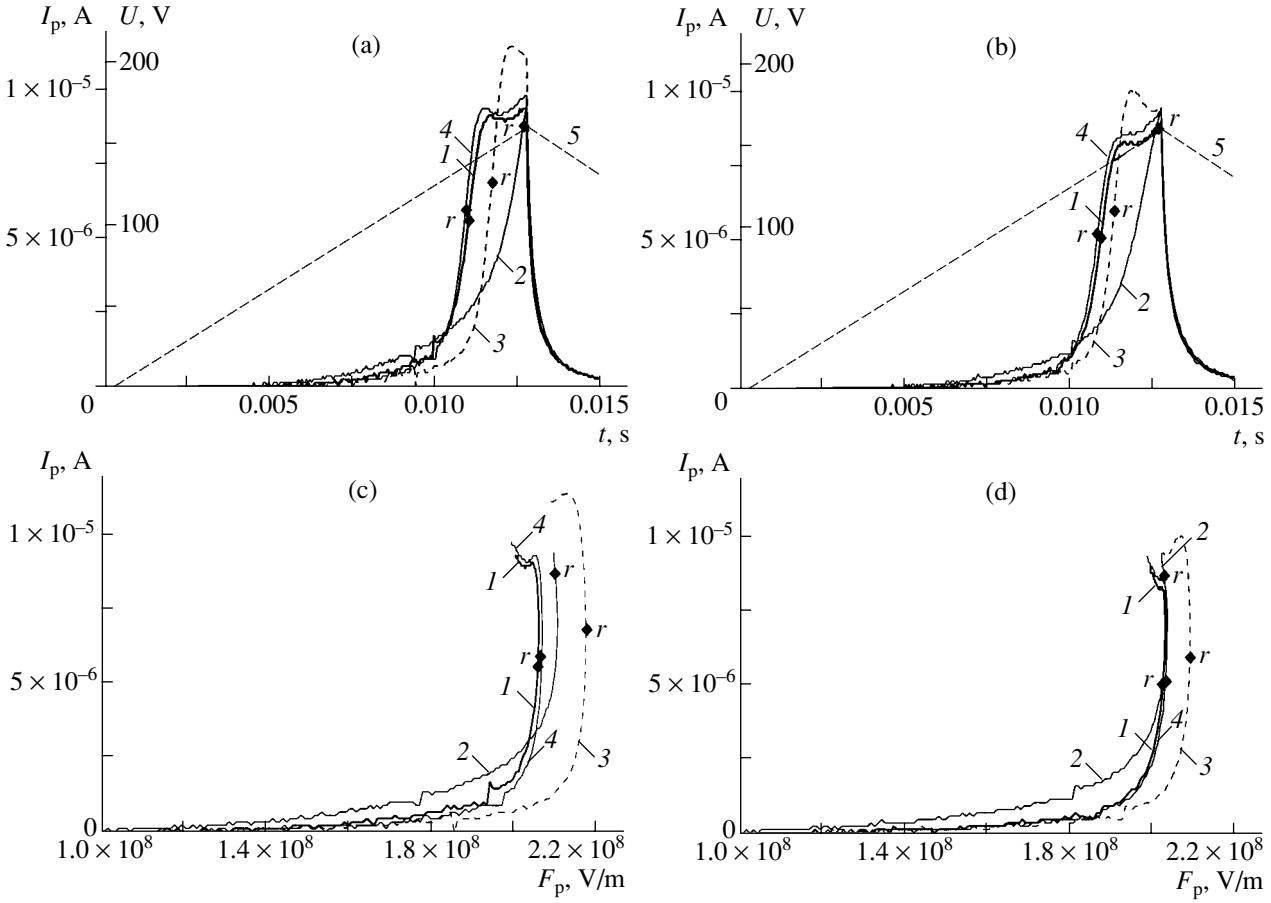

Fig. 2. (a, b) $I_p(t)$ and (c, d) $I_p(F_p)$ under pulsed illumination of the device in the interval between voltage pulses: (a, c) variant (–Al) and (b, d) variant (+Al). (*1*) Without illumination, (*2*) blue illumination, (*3*) red illumination, (*4*) IR illumination, and (*5*) $V(t)$ ($V(t) = U$). $f = 20$ Hz, $T = 100$ s.

(–Al), the slow portion shrinks (Figs. 2a, 2b) and the mean field $F_p(t)$ in the phosphor grows (Figs. 2c, 2d, 3b, 3d). (ii) For red illumination, the current in the initial (fast) portion of $I_p(t)$ decreases, while in the slow portion, the amplitude of the pulse $I_p(t)$ increases (Figs. 2a, 2b). The field $F_p(t)$ grows in both portions (Figs. 2c, 2d, 3b, 3d), the growth being more noticeable for the (–Al) variant. (iii) In the case of IR illumination, the current $I_p(t)$ early in the fast portion slightly declines. Then, the rate of rise of the current $I_p(t)$ grows and its amplitude grows insignificantly (Figs. 2a, 2b). The field $F_p(t)$ grows only slightly in both (±Al) variants (Figs. 2c, 2d, 3b, 3d).

The above variations of the current $I_p(t)$ for the blue and red illuminations become much less pronounced in the second half-period. In the third half-period, they diminish still more and finally become comparable to the measurement accuracy in the fourth half-period (Fig. 4).

For pulsed illumination during the action of the voltage pulses, the variations of the $I_p(t)$ curves are much weaker compared with the no-illumination case. Slight variations are observed only in the portion where $I_p(t)$ grows. Here, the $I_p(t)$ amplitude varies within the measurement accuracy (Fig. 5). Blue illumination also causes a significant rise in the current $I_p(t)$ early in the fast portion, and red illumination also results in a weak growth of $I_p(t)$ compared with the no-illumination case for the (+Al) variant and does not change $I_p(t)$ for the (–Al) variant within the measurement accuracy. For IR illumination, the rate of rise of $I_p(t)$ in the ascending portion is higher than in the no-illumination case (as in the cases with and without illumination in the pause between the voltage pulses; Figs. 2a, 2b), this effect being more significant for the (–Al) variant.

For continuous illumination (excitation) of the TFELE, the effect of the illumination for all the exciting voltage frequencies and spectral ranges mentioned above is absent (all the variations of $I_p(t)$ are within the measurement accuracy).

The results obtained can be explained as follows. The process of ZnS : Mn layer fabrication introduces various structural defects, including those that are the most plausible in terms of thermodynamics, namely,

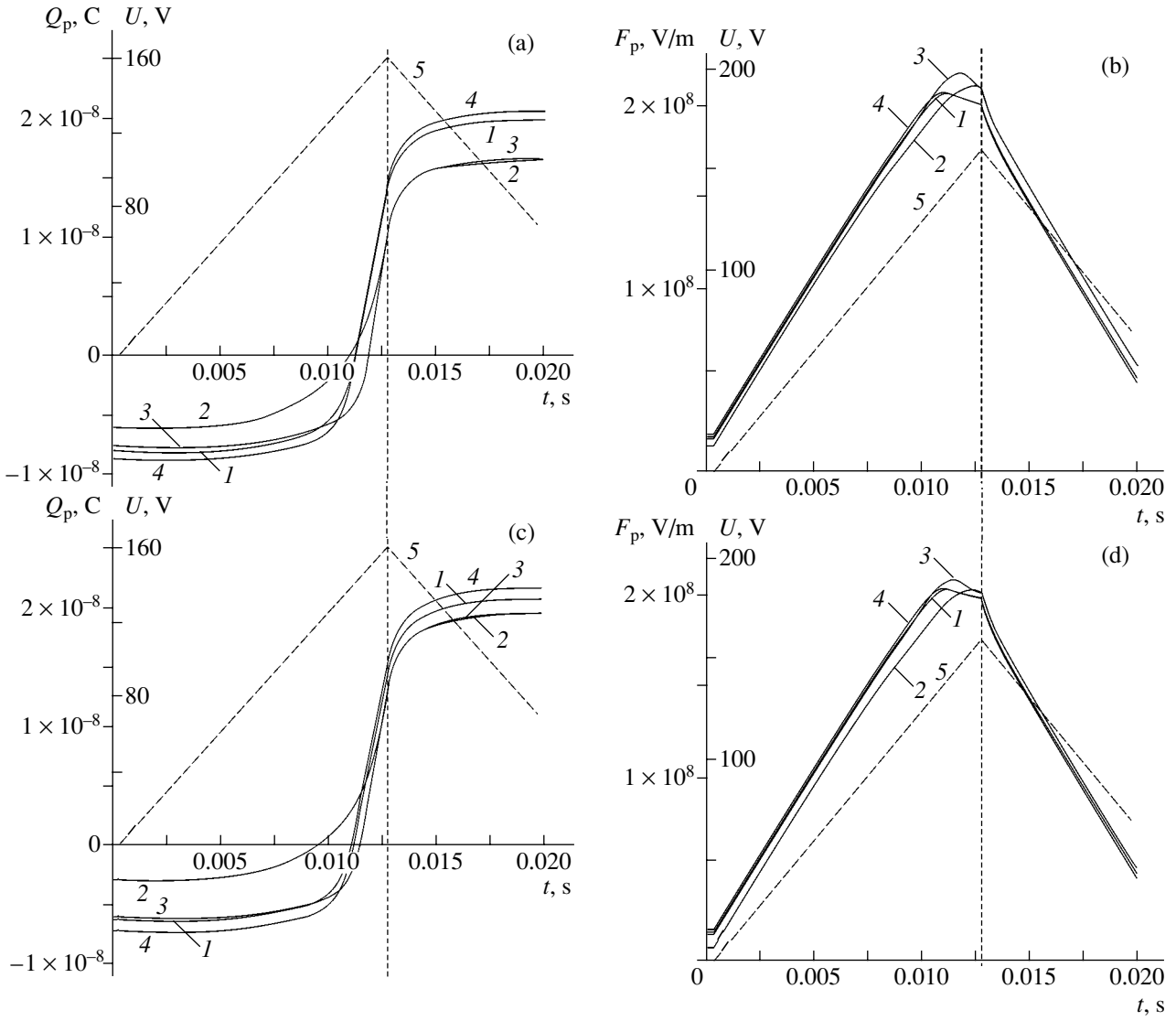

Fig. 3. (a, c) $Q_p(t)$ and (b, d) $F_p(t)$ under pulsed illumination of the device in the interval between voltage pulses: (a, b) variant (–Al) and (c, d) variant (+Al); (*1*)–(*5*) mean the same as in Fig. 2.

zinc and sulfur vacancies [9]. The sulfur concentration must be high because of doping by manganese and increase from the lower to upper electrode according to the growth of the manganese concentration. For the phosphor layer to remain electrically neutral, it must also contain a sufficient concentration of zinc vacancies. Their concentration, conversely, is the highest at the lower insulator–phosphor interface, at least because of a higher sulfur vapor pressure at the beginning of phosphor layer application and possibly because of the presence of defects like singly charged or neutral zinc interstitials (Zn_i^+, Zn_i^0). Since the concentration of Mn is high (0.5 wt%), it can not only substitute zinc at lattice sites in the form of Mn^{2+} ions but also produce Mn interstitials Mn_i. The energy positions of deep centers due to these defects are as follows: for singly charged zinc vacancies V_{Zn}^-, 0.5–0.6 eV above the valence band top [10, 11]; for singly charged sulfur vacancies V_S^+, 0.6–2.0 eV below the conduction band bottom [10, 12]; for doubly charged zinc vacancies V_{Zn}^{2-}, 1.0–1.1 eV above the valence band top [1–4, 11]; for doubly charged sulfur vacancies V_S^{2+}, 1.05–1.3 eV above the valence band top [9, 13]; for neutral sulfur vacancies V_S^0, 0.2–1.05 eV below the conduction band bottom [10, 12]; and for zinc interstitials, 0.2 and 0.1–0.12 eV below the conduction band bottom for Zn_i^+ and Zn_i^0, respectively [12, 13]. Centers due to Mn_i defects apparently lie near the valence band [11]. In addition, centers related to other lattice defects, including defect com-

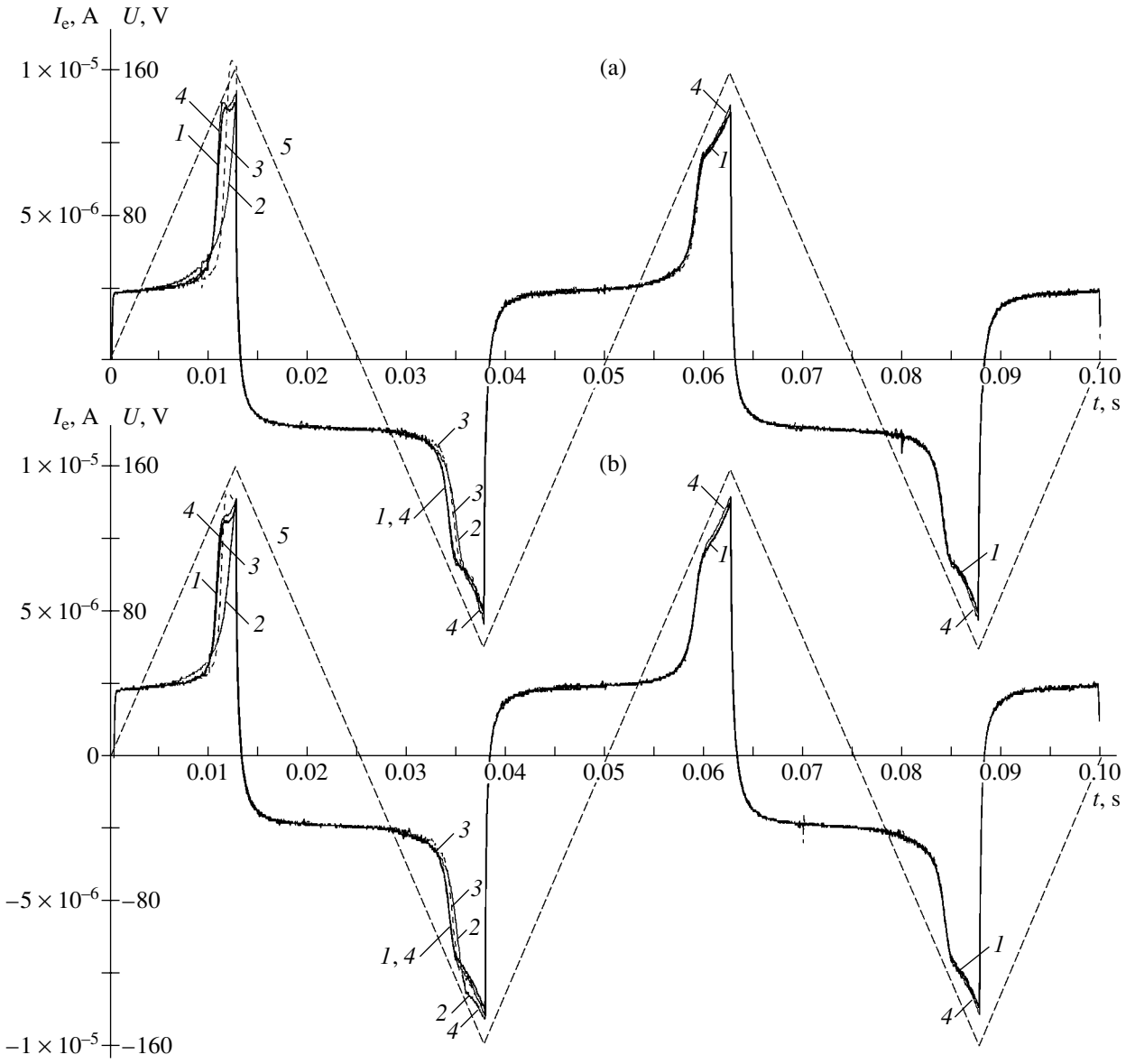

Fig. 4. $I_e(t)$ for two periods of the voltage pulse $V(t)$ under pulsed illumination of the device in the interval between voltage pulse trains. (a) Variant (–Al) and (b) variant (+Al). For (*1*)–(*5*), see Fig. 2.

plexes, may appear under nonequilibrium conditions [8–15].

Since the equilibrium Fermi level in ZnS is slightly above the midgap (hence, the low electron conductivity of the material), deep centers associated with singly charged V_S^+ vacancies (which appear to be responsible for a deep center near the midgap with an energy exceeding the energy of a V_S^{2+} center) and doubly charged zinc vacancies V_{Zn}^{2-} (which produce a deeper center than V_{Zn}^-) seem to be the most plausible under equilibrium conditions. The phosphor structure is polycrystalline; therefore, levels corresponding to these centers are not discrete but have some energy distribution in the energy gap.

When the applied voltage exceeds a threshold value in the active operating mode of the TFELE, the tunnel emission of electrons from surface states at the near-cathode insulator–phosphor interface takes place. The electrons are ballistically accelerated, causing the impact ionization of shallow levels (which, in particular, are due to Zn_i defects); Mn^{2+} luminescence centers (the energy of excitation of these centers is 2.4–2.5 eV); and deep centers, including those associated with zinc and sulfur vacancies [2–4, 6, 7]. The first excitation of the device to the active mode is, however, accomplished at a higher excitation voltage [4, 6]. As a result of the impact ionization of deep centers, whose concentration

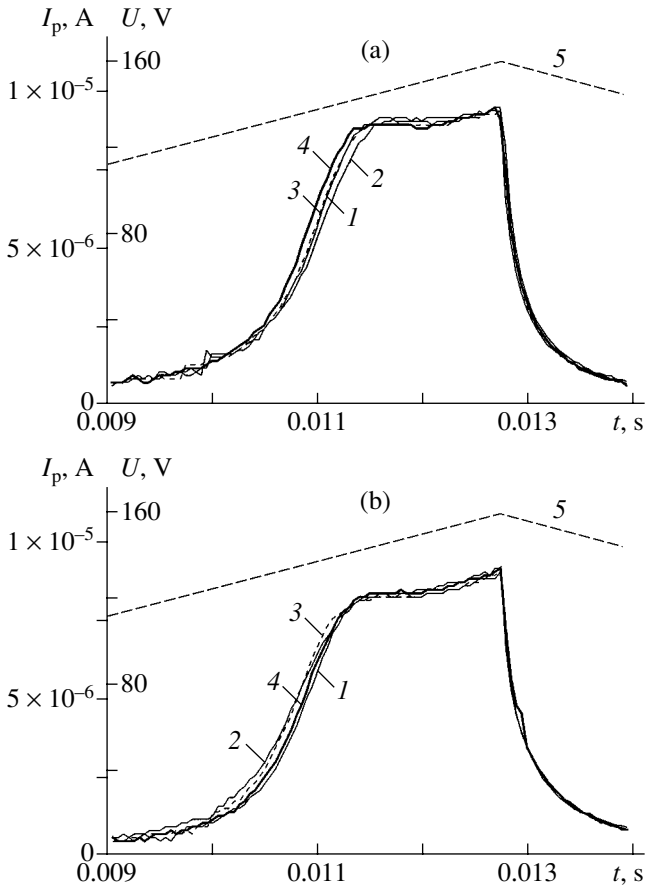

Fig. 5. $I_p(t)$ under pulsed illumination of the device during the action of exciting voltage pulses. (a) Variant (–Al) and (b) variant (+Al). For (*1*)–(*5*), see Fig. 2.

is maximal in the phosphor layer, a positive space charge (PSC) forms in the near-anode region of the phosphor possibly via the reactions

$$V_{Zn}^{2-} \longrightarrow V_{Zn} + e, \quad (1)$$

$$V_S^+ \longrightarrow V_S^{2+} + e, \quad (2)$$

$$V_{Zn}^- \longrightarrow V_{Zn}^0 + e. \quad (3)$$

It is likely that the ionization of V_{Zn}^- levels proceeds at close-to-maximal excitation voltages, since these levels lie near the valence band top. To these voltages, there corresponds that portion of the *I–V* characteristic where $I_p(t)$ rises after the first *S*-shaped region and then falls (Figs. 2c, 2d).

Simultaneously, in the near-cathode region of the phosphor, electrons are captured by V_S^{2+} and V_S^+ deep centers, which have a large capture cross section, and a negative space charge (NSC) forms [8]. Specifically, the latter effect arises as a result of tunneling from surface states:

$$V_S^+ + e \longrightarrow V_S^0, \quad (4)$$

$$V_S^{2+} + e \longrightarrow V_S^+. \quad (5)$$

Over the time interval the device is inactive, the space charges are neutralized by capturing free electrons and holes by the deep centers. However, the relaxation time of these charges is relatively large (from several seconds to several tens of seconds [6]). Consequently, under continuous excitation with a frequency $f \geq 1$ Hz, the space charges do not have time to be neutralized. In the next exciting voltage half-period, the voltage polarity on the electrodes reverses. Accordingly, the PSC in the near-cathode region facilitates the tunnel emission of electrons and the NSC in the near-anode region provides the impact ionization of Mn^{2+} luminescence centers and deep centers associated with zinc and sulfur vacancies at a lesser mean field $F_p(t)$ in the phosphor layer. Eventually, the threshold voltage and the luminescence field of the TFELE decrease compared with the first switching of the device (Fig. 1, curve *1*) [4, 6, 7]. When sufficiently long exciting voltage pulses, e.g., with $T = 100$ s, are separated by a pause of duration $T - 2/f = 100$ s – 100 ms = 99.9 s, the space charges are neutralized by the following scheme:

in the former near-anode region, the PSC is neutralized by capturing free electrons (or tunneling) or their thermal emission from the valence band or surface states:

$$V_{Zn}^0 \longrightarrow V_{Zn}^- + e, \quad (6)$$

$$V_{Zn}^- + e \longrightarrow V_{Zn}^{2-}, \quad (7)$$

$$V_S^{2+} + e \longrightarrow V_S^+; \quad (8)$$

in the former near-cathode region, the NSC is neutralized by tunneling or thermal excitation of the captured electrons into the conduction band by the reaction

$$V_S^0 \longrightarrow V_S^+ + e \quad (9)$$

and reaction (2).

It appears that reaction (6) proceeds much faster than reactions (7) and (8), because the level due to V_{Zn}^0 is located near the valence band top.

Blue (photon energy $h\nu \approx 2.6$ eV) illumination of the device in the interval between the exciting voltage pulses suppresses the PSC neutralization in the former near-anode regions according to reactions (7) and (8) because of the photoionization of V_{Zn}^{2-} and V_S^+ by reactions (1) and (2) and leads to the formation of the PSC in the former near-cathode region via the photoionization of V_{Zn}^{2-}, V_S^+, and V_S^0 by reactions (1), (2), and (9). These effects in combination increase the tunnel emission current, along with the current $I_p(t)$ in the initial

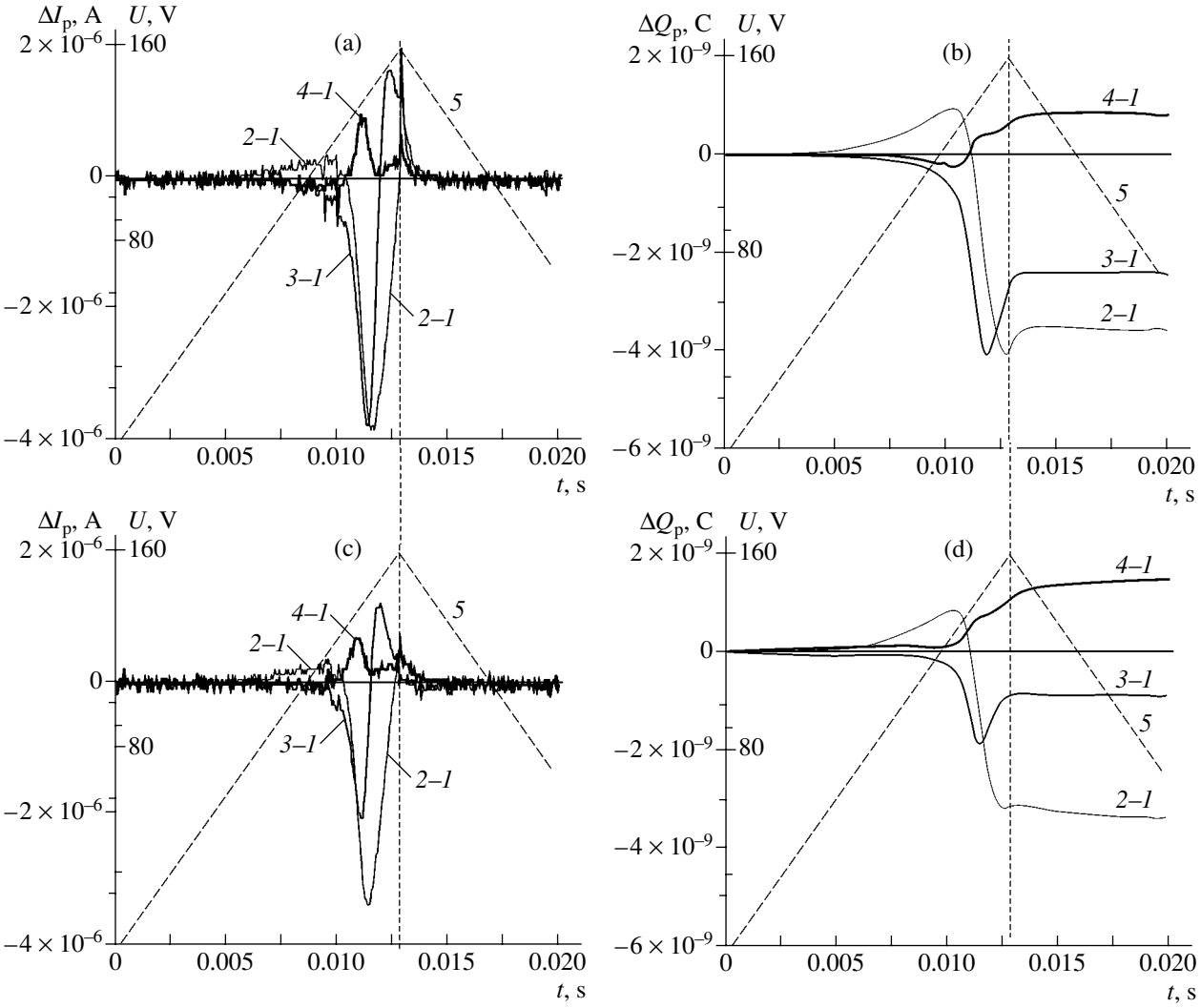

Fig. 6. (a, c) $\Delta I_p(t)$ and (b, d) $\Delta Q_p(t)$. (*2–1*) Blue illumination, (*3–1*) red illumination, (*4–1*) IR illumination, and (*5*) $V(t)$. f = 20 Hz, T = 100 s. (a, b) Variant (–Al) and (c, d) variant (+Al).

portion of the dependence $I_p(t)$ (Figs. 2a, 2b, 6a, 6c), and almost eliminate the portion where $I_p(t)$ slowly rises, which is apparently associated with the ionization of V_{Zn}^{2-} and V_S^+ centers because they are almost entirely depleted. In the initial portion of the $I_p(t)$ rise, the difference $\Delta I_p(t)$ between the currents $I_p(t)$ with and without illumination, as well as the related difference in the charge transferred through the phosphor layer

$$\Delta Q_p(t) = \int_0^t \Delta I_p(t)dt,$$

is positive (Fig. 6). As the voltage $V(t)$ increases, the values of $\Delta I_p(t)$ and $\Delta Q_p(t)$ become negative. In this case, $|\Delta I_p(t)|$, $|\Delta Q_p(t)|$, and the field $F_p(t)$ (Figs. 2c, 2d, 3b, 3d, 6a–6d) are greater in the variant (–Al) possibly because of the greater total concentration of zinc and sulfur vacancies in the phosphor layer at the lower insulator–phosphor interface.

Red ($h\nu \approx 1.9$ eV) illumination in the interval between exciting voltage pulses suppresses NSC neutralization in the former near-cathode region by reaction (9) and favors PSC neutralization and NSC formation in the former near-anode region, since illumination-excited electrons from the valence band are captured by singly charged V_S^+ sulfur vacancies by reaction (4). Because of this, the tunnel emission current, as well as the current $I_p(t)$ and the charge $Q_p(t)$ in the initial portion of the $I_p(t)$ rise, decreases (Figs. 2c, 2d, 3a, 3c) compared with the no-illumination case. Simultaneously, the field $F_p(t)$ grows with increasing $V(t)$ (Figs. 2c, 2d, 3b, 3d). Since the concentration of sulfur vacancies in the phosphor layer is greater at the upper phosphor–insulator interface, the values of $|\Delta I_p(t)|$ and $|\Delta Q_p(t)|$, as well as the rise in the field $F_p(t)$,

are larger for the variant (–Al). At the same time, in the portion of steep rise for this variant, $I_p(t)$ is higher precisely because of higher $F_p(t)$, which favors the impact ionization (in the near-anode region) of the remaining V_{Zn}^{2-} centers and also other centers located in the ZnS energy gap farther from the conduction band bottom.

IR ($h\nu \approx 1.3$ eV) illumination in the interval between exciting voltage pulses contributes to PSC neutralization in the former near-anode region by exciting electrons from the valence band according to reaction (8) and accelerates NSC neutralization in the former near-cathode region by reaction (9). The additional PSC neutralization causes a small decrease in the tunnel emission current, the current $I_p(t)$ (Figs. 2a, 2b, 6a, 6c), and the charge $Q_p(t)$ (Figs. 3a, 5a, 5b) and increases $F_p(t)$ (Figs. 2c, 3b) in the initial portion of $I_p(t)$ growth compared with their values without IR illumination. This is observed in the variant (–Al) alone and is explained by the higher concentration of sulfur vacancies at the upper phosphor–insulator interface. As the voltage $V(t)$ and the field $F_p(t)$ increase, the current $I_p(t)$ and the charge $Q_p(t)$ grow in comparison with their values without IR illumination of the device (Fig. 6). This effect is associated with the impact ionization of V_S^+ formed by the additional IR neutralization (by reaction (8)) of V_S^{2+} sulfur vacancies in the bulk and near-anode region of the phosphor. In the (+Al) variant, the concentration of sulfur vacancies in the phosphor layer is smaller than that of zinc vacancies. Moreover, at the lower insulator–phosphor interface, the concentration of sulfur vacancies is smaller than at the upper interface. Therefore, the additional neutralization of sulfur vacancies affects the initial portions of $I_p(t)$, $Q_p(t)$, and $F_p(t)$ growth insignificantly (Figs. 2d, 3c, 3d). As $V(t)$ rises, the impact ionization of additionally neutralized V_S^+ sulfur vacancies that are present in the near-anode region causes $I_p(t)$ and $Q_p(t)$ to grow (Fig. 6), the growth being somewhat faster than in the variant (–Al) because of the higher concentration of sulfur vacancies at the upper phosphor–insulator interface. The fact that the increments of the $I_e(t)$ and, hence, $I_p(t)$ amplitudes remain the same in the second to the fourth half-periods of the voltage $V(t)$ (Fig. 4) can presumably be explained as follows. For V_S^{2+} centers that are additionally neutralized by the IR illumination to the charge state V_S^+, the time of relaxation into the equilibrium state is the longest compared with the relaxation times for the other centers, because V_S^{2+} centers are the deepest in the ZnS energy gap. This relaxation time far exceeds the time interval between successive TFELE switchings during two exciting voltage periods. As a result, the concentration of the additional V_S^+ centers remains elevated and practically unchanged during the voltage pulse application.

As was mentioned above, under the blue and red illuminations, the current $I_p(t)$ varies to a lesser extent in the second, third, and fourth half-periods of the exciting voltage (Fig. 4). Similar effects were also observed in [1], where the device was exposed to UV radiation, and in [6], where illumination was absent. This is explained by successive charge exchange between V_{Zn}^{2-} and V_S^+ deep centers in the ZnS : Mn energy gap because of impact ionization when the device is switched on and carrier capture when the device is switched off under the condition that the equilibrium concentrations of these centers are constant. As a result, by the third or fourth exciting voltage half-period, the charge state of these centers corresponds to the state typical of the continuous mode of TFELE operation.

The much weaker variation of the current $I_p(t)$ when the pulsed illumination is imposed on the exciting voltage pulses (Fig. 5) is explained by the much shorter illumination time (with $f = 20$ Hz, this time is no more than 12.5 ms for the first voltage half-period) compared with the pause duration (99.9 s) and, accordingly, by the proportional decrease in the number of absorbed photons. Blue illumination results in the photoionization of V_{Zn}^{2-} and V_S^+ centers by reactions (1) and (2), increasing the PSC in the near-cathode region and the current $I_p(t)$ at the early stage of its growth (Fig. 5). This is particularly noticeable in the variant (+Al), presumably because of the higher net concentration of zinc and sulfur vacancies at the lower insulator–phosphor interface. Red illumination causes neutral sulfur vacancies V_S^0 to form by reaction (4). The energy position of the corresponding center is close to the conduction band bottom, and the impact ionization of this center increases the current $I_p(t)$. This increase is greater in the variant (+Al) (Fig. 5b), because, as was already mentioned, the concentration of sulfur vacancies at the upper phosphor–insulator interface is higher. IR illumination, on the one hand, generates holes in the valence band via the capture of excited electrons from the valence band by V_S^{2+} centers (reaction (8)); on the other hand, it raises the concentration of V_S^+ centers, which are a basic source of increasing $I_p(t)$ through impact ionization. Since the mobility of electrons in ZnS is ≈28 times that of holes [15], the latter effect makes a greater contribution to $I_p(t)$ (Fig. 5).

The fact that the illumination has a negligible effect on $I_p(t)$ under the continuous excitation of the device with the illumination intensities and exciting voltage frequencies used is explained as follows. As was noted, the time interval between the active modes of the device is much shorter than the relaxation time of the space charges; therefore, these charges in the near-cathode

and near-anode regions have no time to be neutralized and the number of photons absorbed during the pause and the active regime is three or four orders of magnitude smaller than when the device is excited once and the pause lasts several seconds to several tens of seconds. As a result, the effect of the illumination on the space charges formed by reactions (1)–(4) is weak and the concentration of photogenerated free carriers is small compared with that of free carriers generated by tunnel emission and impact ionization in the device's active mode.

The results obtained allow us to estimate the concentration of deep centers in the ZnS : Mn energy gap. Let it be assumed that, in the case of blue illumination, all centers associated with V_{Zn}^{2-} and V_S^+ are completely depleted during the pause between exciting voltage pulses with $T = 100$ s. Then, the maximal difference $|\Delta Q_p(t)|$ in the charges transferred through the phosphor layer in the active mode with and without the illumination that is estimated in the portion of steep rise (Figs. 6b, 6d; curves *2–1*) is $|\Delta Q_p(t)| \approx 4.0 \times 10^{-9}$ C (variant (–Al)) and $|\Delta Q_p(t)| \approx 4.0 \times 10^{-9}$ C (variant (+Al)) for a TFELE area $S = 2$ mm^2 and a PSC layer thickness close to half the thickness of the phosphor layer (≈0.2 μm). With such values, the net concentration of V_{Zn}^{2-} and V_S^+ centers is $\approx 7.7 \times 10^{16}$ cm^{-3} at the lower insulator–phosphor interface and $\approx 6.2 \times 10^{16}$ cm^{-3} at the upper interface. By order of magnitude, these concentrations equal the expected concentration of zinc and sulfur vacancies in ZnS [9, 12]. The somewhat higher concentration of V_{Zn}^{2-} and V_S^+ centers at the lower interface is due to a greater number of defects in that part of the phosphor film obtained at the early stage of ZnS : Mn growth.

It should be noted that the total charge transferred through the phosphor in the active mode without illumination up to the point the voltage $V(t)$ reaches its highest (amplitude) value amounts to $\approx 2.1 \times 10^{-8}$ C (Figs. 3a, 3c) in both (±Al) variants. Then, a fraction of the charge that is released from V_{Zn}^{2-} and V_S^+ centers is ≈0.23 and ≈0.19 (relative to the total charge) for the (–Al) and (+Al) variants, respectively.

Since red illumination generates, according to reaction (4), the NSC in the near-cathode region, which produces a field preventing the tunnel emission of electrons from surface states, the determination of the V_S^+ sulfur vacancy concentration requires the potential barrier at the insulator–phosphor interface to be known. Since the local field in the near-anode region changes with the mean field in the phosphor layer remaining unchanged, the component of $I_p(t)$ that is due to the impact ionization of V_{Zn}^{2-} centers also changes. Therefore, the fractions of $I_p(t)$ and $Q_p(t)$ that are associated with impact ionization or electron tunneling from V_S^0 centers are difficult to separate exactly.

IR illumination allows us to estimate the concentration of additionally produced V_S^+ vacancies and, hence, the equilibrium concentration of V_S^{2+}. It equals 1.5×10^{16} cm^{-3} with $\Delta Q \approx 1 \times 10^{-9}$ C and a PSC region thickness of ≈0.2 μm. It is interesting to note that the increments of the current, $\Delta I_p(t)$; charge, $\Delta Q_p(t)$; and, accordingly, concentration of V_S^+ centers in the portion where $\Delta I_p(t)$ and $\Delta Q_p(t)$ sharply grow are almost the same for both variants (±Al) (Fig. 6). This may indicate that the structural defects like V_S^{2+} are uniformly distributed across the phosphor layer.

Thus, the blue, red, and IR photoexcitations of a TFELE in the time interval between exciting voltage pulses (in the case when this interval and photon flux density are sufficiently large) confirms the existence of deep levels in the energy band of ZnS : Mn. They lie ≈1.1, ≤1.3, and ≤1.9 eV above the valence band top and can probably be related to doubly charged zinc vacancies, doubly charged sulfur vacancies, and singly charged sulfur vacancies, respectively. The net concentration of centers due to V_{Zn}^{2-} and V_S^+ vacancies varies from 7.7×10^{16} cm^{-3} at the lower insulator–phosphor interface to 6.2×10^{16} cm^{-3} at the upper phosphor–insulator interface. To provide the electroneutrality condition for the phosphor, the concentrations of V_{Zn}^{2-} and V_S^+ vacancies must be $(3–4) \times 10^{16}$ cm^{-3}. The concentration of V_S^{2+} vacancies is about 1.5×10^{16} cm^{-3}. The results of our study suggest that during the TFELE operation, accelerated electrons in the near-anode region ionize (by impact ionization) Mn^{2+} luminescence centers and also deep centers due to V_{Zn}^{2-}, V_S^+, and V_{Zn}^- vacancies, causing the formation of the PSC. In the near-cathode region, deep centers related to V_S^+ and V_S^{2+} vacancies trap free electrons, neutralizing the PSC formed in the preceding operating cycle and forming the NSC. In time intervals between the active states of the device, these space charges are neutralized; the degree of neutralization is the higher, the longer the pause. The photoexcitation of the device by "blue" photons during the pause prevents PSC neutralization in the former near-anode region, thereby increasing the field in the near-cathode region and the tunnel emission current from surface states at the insulator–phosphor interface. As a result, the current component due to the impact ionization of V_{Zn}^{2-} and V_S^+ deep centers decreases in the next operating cycle of the device.

The irradiation by "red" photons decelerates NSC neutralization in the former near-cathode region and favors PSC neutralization. Simultaneously, it generates the NSC in the former near-anode region by transferring radiation-excited electrons from the valence band to a level corresponding to V_S^+. This decreases the field in the near-cathode region and the tunnel emission current. At the same time, the mean field in the phosphor layer that is required for V_S^0 and V_{Zn}^{2-} deep centers in the near-anode region to be ionized in the next operating cycle grows. When the field reaches the value required, the current amplitude becomes larger than in the absence of the illumination.

IR irradiation causes additional V_S^+ sulfur vacancies to form via capturing free electrons that are radiation-excited from the valence band by V_S^{2+} centers. Because of this, in the next operating cycle, the space charges, the field in the near-cathode region, and the tunnel emission current decrease in the variant (–Al), where the sulfur vacancy concentration at the upper phosphor–insulator interface exceeds the concentration of zinc vacancies. As the applied field grows, so does the current amplitude because of the ionization of additional V_S^+ vacancies in the phosphor layer.

The illumination combined with exciting voltage pulses affects the current passing through the device insignificantly, since the number of photons absorbed by the phosphor layer is small (for the photon flux densities and exciting voltage frequencies used). For the same reason, the illumination of the device under the continuous excitation regime does not affect the current passing through the device.

REFERENCES

1. W. E. Howard, O. Sahni, and P. M. Alt, J. Appl. Phys. **53**, 639 (1982).
2. K. W. Yang and S. J. T. Owen, IEEE Trans. Electron Devices **ED-30**, 452 (1983).
3. A. A. Douglas, J. F. Wager, D. C. Morton, *et al.*, J. Appl. Phys. **73**, 296 (1993).
4. K. A. Neyts, D. Corlatan, P. De Visschere, *et al.*, J. Appl. Phys. **75**, 5339 (1994).
5. E. Bringuier, Philos. Mag. **75**, 209 (1997).
6. N. T. Gurin, O. Yu. Sabitov, and A. V. Shlyapin, Zh. Tekh. Fiz. **71** (8), 48 (2001) [Tech. Phys. **46**, 977 (2001)].
7. N. T. Gurin, A. V. Shlyapin, and O. Yu. Sabitov, Zh. Tekh. Fiz. **72** (2), 74 (2002) [Tech. Phys. **47**, 215 (2002)].
8. N. T. Gurin, O. Yu. Sabitov, and A. V. Shlyapin, Pis'ma Zh. Tekh. Fiz. **28** (15), 24 (2002) [Tech. Phys. Lett. **28**, 631 (2002)].
9. *Physics of II–VI Compounds*, Ed. by A. N. Georgobiani and M. K. Sheĭnkman (Nauka, Moscow, 1986).
10. F. Kreger, *Chemistry of Imperfect Crystals* (Mir, Moscow, 1969).
11. N. A. Vlasenko, M. M. Chumachkova, Z. L. Denisova, *et al.*, J. Cryst. Growth **216**, 249 (2000).
12. I. K. Morozova and V. A. Kuznetsova, *Zinc Sulfide: Production and Properties* (Nauka, Moscow, 1987).
13. A. N. Gruzintsev, Doctoral Dissertation (Chernogolovka, 1999).
14. A. N. Georgobiani, A. N. Gruzintsev, Xu Xurong, *et al.*, Neorg. Mater. **35**, 1429 (1999).
15. *Tables of Physical Data: Reference Book*, Ed. by I. K. Kikoin (Atomizdat, Moscow, 1976).

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 4, 2003, pp. 479–491. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 4, 2003, pp. 100–112.

OPTICS, QUANTUM ELECTRONICS

Quantum Yield and Luminous Efficacy of Thin-Film ZnS Electroluminescent Emitters

N. T. Gurin, A. V. Shlyapin, and O. Yu. Sabitov

Ul'yanovsk State University, Ul'yanovsk, 432700 Russia

e-mail: soy@sv.uven.ru

Received June 24, 2002

Abstract—The instantaneous values of the internal quantum yield and luminous efficacy of thin-film electroluminescent emitters are experimentally studied as functions of time, the mean field in the phosphor layer, and the charge passing through this layer. Also, the dependences of the internal quantum yield and luminous efficacy on the exciting voltage amplitude are explored. At exciting voltage frequencies above 10 Hz, the time dependences of the instantaneous quantum yield and luminous efficacy exhibit a dip in the range where the brightness and current through the phosphor layer grow and a peak in the range where the brightness and current decline. The dip and peak are related to the different rates of rise and fall of the brightness and current. © *2003 MAIK "Nauka/Interperiodica".*

The basic parameters characterizing the efficiency of thin-film electroluminescent emitters (TFELEs), internal and external quantum yields and luminous efficacy, are integral quantities. For this reason, physical processes governing their variation with various parameters of TFELE excitation (the shape, frequency, and amplitude of the applied voltage, etc.) cannot be considered in detail. Earlier, we studied the EL kinetics in such devices, including the kinetics of the instantaneous quantum yield, for exciting voltage frequencies in the range 0.1–2.0 Hz. With these frequencies, the adjacent brightness waves do not overlap and the characteristic times of EL are less than a quarter of the exciting voltage period [1]. Of practical interest are, however, exciting voltage frequencies of 50 Hz or higher.

The aim of this work is to study the variation of the instantaneous internal quantum yield and luminous efficacy of TFELEs with time, the mean field in the phosphor layer, and the charge passing through this layer. Also, we were interested in the dependences of the external and internal quantum yields and luminous efficacy on the exciting voltage amplitude in the voltage frequency range 2–500 Hz.

The object of our experiments was metal–insulator–semiconductor–insulator–metal (MISIM) TFELEs, where M is a lower 0.2-μm-thick transparent SnO_2 electrode applied on a glass substrate and an upper 0.15-μm-thick opaque thin-film Al electrode of diameter 1.5 mm; S, a Mn-doped (0.5 wt %) 0.54-μm-thick ZnS : Mn phosphor layer; and I, a 0.15-μm-thick Y_2O_3-stabilized (13 wt %) ZrO_2 insulating layer. The phosphor layer was thermally evaporated *in vacuo* in a quasi-closed volume at a substrate temperature of 250°C with subsequent annealing at 250°C for 1 h, the opaque electrode was also applied by thermal evaporation, and the insulating layers were applied by electron-beam evaporation.

The time variations of the brightness L and current through the TFELE $I_e(t)$ were recorded upon the excitation of the device by an alternating-sign triangular voltage $V(t)$ from a G6-34 voltage generator equipped with a driver amplifier and G5-89 external trigger generator. The maximal amplitude of pulses was V_m = 160 V at a voltage nonlinearity factor of no more than 2%. The exciting voltage (in the pulsed excitation regime) was applied as a train of pulses with a duration equal to two periods of the triangular voltage and a repetition rate f = 2, 10, 50, 200, and 500 Hz. The time T between excitations was varied between 1 and 100 s. In the continuous excitation mode, the exciting voltage frequency was also 2, 10, 50, 200, and 500 Hz. The current I_e was measured with a 0.1- to 10 kΩ-resistor series-connected to the TFELE. The voltage drop across the resistor was no higher than 0.5% of the voltage amplitude. The instantaneous value of the brightness was measured with an FÉU-84-3 photoelectric multiplier. The time dependences of the exciting voltage, current through the device, and instantaneous brightness were recorded with an S9-16 two-channel storage oscilloscope linked to a computer via an interface. For either channel, 2048 points with a given period of discretization and 256 levels of amplitude quantization were measured and stored. Mathematical and graphical processing was accomplished with the Maple V Release4 Version 4.00b and GRAPHER Version 1.06 2-D Graphing System application packages.

The time variations of the mean field in the phosphor $F_p(t)$, as well as that of the current, $I_p(t)$, and charge, $Q_p(t)$, passing through the phosphor layer in the

1063-7842/03/4804-0479$24.00

active mode of operation, were calculated as in [1, 2] for a total capacitance of the device's layers C_{tot} = 986 pF and initial geometrical phosphor capacitance C_p = 250 pF with allowance for the voltage drop across the resistor, which was subtracted from the voltage $V(t)$. The capacitances were found from the total capacitance of the device C_e = 200 pF, which was measured with an E7-14 immitance meter for known geometrical parameters of the device.

The external quantum yield η_{ext} of a TFELE is known to be defined as the ratio of the number of photons emitted by the device to the number of charge carriers that passed through the phosphor layer during luminescence. Under the assumption that the emission from the device's surface is monochromatic and direction-independent, we have [3]

$$\eta_{ext} = K_0\eta_{int} = K_0 A\frac{L_e T}{Q_p} = K_0 A\frac{\int_0^T L(t)dt}{\int_0^T I_p(t)dt}, \quad (1)$$

where K_0 is the coefficient of emission extraction from the device; η_{int} is the internal quantum yield; $A = (\pi S_e q)/(h\nu f_\lambda)$, S_e is the emitting surface area; q is the charge of an electron; $h\nu$ is the photon energy; f_λ is the spectral luminous efficiency of the emission; L_e is the electroluminescence brightness averaged over the exciting voltage period T,

$$L_e = \frac{1}{T}\int_0^T L(t)dt;$$

and Q_p is the charge passed through the device for the period T during the formation of two brightness waves.

The luminous efficacy η_L of a TFELE is defined as the ratio of the optical flux from the device, $\Phi_e = \pi S_e L_e$, to the active power P_p spent to produce this flux [3]:

$$\eta_L = \frac{\pi L_e S_e}{P_p} = \frac{\pi S_e\int_0^T L(t)dt}{d_p\int_0^T I_p(t)F_p(t)dt}, \quad (2)$$

where d_p is the phosphor thickness and

$$P_p = \frac{d_p}{T}\int_0^T I_p(t)F_p(t)dt.$$

As follows from (1) and (2), η_{ext}, η_{int}, and η_L are integral parameters and cannot clarify physical processes governing EL in TFELEs. The instantaneous values of these parameters, $\eta_{ext}(t)$, $\eta_{int}(t)$, and $\eta_L(t)$, bear much more information. Earlier [1], we determined the dependences $\eta_{ext}(t)$ and $\eta_{int}(t)$ in the range of ultralow frequencies 0.1–2.0 Hz. At higher frequencies, the solution of the kinetic equation relating the concentration of excited EL centers (hence, the instantaneous brightness $L(t)$) to the current $I_p(t)$ throughout the brightness wave may be difficult. Therefore, we will use another approach.

As the instantaneous value of the internal quantum yield, we will take the increment of the number of photons emitted from the phosphor layer, $\Delta m(t)$, divided by the increment of the number of charge carriers, $\Delta n(t)$, passing through the layer over a time interval Δt:

$$\eta_{int}(t) = \frac{\Delta m(t)}{\Delta n(t)} = \frac{[dm(t)/dt]\Delta t}{[dn(t)/dt]\Delta t} = A\frac{d\left[\int_0^t L(t)dt\right]/dt}{d\left[\int_0^T I_p(t)dt\right]/dt} = A\frac{L(t)}{I_p(t)}. \quad (3)$$

Similarly, the instantaneous luminous efficacy is defined as the ratio of the increment of the optical flux, $\Delta\Phi_e(t)$, to the increment of the active power, $\Delta P_p(t)$, spent to create this increment:

$$\eta_L(t) = \frac{\Delta\Phi_e(t)}{\Delta P_p(t)} = \frac{\pi S_e\Delta L_e(t)}{\Delta P_p(t)} = \pi S_e\frac{[dL_e(t)/dt]\Delta t}{[dP_p(t)/dt]\Delta t} = \pi S_e\frac{d\left[\int_0^t L_e(t)dt\right]/dt}{d\left[\int_0^t P_p(t)dt\right]/dt} = \frac{\pi S_e}{d_p}\frac{L(t)}{I_p(t)F_p(t)}. \quad (4)$$

In the curves $\eta_{int}(t)$ (see (3)) obtained from the experimental dependences $L(t)$ (Figs. 1a, 1e, 2a, 2e) and $I_e(t)$, as well as from the analytical dependences $I_p(t)$ [1, 2], for the frequencies f = 10 (Figs. 1b, 1f) and 200 Hz (Figs. 2b, 2f) one can separate three portions in accordance with similar portions in the curves $L(t)$, $I_p(t)$, and $F_p(t)$ (Figs. 1, 2). Here, I is the portion where the parameters rapidly grow when the exciting voltage $V(t)$ slightly exceeds the threshold value V_t; II, the portion where the parameters vary in a complicated manner; and III, the portion where the voltage diminishes. For the two given frequencies, the behavior of $\eta_{int}(t)$ is noticeably different (Fig. 3), exhibiting the following features:

(1) Asymmetry for both variants (±Al) in correspondence with the asymmetry of the dependences $L(t)$, $I_p(t)$, and $F_p(t)$ as in [1, 2]. The value of η_{int} is greater for the variant (+Al), since Mn^{2+} impurity ions and lattice defects are nonuniformly distributed across the phosphor layer.

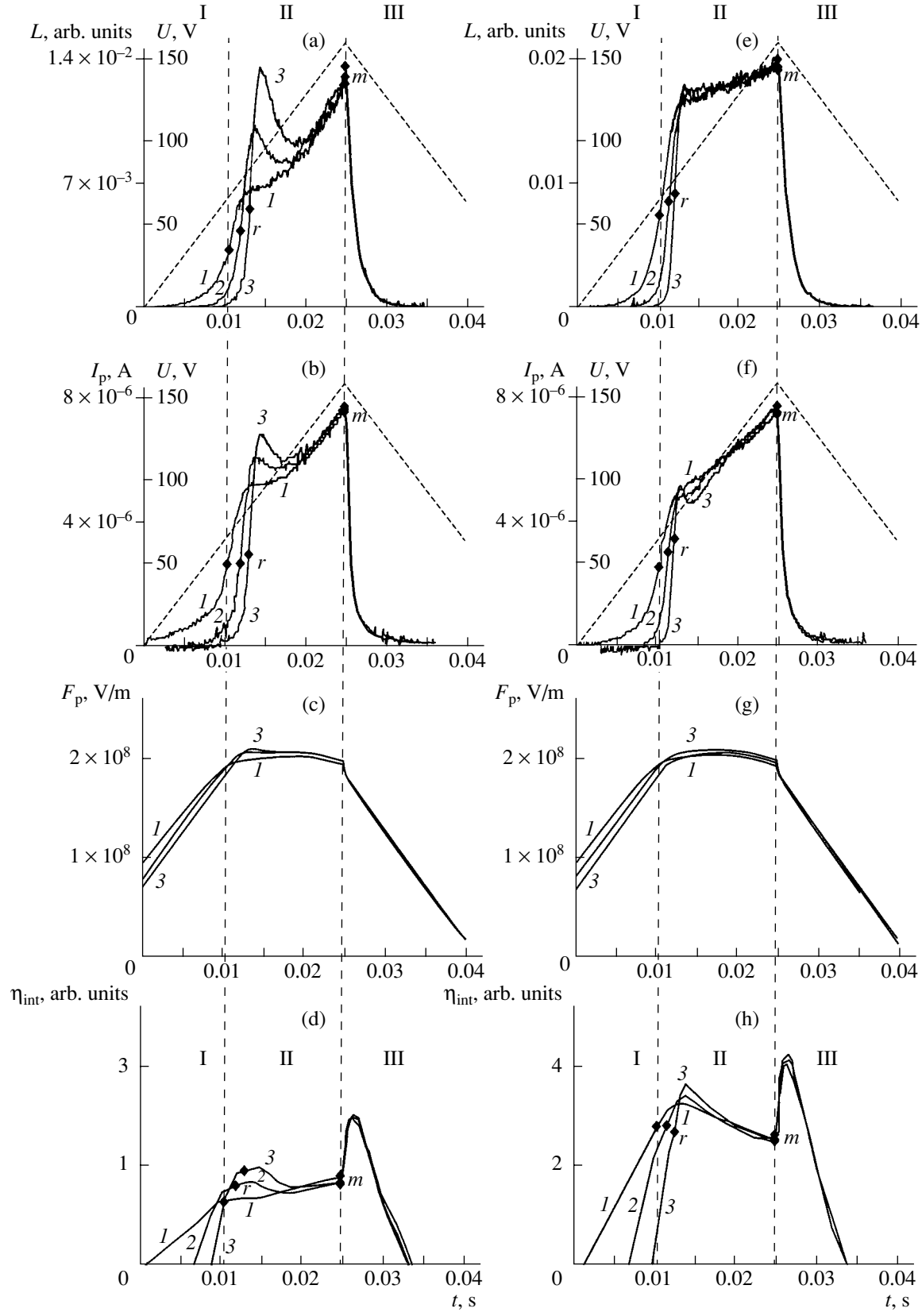

Fig. 1. (a, e) $L(t)$, (b, f) $I_p(t)$, (c, g) $F_p(t)$, and (d, h) $\eta_{int}(t)$ at $f = 10$ Hz. (a–d) Variant (–Al) and (e–h) variant (+Al). (*1*) Continuous excitation and (*2*, *3*) pulsed excitation with $T = 5$ (*2*) and 100 s (*3*). I, portion of fast rise; II, intermediate portion; III, portion of fall. Dashed line in Figs. 1–3, 5, and 7 shows the shape of the voltage $V(t)$. Sample 1.

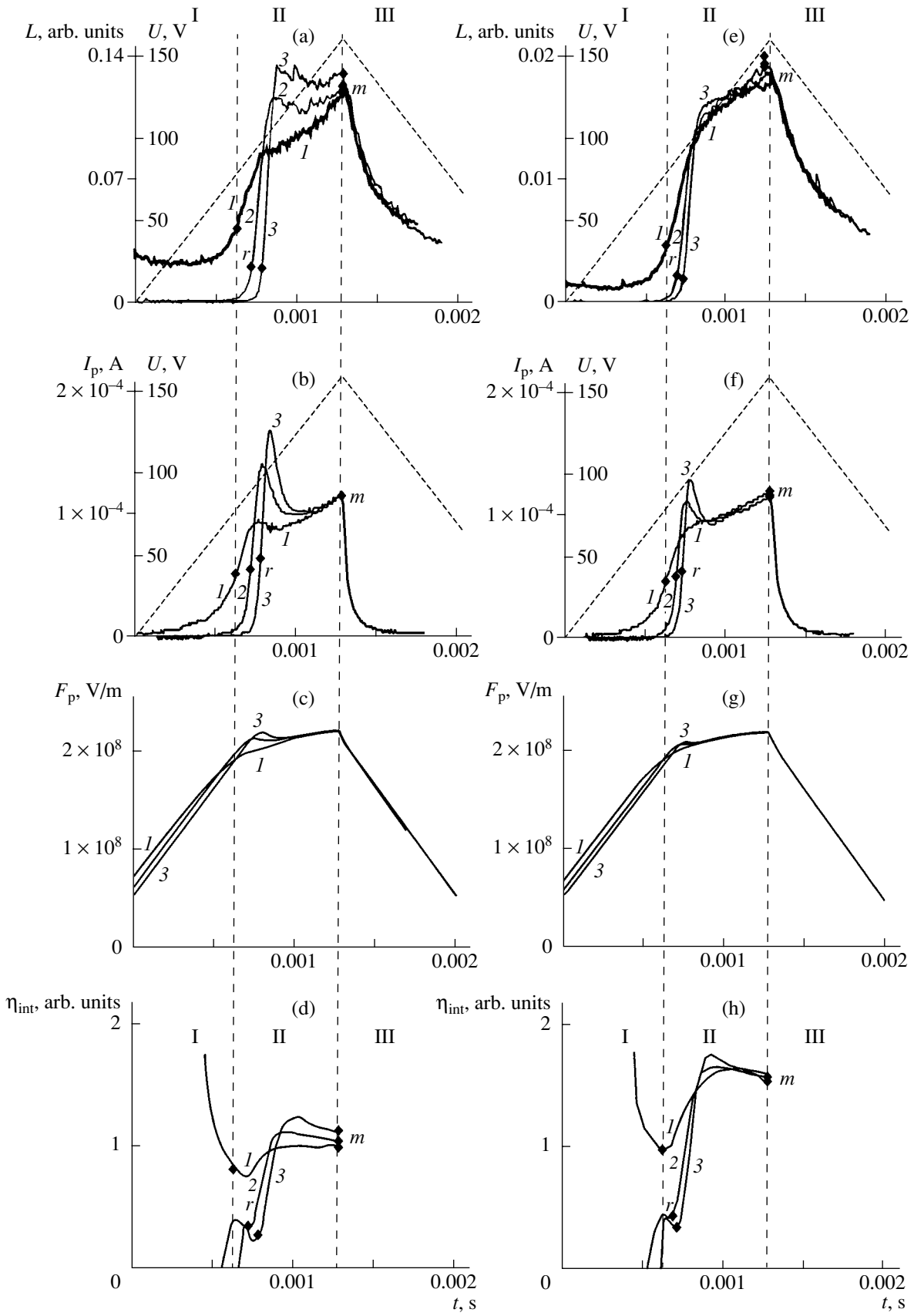

Fig. 2. The same as in Fig. 1 for $f = 200$ Hz. Sample 1.

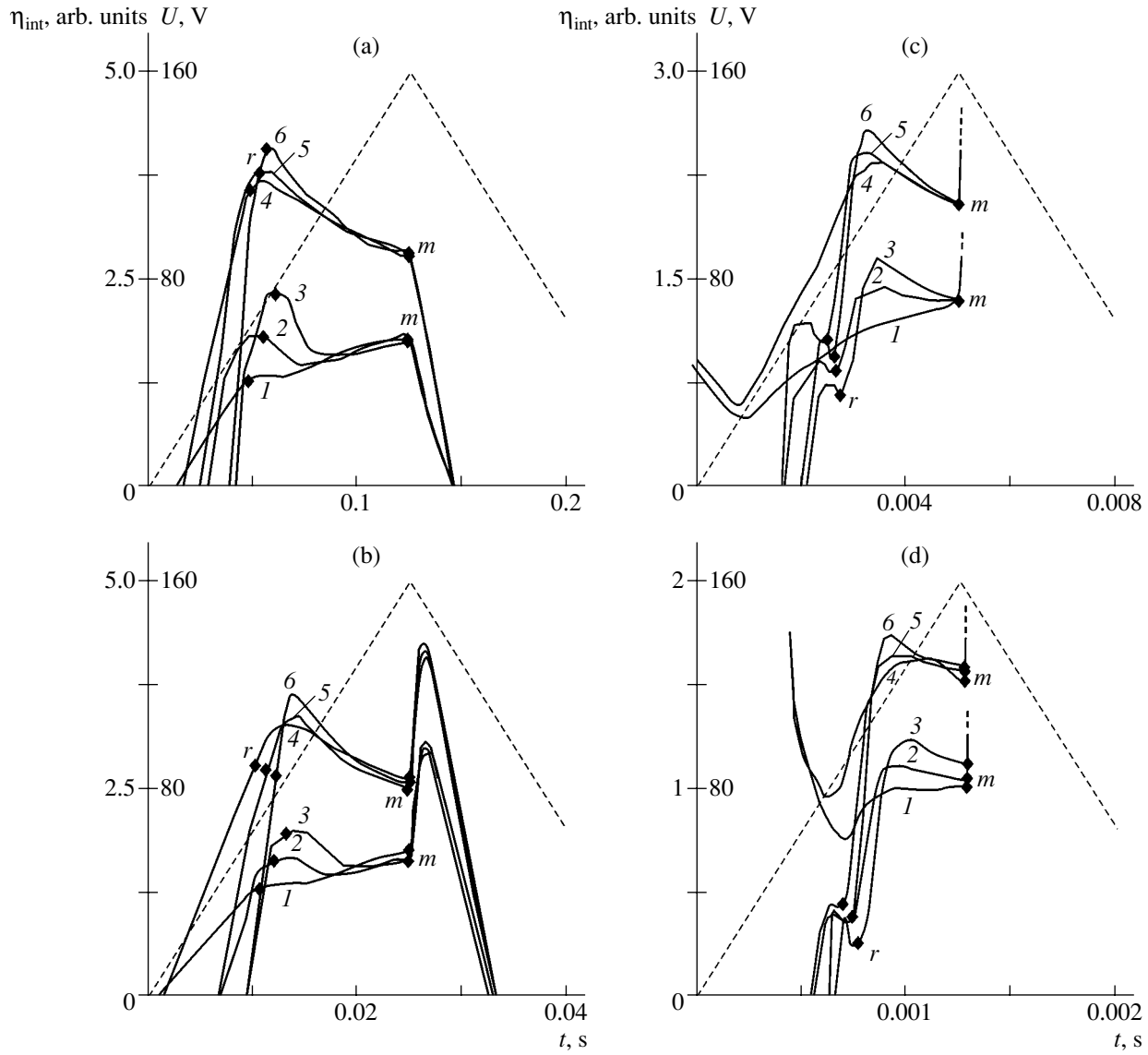

Fig. 3. $\eta_{int}(t)$ for f = 2 (a), 10 (b), 50 (c), and 200 Hz (d). (*1*, *4*) Continuous excitation and (*2*, *3*, *5*, *6*) pulsed excitation with T = (*2*, *5*) 5 and (*3*, *6*) 100 s. (*1*–*3*) Variant (–Al) and (*4*–*6*) variant (+Al). Sample 1.

(2) In portion I, $\eta_{int}(t)$ monotonically grows at frequencies f = 2 and 10 Hz (Figs. 3a, 3b). In this portion, the dependence $F_p(t)$ is also close to linear (Figs. 1c, 1g).

(3) With an increase in the period T of the pulsed excitation regime, the beginning of the curve $\eta_{int}(t)$ in portion I shifts in time (Fig. 3) in accordance with the shift of the curves $L(t)$ and $I_p(t)$ (Figs. 1, 2). As in [1, 2], this is explained by an increase in the threshold voltage V_t because of a decrease in the residual field $F_p(t)$ in the phosphor layer (Figs. 1c, 1g, 2c, 2g). The decrease in $F_p(t)$ is due to polarization charges on states at the phosphor–insulator interface and also to space charges in the phosphor layer, which are neutralized as T increases [1, 2].

(4) In the continuous excitation mode, at frequencies of 50 Hz or higher, brightness waves overlap (Figs. 2a, 2e). This causes the overlap of adjacent voltage half-periods in portion I of the dependence $\eta_{int}(t)$ (Figs. 2d, 2h, 3c, 3d).

(5) At frequencies of 50 Hz or higher, portion I exhibits a dip (Figs. 3c, 3d), which becomes deeper when continuous excitation changes to the pulsed excitation regime and T increases.

(6) In portion II, the dependence $\eta_{int}(t)$ is different for the variants (–Al) and (+Al) at f = 2, 10, and 50 Hz (Figs. 3a–3c). Namely, under continuous excitation, the curve $\eta_{int}(t)$ slightly increases in the variant (–Al) and declines in the variant (+Al). At f = 200 Hz under continuous excitation (Fig. 3d), the curve $\eta_{int}(t)$ in this por-

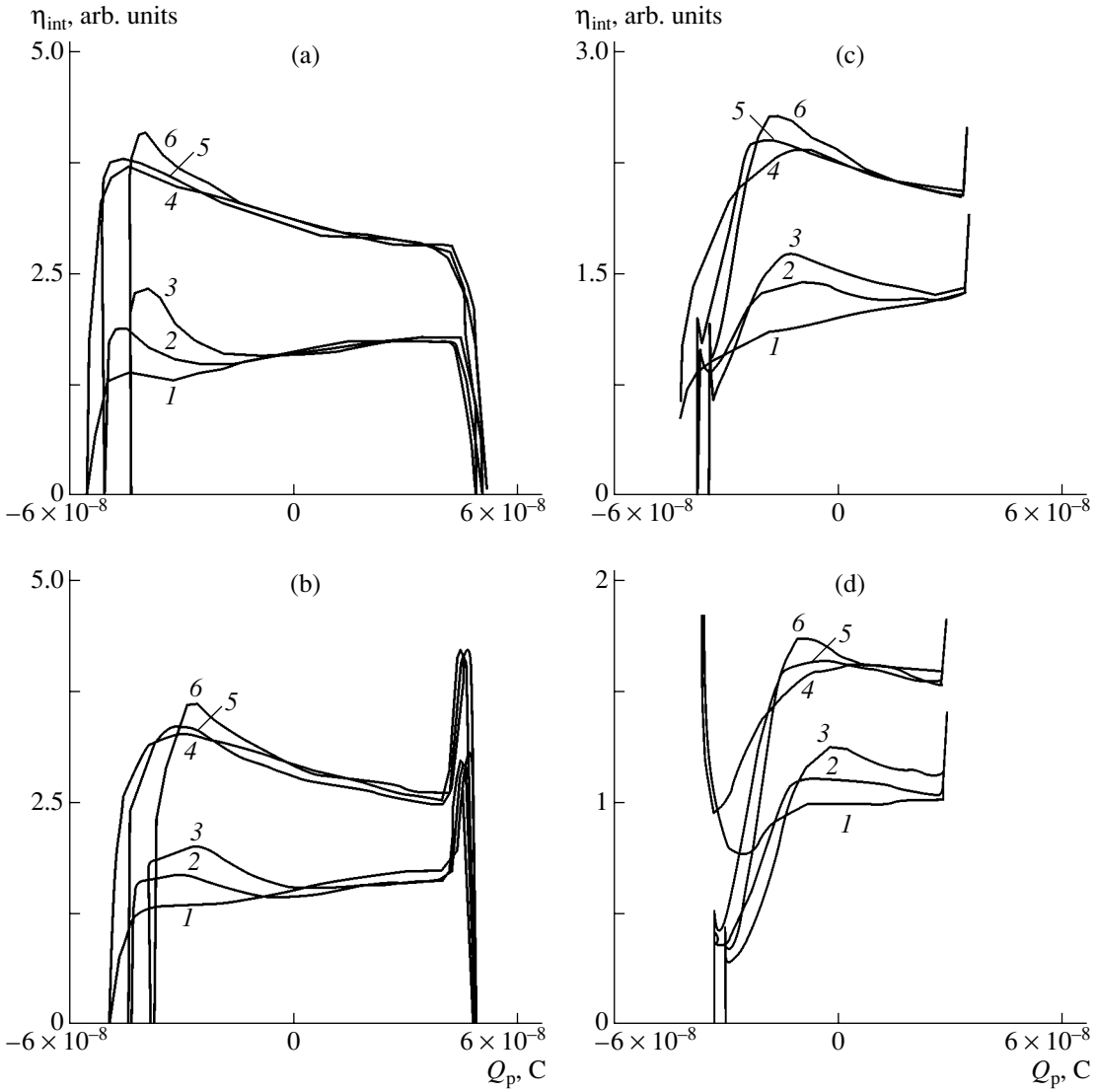

Fig. 4. $\eta_{int}(Q_p)$. The frequencies, excitation conditions, and variants are the same as in Fig. 2. Sample 1.

tion tends toward saturation and at still higher frequencies also falls in the variant (–Al).

(7) At f = 2, 10, 50, and 200 Hz, portion II of the curve $\eta_{int}(t)$ shows a peak in the variant (–Al) in the case of pulsed excitation. A similar peak is observed in the variant (+Al) under continuous excitation as in the previously obtained data for f = 2 Hz [1] (Figs. 1d, 1h, 2d, 2h, 3). The height of this peak grows with T in the variant (+Al).

(8) In portion III, the form of the curves $\eta_{int}(t)$ is frequency dependent. Namely, at f = 2 Hz, η_{int} falls (Fig. 3a); at f = 10 Hz or higher, an additional peak is observed in this portion (Figs. 1d, 1h, 3b). For the frequencies 50 and 200 Hz, this additional peak is shown incompletely because of its large amplitude and long fall time, which extends beyond the figures (Figs. 2d, 2h, 3c, 3d). Its appearance is associated with the fact [1] that the brightness $L(t)$ drops more slowly than the current $I_p(t)$ (Figs. 2a, 2b, 2e, 2f).

(9) In portions I and II, as the frequency grows, η_{int} decreases at the same voltage values for both variants (Fig. 3).

The dependences of η_{int} on the charge Q_p passing through the TFELE during the formation of brightness waves (Fig. 4) are similar to the curves $\eta_{int}(t)$ (Fig. 3). This is because the dependence $Q_p(t)$ is almost linear in portions I and II and varies only slightly in portion III (Fig. 5).

The dependence of η_{int} on the mean field F_p in the phosphor layer (Fig. 6) is nearly linear in portion I for f = 2 and 10 Hz. At f = 50 Hz or higher, it exhibits a dip like the curves $\eta_{int}(t)$ and $\eta_{int}(Q_p)$ at f = 50 Hz. This is explained by the nearly linear dependence $F_p(t)$ in this portion (Figs. 1c, 1d, 2c, 2d). In portion II, however, the

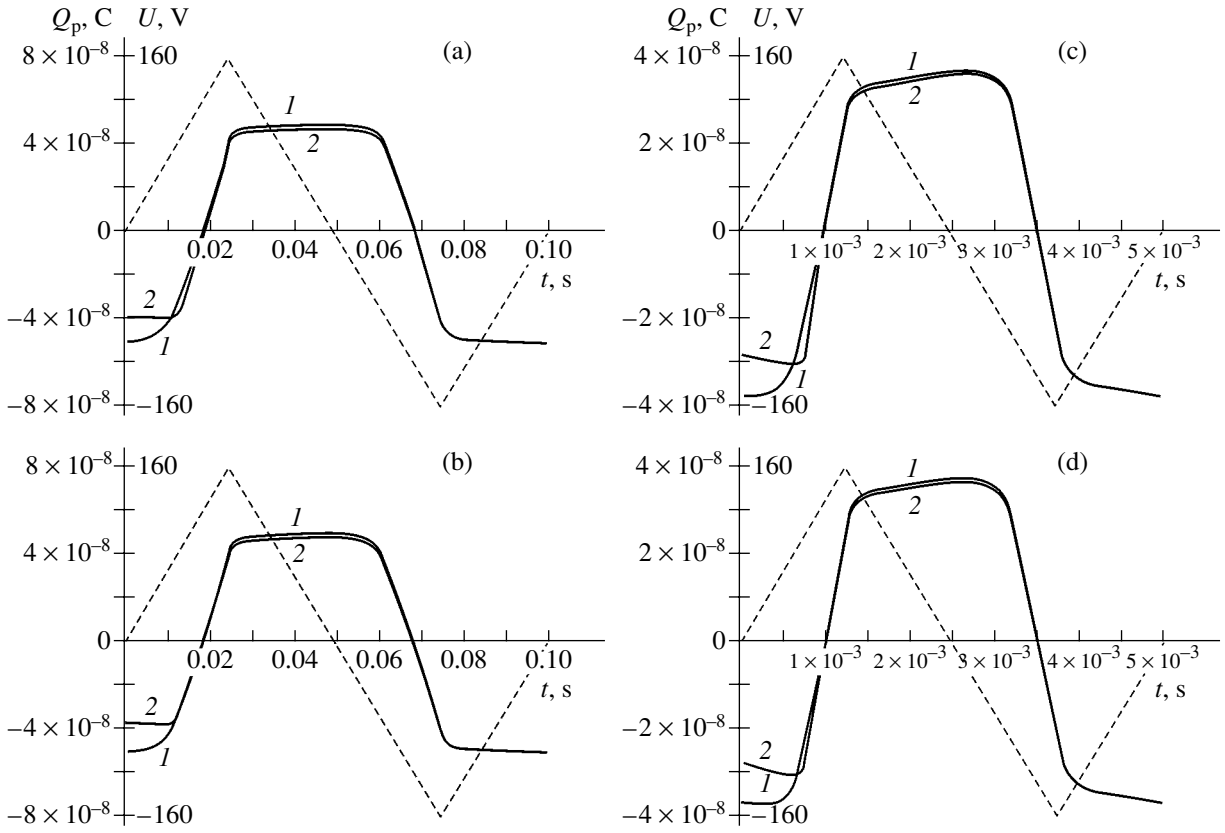

Fig. 5. $Q_p(t)$ at $f = 10$ (a, b) and 200 Hz (c, d). (a, c) Variant (–Al) and (b, d) variant (+Al). (*1*) Continuous excitation and (*2*) pulsed excitation with $T = 100$ s. Sample 1.

curves $\eta_{int}(F_p)$ are highly irregular, which reflects the different behavior of $F_p(t)$ in this portion depending on the variant (±Al) and frequency (Figs. 1c, 1d, 2c, 2d). This situation is the result of space charges forming in the near-anode and near-cathode regions of the phosphor [1, 2].

As follows from (3) and (4), the dependence $\eta_L(t)$ differs from the dependence $\eta_{int}(t)$ by the factor $F_p(t)$ in the denominator. Then, the instantaneous active power $P_p(t)$ can be fairly accurately approximated by the relationship [3]

$$P_p(t) = I_p(t)F_p(t)d_p. \quad (5)$$

In view of the linear growth of $F_p(t)$ in portion I and the relatively weak dependence $F_p(t)$ in portion II (Figs. 1c, 1g, 2c, 2g), the form of the dependence $P_p(t)$ turns out to be similar to that of the dependence $I_p(t)$ (Figs. 7b, 7d). Therefore, the dependences $\eta_{int}(t)$ and $\eta_L(t)$ are similar in shape (Figs. 7e, 7f). As follows from Figs. 7e and 7f, at frequencies $f \le 10$ Hz, the values of η_{int} and η_L for sample 2 reach a maximum at the point *r*, which separates the regions of fast and slow growth of $I_p(t)$ and $L(t)$. Above this point, the rate of rise of these curves declines [1, 2] even if the amplitude (maximal value) V_m of the exciting voltage increases further. However, this maximum in the curve $\eta_L(t)$ slightly lowers with increasing V_m (Fig. 7f), because the field $F_p(t)$ increases at the point *r* with V_m (Fig. 7c).

The values of the internal quantum yield η_{int} and luminous efficacy η_L in the dependences $\eta_{int}(V_m)$ and $\eta_L(V_m)$, which were found from Fig. 7 by formulas (1) and (2), are greater for the variant (+Al) (Figs. 8d, 8e), which is due to different values of L_e in the variants (+Al) and (–Al) (Fig. 8c). At the same time, the half-period-averaged values of the current I_p passing through the phosphor layer,

$$I_p = \frac{2}{T}\int_0^{T/2} I_p(t)dt, \quad (6)$$

and power P_p,

$$P_p = \frac{2}{T}\int_0^{T/2} P_p(t)dt, \quad (7)$$

coincide within the accuracy of measurement and calculation for both variants (Figs. 8a, 8b). The brightness–voltage characteristic of the TFELE, $L_e(V_m)$, is typical of this device: it has a portion where the growth of L_e slows down (for V_m above 125 V; Fig. 8c) or satu-

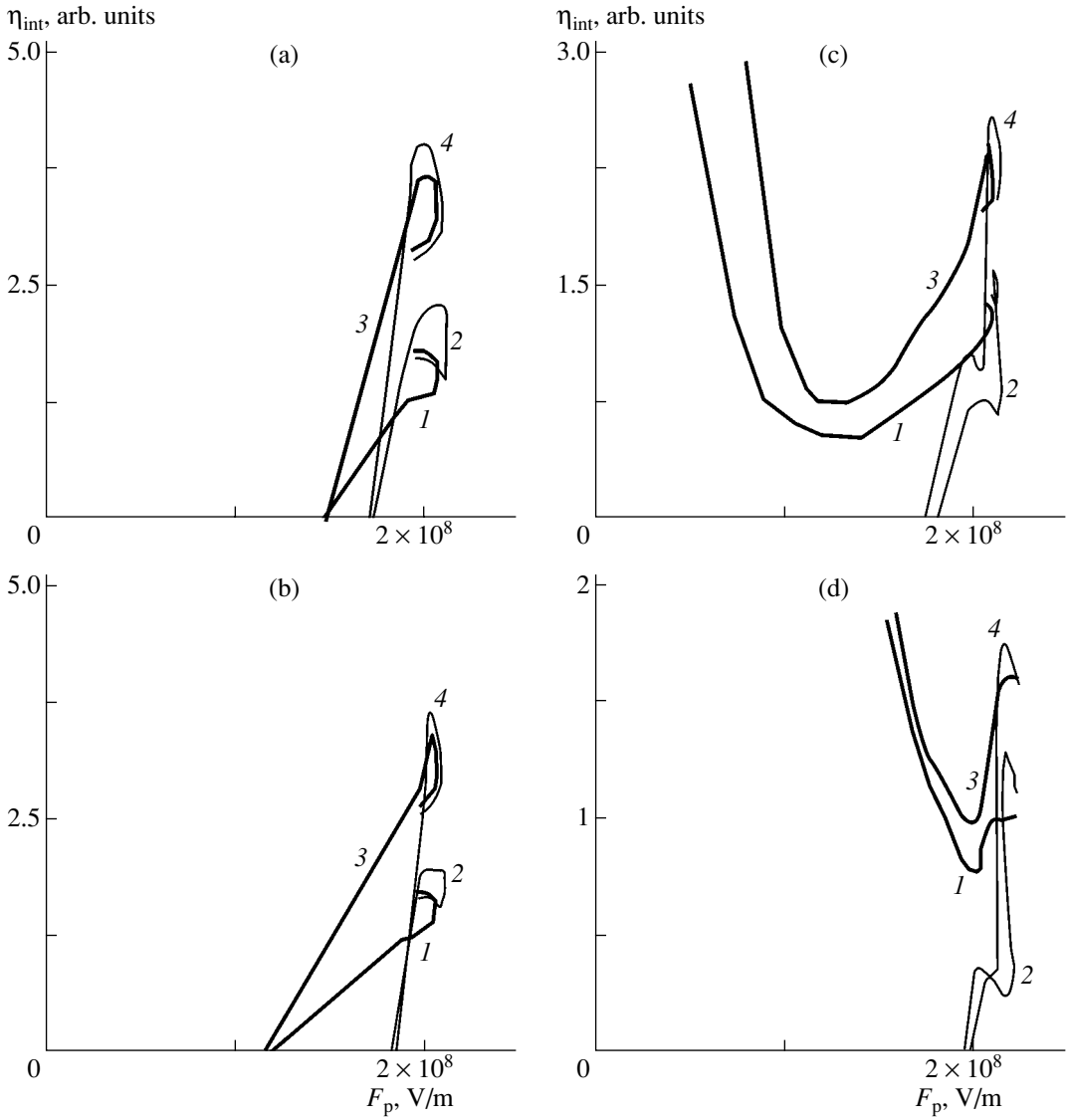

Fig. 6. $\eta_{int}(F_p)$ for the same frequencies as in Fig. 3. (*1*, *3*) Continuous excitation and (*2*, *4*) pulsed excitation with T = 100 s. (*1*, *2*) Variant (–Al) and (*3*, *4*) variant (+Al). Sample 1.

rates (in the semilogarithmic scale [4]; Fig. 8f). In this portion, the curves $I_p(V_m)$ and $P_p(V_m)$ are similar to the curve $L_e(V_m)$ (Figs. 8a–8c). Therefore, at V_m above 125 V, η_{int} depends on V_m only slightly, reaching a maximum at $V_m \approx 135$ V. Since F_p grows with V_m in portion II (Fig. 7b), the curves $\eta_L(V_m)$ exhibit a more distinct fall at $V_m \geq 125$–130 V (Fig. 8e). However, from these dependences $\eta_{int}(V_m)$ and $\eta_L(V_m)$, one cannot elucidate the physical mechanisms governing the EL kinetics, particularly because the portion where the brightness $L(t)$ falls (Figs. 1a, 1e, 2a, 2e) makes a considerable contribution to the mean brightness at $f \geq 10$ Hz.

The results obtained can be explained as follows.

The behavior of the curves $\eta_{int}(t)$ in portions I and II can be explained in the same way as in [1, 4]. The values of $L(t)$ and $\eta_{int}(t)$ are related to the concentration $N^*(t)$ of excited luminescence centers as [5, 6]

$$L(t) = \frac{K_0 h\nu f_\lambda}{\pi} \frac{\eta_{int}(t) N^*(t) d_p(t)}{\tau^*}, \quad (8)$$

where

$$\eta_{int}(t) = N_1(t) P_r(t); \quad (9)$$

$N_1(t)$ is the number of luminescence centers excited by one electron that passed through the phosphor layer,

$$N_1(t) = d_p(t) \sigma N(x, t); \quad (10)$$

$d_p(t)$ is the effective phosphor thickness within which the impact excitation of luminescence centers takes place; σ is the impact excitation cross section; $N(x, t)$ is the distribution of unexcited luminescence centers

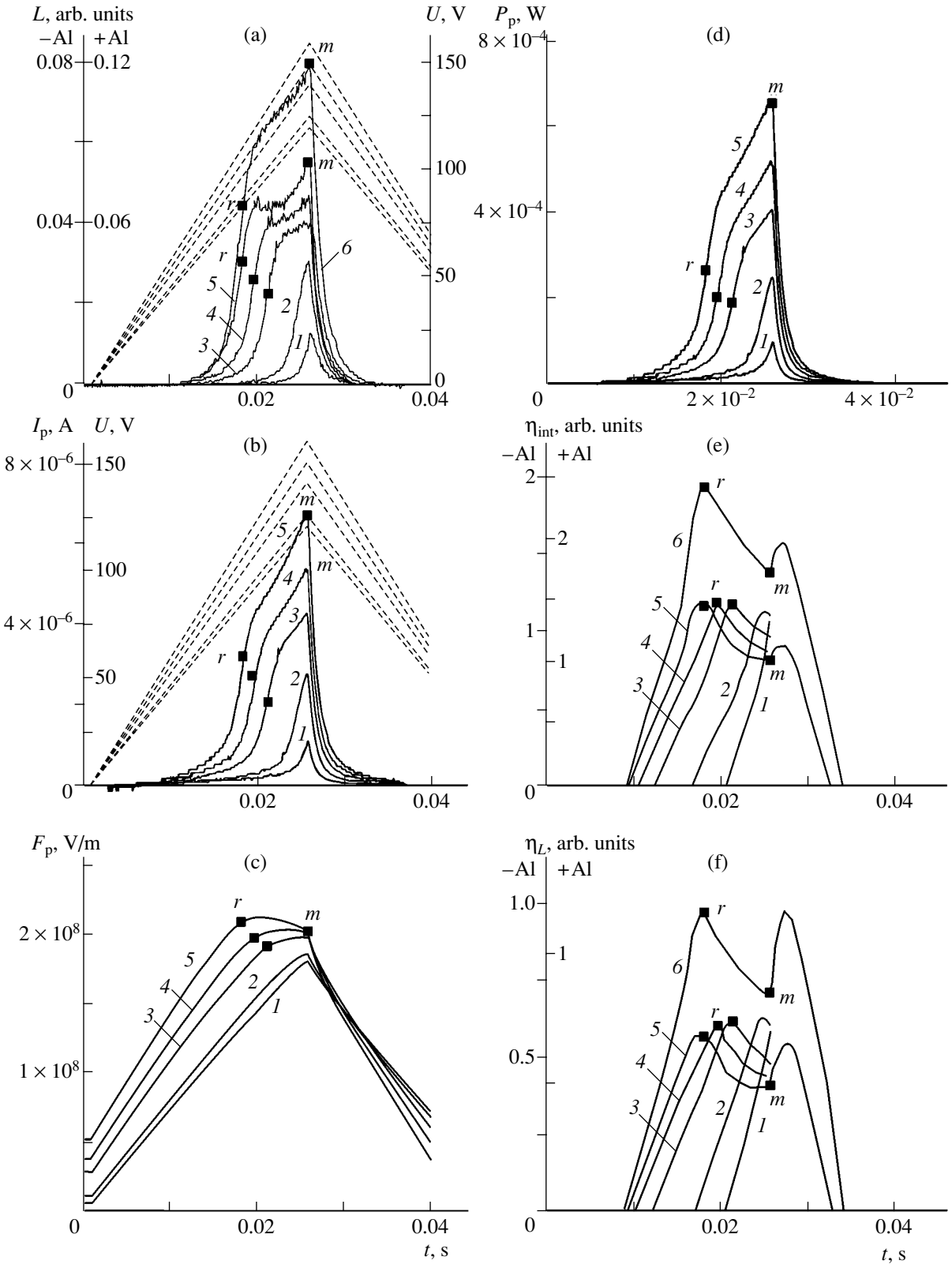

Fig. 7. Formation of the dependences $\eta_{int}(t)$ and $\eta_L(t)$ at various V_m and $f = 10$ Hz for pulsed excitation with $T = 1$ s. (a) $L(t)$, (b) $I_p(t)$, (c) $F_p(t)$, (d) $P_p(t)$, (e) $\eta_{int}(t)$, and (f) $\eta_L(t)$. V_m = 120 (*1*), 125 (*2*), 140 (*3*), 150 (*4*), and 160 V (*5*, *6*). (*1–5*) Variant (–Al) and (*6*) variant (+Al). Sample 2.

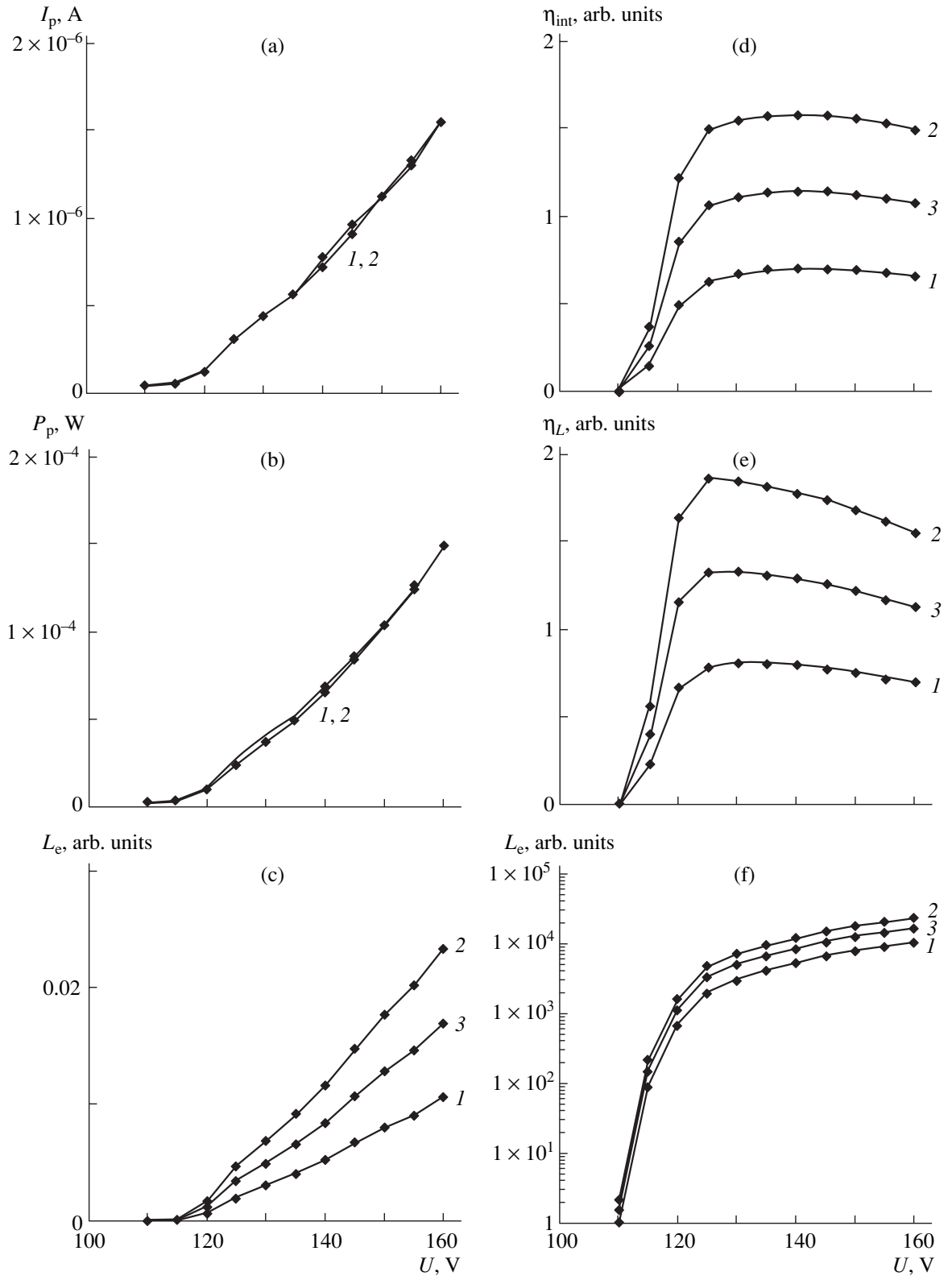

Fig. 8. (a) $I_{pe}(V_m)$, (b) $P_{pe}(V_m)$, (c) $L_e(V_m)$, (d) $\eta_{int}(V_m)$, (e) $\eta_L(V_m)$, and (f) $\log L_e(V_m)$ at $f = 10$ Hz under pulsed excitation with $T = 1$ s for the voltage half-period in the variant (*1*) (–Al) and (*2*) (+Al). (*3*) Values averaged over the voltage period. Sample 2.

across the phosphor layer; $P_r(t)$ is the probability of radiative recombination of luminescence centers,

$$P_r(t) = \frac{\tau^*}{\tau_r}; \quad (11)$$

τ^* is the lifetime of excited luminescence centers; and τ_r is the time constant of excited center relaxation via radiative transitions to the ground state.

The concentration $N^*(t)$ of excited luminescence centers is proportional to the probability of excitation per unit time $\alpha(t)$, which is generally expressed as [4]

$$\alpha(t) = \sigma n(t) v(t) = \frac{\sigma I_p(t)}{qS_e}, \quad (12)$$

where $n(t)$ and $v(t)$ are, respectively, the concentration and velocity of free electrons with an energy sufficient for exciting luminescence centers.

Thus, the rapid growth of $\eta_{int}(t)$ in portion I at $f = 2$ and 10 Hz (Figs. 1d, 1h, 3a, 3b) is associated with the growth of $F_p(t)$ (Figs. 1c, 1g), the energy of accelerated electrons, $\alpha(t)$ (see (12)), and $N_1(t)$ (see (10)).

The dip in portion I of the curve $\eta_{int}(t)$ at frequencies of 50 Hz and higher (Figs. 2d, 2h, 3c, 3d), as well as in the curves $\eta_{int}(Q_p)$ (Figs. 4c, 4d) and $\eta_{int}(F_p)$ (Figs. 6c, 6d), is explained by the changed relationship between the rates of rise of the brightness $L(t)$ and current $I_p(t)$ in this portion. At $f = 2$ and 10 Hz, the rates of rise of $L(t)$ and $I_p(t)$ are nearly the same (in accordance with data obtained in [1]). Also, $\eta_{int}(Q_p)$ and $\eta_{int}(F_p)$ also grow in portion I (Figs. 1d, 1h, 3a, 3b, 4a, 4b, 6a, 6b). At higher frequencies, the current $I_p(t)$ rises faster than $L(t)$ up to the point r (cf. Figs. 2b, 2f and 2a, 2e). As a result, $\eta_{int}(t)$ decreases, reaching a minimum at the point r (Figs. 2d, 2h, 3c, 3d). In going over the point r, the growth of $I_p(t)$ slows down and $\eta_{int}(t)$, as well as $\eta_{int}(Q_p)$ and $\eta_{int}(F_p)$, increases. Under continuous excitation at frequencies of 50 Hz or higher, adjacent brightness waves overlap. Therefore, in portion I, Mn^{2+} luminescence centers excited in both the previous and current voltage half-periods relax, with their concentrations decreasing and increasing, respectively. Such a relaxation causes the minimum in the curve $\eta_{int}(t)$ (Figs. 2d, 2h, 3c, 3d). This minimum shades the dip (including in the curves $\eta_{int}(Q_p)$ and $\eta_{int}(F_p)$; Figs. 4c, 4d, 6c, 6d), which appears when the relationship between the growth rates of $I_p(t)$ and $L(t)$ changes.

In portion II, the excitation of Mn^{2+} centers is characterized by a weak variation of the mean field $F_p(t)$ (Figs. 1c, 1g, 2c, 2g). Therefore, $\eta_{int}(t)$ varies under the condition that $P_r(t)$ is constant and the mechanism of direct impact excitation of luminescence centers with the distribution $N(x, t)$ in a phosphor layer of effective thickness $d_p(t)$ (see (9) and (10)) holds. As was mentioned above, the distribution $N(x, t)$ is nonuniform: the concentration of the centers is higher at the upper electrode (Al). The value of $d_p(t)$ changes when deep centers at the anode and cathode ionize and exchange charge. This results in the formation of space charge fields at the electrodes. The ionization of deep centers at the anode also leads to the dissipation of the energy of accelerated electrons by these centers; hence, $n(t)$, $v(t)$, and, accordingly, $\alpha(t)$ decrease [4]. Eventually, the behavior of $\eta_{int}(t)$ in portion II is governed by the decrease in $d_p(t)$ and $\alpha(t)$, as well as by the change in $N(x, t)$. The distribution $N(x, t)$ is the variation of the Mn^{2+} initial distribution $N(x)$ across the phosphor layer with time because of the decrease in the thickness $d_p(t)$ of the region where luminescent centers are excited and the shift of this region toward the anode. For the variant (+Al), all these factors in portion II become less significant and $\eta_{int}(t)$ drops (Figs. 1h, 2h, 3). For the variant (–Al), a decrease in $d_p(t)$ and $\alpha(t)$ may be compensated for by an increase in $N(x, t)$. Therefore, $\eta_{int}(t)$ may slightly grow (Figs. 1d, 3a, 3b) for same samples and decline for others (Fig. 7e).

In the case of pulsed excitation of the TFELE, the space charges relax and the fields of these charges in the near-anode and near-cathode regions of the phosphor decrease over the time interval between voltage pulses [1, 2]. The field decrease is the stronger, the larger T is. As a result, the threshold value V_t of the device and the mean field $F_p(t)$ in the phosphor rise in the next excitation cycle, including in portion II (Figs. 1, 2). This causes the energy of accelerated electrons and the probability $\alpha(t)$ of excitation of luminescence centers (see (12)) to increase and also results in the appearance and/or growth of the peak in the dependence $\eta_{int}(t)$ in portion II (Figs. 1d, 1h, 2d, 2h, 3).

The decrease in $\eta_{int}(t)$ with increasing frequency f in portion II for the same values of $V(t)$ is explained as follows.

The kinetic equation for the rate of change of the concentration of excited luminescence centers has the form [1]

$$\frac{dN^*(t)}{dt} = \alpha(t)[N(t) - N^*(t)] - \frac{N^*(t)}{\tau^*}. \quad (13)$$

If σ is constant, $\alpha(t)$ is independent of $N^*(t)$, τ^* is independent of time t, $\alpha(t) \ll 1/\tau^*$, and the current $I_p(t)$ in portion III is approximated as

$$I_p(t) = \frac{I_{pm}}{2}(e^{-t/\tau_4} + e^{-t/\tau_5}) \quad (14)$$

(where τ_4 and τ_5 are the time constants of $I_p(t)$ fall), the solution to Eq. (13) that contains the value I_{pm} at the point m (Figs. 1b, 1f, 2b, 2f) was shown [1] to yield (in view of expressions (8) and (12)) the following expression for $L(t)$ in portion III:

$$L(t) = \frac{A\eta_{int}(t)N_1(t)}{qS_e}\left\{\frac{I_{pm}}{2}\left(\frac{\tau_4}{\tau_4 - \tau^*}e^{-t/\tau_4} + \frac{\tau_5}{\tau_5 - \tau^*}e^{-t/\tau_5}\right)\right.$$

Table

f, Hz	2		10		50		200	
Variant	–Al	+Al	–Al	+Al	–Al	+Al	–Al	+Al
τ_4, ms	2.79	2.08	0.35	0.574	0.088	0.055	0.0199	0.0166
τ_5, ms	10.8	9.4	2.05	2.96	0.45	0.397	0.111	0.105
B	0.67	0.609	0.0877	0.25	0.143	0.0003	7.95×10^{-7}	5.35×10^{-5}
C	0.36	0.399	0.0103	0.0089	0.591	0.717	0.81	0.76
D	3.26×10^{-5}	4.15×10^{-5}	0.914	0.727	0.277	0.269	0.27	0.34

$$+\left[I_{\mathrm{pr}}\left(-\frac{\tau_2}{\tau_2-\tau^*}e^{-t_m/\tau_2}+\frac{\tau_3}{\tau_3-\tau^*}e^{t_m/\tau_3}+\frac{\tau_1}{\tau_1+\tau^*}+\frac{\tau_2}{\tau_2-\tau^*}-\frac{\tau_3}{\tau_3+\tau^*}\right)-\frac{I_{pm}}{2}\left(\frac{\tau_4}{\tau_4-\tau^*}+\frac{\tau_5}{\tau_5-\tau^*}\right)\right]e^{-t/\tau^*}\Bigg\}. \quad (15)$$

Here, τ_1, τ_2, and τ_3 are the respective time constants of $I_p(t)$ growth in portion I up to the point *r* and in portion II before and after the point where the rate of current growth changes sign; t_m is the time corresponding to the value V_m; and I_{pr} is the value of the current I_p at the point *r*.

For $\tau^* = 1.4$ ms, the inequality $\alpha(t) \ll 1/\tau^*$ holds at $f < 200$ Hz. Indeed, since $\sigma = (2–4) \times 10^{-16}$ cm^2 [1, 5] and $S_e = 2$ mm^2, then, according to (12), $\alpha(t) \approx 5$ s^{-1} for $f = 50$ Hz and $I_{pm} \approx 5 \times 10^{-5}$ A and $\alpha(t) \approx 15$ s^{-1} for $f = 200$ Hz and $I_{pm} \approx 1.5 \times 10^{-4}$ A. At the same time, $1/\tau^* = 714$ s^{-1}.

As follows from Figs. 1 and 2, the current $I_p(t)$ drops almost to zero when the mean field $F_p(t)$ diminishes to ~10^8 V/m, which roughly equals the luminescence threshold; that is, the ionization of Mn^{2+} luminescence centers continues throughout the $I_p(t)$ fall. This points to the need to solve kinetic equation (13), which involves the term $\alpha(t)[N - N^*(t)]$ responsible for generation, in portion III (the portion of fall) for the frequency range considered above, as in the case of ultralow frequencies [1].

At $f < 2$ Hz, $\tau_1 \ldots \tau_5 \gg \tau^*$ and the dependence $L(t)$ is similar to the dependence $I_p(t)$ [1]. However, even at $f = 2$ Hz, τ_4 becomes comparable to τ^* [1]. At $f \geq 10$ Hz, the same is true for the other time constants. At high frequencies, $\tau_1 \ldots \tau_5$ become less than τ^*. This decreases the preexponentials in (15) and slows down the growth of $L(t)$ in comparison with that of $I_p(t)$ with increasing frequency. Physically, this means that the relaxation of excited Mn^{2+} luminescence centers is slower than the variation of $I_p(t)$ and that the concentration $[N - N^*(t)]$ of unexcited centers in Eq. (13) decreases with increasing frequency *f*. Under continuous excitation of the device, the decrease in $\eta_{int}(t)$ at $f \geq 50$ Hz is more appreciable because adjacent brightness waves overlap (Figs. 2a, 2e).

The similar variation of $\eta_{int}(t)$ with frequency in portion II (Fig. 3) under continuous excitation and under pulsed excitation with different *T* confirms the fact that $d_p(t)$, $N(x, t)$, and $F_p(t)$ vary with frequency insignificantly. The dependence $F_p(t)$ in portion II is also weak at different frequencies (Figs. 1c, 1g, 2c, 2g).

The explanation of the behavior of the curve $\eta_{int}(t)$ in portion II, including at various V_m (Figs. 1, 2, 3, 7), as well as solution (15) to Eq. (13), is based on the assumption that the mechanism of direct impact excitation of single Mn^{2+} centers holds and that σ, P_r, and, hence, τ^* appearing in (9)–(11) remain constant. The constancy of τ^* unambiguously characterizes the constancy of the relationship between radiative and nonradiative recombination centers [7, 8], as well as the absence of luminescence centers of another type.

The data obtained validate the above assumptions. For example, approximation (14) of portions III, where the curves $I_p(t)$ fall, fits the values of $I_p(t)$ obtained from the experimental dependence $I_e(t)$ within 0.4%. Taking into account that the exponentials in expression (15) for $L(t)$ vary in portion III much more than the preexponentials, we represent (15) in the form

$$L(t) \approx Be^{-t/\tau_4} + Ce^{-t/\tau_5} + De^{-t/\tau^*}, \quad (16)$$

where *B*, *C*, and *D* are constants.

Approximation (16) of $L(t)$ in portion III, where the curve falls, is accurate to within 0.1% for the range of fall of two decades (one hundred times), which is practically the most important, at $f = 2$, 10, and 50 Hz and $\tau^* = 1.4$ ms. At $f \geq 200$ Hz, the agreement breaks possibly because the assumptions used when solving Eq. (13) are not quite correct.

As follows from the table, τ_4 and τ_5 vary almost inversely proportionally to *f* and so rapidly decreases with increasing *f*. This time constant apparently characterizes the capture of free charge carriers by volume and surface centers when the mean field in the phosphor layer drops to the threshold value or below.

The f dependences of the coefficients B, C, and D suggest that the slowly decaying component of $I_p(t)$ with the time constant τ_5 and the lifetime τ^* of luminescence centers play an increasingly important role in the fall of $I_p(t)$ and $L(t)$, respectively, as f grows.

The fact that the branch where $L(t)$ falls remains unchanged at various V_m (Figs. 7a, 7b) also indicates the constancy of the parameters that describe the excitation mechanisms, as well as the mechanisms of radiative and nonradiative recombinations of luminescence centers, at greatly differing excitation levels. The branches where $L(t)$ and $I_p(t)$ fall also do not change when the continuous excitation is changed to the pulsed excitation with different T (Figs. 1a, 1b, 1e, 1f, 2a, 2b, 2e, 2f). This is additional evidence that the charge state of defects in the phosphor layer has a minor effect on the excitation and relaxation of Mn^{2+} luminescence centers.

It should be noted that the product of exponential functions used in [9] to approximate brightness fall does not agree with the experimental dependence $L(t)$, including in that range of f where $\tau^* \gg \tau_4$ and τ_5. This may indicate that, with the concentration of Mn^{2+} centers in the phosphor layer used in this work, the concentration quenching of luminescence is absent.

Thus, the dependences of the instantaneous values of the internal quantum yield η_{int} and luminous efficacy η_L on time t and exciting voltage amplitude V_m in combination with other electrical and illumination engineering characteristics show that at exciting voltage frequencies $f \le 10$ Hz, instantaneous values $\eta_{int}(t)$ and $\eta_L(t)$ rapidly grow in the portion where the brightness $L(t)$ and the current $I_p(t)$ passing through the phosphor layer also steeply grow. This may be associated with an increase in the number of luminescence centers excited by one electron passing through the phosphor layer when the mean field in the phosphor increases. At $f >$ 10 Hz, the portion where $\eta_{int}(t)$ grows shows a dip, which appears when the current I_p rises faster than the instantaneous brightness $L(t)$. The dip becomes more pronounced in the pulsed excitation mode especially when the time interval T between the pulses widens. This effect is due to the enhanced rate of growth of $I_p(t)$ when the space charge fields in the near-anode and near-cathode regions of the phosphor layer are compensated for in the pause between the pulses and the threshold field of luminescence increases. As the current $I_p(t)$ and brightness $L(t)$ grow further (more slowly), the dependences $\eta_{int}(t)$ and $\eta_L(t)$ in both the continuous and pulsed excitation modes are apparently governed by the decrease in the effective phosphor thickness $d_p(t)$ within which the ionization of Mn^{2+} luminescence centers occurs and in the probability of excitation of these centers per unit time $\alpha(t)$. One more characteristic controlling the behavior of $\eta_{int}(t)$ and $\eta_L(t)$ is the variation of the initial distribution of luminescence centers across the phosphor layer $N(x)$ with time; i.e., the distribution $N(x, t)$. The initial distribution changes because the effective phosphor thickness $d_p(t)$ (the thickness of the ionization region) diminishes and shifts toward the anode. Eventually, according to the frequency f, the curves $\eta_{int}(t)$ and $\eta_L(t)$ in this portion may reach a maximum (its height grows with T in the pulsed excitation mode) and then fall, have a plateau (where η_{int} and η_L are time independent), or exhibit a region where η_{int} and η_L grow with time.

The integral characteristics $\eta_{int}(V_m)$ and $\eta_L(V_m)$ do not allow the physical mechanisms responsible for electroluminescence to be considered in detail. The reason is that, at $f \ge 10$ Hz, the branch where the brightness falls because of the relaxation of excited Mn^{2+} centers contributes significantly to the mean brightness, which is necessary to determine these processes. The portion of brightness fall lasts long after the fall of the current through the phosphor and masks processes responsible for the excitation of the centers.

For the concentration of luminescence centers used in this work, the analytical solution to the kinetic equation describing the change of this concentration with time yields a fairly accurate approximation of the curve of $L(t)$ fall at frequencies $f < 200$ Hz under various exciting voltage amplitudes V_m. In this case, the fall of the current $I_p(t)$ through the phosphor layer is approximated by the sum of two exponential functions under the assumption of simple impact ionization of single Mn^{2+} centers, which subsequently relax without changing the probability of radiative transitions.

REFERENCES

1. N. T. Gurin, O. Yu. Sabitov, and A. V. Shlyapin, Zh. Tekh. Fiz. **72** (2), 74 (2002) [Tech. Phys. **47**, 215 (2002)].
2. N. T. Gurin, O. Yu. Sabitov, and A. V. Shlyapin, Zh. Tekh. Fiz. **71** (8), 48 (2001) [Tech. Phys. **46**, 977 (2001)].
3. N. T. Gurin, Zh. Tekh. Fiz. **66** (5), 77 (1996) [Tech. Phys. **41**, 448 (1996)].
4. N. T. Gurin, A. V. Shlyapin, and O. Yu. Sabitov, Pis'ma Zh. Tekh. Fiz. **27** (22), 52 (2001) [Tech. Phys. Lett. **27**, 956 (2001)].
5. *Electroluminescence Sources of Light*, Ed. by I. K. Vereshchagin (Énergoatomizdat, Moscow, 1990).
6. N. T. Gurin, O. Yu. Sabitov, A. V. Shlyapin, *et al.*, Pis'ma Zh. Tekh. Fiz. **27** (4), 12 (2001) [Tech. Phys. Lett. **27**, 138 (2001)].
7. *Polycrystalline Semiconductors*, Ed. by G. Harbeke (Springer, Berlin, 1985; Mir, Moscow, 1989).
8. H. Xian, P. Benalloul, C. Berthou, *et al.*, Jpn. J. Appl. Phys. **33**, 5801 (1994).
9. P. De Visschere, K. Neyts, D. Corlatan, *et al.*, J. Lumin. **65**, 211 (1995).

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 4, 2003, pp. 492–495. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 4, 2003, pp. 113–116.

RADIOPHYSICS

Magnetic Modes in a Resonator Built around a Circular Evanescent Waveguide and Coaxial Line

Yu. G. Makeev and A. P. Motornenko

Usikov Institute of Radiophysics and Electronics, National Academy of Sciences of Ukraine, ul. Akad. Proskura 12, Kharkov, 61085 Ukraine

e-mail: briz@ire.kharkov.ua

Received June 13, 2002

Abstract—The eigenfrequencies of magnetic modes in a resonator made up of two segments of a circular evanescent waveguide and a coaxial line are found by solving an electrodynamic problem. For the fundamental $H_{11\delta}$ mode, the performance of the resonator is studied analytically and experimentally over a wide range of its parameters. The feasibility of a high-power microwave semiconductor oscillator based on the resonator under study is demonstrated. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Resonators built around evanescent waveguides partially filled with a dielectric (known as waveguide dielectric resonators, WDRs [1]) have a sparse eigenmode spectrum, small weight and dimensions, and an appreciably high intrinsic Q factor. WDRs are used in measuring insulator parameters, as well as in microwave frequency-selective devices and oscillators. WDR-based devices and instruments are readily matched to both waveguides and microstrip lines. WDRs axisymmetrically filled with a multilayer (specifically, two-layer) dielectric offer improved performance.

A two-layer WDR is easily transformed into a waveguide coaxial resonator (WCR) (segments of a circular evanescent waveguide placed on both sides of a coaxial line) [2] if the inner dielectric layer in the former is replaced by a metal layer. The scattering of electromagnetic waves incident from the side of the circular waveguide on various metal-dielectric irregularities (including on a coaxial line segment) was studied in [3–5]. In those works and also in [6], which is devoted to studying the eigenmode spectrum of a coaxial waveguide resonator (CWR) (two coaxial segments attached to a circular waveguide on both sides), axisymmetric H and E modes were investigated. We are not aware of studies where the resonance properties of WCRs of the given design are investigated.

A number of WCR features were experimentally studied in our previous work [7]. It was shown that in such resonators, electromagnetic waveguide eigenmodes common to a coaxial line may coexist with *TEM* modes, which are typical of coaxial resonators.

In this work, we consider $H_{mn\delta}$ magnetic modes in a WCR that are coupled with waveguide modes in a coaxial line.

STATEMENT AND SOLUTION OF THE BOUNDARY-VALUE PROBLEM

The spectrum of $H_{mn\delta}$ magnetic eigenmodes in a WCR was sought by the method of domains using the projection procedure and vector eigenfunctions to derive a set of linear algebraic equations.

The resonator being analyzed (Fig. 1) is conventionally divided into four domains: I, the domain occupied by the inner metal conductor of a coaxial segment; II, the domain occupied by the insulating layer filling the coaxial segment; and III and IV, the semi-infinite domains of the circular waveguides adjacent to the coaxial segment on both its sides.

When solving the problem, we assume that dielectric losses are negligibly small and the metal surfaces are of infinite conductivity. Electromagnetic fields in each of the domains will be described in terms of the Hertz magnetic vector, which vanishes in domain I because of the absence of the electromagnetic field in it.

In domain II, the Hertz magnetic vector takes the form

$$\Pi_z^{\mathrm{II}} = \mathbf{z}_0\Bigg\{\sum_{m,n} A_{mn}\Bigg[J_m(\xi_{mn}r) - \frac{J'_m(\xi_{mn}b)}{N'_m(\xi_{mn}b)}N_m(\xi_{mn}r)\Bigg] \times \cos(\beta_{mn}z)e^{-imp} + \Bigg[B_{mn}\ J_m(\xi_{mn}r) - \frac{J'_m(\xi_{mn}b)}{N'_m(\xi_{mn}b)}N_m(\xi_{mn}r)\Bigg]\sin(\beta_{mn}z)e^{-imp}\Bigg\};$$

in domains III and IV,

$$\Pi_z^{\mathrm{III}} = \mathbf{z}_0\sum_{m,n} C_{mn}J_m(\xi_{mn}^{(1)}r)e^{-imp}e^{\gamma_{mn}(z+h/2)};$$

1063-7842/03/4804-0492$24.00

$$\mathbf{\Pi}_z^{\mathrm{IV}} = \mathbf{z}_0 \sum_{m,n} D_{mn} J_m(\xi_{mn}^{(1)} r) e^{-imp} e^{-\gamma_{mn}(z-h/2)}.$$

Here, A_{mn}, B_{mn}, C_{mn}, and D_{mn} are the electromagnetic field amplitudes in domains II–IV; $J_m(\xi_{mn}r)$, $N_m(\xi_{mn}r)$ and $J'_m(\xi_{mn}b)$, $N'_m(\xi_{mn}b)$ are the mth-order Bessel functions of the first and second kind and their derivatives; β_{mn}, γ_{mn} and ξ_{mn}, $\xi_{mn}^{(1)}$ are the longitudinal and transverse wavenumbers of the coaxial and circular waveguide, respectively; and h is the length of the coaxial segment.

The next step in the solution of the eigenmode problem for the resonator under study is joining the shear components of the electric and magnetic fields at the boundaries between the domains. The boundary conditions used in joining the electric fields at the II–III and II–IV boundaries are written as

$$\mathbf{r}_0 E_r^{\mathrm{III(IV)}} + \boldsymbol{\varphi}_0 E_\varphi^{\mathrm{III(IV)}} = \mathbf{r}_0 E_r^{\mathrm{II}} + \boldsymbol{\varphi} E_\varphi^{\mathrm{II}} \big| z = \pm h/2,$$

where $\mathbf{r}_0$ and $\boldsymbol{\varphi}_0$ are the unit vectors along the r and φ coordinate axes, respectively.

For the magnetic fields, the boundary conditions are written in a similar way. The shear components of the electric field at the end faces of the metal rod (domain I) were set equal to zero. The fulfillment of the boundary conditions yields a set of four functional equations involving the amplitudes A_{mn}, B_{mn}, C_{mn}, and D_{mn} of the electromagnetic fields. As was noted above, the reduction of this set to the set of linear algebraic equations was accomplished with the projection technique. Equations obtained by joining the electric fields were scalarly multiplied by the vector eigenfunction $\boldsymbol{\varphi}^{(e)}$ of the electric field of the coaxial line:

$$\boldsymbol{\phi}^{(e)}(r,\varphi) = \left\{ \mathbf{r}_0 \frac{im}{r}\left[J_m(\xi_{mk}r) - \frac{J'_m(\xi_{mk}b)}{N'_m(\xi_{mk}b)} N_m(\xi_{mk}r) \right] \right.$$

$$\left. - \boldsymbol{\psi}_0 \xi_{mk}\left[J'_m(\xi_{mk}r) - \frac{J'_m(\xi_{mk}b)}{N'_m(\xi_{mk}r)} N'_m(\xi_{mk}r) \right] \right\} e^{im\varphi}.$$

The resultant product was integrated over the resonator cross section. Equations obtained by joining the magnetic fields were multiplied by the vector eigenfunction $\boldsymbol{\phi}^{(h)}(r,\varphi)$ of the magnetic field of the circular waveguide:

$$\boldsymbol{\phi}^{(h)}(r,\varphi) = \left[\mathbf{r}_0 \xi_{mn}^{(1)} J'_m\left(\xi_{mn}^{(1)} r\right) + \boldsymbol{\varphi}_0 \frac{im}{r} J_m(\xi_{mn}^{(1)} r) \right] e^{im\varphi}.$$

The eigenfunctions were selected so as to satisfy the orthogonality condition

$$\int_S \boldsymbol{\phi}_{mn}^{(e)} \boldsymbol{\phi}_{kp}^{(e^*)} dS = N_{mn}^e \delta_{mk} \delta_{np},$$

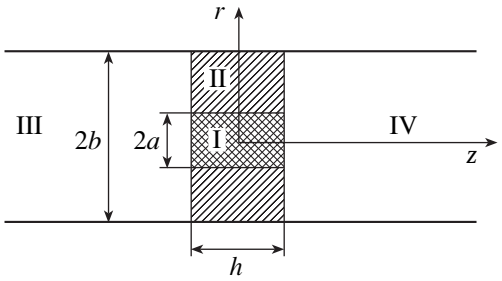

Fig. 1. Design of a waveguide coaxial resonator.

$$\int_S \boldsymbol{\phi}_{mn}^{(h)} \boldsymbol{\phi}_{kp}^{(h^*)} dS = N_{mn}^h \delta_{mk} \delta_{np},$$

where δ_{mk} and δ_{np} are the Kronecker symbols and N_{mn}^e and N_{mn}^h are the norms of the electric and magnetic eigenfunctions of the coaxial and circular waveguide, respectively.

Then, we checked that the condition of energy finiteness in each of the domains is met for any finite volume V:

$$\int_V (\varepsilon |\mathbf{E}|^2 + \mu |\mathbf{H}|^2) dV < \infty.$$

This condition virtually specifies the field at metal fins [8]. As follows from the statement of the problem, the electromagnetic field is absent in domain I and has no singularities in domains III and IV. One can show that the energy stored in domain II tends to zero at $a \longrightarrow b$ or to the energy stored in the circular waveguide of permittivity ε when $a \longrightarrow 0$.

The fulfillment of the boundary conditions yields two sets of linear algebraic equations of the second kind:

$$A_{mk} R_{mk} \cos\left(\beta_{mk}\frac{h}{2}\right) - \sum_{n=1}^{\infty} A_{mn} \frac{\beta_{mn}}{\gamma_{mn}} \left(\sin\left(\beta_{mn}\frac{h}{2}\right) \right) Q_{mk} = 0, \quad (1)$$

$$B_{mk} R_{mk} \sin\left(\beta_{mk}\frac{h}{2}\right) + \sum_{n=1}^{\infty} B_{mn} \frac{\beta_{mn}}{\gamma_{mn}} \left(\cos\left(\beta_{mn}\frac{h}{2}\right) \right) Q_{mk} = 0, \quad (2)$$

where

$$R_{mk} = \frac{\xi_{mk}\xi_{mk}^{(1)}}{\xi_{mk}^2 - \xi_{mk}^{(1)^2}} \{ \xi_{mk}^{(1)} a J_m(\xi_{mk}^{(1)} a) N'_m(\xi_{mk} a)$$

$$- \xi_{mk} a J'_m(\xi_{mk}^{(1)} a) N_m(\xi_{mk} a) - \xi_{mk}^{(1)} b J_m(\xi_{mk}^{(1)} b) N'_m(\xi_{mk} b) \},$$

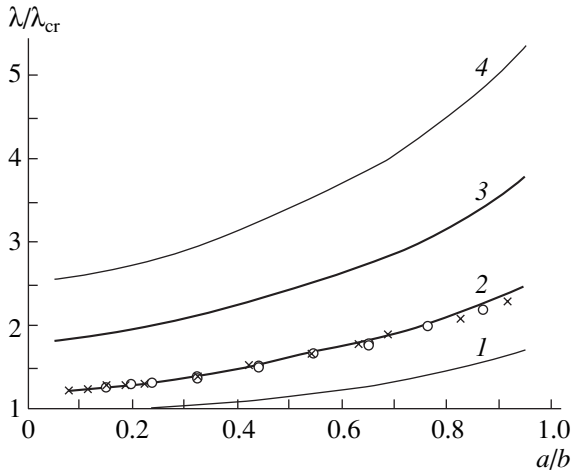

Fig. 2. Normalized resonance wavelength λ/λ_{cr} vs. a/b.

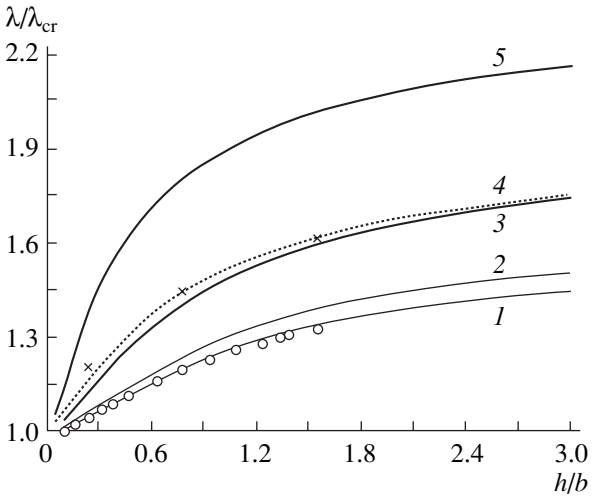

Fig. 3. Normalized resonance wavelength λ/λ_{cr} vs. h/b.

$$Q_{mk} = \sum_{p=1}^{\infty} \frac{2\xi_{mk}^2}{\gamma_{mp}} \frac{a}{b} \frac{\xi_{mp}^{(1)}\xi_{mk}^{(1)}}{\xi_{mk}^2 - \xi_{mp}^{(1)^2}} \frac{\xi_{mp}^2 a}{\xi_{mp}^2 - \xi_{mk}^{(1)^2}}$$

$$\times \frac{1}{(\xi_{mp}^{(1)^2} b^2 - m^2)} \frac{J_m'(\xi_{mk}^{(1)} a) J_m'(\xi_{mp}^{(1)} a)}{J_m^2(\xi_{mp}^{(1)} b)}$$

$$\times \left[J_m(\xi_{mp} a) - \frac{J_m'(\xi_{mp} a)}{N_m'(\xi_{mp} a)} N_m(\xi_{mp} a) \right]$$

$$\times \left[J_m(\xi_{mk} a) - \frac{J_m'(\xi_{mk} a)}{N_m'(\xi_{mk} a)} N_m(\xi_{mk} a) \right].$$

The initial equations were transformed into two uncoupled sets of $2m$ equations. This is because the resonator is uniform in φ. Thus, the 3D problem is reduced to $2m$ 2D problems. Equation (1) describes $H_{mn\delta}$ modes where δ (the number of field variations along the z axis) = 1, 3, 5, …; Eq. (2), $H_{mn\delta}$ modes where δ = 2, 4, 6, …. The resonance frequencies of symmetric and asymmetric $H_{mn\delta}$ modes are found by equating the determinant of the set of Eqs. (1) and (2) to zero.

ANALYSIS OF THE EQUATIONS

Equations (1) and (2) allow one to find the complete spectrum of magnetic modes in the resonator studied. The basic lowest frequency magnetic mode in this resonator is the $H_{11\delta}$ mode. Therefore, in subsequent analysis, we will consider this mode alone. In this case, Eqs. (1) and (2) take the simple form

$$\cos\left(\beta_{11}\frac{h}{2}\right) - \frac{\beta_{11}Q_{11}}{\gamma_{11}R_{11}}\sin\left(\beta_{11}\frac{h}{2}\right) = 0, \quad (3)$$

$$\sin\left(\beta_{11}\frac{h}{2}\right) + \frac{\beta_{11}Q_{11}}{\gamma_{11}R_{11}}\cos\left(\beta_{11}\frac{h}{2}\right) = 0. \quad (4)$$

From Eqs. (3) and (4), one can determine the resonance frequencies for the H_{111} and H_{112} modes, respectively.

Figure 2 shows (for the H_{111} mode) the dependences of the normalized resonance wavelength λ/λ_{cr} ($\lambda_{cr} = 3.41b$ is the critical wavelength for an empty circular waveguide) on a/b for several permittivities ε of a dielectric filling the coaxial line (ε = 1.0, 2.1, 5.0, and 10, respectively for curves *1–4*). The calculation was made in the one-mode approximation for $h/b = 1.34$.

Figure 3 illustrates (for the same H_{111} mode) the λ/λ_{cr} vs. h/b curves for a/b = 0.19 (*1*), 0.25 (*2*), 0.45 (*3*, *4*), and 0.75 (*5*) with ε = 2.1. The continuous curves were calculated in the one-mode approximation, while dashed curve *4* takes into account 20 modes ($k = n = p = 20$).

We studied two WCRs designed for the centimeter- and millimeter-wave ranges, respectively. The former had a circular waveguide with inner diameter 13 mm and metal-dielectric elements made of FT-4 Teflon (ε = 2.1) with brass cylinders of different diameters and lengths ($h/b = 1.34$). This resonator generated resonance oscillations in the 5.5–11.0 GHz frequency band. Experimental data for this resonator are marked by crosses in Fig. 2. The millimeter-wave resonator represented a segment of a circular waveguide of inner diameter 4.62 mm in which metal-dielectric elements with internal conductors of different inner diameter were placed (the length of the conductors was the same). Experimental data for this resonator are marked by circles in Fig. 2. Its resonance frequencies lie in the range 17–30 GHz.

In Fig. 3, circles in curve *1* and crosses in curves *3* and *4* display measured resonance frequencies of the centimeter-wave WCR. It is seen that as the length h/b of the coaxial line shortens, the error in λ/λ_{cr} calculated in the one-mode approximation grows. Also, the error

increases with the parameter a/b. From the analytical and experimental curves in Figs. 2 and 3, one can conclude that the one-mode approximation is valid over a wide range of a/b for $h/b \geq 1$. In this case, the error in determining λ does not exceed 2.5%.

CONCLUSION

Thus, we solved the problem of magnetic eigenmodes in a resonator made up of a coaxial line and two circular evanescent waveguides. Two sets of linear algebraic equations were derived, which imply that the eigenmodes of WCRs with an even and odd number of variations along the z axis are not coupled to each other, as in WDRs. Modes differing in number of variations with respect to the azimuth angle φ are also uncoupled. The resonance frequencies of the resonator are found by equating the determinant of the set of equations to zero. The error in calculating the eigenfrequency of the fundamental mode in the one-mode approximation is $\leq 2.5\%$.

Note that, using this resonator matched to a microstrip input, we succeeded in fabricating a 3-cm-range microwave semiconductor oscillator based on an AA725A Gunn diode. The oscillator thus designed had an output power higher than a WDR-based oscillator. The increased power of the WCR-based oscillator is supposedly associated with improved diode–resonator matching. The resonator suggested in this work may find wide application in frequency-selective devices used in microwave technology because of the possibility of electronic and mechanical tuning. On the basis of such resonators, techniques for measuring the parameters of both insulating materials and metal conductors can be developed.

REFERENCES

1. M. E. Il'chenko, V. F. Vzyatyshev, and L. G. Gassanov, *Dielectric Resonators*, Ed. by M. E. Il'chenko (Radio i svyaz', Moscow, 1989).
2. R. I. Belous, Yu. G. Makeev, and A. P. Motornenko, Izv. Vyssh. Uchebn. Zaved. Radioélektronik. **40** (2), 13 (1997).
3. A. A. Kirilenko, S. A. Masalov, V. P. Shestopalov, *et al.*, Preprint No. 37, IRÉ (Institute of Radioelectronics, Kharkov, 1974).
4. V. B. Budanov, V. P. Shestopalov, and V. F. Shinkarenko, Preprint No. 49, IRÉ (Institute of Radioelectronics, Kharkov, 1975).
5. A. A. Kirilenko, S. P. Senkevich, and I. S. Tsakanyan, Preprint No. 90-3, IRÉ (Institute of Radioelectronics, Kharkov, 1990).
6. Yu. K. Sirenko, V. P. Shestopalov, and N. P. Yashina, Radiotekh. Élektron. (Moscow) **32**, 535 (1987).
7. Yu. G. Makeev and A. P. Motornenko, in *Proceedings of Symposium on Physics and Engineering of Millimeter and Submillimeter Waves, Kharkov, 2001*, Vol. 2, pp. 708–709.
8. V. P. Shestopalov, A. A. Kirilenko, and L. A. Rud', *Waveguide Inhomogeneities* (Naukova Dumka, Kiev, 1976), Vol. 2.

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 4, 2003, pp. 496–502. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 4, 2003, pp. 117–123.

SURFACES, ELECTRON AND ION EMISSION

Cobalt Redistribution over the Surface of Inhomogeneous Cobalt–Copper Alloy Films

A. I. Stognij, S. V. Koriakin, and N. N. Novitskii

Institute of Solid-State and Semiconductor Physics, National Academy of Sciences of Belarus, ul. Brovki 17, Minsk, 220072 Belarus

e-mail: stognij@ifttp.bas-net.by

Received September 13, 2001; in final form, March 15, 2002

Abstract—The surfaces of electrodeposited 1-μm-thick Co_xCu_{100-x} (x = 8, 11, and 20 at. %) films and also of 0.2-μm-thick films obtained by sputtering targets made of the electrodeposited films with an argon ion beam are analyzed by atomic force microscopy, scanning electron microscopy, and X-ray photoelectron spectroscopy (XPS). XPS data indicate that cobalt is absent on the surface of the electrodeposited films but is present in the bulk and on the surface of the sputtered films. The difference in the XPS spectra of copper in the electrodeposited and sputtered films of the same composition is less significant. The data obtained are explained within the framework of a qualitative model according to which subgrains of the basic (copper) component coalesce into large clusters, which subsequently take on a regular oval shape on the free surface. This process favors cobalt atom migration from the free surface to near-surface voids. High-energy particles existing in the flux of the target sputtering products bombard the growth front of the ion-sputtered films, causing the fastest sputtered cobalt atoms to penetrate into the copper matrix as point defects. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Questions related to ordering on the surface of homogeneous metal alloys have received much attention owing to the great practical significance of these alloys and the fundamental importance of ordering processes [1–4]. For example, it was shown [1, 2] that, in homogeneous metallic alloys that have components with different chemical activities, the more inert component may form a developed porous surface structure as a result of its redistribution over the surface after removing the chemically active component, e.g., by selective dissolution in an electrolyte. Unfortunately, the composition of this porous structure was not analyzed in the works cited. The authors of [3, 4] considered melting during the deposition of chromium on the iron surface and, conversely, of iron on the chromium surface in a high vacuum. The element redistribution at the interface after the deposition was shown to play an important role in the formation of multilayer iron–chromium structures and govern their magnetic properties.

Thin films of inhomogeneous cobalt–copper magnetic alloys have also been studied extensively [5]; however, data on their surface composition are scarce. Here, attention has mainly been focused on structural ordering, which has been considered for ultrathin cobalt films on the surface of thicker copper films in multilayer structures [6] or for ultrathin cobalt films on the single-crystalline copper surface [7]. This work, in which we analyze the surface and subsurface layers of cobalt–copper thin films, is an extension of our previous investigation [8, 9]. Comparing the surface states of films obtained under equilibrium conditions of electrodeposition with those prepared by the nonequilibrium ion sputtering of the electrodeposited films as targets, we make an attempt to explain the process of cobalt ordering on the surface.

EXPERIMENTAL

Co_xCo_{100-x} ($6 \le x \le 35$) 1-μm-thick films were electrolytically deposited according to the technique described in [10]. Similarly to [10], we controlled the film composition by varying the cobalt sulfate content in the electrolyte, all other things being equal. Then, some of the electrodeposited films (EDFs) were used as targets for ion sputtering. The ion sputtering was performed with the setup described in [11]. The argon ion beam had an energy of 1 keV and a current density of 0.25 mA/cm^2. The ultimate pressure was no higher than 10^{-4} Pa, and the working pressure was no higher than 2×10^{-2} Pa. The sputtering products were films about 0.2 μm thick (hereafter, ion-beam films (IBFs)). EDF and IBF substrates measuring 20×20 mm were made of one piece of 50-μm-thick rolled copper foil. Immediately before the deposition, the substrate surface was etched in a 5% hydrochloric acid solution and then sequentially cleaned with oxygen and argon ion beams. In both cleaning modes, the ion current density was 0.2 mA/cm^2 and the beam energy was 400 eV. Visual inspection was made with an NU-2 optical microscope (Germany) with a magnification of up to 600. To examine the surfaces and to determine the volume composition of the films, we used a Nanolab-7 scanning electron microscope (SEM) equipped with a System 810-

1063-7842/03/4804-0496$24.00

500 energy dispersion analyzer (Great Britain). The surfaces were also studied with a Femtoskan-001 atomic force microscope (AFM) (Moscow State University, Russia) equipped with a Park Scientific (USA) silicon cantilever and operating in the contact mode. Surface layers thinner than 5 nm were analyzed by an XPS spectrometer with a magnesium cathode (MgK_α radiation with a photon energy of 1253.6 eV) as an X-ray source. The energy scale was calibrated against the C(1*s*) line and additionally against the O(1*s*), Co(2*p*), and Cu(2*p*) lines. In this case, a 5×5-mm surface area of the samples was preliminary cleaned by an argon ion beam of energy 500 eV and current density 10 μA in the spectrometer for 5 min. When analyzing near-surface layers, we cleaned the sample surface for a longer time.

EXPERIMENTAL RESULTS

The surface analysis of the EDFs of various composition demonstrated that only the films with $8 \le x \le 20$ at. % Co can be used as targets. Those with a cobalt content $x < 8$ at. % did not have metallic luster and consisted of irregularly shaped aggregates. Figure 1a shows a typical SEM micrograph of the Co_6Cu_{94} EDF. The surface is seen to be discontinuous with agglomerates as large as 1 μm. The discontinuous surface of the EDFs with a cobalt content $x < 8$ at. % does not allow for surface analysis with XPS and AFM. Moreover, because of the uncertain thickness of these films, we cannot specify ion sputtering conditions. The EDFs with a cobalt content $x \ge 8$ at. % had a continuous surface with characteristic metallic luster. Figure 1b shows a typical SEM surface micrograph from an EDF with the cobalt content $x = 8$ at. %. The surface is fairly continuous and consists of uniformly distributed valleys and ridges with a small number of regularly shaped inhomogeneities. Their size is much smaller than that of the inhomogeneities on the surface of the EDFs with a cobalt content $x < 8$ at. %. As the cobalt content increases from 8 to 35 at. %, the surface of the EDFs becomes glossy, the surface color changes from reddish to white, and the number and size of large surface irregularities decrease. However, during the ion sputtering of the EDFs with $x > 20$ at. %, we observed cracking and swelling in several film areas to the extent they peel off from the substrates. As a result, the copper substrate makes an uncontrollable contribution to the composition of the sputtered material flux. Therefore, films with $20 < x < 35$ at. % are not considered below.

The Cu(2*p*) and Co(2*p*) lines of the XPS spectra taken from the surfaces of the Co_8Cu_{92}, $Co_{11}Cu_{89}$, and $Co_{20}Cu_{80}$ EDFs are shown in Fig. 2. The alloy composition was determined by microprobe analysis. The Cu(2*p*) line of the Co_8Cu_{92} film is the weakest compared to the Cu(2*p*) lines of the other alloys, although the volume concentration of copper in the Co_8Cu_{92} film is the highest. The additional ion cleaning of the surfaces for 10 min noticeably increased the intensity of the Cu(2*p*) line only for the $Co_{11}Cu_{89}$ EDF (Fig. 2c), leaving the shapes of the spectra almost unchanged. This means that the bulk and surface copper contents are different. However, an increase in the surface concentration of copper with as its bulk content decreases may also be caused by a relative increase in the surface density of the EDFs. This assumption agrees with the fact that the surface smoothness and homogeneity improve with increasing cobalt content, as follows from the SEM images and visual inspection. Furthermore, the intensities of the C(1*s*) and O(1*s*) lines taken from the surface of the as-deposited Co_8Cu_{92} sample were the highest in comparison with the other samples and changed only slightly after ion cleaning of various duration, whereas those for the $Co_{11}Cu_{89}$ and $Co_{20}Cu_{80}$ samples decreased to constant values after the ion cleaning. This finding also evidences a higher surface density of the EDFs with a high cobalt content. The uniform cobalt distribution across EDFs of the given compositions, which was found earlier [8–10], along with the assumption that the cobalt distribution over the surface remains constant, allowed us to suppose that the Co(2*p*) line in the XPS spectra from the sample surfaces studied also has specific features caused by the

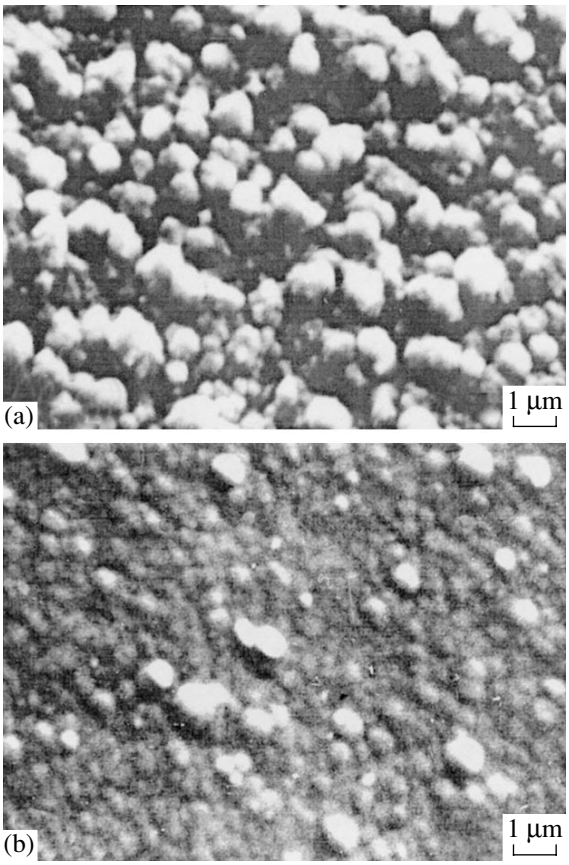

Fig. 1. SEM images of the surfaces of electrodeposited (a) Co_6Cu_{94} and (b) Co_8Cu_{92} films.

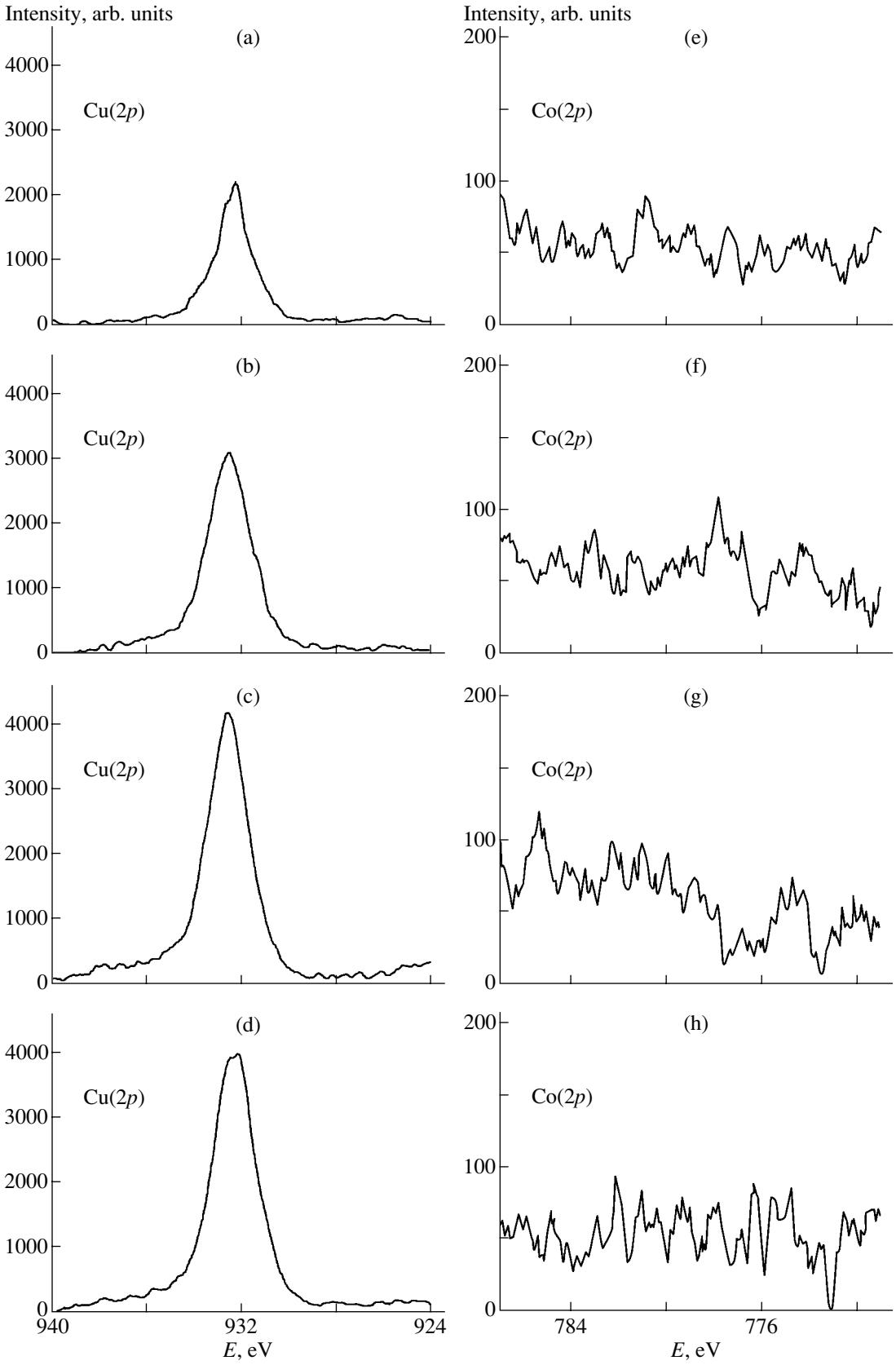

Fig. 2. XPS spectra from the electrodeposited films. The Cu(2*p*) line for the films (a) Co_8Cu_{92}, (b) $Co_{11}Cu_{89}$, (c) $Co_{11}Cu_{89}$ after additional ion cleaning, and (d) $Co_{20}Cu_{80}$. The Co(2*p*) line for the films (e) Co_8Cu_{92}, (f) $Co_{11}Cu_{89}$, (g) $Co_{11}Cu_{89}$ after additional ion cleaning, and (h) $Co_{20}Cu_{80}$.

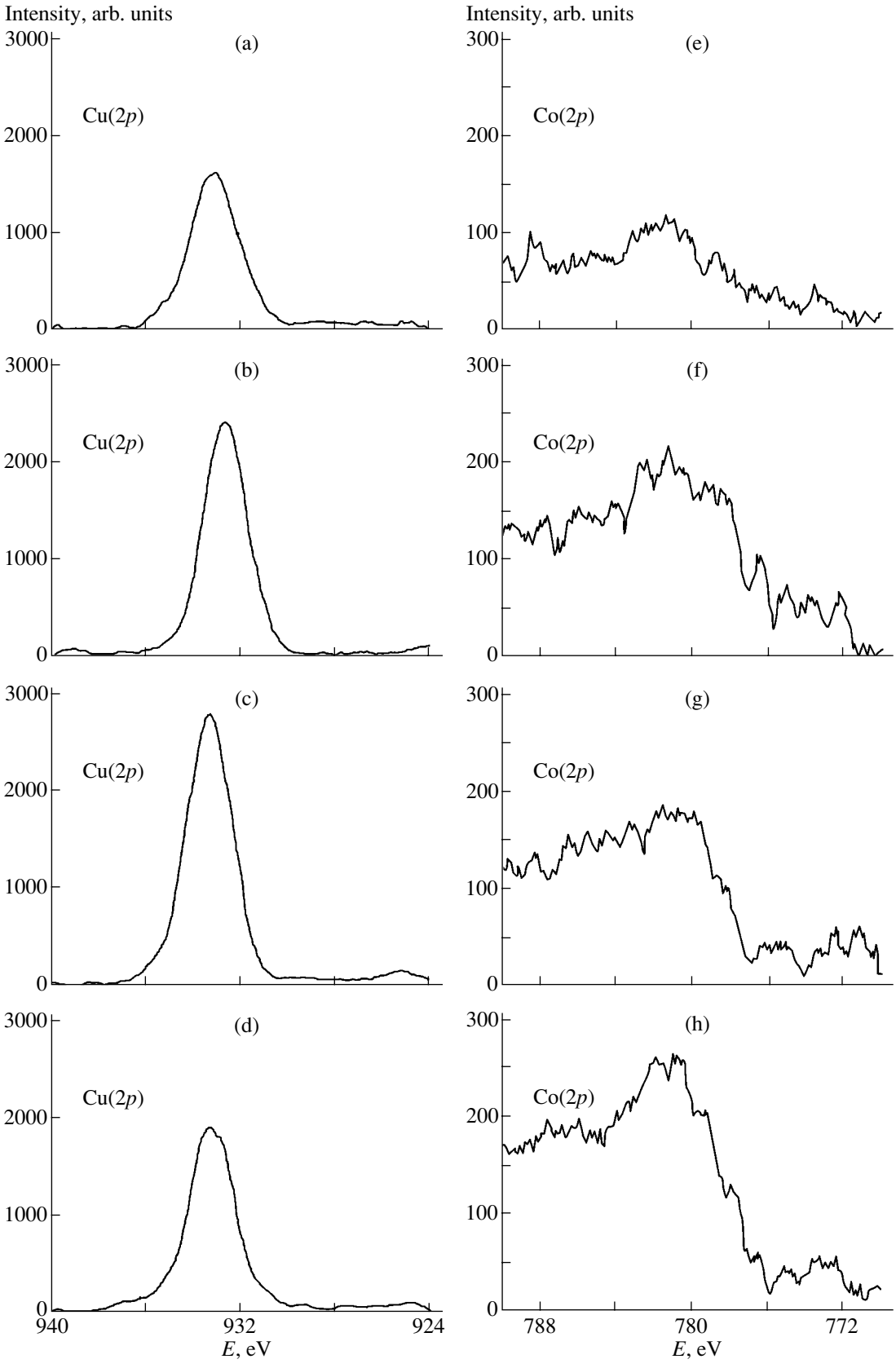

Fig. 3. XPS spectra from the ion-sputtered films. Cu(2*p*) line for the films (a) Co_8Cu_{92}, (b) $Co_{11}Cu_{89}$, (c) $Co_{11}Cu_{89}$ after additional ion cleaning, and (d) $Co_{20}Cu_{80}$. Co(2*p*) line for the films (e) Co_8Cu_{92}, (f) $Co_{11}Cu_{89}$, (g) $Co_{11}Cu_{89}$ after additional ion cleaning, and (h) $Co_{20}Cu_{80}$.

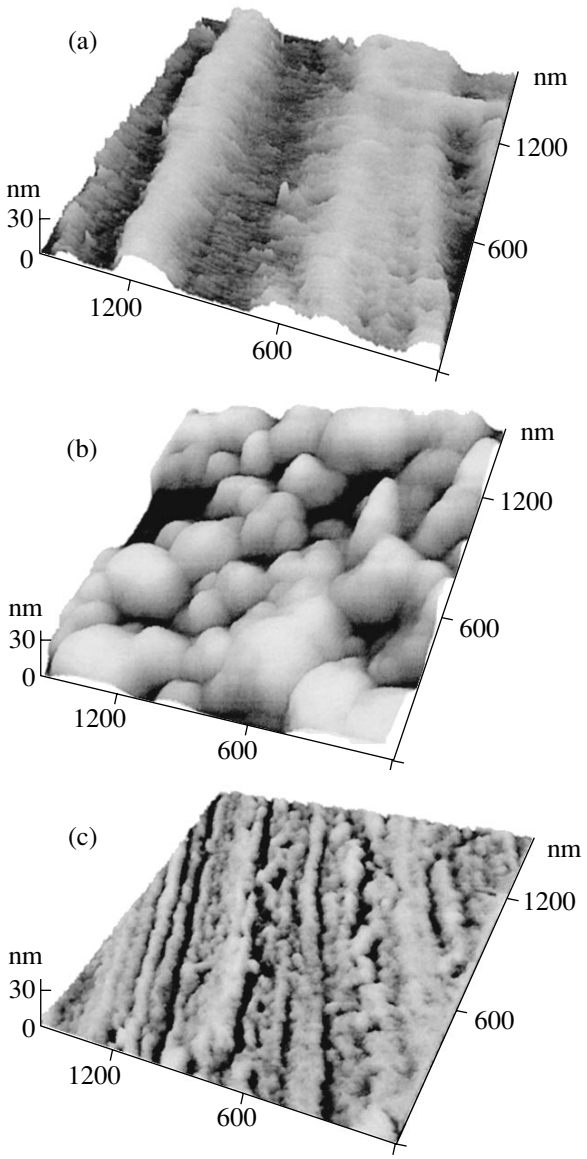

Fig. 4. AFM images of the surfaces of the (a) copper substrate, (b) electrodeposited $Co_{11}Cu_{89}$ films, and (c) ion-sputtered $Co_{11}Cu_{89}$ films.

specific variation of the surface component concentration with bulk copper concentration. However, we failed to resolve this line in the XPS spectra taken both from the samples with various Co contents (Figs. 2e, 2g) and after ion cleaning of various durations (Fig. 2). As follows from these spectra, cobalt is absent in surface layers of thickness less than 5 nm in both the as-deposited and ion-etched samples.

According to the microprobe analysis data, IBF samples about 0.2 μm thick had the same composition as the EDFs used as targets for sputtering. All the IBFs had identically smooth surfaces with metallic luster, although those with a high cobalt content in the bulk were more whitish. The XPS spectra from the IBF surfaces are given in Fig. 3. As for the EDFs, the Cu(2*p*) line from the Co_6Cu_{94} surface (Fig. 3a) has a lower intensity compared with the equally intense Cu(2*p*) lines from the surfaces of $Cu_{11}Co_{89}$ (Fig. 3b) and $Co_{20}Cu_{80}$ (Fig. 3d). The additional ion cleaning of the Co_6Cu_{94} surface changed the shape and intensity of the Cu(2*p*) line insignificantly. For all the IBFs, the intensities of the carbon and oxygen lines were maximum for the as-deposited samples, decreased two or three times after ion cleaning for several minutes, and then leveled off. The leveling off indicates that most of the organic impurities are located directly on the surface and that their appearance is caused mainly by the storage conditions. The rest of the impurities, which have a constant concentration over the thickness, are residual impurities in the vacuum chamber during sputtering and those present in the targets. The variation of the Cu(2*p*) line of the XPS spectra with the bulk composition of the films suggests that, as for the EDFs, the surface density of copper in the $Co_{11}Cu_{89}$ and $Co_{20}Cu_{80}$ films is higher than in the Co_8Cu_{92}. However, unlike the EDFs, the Co(2*p*) lines in the XPS spectra from the IBF surfaces are resolved for all the samples (Figs. 3e–3h). The higher the bulk cobalt content, the higher the intensity of the Co(2*p*) line (Figs. 3e, 3f, 3h). The XPS spectrum of cobalt depends only slightly on the ion cleaning duration (Fig. 3g). It has a complex shape for all the IBFs and contains a number of peaks, which are shifted toward higher binding energies with respect to the value for metallic cobalt (778 eV), indicating the inhomogeneous environment of cobalt atoms in the surface layer [12].

AFM images taken of characteristic regions on the substrate and on the electrodeposited and sputtered $Co_{11}Cu_{89}$ films are shown in Fig. 4. The texture of the substrate surface, which is associated with rolling during the manufacture of the copper foil (Fig. 4a), is not observed on the EDF surface (Fig. 4b). The IBF surface (Fig. 4c) is more uniform along the rolling direction than across it. It is seen that even an IBF as thin as ≈0.2 μm smoothes out the substrate surface relief height from an initial 30 nm to less than 10 nm. The EDF surface consists of relatively large islands, most of which have a smooth shape and a height of several tens of nanometers; however, their transverse sizes vary. Between the islands, one can observe shallow small-area and rare deep large (100 × 100 nm) depressions. Such a surface appearance is typical of films prepared by electrodeposition. The large depressions may be related to lower density regions or dead-end pores in electrodeposited films [13].

DISCUSSION

The experimental data indicate differences in the bulk and surface ordering processes in inhomogeneous CoCu alloys. Cobalt is fairly uniformly distributed only over the volume of the EDFs, including the ≈0.1-μm-thick surface layer, which agrees with the early results

[8–10]. Directly on the surface of the EDFs (Figs. 2e, 2f), cobalt is absent, according to the XPS data. During the continuous ion sputtering of the films (at a rate of 10–30 nm/min in our case), cobalt has no time to be redistributed over the surface and is sputtered together with the copper matrix, its portion in the sputtered flux being proportional to its bulk concentration. Therefore, the compositions of the IBFs and their targets made of the corresponding EDFs coincide. When ion cleaning in the XPS spectrometer or ion sputtering in the ion-beam setup is interrupted, a cobalt deficiency on the EDF surface is observed again (Figs. 2g, 2h). Our earlier investigation of CoCu films by conversion electron Mössbauer spectroscopy [9] showed that a resonance iron atom was detected only in the pure copper or mixed cobalt–copper environment in a surface layer 0.15 μm thick. Therefore, we can assume that cobalt near the free surface of the films is also in the form of small nanoclusters distributed over the copper matrix surface. If the islands of regular shape (Fig. 4b) are considered as the free surface of the copper matrix, which forms during the EDF growth via the coalescence of 10-nm subgrains, as was shown earlier in [10], we may assume that cobalt atoms, migrating over the surface, occupy energetically more favorable positions near the lower density regions (surface voids) rather than directly on the free surface of the copper matrix (if it is additionally taken into account that bulk copper and cobalt do not form alloys). During the ion sputtering, the surface is exposed to an ion-beam plasma and bombarding ions and ceases to be free. Simultaneously, conditions for the surface redistribution of cobalt change. As a result, usual layer-by-layer sputtering takes place; that is, both copper and cobalt are sputtered in amounts proportional to their bulk concentrations over long time intervals. After the termination of ion bombardment, the free surface takes on an appearance close to the original, which was also shown in [8]. In the case of EDFs, the surface deficiency of cobalt exists both before (Figs. 3e, 3f) and after (Figs. 3g, 3h) the ion bombardment. This means that this deficiency can be regarded as a recoverable quantity that is caused by ordering processes on the surface of the inhomogeneous alloys rather than by the application of XPS to examine the surfaces [12].

To explain the difference between the surface condition of EDFs and IBFs of the same composition that are obtained on identical substrates, one should take into account that the deposition conditions for IBFs are more nonequilibrium than for EDFs. One reason for this nonequilibrium is the form of the energy distribution function for the flux of sputtered cobalt and copper atoms. This flux can be conventionally divided into two parts: the main part, consisting of sputtered atoms with a mean energy approximately equal to the heat of evaporation (4.43 eV/atom for cobalt and 3.52 eV/atom for copper), and the high-energy part, consisting of atoms with energies of 40–200 eV [14]. The energy of atoms in the high-energy tail of the distribution function turns out to be sufficient for the "self-irradiation" of the IBF growth surface. According to estimates made with the SRIM2000 program (www.srim.org), cobalt atoms with an energy of 50 eV have a range of 0.4 ± 0.1 nm in a massive copper matrix and generate 0.6 vacancies/ion; for those with an energy of 100 eV, the range is 0.5 ± 0.1 nm and the yield is 1.4 vacancies/ion; and for those with an energy of 150 eV, these parameters equal 0.5 ± 0.2 nm and 2.2 vacancies/ion. Hence, some cobalt atoms from the high-energy tail can penetrate into the copper matrix during deposition to form point defects. Naturally, the concentration of such cobalt atoms in the surface layer is significantly lower than in the bulk and their XPS spectra differ substantially from the XPS spectrum from the pure cobalt surface [12] because of the higher density copper environment, which agrees with Figs. 3e–3h. The existence of the self-irradiation mode is indirectly supported by the fact that depressions on the substrate surface are covered more readily than asperities during the IBF growth and that the IBFs are more uniform along the rolling direction (Fig. 4c). These specific features indicate the anisotropy of the deposition process, which may arise when energetic sputtered atoms from the tail strike the surface and favor the surface migration of adatoms along the rolling direction. Irradiation upon sputtering also results in the formation of finer and more closely spaced islands on IBFs as compared with EDFs of the same composition.

CONCLUSION

Inhomogeneous CoCu alloy films prepared by electrodeposition and ion-beam sputtering show a significant difference in cobalt distribution, which should be taken into account when analyzing the properties of the alloy. The redistribution of unbound cobalt over the free surface of the copper matrix during electrodeposition and the generation of point defects due to cobalt atoms in copper grains upon ion-beam sputtering should be taken into consideration when studying the formation of interfaces in inhomogeneous CoCu alloy films, especially in the production of multilayer structures.

ACKNOWLEDGMENTS

We thank V.M. Fedosyuk and T.A. Tochitskii for encouragement and valuable discussions and also O.M. Stukalov and K.N. Kasparov for assistance in the experiments.

REFERENCES

1. M. Stratmann and M. Rohwerder, Nature **410**, 420 (2001).
2. J. Erlebacger, M. J. Aziz, A. Karma, *et al.*, Nature **410**, 450 (2001).
3. A. Davies, J. A. Stroscio, D. T. Priece, *et al.*, Phys. Rev. Lett. **76**, 4175 (1996).

4. Y. J. Choi, I. C. Jeong, J.-Y. Park, *et al.*, Phys. Rev. B **59**, 10918 (1999).
5. A. P. Keefe, O. I. Kasyutich, W. Schwarzacher, *et al.*, Appl. Phys. Lett. **73**, 1002 (1998).
6. M. Shima, L. Salamanca-Riba, T. P. Moffat, *et al.*, J. Magn. Magn. Mater. **198–199**, 52 (1999).
7. R. Allenspach, A. Bischof, and U. Dürig, Surf. Sci. **381**, L573 (1997).
8. A. I. Stognij and S. V. Koriakin, Poverkhnost: Fiz. Khim. Mekh., No. 6, 74 (2001).
9. A. I. Stognij, S. V. Koriakin, and V. A. Virchenko, Zh. Tekh. Fiz. **71** (6), 87 (2001) [Tech. Phys. **46**, 729 (2001)].
10. T. A. Tochitskii, G. A. Jones, H. J. Blythe, *et al.*, J. Magn. Magn. Mater. **224**, 221 (2001).
11. A. I. Stognij, V. T. Svirin, S. D. Tushina, *et al.*, Prib. Tekh. Éksp., No. 3, 151 (2001).
12. *Practical Surface Analysis by Auger and X-ray Photoelectron Spectroscopy*, Ed. by D. Briggs and M. P. Seach (Wiley, New York, 1983).
13. P. G. Cheremskoĭ, V. V. Slezov, and V. I. Betekhtin, *Pores in Solids* (Énergoatomizdat, Moscow, 1990).
14. M. V. Thompson, Nucl. Instrum. Methods Phys. Res. B **18**, 411 (1987).

Translated by K. Shakhlevich

Technical Physics, Vol. 48, No. 4, 2003, pp. 503–507. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 4, 2003, pp. 124–128.

SURFACES, ELECTRON AND ION EMISSION

Properties of a Metal-Ceramic Cathode

Yu. A. Kotov, S. Yu. Sokovnin, and M. E. Balezin

Institute of Electrophysics, Ural Division, Russian Academy of Sciences, ul. Komsomol'skaya 34, Yekaterinburg, 620219 Russia

e-mail: sokovnin@iep.uran.ru

Received January 23, 2002

Abstract—The characteristics of a metal-ceramic (MC) cathode designed for nanosecond electron accelerators are studied in relation to the size of metal particles, their density on the ceramic surface, and the type of ceramics. A high emissivity of the MC cathode at a moderate electric field in the diode tube is demonstrated. This allows one to significantly sharpen the electron beam and increase its current (power). The possibility of controlling the emissivity of the MC cathode by varying the composition of the MC plate is established. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The advent of powerful nanosecond generators with a pulse repetition rate of up to several kilohertz and a service life of 10^{10}–10^{11} pulses [1, 2] has motivated the design of electron accelerators that are promising for applications [3, 4].

These accelerators require inexpensive and durable cathodes with reproducible parameters. In [5], we described an MC cathode that is a 2-mm-thick pellet with a diameter of 12 mm made of oxide nanoceramics incorporating metal particles that are fairly uniformly distributed over its volume, with some of them lying at the surface. Because of the great difference in the temperature coefficients of linear expansion of ceramics and metals, the postsynthesis cooling of this system leaves microvoids around the metal particles, which are likely to serve as gas sources during the formation of a plasma on the MC cathode surface.

Early studies [6, 7] revealed that this cathode offers high emissivity (does not limit the power of an accelerator and ensures a rapid rise of the current at a relatively small emitting area), allows for a highly uniform electron density distribution on the anode, and sharpens the accelerator power (as compared with the metal-dielectric cathode of the same size with ten emitting centers that was previously used in our accelerator [4]).

In this study, we consider the characteristics of the MC cathode in relation to the size D of metal particles, their density n on the surface, and the type of ceramics.

EXPERIMENTAL

We used a URT-0.5 nanosecond accelerator [4] with an accelerating voltage $U \le 500$ kV, pulse half-height width $t_p \approx 50$ ns, first-transition duration $\tau_{10.1–0.9} = 46$ ns, and pulse repetition rate $f \le 200$ Hz. An AVR-50 pump reduced the pressure in the diode tube chamber to 10^{-1}–10^{-2} Pa without freezing-out oil vapors.

The design of the cathode unit was similar to that described in [6]: the MC plate was placed edgeways as before but mounted on the cathode holder with a collet. Figure 1 shows stainless steel particles on the MC plate surface.

The measuring system was the same as in the experiment [6]. We also recorded the integral distribution of the plasma glow over the MC plate by means of a Zenit-E open-shutter photographic camera on an RF-3 photographic film through a transparent window (made of Sol organic glass) in the vacuum chamber wall opposite the plate.

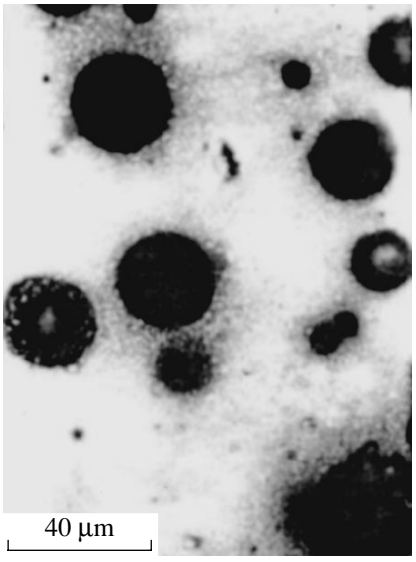

Fig. 1. Typical view of the MC plate surface.

1063-7842/03/4804-0503$24.00 © 2003 MAIK "Nauka/Interperiodica"

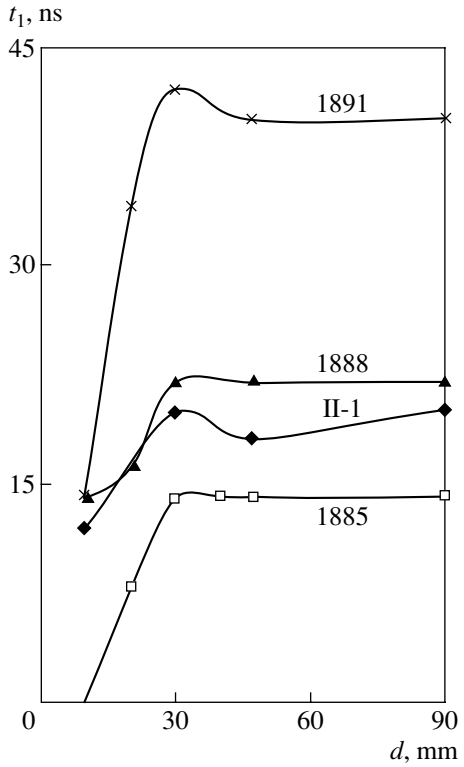

Fig. 2. Time delay between the voltage pulse arrival and the appearance of the current pulse for MC cathodes of different composition vs. the anode–cathode spacing. The figures by the curves are cathode numbers.

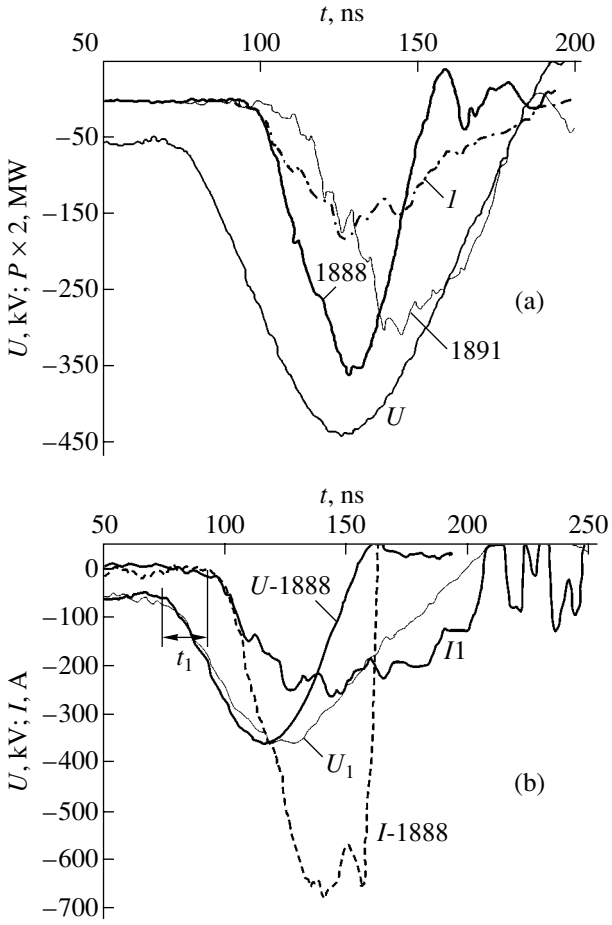

Fig. 3. (a) Voltage U for cathode no. 1891 and the electron beam power for different cathodes and (b) the voltages and current for the MDM cathode (*1*) and MC cathode no. 1888. d = 47 mm.

RESULTS AND DISCUSSION

The parameters of the MC plates studied are listed in Table 1. For all the plates, the diode current almost always has a delay t_1 relative to the voltage (Figs. 2, 3b). The value of t_1 varies with the cathode–anode spacing d (the distance between the edge of the ceramic pellet of height h and the anode plane [6]) in the same way for cathodes of all types: namely, the delay time increases with $d \le 30$ mm and remains practically constant at $d \ge 30$ mm. Such behavior indicates the presence of the total voltage effect [6], which is due to both the shape of the electrodes and the desorption of the gas (especially from microvoids).

Table 1. Characteristics of the MC cathodes under study

Cathode no.	Ceramics	D, μm	n, 1/cm^2
1885			
1886	Al_2O_3	9	1900
1887			
1888	Al_2O_3	9	40 000
1889			
1890			
II-1	Al_2O_3	26	4700
1891			
1892	TiO_2	9	4020
1893			

For Al_2O_3 ceramics, the value of t_1 grows only slightly with the density of the particles on the MC cathode surface (1885 and 1889 in Table 2) and is almost independent of their size (1885, II-1). When Al_2O_3 ceramics (relative permittivity ε = 9.6) is replaced by TiO_2 (ε = 170), t_1 increases nearly twofold (II-1 and 1891).

Therein lies the radical distinction of the process discussed from a discharge over the dielectric surface in a vacuum [8]. In the latter case, both the firing voltage and the current delay decrease with an increase in the permittivity and height of the dielectric plate in the vacuum gap.

By properly choosing the time moment of connecting a load during the rapid growth of the impedance of a current interruptor (of any type), one can improve the operating conditions for the interruptor and, thus, raise

the output power [9]. Since the current break time for the semiconductor interruptor used in the experiment is no more than $\tau = 30$ ns, the output power first grows with t_1 and then (when $t_1 > \tau$) lowers (Table 2). In the latter case, the power peaks at the trailing edge of the voltage pulse (Fig. 3a, 1891).

The efficiency of energy transfer varies nearly linearly with t_1 (Table 2). However, in case of the MC plate no. 1891, where $t_1 > \tau$, the efficiency grows because of an increase in the fraction of low-energy electrons forming at the trailing edge of the voltage pulse.

The maximal power P_m of the accelerator, which was observed for the MC cathodes nos. 1888–1890 (Table 1), was found to be twice as high as the P_m value for the metal–dielectric–metal (MDM) cathode (Fig. 3a, Table 2). With an MC cathode, the rate of current rise appreciably exceeds that achievable with an MDM cathode [7]. In our experiment, at $d = 47$ mm, the mean rate of current rise di/dt (from its beginning to a voltage level of 0.9) comprised 0.84 and 1.7×10^{10} A/s and the amplitude of the current equaled 177 and 358 A for the MDM and MC 1888 cathodes, respectively. In parallel with this, the half-height duration of the voltage pulses shortens from 68 to 50 ns and that of the current pulses, from 92 to 44 ns for the MDM and MC 1888 cathodes, respectively.

From the close coincidence of the rates of rise and amplitudes of the voltage in both cases (Fig. 3b), we conclude that the output power of the accelerator may be not only sharpened but also raised by using an MC cathode. This fact may be explained by a higher emissive capacity of an MC cathode combined with a higher rate of emissivity increase due to a rapidly expanding emitting plasma surface area and/or a high density of the plasma.

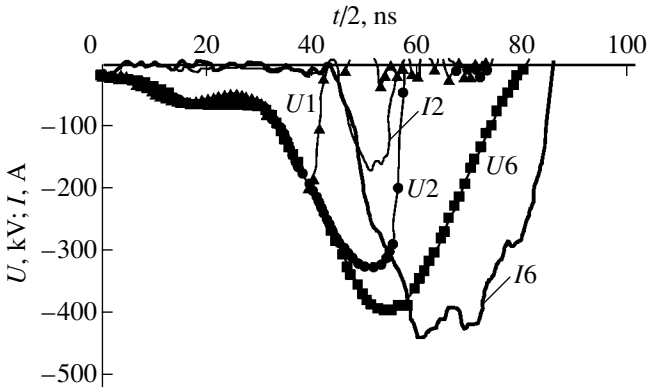

Fig. 4. Accelerating voltages ($U1$, $U2$, and $U6$) and anode currents ($I2$ and $I6$) for a different number of sections in the vacuum insulator.

In order to observe the MC cathode operation without the arc stage, which complicates the interpretation of the experimental data, we cut the voltage pulse by breaking down the vacuum insulator [10] (by changing the number of its sections, Fig. 4). Under these conditions, the photographic film detects only a weak glow at the points where the MC plate contacts the cathode holder even if 30 pulses per frame are recorded. We thus may conclude that the plasma density on the MC plate surface is low and the surface discharge is always initiated at the same points of contact between the MC plate and the cathode holder.

The data obtained suggest that a low-density plasma appears almost simultaneously over a considerable area of the plate surface, which explains the high emissive

Table 2. Experimental results for $d = 47$ mm

Cathode no.	t_1, ns	Power, MW	I, A	Beam energy, J	Efficiency of energy transfer to the beam, %
1891	40	155	428	6.73	30.3
1889	22	181	581	5.65	25.5
II-1	18	125	330	5.24	23.6
1885	14	137	556	4.97	22.4
MDM	14	91	251	4.33	19.5

Table 3. Experimental and calculated parameters for $d = 47$ mm

Sample no.	t_1, ns	U_{i0}, kV	E_{i0}, kV/cm	E_1, kV/cm	$U0$	R_{max}, Ω
1885	14	149	4.93	2669	9.56	985
II-1	18	200	6.64	790	2.10	1106
1890	22	246	8.14	960	2.08	1037
1891	40	376	0.77	287	0.37	2227

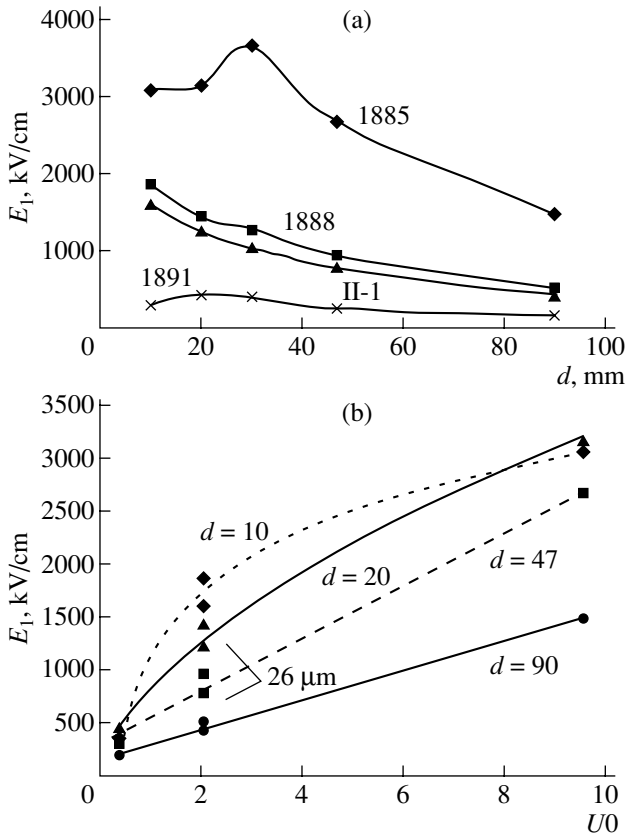

Fig. 5. Field strength E_1 in a microvoid vs. (a) the spacing d and (b) the dimensionless voltage $U0$ at different d.

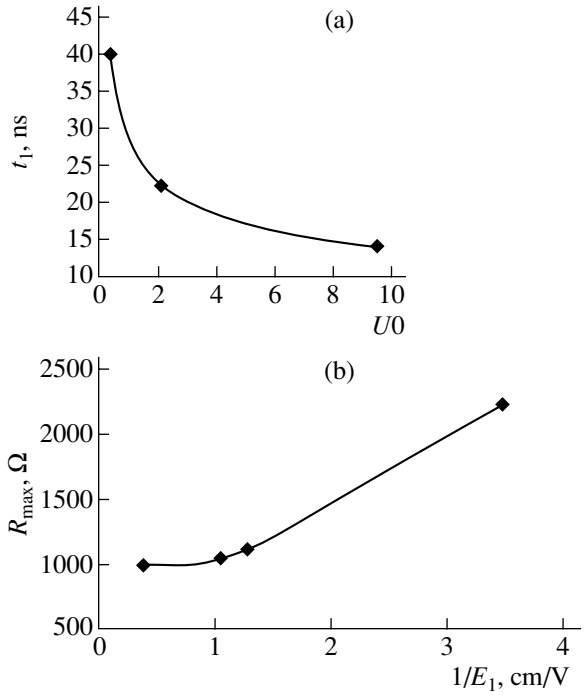

Fig. 6. (a) Delay time t_1 vs. the dimensionless voltage $U0$ and (b) the impedance R_{max} for the voltage peak vs. the field strength E_1 in a microvoid. d = 47 mm.

capacity of the cathode, as well as its long service life and reproducible characteristics. This large-area formation of the plasma is apparently caused by the breakdowns of micropores around the metal particles in the MC plate.

The delay time t_1 calculated on the assumption that the metal particles have the same diameter D (Table 3) appears to be a linear function of the initial (at the instant of current appearance in the diode) electric field strength E_{io} averaged over the MC plate surface. This is hardly probable, and the hypothesis [6] for the decisive role of processes occurring in micropores rather than on the MC plate surface is thus substantiated. As follows from Table 3, for the same diameter D of the particles, t_1 varies inversely with the average field strength E_1 in a pore of width δ ($E_1 \approx E_{i0}/(n^{0.5}\delta)$), which is in line with the present-day concepts.

It should be noted that, at a given field strength, the current delay in our experiment is on the same order of magnitude as that in the case of explosive emission [11], where $t_1 \approx E^4$. In our case, however, $t_1 \approx E^2$. Physical reasons behind this difference are still to be clarified.

Let us take the dimensionless voltage $U0 \approx 1/(\varepsilon n^{0.5} D)$ arising when a metal particle is charged by the displacement currents at U_{i0} (with its size and capacitance taken into consideration) as a parameter and plot the average field strength in a micropore E_1 versus $U0$ (Fig. 5b). It is seen that, as d decreases, the nearly linear dependence transforms into logarithmic, which seems to be explicable in terms of the total voltage effect.

In essence, the parameter $U0$ is a generalized characteristic of the emissive properties of an MC plate. The introduction of $U0$ helps in understanding why cathodes with substantially different properties may have close values of t_1. For example, for the cathodes II-1 and 1890, $U0$ = 2.08 and 2.10, respectively. The values of t_1 are also close (Table 3).

Thus, the MC cathode properties can be predicted based on the composition of an MC plate, the parameter $U0$, and the data obtained (Fig. 6). In particular, the increase in the particle diameter in the plate II-1 to D = 50 μm yields $t_1 \approx 28$ ns.

An important parameter defining the cathode properties is the diode impedance. It is desirable to be able to estimate the impedance at the instant of the peak voltage R_{max}, when the beam energy is maximum. Table 3 implies that R_{max} varies inversely with $U0$; however, the dependence of R_{max} on $1/E_1$ seems to be more evident (Fig. 6b). According to this curve, the increase in the diameter of the particles in the plate II-1 to D = 50 μm gives E_1 = 590 kV/cm and $R_{max} \approx 1300$ Ω; i.e., the emissive capacity of the cathode degrades.

The aforesaid substantiates the view that, before the onset of the current, the electric field distribution is

purely capacitive and that the ceramic material and the amount and size of the metal particles play an essential role. This means that most particles contained in the MC plate are involved in the process.

CONCLUSIONS

MC cathodes studied in this work are demonstrated to have a number of features that make them promising for industrial accelerators.

The high emissive capacity of the cathodes at a relatively low electric field in the diode tube makes it possible to greatly sharpen and increase the current (beam power), i.e., to reduce the low-energy component of the beam spectrum without using any special facilities and, hence, to improve the efficiency of beam extraction.

There appears the possibility of controlling the emissivity of an MC cathode by varying the MC plate parameters (the size and surface density of metal particles, as well as the ceramics permittivity).

In contrast to a discharge over the dielectric surface in a vacuum [8], the voltage U_{io} corresponding to the appearance of the current grows with increasing permittivity and decreasing height h of the MC plate.

The long-term exploitation (10^8 pulses or more) does not change the MC cathode characteristics: they remain highly reproducible from pulse to pulse and from sample to sample of the same type.

ACKNOWLEDGMENTS

We are grateful to I.V. Beketov, O.M. Samatov, A.M. Murzakaev, V.R. Khrustov, and V.M. Tel'nova for the preparation of the MC plates. We thank V.I. Kozhevnikov for valuable assistance in the experiment.

This work was partially supported by the Promising Materials Ural Scientific and Educational Center within the framework of the Civilian Research and Development Foundation (CRDF) (grant no. REC-005).

REFERENCES

1. Yu. A. Kotov, G. A. Mesyats, S. N. Rukin, *et al.*, in *Digest of the 9th IEEE Pulsed Power Conference, Albuquerque, 1993*, Vol. 1, p. 134.
2. Yu. A. Kotov, G. A. Mesyats, S. R. Korzhenevskiĭ, *et al.*, in *Proceedings of the 10th IEEE Pulsed Power Conference, Santa Fe, New Mexico, 1995*, p. 1231.
3. Yu. A. Kotov and S. Yu. Sokovnin, Prib. Tekh. Éksp., No. 4, 84 (1997).
4. Yu. A. Kotov, S. Yu. Sokovnin, and M. E. Balezin, Prib. Tekh. Éksp., No. 2, 112 (2000).
5. Yu. A. Kotov, S. Yu. Sokovnin, and M. E. Balezin, RF Patent No. 2158982.
6. Yu. A. Kotov, E. A. Litvinov, S. Yu. Sokovnin, *et al.*, Dokl. Akad. Nauk **370**, 332 (2000) [Dokl. Phys. **45**, 18 (2000)].
7. Yu. A. Kotov, S. Yu. Sokovnin, and M. E. Balezin, in *Proceedings of the 12th Symposium on High Current Electronics, Tomsk, 2000*, p. 38.
8. S. P. Bugaev, V. V. Kremnev, Yu. I. Terent'ev, *et al.*, Zh. Tekh. Fiz. **41**, 1958 (1971) [Sov. Phys. Tech. Phys. **16**, 1547 (1971)].
9. Yu. A. Kotov and A. V. Luchinskiĭ, *Physics and Technology of Powerful Pulsed Systems*, Ed. by E. P. Velikhov (Énergoatomizdat, Moscow, 1986), pp. 189–210.
10. Yu. A. Kotov, A. L. Filatov, S. Yu. Sokovnin, *et al.*, Prib. Tekh. Éksp., No. 2, 138 (1986).
11. G. A. Mesyats, *Ectons* (Nauka, Ekaterinburg, 1994), Part 2.

Translated by A. Sidorova

Technical Physics, Vol. 48, No. 4, 2003, pp. 508–513. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 4, 2003, pp. 129–134.

**SURFACES,
ELECTRON AND ION EMISSION**

X-ray Reflectometry and Its Application to Studying the Laser Evaporation of an Oxide Film from the Silicon Surface

A. P. Petrakov

Syktyvkar State University, Syktyvkar, 167001 Komi Republic, Russia

e-mail: petrakov@ssu.komi.com

Received June 19, 2002; in final form, October 2, 2002

Abstract—An analysis of the published data on X-ray reflectometry is carried out. The potentialities of X-ray reflectometry are demonstrated with the laser evaporation of an oxide film from a silicon surface. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The method of X-ray reflectometry is based on measuring the reflection of X rays from the material surface near the critical angle ϑ_{cr} of total external reflection (TER) and can be used for determining the geometrical and physical parameters of the surface [1]. The angular range of the measurements covers several tens of angular minutes, and the half-width of the incident radiation is within several angular seconds. Such experimental conditions impose stringent requirements on related equipment. Reflectometers designed around diffractometers have gained most recognition [2–5]. A weakly divergent beam is provided by multiple-reflection slit monochromators.

Reflectometric curves are divided into integral and differential. To obtain the former, the sample is placed so that it halves the beam. Such an angular position is taken to be the initial (zero) position. Then, a wide-aperture detector, rotating with an angular velocity twice that of the sample, records the angular dependence of the reflected intensity.

To obtain a differential curve, a crystal analyzer is placed between the sample and detector at the Bragg angle to one of a set of planes. The detector records the X-ray intensity vs. angle of analyzer rotation with the angle of sample rotation fixed. One can also take the dependence of the X-ray intensity on the angle of sample rotation with the angle of analyzer rotation fixed.

In the integral curves, the reflection coefficient first grows with the angle from 0.5 almost to unity and then sharply drops nearly to zero. The initial rise in the reflection coefficient is due to an increase in the beam area on the sample: since the angle is less than the critical TER angle, the radiation reflects almost completely. The radiation is incident at an angle near to critical. The critical angle is that at which the reflection coefficient decreases twofold [6].

A differential curve has a specular peak and anomalous (so-called Yoneda [7, 8]) peak. The angular position of the latter does not depend on the angle of incidence and equals $\approx 0.9\vartheta_{cr}$ [1].

The shapes of integral and differential curves are related to the microrelief of the sample surface. A rough surface diminishes the maximal reflection coefficient in the integral curves and their steepness in the vicinity of the critical angle [9] and also shifts the critical angle toward smaller angles. In the differential curves, the intensity of the specular peak drops, while that of the anomalous peak grows [10]. Surface roughness is characterized by the rms roughness height σ and correlation length. In the first approximation, the reflection coefficient obtained from the specular peak and the value of σ are related through the Debye–Waller factor [11]:

$$R = R_0 \exp[-(4\pi\sigma\cos\vartheta/\lambda)^2],$$

where R_0 is the reflection coefficients calculated by the Fresnel formula.

With the reflectometry method, one can measure, specifically, the density and thickness of surface layers and thin films. The density is found by the formula [12]

$$\rho = \rho_m(\vartheta_{exp}/\vartheta_{cr})^2,$$

where ρ_m is the density of a massive sample, ϑ_{exp} is the critical angle found experimentally, and ϑ_{cr} is the calculated critical angle.

The critical angle is calculated by the formula $\vartheta_{cr} = (|\chi_{hr}|)^{1/2}$, where χ_{hr} is the real part of the polarizability.

The film density is determined from the interference of X rays scattered by the upper and lower boundaries of the film. As a result of interference, the integral curve exhibits maxima and minima near the region where the reflection coefficient drops. The number N of an inter-

1063-7842/03/4804-0508$24.00

ference maximum or minimum is related to the thickness L as

$$N = M + (2L/\lambda)(\vartheta_N^2 - \vartheta_{\exp}^2)^{1/2},$$

where M is the phase constant and ϑ_N is the angular position of the Nth maximum or minimum.

The thickness is found from the slope of the N vs. $(\vartheta_N^2 - \vartheta_{\exp}^2)^{1/2}$ straight line. Interference maxima (minima) are counted from the first (initial) maximum (minimum) that is the nearest to the critical angle: $N = N_i$ and $N = N_i + 0.5$ for maxima and minima, respectively, where N_i is an integer.

X-ray reflectometry diagnoses thin surface layers, since the penetration depth of X rays is tens of angströms for smaller-than-critical angles of incidence and several tenths of a micrometer for larger-than-critical angles of incidence.

If X rays strike a crystalline object at a small angle, they may be diffraction-reflected from planes that are normal to the surface [9]. A specularly reflected (diffracted) wave carries information on the crystal structure of a film applied on a crystalline substrate [14]. The aim of this work is to generalize the published data on X-ray reflectometry and illustrate its potentialities using laser decontamination of a single-crystal silicon surface as an example.

THEORY OF THE METHOD

A schematic diagram of an X-ray reflectometer is depicted in Fig. 1. The arrows show the direction of X rays. The integral curve is taken without a crystal analyzer. Its shape does not depend on whether the sample is amorphous or crystalline nor on its orientation (if the sample is crystalline). The reflection coefficient is given by the Fresnel formula [9]

$$R = \left|\{\sin\vartheta_1 - [\sin^2\vartheta_1 + 2(n-1)]^{1/2}\}/\{\sin\vartheta_1 + [\sin^2\vartheta_1 + 2(n-1)]^{1/2}\}\right|^2,$$

where ϑ_1 is the glancing angle of incidence and n is the refractive index. For glancing angles of incidence, $\sin\vartheta \approx \vartheta$ because of their smallness; therefore, the difference between σ and π polarizations can be neglected.

Reflection from a multilayer medium is described by the Parratt relationship [15]

$$R_{n-1,n} = a_{n-1}^4(R_{n,n+1} + F_{n-1,n})/(R_{n,n+1}F_{n-1,n} + 1),$$

where $a_n = \exp(-i\pi f_n d_n/\lambda)$, $R_{n,n+1} = a_n^2(E_n^R/E_n)$, $F_{n-1,n} = (f_{n-1} - f_n)/(f_{n-1} + f_n)$, $f_n = [\sin^2\vartheta_1 + 2(n_n - 1)]^{1/2}$, d_n is the thickness of an nth layer, E_n^R and E_n are the amplitudes of the reflected and transmitted fields in the nth layer, and n is the refractive index of the nth layer.

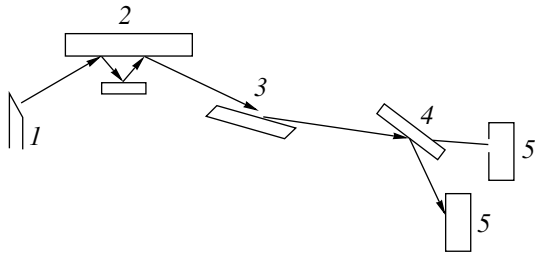

Fig. 1. X-ray reflectometer. (*1*) X-ray tube; (*2*) crystal monochromator; (*3*) sample; (*4*) crystal analyzer; and (*5*) detector.

A distinct interference pattern is observed for a layer thickness of several tens to several thousands of angströms. The upper limit of the homogeneous film thickness can be estimated from the relationship [16]

$$d \le \lambda/2\delta\vartheta,$$

where $\delta\vartheta$ is the angular resolution. The lower limit of the thickness is determined with the formula

$$d \ge N\lambda\{[(\vartheta_{cr}^u)^{-4} + (\vartheta_{cr}^l)^{-4}]/4I_0\tau\}^{1/4},$$

where N is the order of an interference peak; ϑ_{cr}^u and ϑ_{cr}^l are the critical angles of reflection from the upper and lower boundaries of the film, respectively; I_0 is the intensity of primary X-ray radiation; and τ is the exposure time.

The integral curves make it possible to detect foreign spacers between layers of a homogeneous material [17]. Such a possibility has been demonstrated with the time-separated thermal evaporations of two aluminum layers.

In terms of the Born approximation of a distorted wave, the intensity of small-angle scattering by a rough surface can be divided into two, specular and diffuse (Yoneda), components [18]. The former component is given by

$$I_{sp} = I_0|R_r|^2,$$

where I_0 is the intensity of incident radiation,

$$R_r \approx R\left(1 - q_z q_z^g \int_{-\infty}^{\infty} dz\omega z^2/2\right),$$

$q_z = 2k_0\sin\vartheta_1$, $q_z^g = 2k_0\sin\vartheta_g$, k_0 is the wave vector of the incident wave, ϑ_g is the glancing angle of reflection, $\cos\vartheta_g = \cos\vartheta_1/n$, and $\omega = [\exp(-z^2/2\sigma^2)][\sigma(2\pi)^{1/2}]^{-1}$. The coordinates x and y lie in the surface plane, and the z coordinate runs normally to the surface. The coefficient R of reflection from a smooth surface is found by the formula

$$|R| = \left|(1 - n^2)/4\sin^2\vartheta_1\right|.$$

The diffuse component of reflection is given by

$$I_{\text{dif}} = I_0(\Delta\Omega/A)(d\sigma/d\Omega)_{\text{dif}},$$

where $(d\sigma/d\Omega)_{\text{dif}} = (L_xL_y)|T(k_1)|^2|T(k_2)|^2S(q_g)|k_0^2(1 - n^2)|^2/16\pi$, $\Delta\Omega$ is the constant angle of detection of the scattered radiation, A is the beam area, (L_xL_y) is the surface area of the sample, $T(k_i) = 2\sin\vartheta_i/(\sin\vartheta_i + n\sin\vartheta_g)$, and ϑ_2 is the grazing angle of refraction.

The value of $S(q_g)$ is found by the formula

$$S(q_g) = \{\exp[-[(q_z^g)^2 + (q_z^{g*})^2]\sigma^2/2]/|q_z^g|^2\}$$
$$\times \iint dxdy\{\exp[|q_z^g|^2C(xy) - 1]\exp i(q_xx + q_yy)\},$$

where integration is over the surface irradiated and q_z^g is the projection of the refraction wave vector.

In the same approximation, the problem of small-angle scattering by the rough surface of a multilayer medium has been solved [19]. It should be noted that integral curves of small-angle X-ray scattering exhibit a number of peaks whose angular positions are defined by the Bragg formula. Here, the period of the structure plays the role of interplanar spacing. X-ray reflectometry combines well with high-resolution X-ray techniques for studying multilayer structures [20].

The angular distribution of small-angle scattering by a rough surface can be calculated within the model of the inhomogeneous transition layer [1]. Under the assumption that the electron density in a layer is a random fluctuation of the radius vector **r**, the permittivity is represented in the form

$$\varepsilon(\mathbf{r}) = \varepsilon_0(z) + \delta\varepsilon(\mathbf{r}),$$

where $\varepsilon_0(z) = \langle\varepsilon(\mathbf{r})\rangle$ is the permittivity averaged over density fluctuations (the z axis is directed inward to the crystal) and $\delta\varepsilon(\mathbf{r})$ is the permittivity fluctuation.

Let a plane wave with a wave vector $\tau_0 = \{\tau_\parallel, \tau_z\}$ be incident on a crystal. Here, $|\tau_\parallel| = \tau\sin\Theta_0$, $\tau = \omega/c$, Θ_0 is the angle between the incident beam and the normal to the surface, and $\varepsilon_0(z) = 1 + [\exp(z/a)]/[1 + \exp(z/a)]$. The scattering intensity is found by the formula

$$I = I_0\tau^4|t_0|^2\int_{-\infty}^{\infty} dk_\parallel \exp(2\tau_z^\parallel z)|t/2i\tau_z|^2K(s),$$

where $t_0 = t(\Theta_0)$ and t is the refracted wave amplitude.

In the simplest case of plane waves, $K(s)$ can be approximated by the expression

$$K(s) = |\chi|^2/(\tau_{0z} + \tau_z)^2\{(L/2\pi^{1/2})[1 - \exp(\sigma^2(\tau_{0z} + \tau_z)^2)]$$
$$\times \exp(-sL^2/4) + [[\exp(-\sigma^2(\tau_{0z} + \tau_z)^2/2)\sinh\pi a(\tau_{0z} + \tau_z)$$
$$- \pi a(\tau_{0z} + \tau_z)]/\sinh\pi a(\tau_{0z} + \tau_z)]^2\delta(s)\},$$

where the vector $\mathbf{s} = \boldsymbol{\tau}_\parallel - \mathbf{k}_\parallel$; $\mathbf{k}$ is the wave vector of the refracted wave; $L = 1/(1 + \sigma^2(\tau_{0z} + \tau_z)^2)^{1/2}$; l is the correlation length of inhomogeneities; and σ is the variance, which is related to the roughness height d by the relationship $\sigma = d/2$. In the expression for $K(s)$, the first term stands for the Yoneda scattering intensity and the second one describes the intensity of the specular component.

The surface roughness is characterized by the parameter $\sigma(\tau_{0z} + \tau_z)$, which defines the phase difference between the waves reflected from the upper and lower boundaries of the transition layer. The inhomogeneity of the transition layer may also be due to bulk defects and not only to surface roughness. These defects can be studied with this method only if the surface is smooth (finish class 14 or higher) and the roughness can be neglected. If the surface meets this requirement, defects caused, e.g., by ion implantation can be studied [21]. When recording the differential curve, one should keep in mind that its shape may depend on the distance from the illuminated surface area of the sample to its edge [22].

The dependence of small-angle scattering on the electron density fluctuation in the surface layer was confirmed in [23]. In that work, the theoretical formulas were verified by taking experimental dependences of the small-angle scattering of As ions on the irradiation dose during As implantation into the silicon surface of finish class 14.

In [24], the surface roughness was described by postulating the presence of a transition layer whose density varies with depth. The TER of X rays was treated in terms of the Darwin dynamic theory. The surface layer is split into elementary layers that are parallel to the surface. The incident wave partially passes through an elementary layer and partially reflects in the specular direction. The wave reflected from lower layers is partially reflected by upper layers in the direction of the primary beam.

Let us introduce the critical parameter

$$S_0 = 2(t\lambda)^{1/2}/3\vartheta_{\text{cr}}^{3/2},$$

where t is the transition layer thickness. If the lateral sizes of surface irregularities are less than S_0, scattering due to surface roughness can be described in the approximation of the transition layer, whose density grows with depth. Otherwise, macroirregularities occur and reflection is described in terms of geometrical optics. Theoretical calculations are confirmed by experimental integral curves.

The roughness can be taken into consideration by introducing the averaged-density transition layer [25]. This approximation applies when the distribution of surface asperity heights is normal. As model objects, metal (Cr, Cu, and Ni) films evaporated *in vacuo* were

used. Analytical and experimental integral curves were in good agreement.

In [26], diffuse scattering from a surface is explained based on the theory where the surface is viewed as statistically distributed microareas.

The angular position of the diffuse peak depends on the critical angle of a scattering material. Therefore, it was suggested [27] that composites consisting of a dissimilar layer and substrate must exhibit two diffuse peaks. This supposition was substantiated by the shape of the differential curve for the scattering of Mo$K\alpha_1$ radiation by a ≈50-Å-thick organic film on a glass substrate.

The Yoneda effect is also observed for ultrasoft X-ray radiation [28]. According to the Rayleigh criterion $h \le \lambda/(8\vartheta)$, where h is the height of surface irregularities and ϑ is the angle of scattering, the smaller the wavelength, the greater the diffuse scattering component. In [28], small-angle scattering by the face of hexagonal boron nitride that is parallel to the **c** axis was studied as a function of wavelength in the range 50–160 Å (the grazing angle of incidence was larger than critical). It turned out that the diffuse peak may both increase and decrease with wavelength relative to the specular peak.

Of special interest is the absence of the specular peak at several wavelengths. This was assumed to be associated with the presence of the transition layer, which provides a smooth, rather than stepwise, change in the permittivity at the vacuum–material interface.

The fit of the model integral X-ray scattering curves to the experimental ones allows one not only to refine the film thickness but also to estimate the electron density profile in the film [29]. The reflected intensity is calculated by the formula

$$I = \rho_0^{-2}\left|\int (d\rho/dz)\exp(iQz)dz\right|^2,$$

where ρ_0 and ρ are the electron densities in the film and substrate, respectively; $\Theta = (4\pi\sin\vartheta)/\lambda$; and z is the coordinate of the substrate plane.

The recovery of the density profile for a layered system from integral curves is a challenge. This problem cannot be solved directly, except for the simplest cases. To solve the inverse problem, the profile is expanded over a set of given functions. Then, the calculated curve of scattering by a model object is fitted to the experimental curve by finding the appropriate values of independent expansion parameters [30].

EXPERIMENTAL

We studied the (111)Si wafer surface. The wafers were irradiated by a cw ruby laser. The parameters of the radiation were the pulse duration 0.5 ms, energy density 18 J/cm^2, and laser beam diameter 8 mm. Surface areas measuring 25 × 8 mm were successively exposed to laser pulses as the sample was moved after each pulse so that the overlap of two neighboring pulses was 27 mm^2.

The integral and differential curves of TER for Cu$K\alpha_1$ radiation were recorded before and after the irradiation. The incident radiation was monochromatized by means of a slit silicon monochromator with triple reflection from the (111) plane. The angular distribution of the radiation scattered by the sample in the differential scheme was analyzed by a silicon analyzer with a single reflection from the (111) plane.

The integral curves were taken for angles α of sample rotation from 0 to 25'. The differential curves were taken for scan angles ϑ of the analyzer from 0 to 50' with the angle of sample rotation fixed (14').

RESULTS AND DISCUSSION

Figure 2 shows experimental and analytical integral curves. The reflection coefficient first increases from 0.5 to almost 0.9 and then drops, the drop being sharper for the irradiated sample. The differential curves (Fig. 3) have one maximum at $\vartheta = 2\alpha$, which is more intense and more narrow for the irradiated sample.

In the integral scheme, the surface to be studied is positioned in the object holder parallel to the incident beam, shutting out half of it (zero position). If the sample is positioned correctly, half the radiation in the zero position falls directly into the detector and the other half is shut out by the sample face that is normal to the surface. As the sample is rotated, the intensity of the radiation picked up by the detector grows, since that part of the beam striking this face decreases, while the other (incident on the sample surface) increases. At angles smaller than critical, the radiation is almost completely reflected from the surface and falls into the detector. At near-critical angles of rotation, the reflected intensity drastically drops. The ideal integral curve can be recorded only if an infinitely narrow rectangular beam is reflected by a perfectly smooth nonabsorbing

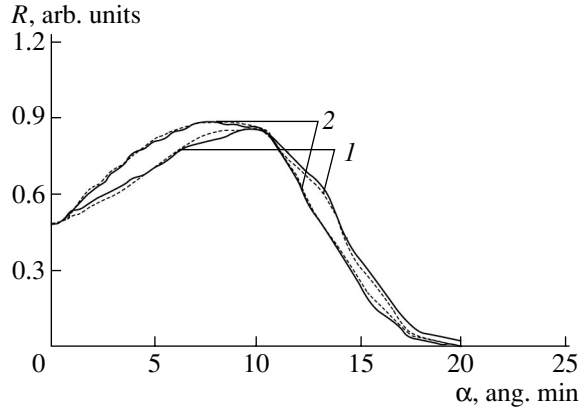

Fig. 2. Integral curves of TER. Solid lines, experiment; dashed lines, theory. (*1*) Before and (*2*) after the laser irradiation.

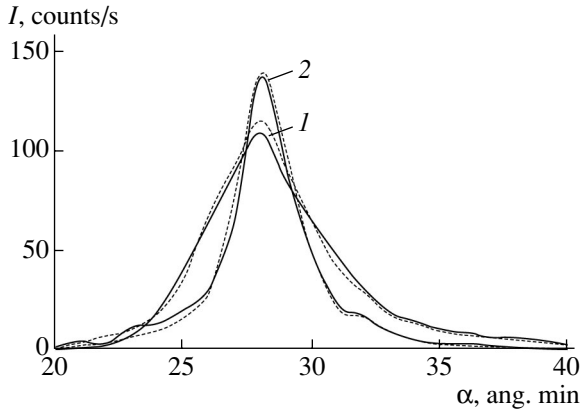

Fig. 3. Differential curves of TER. Solid lines, experiment; dashed lines, theory. (*1*) Before and (*2*) after the laser irradiation.

surface. Actually, however, the primary beam is bell-shaped and, therefore, the intensity near the critical angle smoothly declines by half.

The critical angle determined from the angular position of the half-intensity point equals 14.15' before the irradiation and 13.48' after the irradiation. This difference is due to the presence of an oxide film on the silicon surface, which evaporates under the action of the laser beam. This is also confirmed by calculating the surface layer density from the values of the critical angle. Before the laser irradiation, the density is ≈ 2.65 g/cm^3, which is close to the density of SiO_2; after the irradiation, it is equal to ≈ 2.42 g/cm^3, i.e., to the density of Si.

The thickness of the oxide film found by fitting the theoretical curves to the experimental ones with the procedure described in [10] is (55 ± 5) Å. The presence of the film must give rise to interference maxima. The spacing between the maxima is inversely proportional to the thickness of the film. The absence of the maxima means that the resolution of the reflectometer is insufficiently high to detect interference from such a thin film. Further studies showed that well-resolved interference is observed for film thicknesses between 150 and 1500 Å.

Note that the correct fit of theoretical curves to those found experimentally is possible only if the rms height of surface asperities is known. Relevant data were obtained by fitting the theoretical specular component to the experimental specular component in diffraction curves as in [10]. Before the irradiation, the roughness was higher, (12 ± 2) Å, than after the irradiation, (8 ± 1) Å. Thus, the irradiation of the silicon surface by laser pulses with the parameters listed above removes the oxide film and reduces the surface roughness.

The correlation length, which characterized the surface roughness, has an effect on the diffuse component of the differential curve. To extract reliable information on the correlation length from a differential curve, it is necessary to have a powerful X-ray source. In our experiments, where the usual X-ray tube was combined with the triple-reflection monochromator, the diffuse scattering intensity was close to the background; therefore, we failed to determine the correlation length.

From the published data cited and our experiment, it follows that surface roughness affects the shape of integral and differential curves of TER. The irradiation of the silicon surface by millisecond laser pulses with an appropriate energy removes the oxide film. The evaporation of the film reduces the surface roughness.

REFERENCES

1. A. V. Andreev, Usp. Fiz. Nauk **145**, 113 (1985) [Sov. Phys. Usp. **28**, 70 (1985)].
2. A. P. Petrakov and E. A. Golubev, Poverkhnost: Fiz. Khim. Mekh., No. 9, 15 (2000).
3. V. A. Bushuev and A. P. Petrakov, Fiz. Tverd. Tela (St. Petersburg) **35**, 355 (1993) [Phys. Solid State **35**, 181 (1993)].
4. V. A. Bushuev and A. P. Petrakov, Zh. Tekh. Fiz. **70** (5), 28 (2000) [Tech. Phys. **45**, 613 (2000)].
5. A. A. Lomov, V. A. Bushuev, and V. A. Karavanskiĭ, Kristallografiya **45**, 915 (2000) [Crystallogr. Rep. **45**, 842 (2000)].
6. O. G. Guentert, J. Appl. Phys. **36**, 1361 (1965).
7. Y. Yoneda, Phys. Rev. **131**, 359 (1963).
8. J. B. Bindell and N. Wainfan, J. Appl. Phys. **3**, 503 (1970).
9. M. A. Andreeva, S. F. Borisova, and S. A. Stepanov, Poverkhnost: Fiz. Khim. Mekh., No. 4, 5 (1985).
10. V. A. Bushuev and A. P. Petrakov, in *Proceedings of Conference of Application of X-rays, Synchrotron Radiation, Neutrons and Electrons for Materials Study, Dubna, Russia, 1997*, Vol. 2, pp. 188–193.
11. F. E. Christensen, Rev. Phys. Appl. **23**, 1701 (1988).
12. V. M. Sinaĭskiĭ and V. I. Sidenko, Prib. Tekh. Éksp., No. 6, 5 (1974).
13. A. M. Afanas'ev, P. A. Aleksandrov, and R. I. Imamov, *X-ray Diffraction Diagnostics of Submicron Layers* (Nauka, Moscow, 1989).
14. P. A. Aleksandrov and S. A. Stepanov, Poverkhnost: Fiz. Khim. Mekh., No. 6, 117 (1986).
15. L. G. Parratt, Phys. Rev. **95**, 359 (1954).
16. S. V. Sinitsyn, Prib. Tekh. Éksp., No. 1, 203 (1994).
17. V. A. Nachinov, L. G. Eliseenko, and V. N. Dostovalov, Opt. Spektrosk. **53**, 546 (1982) [Opt. Spectrosc. **53**, 321 (1982)].
18. S. K. Sinha, E. B. Sirota, S. Garoff, *et al.*, Phys. Rev. B **38**, 2297 (1988).
19. V. Holy and T. Baumbach, Phys. Rev. B **49**, 10668 (1994).

20. A. A. Darhuber, P. Schittenhelm, V. Holy, *et al.*, Phys. Rev. B **55**, 15652 (1997).
21. R. M. Imamov, E. A. Kondrashkina, P. A. Aleksandrov, *et al.*, Poverkhnost: Fiz. Khim. Mekh., No. 3, 41 (1987).
22. É. K. Kov'ev and Yu. A. Matveev, Fiz. Tverd. Tela (Leningrad) **23**, 587 (1981) [Sov. Phys. Solid State **23**, 331 (1981)].
23. P. V. Petrashen', É. K. Kov'ev, F. N. Chukhovskiĭ, *et al.*, Fiz. Tverd. Tela (Leningrad) **25**, 1211 (1983) [Sov. Phys. Solid State **25**, 695 (1983)].
24. L. A. Smirnov, T. D. Sotnikova, S. B. Anokhin, and B. Z. Taibin, Opt. Spektrosk. **46**, 593 (1979) [Opt. Spectrosc. **46**, 329 (1979)].
25. L. A. Smirnov and S. B. Anokhin, Opt. Spektrosk. **48**, 574 (1980) [Opt. Spectrosc. **48**, 315 (1980)].
26. B. M. Rovinskiĭ, V. M. Sinaĭskiĭ, and V. I. Sidenko, Fiz. Tverd. Tela (Leningrad) **14**, 409 (1972) [Sov. Phys. Solid State **14**, 340 (1972)].
27. S. R. Sarkisov and É. R. Sarkisov, Pis'ma Zh. Tekh. Fiz. **16** (16), 87 (1990) [Sov. Tech. Phys. Lett. **16**, 637 (1990)].
28. E. O. Filatova and T. A. Blagoveshchenskaya, Fiz. Tverd. Tela (St. Petersburg) **33**, 2321 (1991) [Sov. Phys. Solid State **33**, 1307 (1991)].
29. Yu. M. L'vov and G. Dekher, Kristallografiya **39**, 696 (1994) [Crystallogr. Rep. **39**, 628 (1994)].
30. I. I. Samoĭlenko, O. V. Konovalov, L. A. Feĭgin, *et al.*, Kristallografiya **44**, 347 (1999) [Crystallogr. Rep. **44**, 310 (1999)].

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 4, 2003, pp. 514–515. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 4, 2003, pp. 135–136.

BRIEF COMMUNICATIONS

Dependence of Water Viscosity on Temperature and Pressure

E. R. Likhachev

Voronezh State University, Universitetskaya pl. 1, Voronezh, 394006 Russia

e-mail: phssd18@main.vsu.ru

Received May 14, 2002; in final form, September 2, 2002

Abstract—In the first approximation, the following formula is derived for water viscosity as a function of temperature and pressure:

$$\eta = \eta_0 \exp\left[ap + \frac{E - bp}{R(T - \theta - cp)}\right],$$

where p is the pressure and E, η_0, θ, a, b, and c are constants. A formula obtained in the second-order approximation contains terms quadratic in pressure and temperature-dependent coefficients outside pressure terms.

In [1], the temperature dependence of the viscosity of liquids is given by

$$\eta = \eta_0 \exp\left[\frac{E}{R(T + T_0)}\right], \quad (1)$$

where T_0 is a correction for the nonideality of the liquid.

Since for water $T_0 < 0$, we hereafter use the parameter $\theta = -T_0$.

Detailed tabulated values of water viscosity as a function of temperature [2, 3] were obtained from experimental data. It is therefore of interest to derive a general formula for water viscosity as a function of temperature and pressure.

In this work, the temperature dependence of the water viscosity was calculated by formula (1) for different temperatures using data from the reference books [2, 3]. The results of these calculations are summarized in the table, which gives, along with the parameters E, η_0, and θ, the temperature intervals covered, the number of the values of η used, and the relative error δ. It turned out that, at low and medium pressures (1–250 bar), formula (1) fits the tabulated data well in a temperature interval from the melting point to a pressure-dependent temperature in the range 483–523 K. At higher temperatures, the viscosity deviates from the calculated curve toward lower values the greater, the higher the temperature is. This deviation is caused by the gradual transition from the liquid to gas phase. As follows from the tables for viscosity [2, 3], at higher-than-critical pressures, the liquid–gas transition is continuous. At lower pressures, this transition is interrupted, because the liquid starts boiling when the pressure of the gas being in equilibrium with the liquid equals the external pressure. At above-critical pressures, the liquid entirely turns into the gas before boiling and the transition, therefore, proceeds continuously.

From our tabulated values, one can find a general dependence of the viscosity on temperature and pressure. In the first approximation, this dependence for low and medium pressures is given by the formula

$$\eta = \eta_0 \exp\left[ap + \frac{E - bp}{R(T - \theta - cp)}\right]. \quad (2)$$

For $E = 4.753$ kJ/mol, $\eta_0 = 2.4055 \times 10^{-5}$ Pa s, $\theta = 139.7$ K, $a = 4.42 \times 10^{-4}$ bar^{-1}, $b = 9.565 \times 10^{-4}$ kJ/(mol bar), and $c = 1.24 \times 10^{-2}$ K/bar, the viscosity values calculated by formula (2) over a temperature interval of 273–463 K and at pressures of 1–250 bar are in good agreement with the tabulated data.

Formula (2) explains the fact that, with rising pressure, water viscosity declines at temperatures of 273–303 K and grows at higher temperatures. As is seen from (2), the viscosity varies with pressure, because the energy $E' = E - bp$ diminishes with pressure, while the quantities $\eta_0' = \eta_0 \exp(ap)$ and $\theta' = \theta + cp$ increase. At low temperatures, the former factor prevails; at high temperatures, the increase in η_0' and θ' becomes dominant. Physically, the increase in θ' is related to that in the water density. As the density grows, the system is more and more disturbed from equilibrium, causing [1] θ' to increase. The table also gives the results for heavy water calculated with data in [3]. Heavy water is denser than light water; accordingly, η_0 and θ for the former are larger. The slight decline of the energy E with rising pressure may be explained by an increase in θ if it is assumed [1] that the potential barrier height linearly

Table

Liquid	*P*, bar	*T*, K	*n*	η_0, 10^{-5} Pa s	*E*, KJ/mol	θ, K	δ, %
Light water	1	273–363	10	2.4152	4.7428	139.86	0.0046
	60	273–493	23	2.4638	4.703	140.3	0.030
	100	273–503	24	2.5124	4.659	140.9	0.026
	150	273–513	25	2.5702	4.608	141.6	0.033
	210	273–513	25	2.6484	4.539	142.6	0.037
	250	273–513	25	2.7042	4.491	143.3	0.042
	300	273–533	27	2.7830	4.419	144.5	0.063
	500	273–553	29	3.0816	4.181	148.3	0.12
	800	273–573	31	3.5940	3.823	154.4	0.24
Heavy water	1	277–373	13	3.175	4.234	155.0	0.61

varies with temperature. In this case, formula (1) takes the form

$$\eta = \eta_0 \exp\left[\frac{U_0 - \beta T}{R(T-\theta)}\right] = \eta_0' \exp\left[\frac{E}{R(T-\theta)}\right], \quad (3)$$

where $\eta_0' = \eta_0 \exp(-\beta/R)$ and $E = U_0 - \beta\theta$.

It follows that, as θ increases, the energy E diminishes. This may explain why the energy E of heavy water is lower than that of light water. However, from (3) it follows that, at constant U_0 and β, the viscosity always grows with p (and, correspondingly, θ). Therefore, to explain the lowering of water viscosity with rising pressure at low temperatures, one should assume that U_0 or β (or both) vary with pressure accordingly. One reason why the potential barrier height varies with pressure may be a change in the intermolecular spacing. As this spacing shrinks with pressure, the repulsion forces between molecules grow, facilitating the transition of molecules to adjacent "holes." It should be noted, however, that, with rising pressure, the probability of hole formation drops. Therefore, for the barrier to lower with rising pressure, it is necessary that the former factor prevail. Such conditions are likely to exist at low temperatures, where immobile holes (due to structural defects), whose concentration is independent of pressure, dominate.

At high pressures, considerable deviations from the viscosity curve given by formula (1) appear not only at temperatures of liquid–gas transition but also at those close to the melting point. These deviations add to the average error δ, as is seen from the table. The calculation results show that, at high pressures, where pressure corrections increase, terms quadratic in pressure and temperature-dependent coefficients outside pressure terms become essential. In the second-order approximation, the formula for water viscosity takes the form

$$\eta = \eta_0 \exp\left[(a + a_1 T)p + (a_2 - a_3 T)p^2 + \frac{E - (b + b_1 T)p}{R(T - \theta - (c + c_1 T)p)}\right]. \quad (4)$$

Formula (4) agrees well with the data given in the table over the pressure range 1–800 bar with the following values of the constants: $E = 4.753$ kJ/mol, $\eta_0 = 2.4055 \times 10^{-5}$ Pa s, $\theta = 139.7$ K, $a = 2.547 \times 10^{-4}$ bar^{-1}, $a_1 = 6.42 \times 10^{-7}$ K^{-1} bar^{-1}, $a_2 = 7.967 \times 10^{-8}$ bar^{-2}, $a_3 = 1.16 \times 10^{-10}$ K^{-1} bar^{-2}, $b = 2.795 \times 10^{-4}$ kJ/(mol bar), $b_1 = 2.48 \times 10^{-6}$ kJ/(mol K bar), $c = -4.85 \times 10^{-3}$ K/bar, and $c_1 = 6.32 \times 10^{-5}$ bar^{-1}.

In conclusion, it should be noted that, using a greater number of variable parameters, one can arrive at expressions that could fit the data in the table better than formula (4). However, this makes little sense, because experimental data on viscosity, as well as the data in the table, are not accurate enough.

REFERENCES

1. R. L. Fogel'son and E. R. Likhachev, Zh. Tekh. Fiz. **71** (8), 128 (2001) [Tech. Phys. **46**, 1056 (2001)].
2. M. P. Vukalovich, S. L. Rivkin, and A. A. Aleksandrov, *Tables of Thermal Properties of Water and Steam* (Izd. Standartov, Moscow, 1969).
3. N. B. Vargaftik, *Tables of Thermophysical Properties of Liquids and Gases* (Fizmat, Moscow, 1972; Halsted, New York, 1975).

Translated by B. Kalinin

Technical Physics, Vol. 48, No. 4, 2003, pp. 516–518. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 4, 2003, pp. 137–139.

BRIEF COMMUNICATIONS

"Gravid" Shells

E. Rapis

Laboratory of Applied Physics, Tel-Aviv University, Ramat-Aviv, 64239 Israel

Received September 9, 2002

Abstract—An explanation for an intriguing natural phenomenon, the formation of "gravid" shells, is given. Our experiments show that it is unrelated to mollusk breeding but reflects the self-assembly of dissipative nanostructured protein films under ordinary nonequilibrium conditions. This point concerns several important aspects of biology, nanotechnology, and mineralogy. © *2003 MAIK "Nauka/Interperiodica".*

Have you ever seen gravid shells? What other name could be given to specimens of shells whose interior cavities contain one or many small shells that look like their young (Fig. 1)? They are easy to find: it is only necessary to look carefully at a coastal strip. One can find many of them; therefore, they are of certain interest as an object of investigation. The "young" of shells are totally different from one another. Some of them, already convoluted, fully copy the shape of their "parents." They are located in the proper place, that is, in the shell interior, which is similar to an abdominal cavity (Fig. 1). Others are shaped into numerous twisted or extended tubes and tubules that are stratified on the inner (and sometimes outer) surface of the shell (Fig. 2). But what does it mean in reality?

In viewing the inner surface of screw shells under a microscope, we managed to discover a number of features that can be considered as early phases in the development of the specimens described. For example, a thin fibrous cancellation with alternating dark and bright zones can be discerned. Brown-colored compacts occur extensively, which are similar in shape and size and have expanding endings ("heads"). They appear as small equispaced rods or tubules, resembling larvae, and are arranged symmetrically about a thin filament at an angle of 30°–40° to it (like leaves on a tree). In many cases, voids at the former places of such larvae can be detected. Areas of regular cells (like honeycombs) or rectangular lunules with recessed bright centers separated by darker interstices are still more convincing evidence that such a phenomenon does exist. Also, laminated films are often observed.

Thus, we are dealing with manifestations of a nonstochastic process. It can be said with confidence that they are early stages of system structuring with the appearance of helical, tubular, and film fibrous forms. These are firmly bonded to the solid surface of a "maternal" shell, are three-dimensional, and exhibit macroscopic periodic density variations. The last effect shows up not only in the alternation of bright and dark areas but also in the presence of regular equispaced holes in the "bony" medium (Fig. 3).

Thus, we observe the phenomena of discontinuous bilateral, helical, and chiral symmetry and asymmetry; angular momentum; helical motion; adhesion to a solid substrate; macroscopic three-dimensionality of the structures; and other features typical of the self-organization of matter. The basic properties of this phenomenology, such as coherence, cooperation, synchronism, nucleation, and self-similarity, are evident.

Of special interest is the fact that similar phenomena are experimentally observed at the final stage of assem-

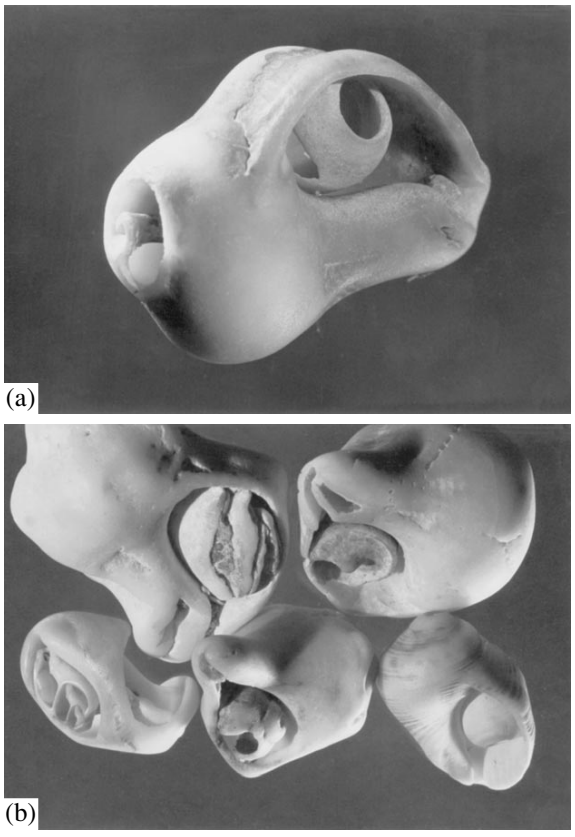

Fig. 1. Gravid shells with small youngs inside. Magnification ×1.5.

1063-7842/03/4804-0516$24.00

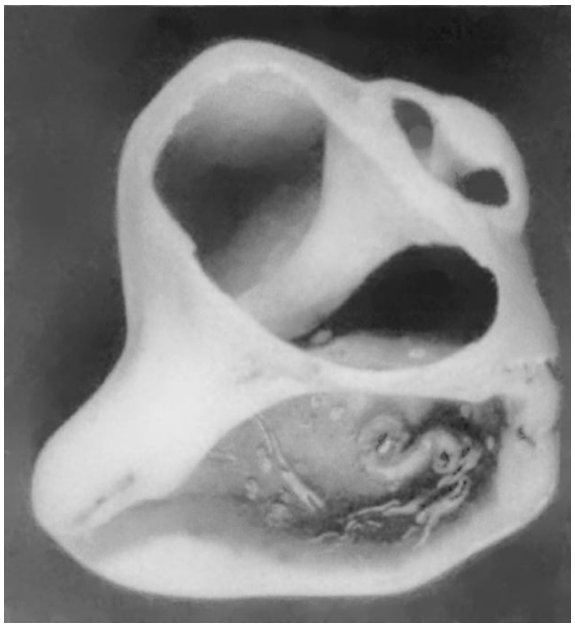

Fig. 2. Tubules inside a helical shell. Magnification ×1.2.

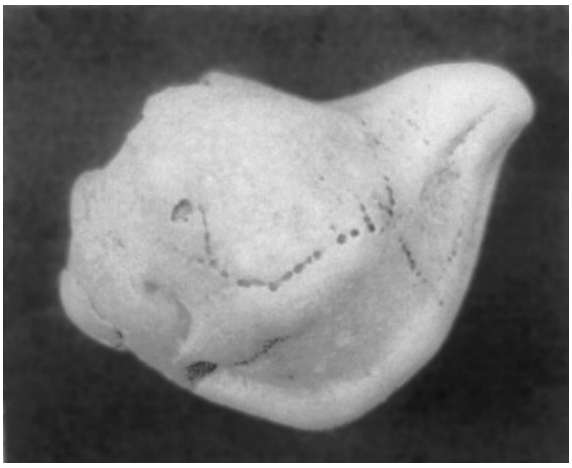

Fig. 3. Circular voids indicating the larva positions. Magnification ×1.2.

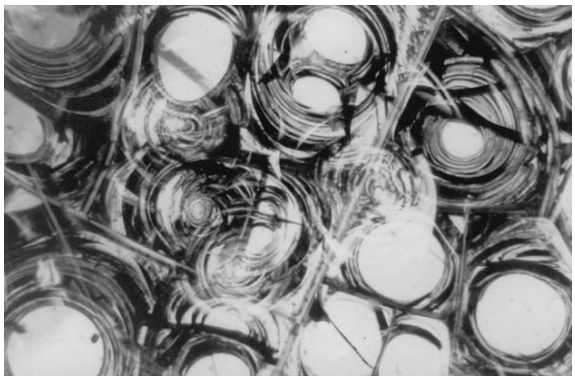

Fig. 4. Nonequilibrium film of the protos protein (lysozyme–water system). Geometrically similar blocks (cells) with nuclei are seen.

bling nonequilibrium protein films *in vitro* [1–4]. The films are divided into cells with helically twisted tubes, which serve as shell nuclei, etc. (Figs. 4, 5).

The question arises as to whether the similarity between the features in the phenomenon of gravid shells and the protein properties upon self-assembly discovered in the *in vitro* experiments is accidental. We have good grounds to believe that this identity is regular and reflects the fundamental capability of protein to autonomously self-organize and make other substances, specifically, calcites, take its shape.

It is known that when forming shells, mollusks secrete a colloidal organic mass, which is mineralized during condensation and aggregation. Our observations show that among all organic ingredients (protein, DNA, RNA, ATPase, etc.), only a protein solution condensed *in vitro* acquires the above properties [4–6].

Our experiments where a protein solution combined with calcium confirmed that the mineralization of the protein film favors the stabilization process but does not radically change the protein self-assembly and symmetry properties. These findings are in complete agreement with those obtained by Belcher *et al*. [7], who demonstrated that "a small amount of protein controls the growth of calcium crystals, facilitating the shell formation." This seems to be one more argument supporting the capability of protein for autonomous and competitive self-organization [5, 6].

Based on the aforesaid, we assume that shells both in the interior and on the surface of a maternal individual, being minerals of specific type, exemplify the natural self-assembly of dissipative protein nanostructures combined with other organic substances and minerals. In these structures, the same kinds of symmetry typical of protein nonequilibrium forms are observed from the nano- to macrolevel [6]. This substantiates the view that nanoscience and material self-assembly are of great significance in the problem of protein self-organization [8–11].

In essence, it is meaningless to describe the formation of a new biomineral shell frame on the existing exoskeleton of a mollusk using the notion of gravidity, as adopted for living creatures, since the process described has nothing to do with the known forms of mollusk breeding [12].

Yet it is difficult to get rid of the idea that such an intriguing and unexpected similarity of these phenomena reflects to some extent the fundamental unity of processes involved in the self-organization of matter, especially in going from living organisms to inanimate matter.

If so, it becomes clear that the corpus of a mollusk and its organic constituents, on the one hand, and proteins of the environment, on the other, can form nano-

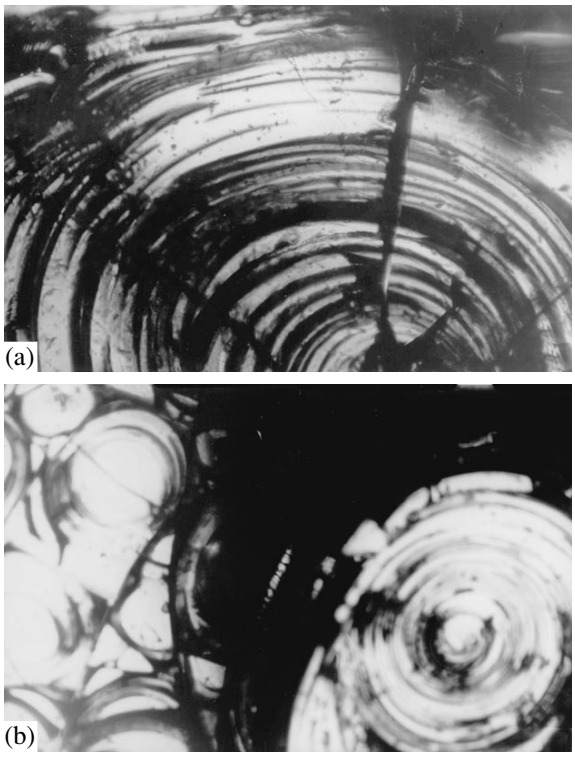

Fig. 5. Division of a nonequilibrium protein film into cells with shell-like nuclei by large rectilinear and helical defects. (a, b) 3D vortices of opposite rotation indicate the nucleation process. The images were obtained with an MIN-8 optical polarizing microscope (×200).

structured films, turning into biomineral helical forms (shells). Then, there appears the exoskeleton of a mollusk in the former case and a similar ("filial") form in the latter.

The structural similarity between such individuals of quite different nature (both unrelated to mollusk breeding) seems to reflect the same mechanisms underlying the self-assembly and mineralization of protein films.

The point raised in this article is worth studying more thoroughly, since it is not only of theoretical interest. The natural phenomenon of protein self-assembly may be of practical importance in biology, nanotechnology, and mineralogy.

ACKNOWLEDGMENTS

The author is grateful to Profs. M. Amusia, E. Braudo, V. Volkov, V. Buravtsev, A. Zaikin, M. Klinger, S. Moiseev, L. Manevich, M. Safro, Yu. Neeman, and I. Prigogine for their encouragement, fruitful discussion, and valuable advice.

REFERENCES

1. E. Rapis and G. Gassanova, Zh. Tekh. Fiz. **61** (4), 62 (1991) [Sov. Phys. Tech. Phys. **46**, 406 (1991)].
2. E. Rapis, Pis'ma Zh. Tekh. Fiz. **21**, 13 (1995) [Tech. Phys. Lett. **21**, 321 (1995)].
3. E. Rapis, Pis'ma Zh. Tekh. Fiz. **23**, 28 (1997) [Tech. Phys. Lett. **23**, 263 (1997)].
4. E. Rapis, Zh. Tekh. Fiz. **70** (1), 122 (2000) [Tech. Phys. **45**, 121 (2000)].
5. E. Rapis, Zh. Tekh. Fiz. **71** (10), 104 (2001) [Tech. Phys. **46**, 1307 (2001)].
6. E. Rapis, in *Book of Abstracts of the 4th Interdisciplinary Symmetry Congress ISIS, Technion, Haifa, 1998*, pp. 119–120.
7. A. Belcher *et al.*, Nature **381**, 56 (1996).
8. B. Parsegian, Science **270**, 1157 (1995).
9. S. Stupp *et al.*, Science **276**, 384 (1997).
10. C. Macllwain, Nature **405**, 730 (2000).
11. B. Parkinson, Science **270**, 1157 (1995).
12. D. Raup and S. Stanley, *Principles of Paleontology* (Freeman, San Francisco, 1971), p. 369.

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 4, 2003, pp. 519–522. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 4, 2003, pp. 140–142.

BRIEF COMMUNICATIONS

Effect of the Coulomb Blockade of Cooper Pairs on the Dynamic Properties of Small-Size Josephson Junctions

I. N. Askerzade

Institute of Physics, Academy of Sciences of Azerbaijan, Baku, 370143 Azerbaijan
Physics Department, Ankara University, 06100 Tandogan, Ankara, Turkey
e-mail: solstphs@lan.ab.az

Received November 29, 2001; in final form, August 13, 2002

Abstract—The dynamics of Cooper pair tunneling under Coulomb blockade in small-size Josephson junctions is studied in terms of the resistive model. A relationship between the delay time and temperature fluctuations of the Coulomb blockade edge and the rate of rise of the voltage across the junction is derived. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Quantum effects in small Josephson junctions have been the subject of much investigation in recent years [1, 2]. An important manifestation of quantum fluctuations is the microscopic quantum tunneling of an effective particle, which reflects its behavior in going over the potential barrier in phase coordinates. It is known that the switching dynamics of a usual Josephson junction is similar to the motion of a particle in the field of the "washing-board"-like potential [3]

$$U(\phi) = -E_J(\cos\phi + \varepsilon\phi), \quad (1)$$

where ε is the displacement current in terms of the critical Josephson current I_c, ϕ is the Josephson phase, and $E_J = \hbar I_c/2e$ is the Josephson energy.

The lifetime τ_l of the metastable state of a Josephson junction with current is given by [3]

$$\tau_l = \frac{2\pi}{\omega_A}\exp\left(\frac{U_0}{kT}\right), \quad (2)$$

where $U_0 = E_J\{-\pi\varepsilon + 2(\varepsilon\sin\varepsilon + (1-\varepsilon^2)^{1/2})\}$ is the potential barrier height and ω_A is the frequency of attempts, which depends on the McCamber capacitance parameter

$$\beta = \frac{2e}{\hbar}I_c R_N^2 C.$$

According to formula (2), quantum fluctuations become essential at low temperatures. The potential barrier height U_0 lowers to a value that is comparable to the plasma oscillation energy $\hbar\omega_p$. In this case, it is natural to suppose macroscopic tunneling through this barrier.

Another manifestation of the quantum fluctuation effect at low temperatures is Coulomb blockade in small Josephson junctions [4, 5]. Coulomb blockade appears when [3]

$$\min\lfloor \hbar\omega_p, \hbar\omega_c \rfloor \gg E_J, \quad (3)$$

where $\omega_p = (2eI_c/\hbar C)^{1/2}$ is the plasma frequency, $\omega_c = 2eI_cR_N/\hbar$ is the characteristic frequency of a Josephson junction with a critical current I_c, normal resistance R_N, and capacitance C.

In terms of resistance, condition (3) is recast as

$$R_N > R_Q; \quad R_Q = \hbar/4e^2, \quad (4)$$

where $R_Q = 1$ qΩ is the quantum unit of resistance.

The I–V characteristic of small Josephson junctions shows that at voltages below the Coulomb blockade edge $V_0 = e/C$ (i.e., at $V < V_0$), the capacitor made up of the junction electrodes is charged (the blockade condition) [3]. The mean value of the supercurrent in this case is zero. When the charge of the capacitor approaches an odd number of elementary (electron) charges, the supercurrent becomes nonzero and the electrodes exchange a Cooper pair. In the presence of thermal fluctuations, the transfer of the Cooper pair occurs somewhat earlier; that is, when the equality $V(t) = e/C$ is not strict. In other words, the Coulomb blockade edge "diffuses." In this work, we consider the effect of Coulomb blockade on the tunneling dynamics of a Cooper pair and the effect of thermal fluctuations on the Coulomb blockade edge.

1063-7842/03/4804-0519$24.00

BASIC EQUATIONS

For junctions with parameters satisfying condition (4) and at low voltages $eV < \Delta$, where Δ is the energy gap of a superconductor, the Hamiltonian has the form

$$H = \hat{Q}^2/2C + E_c(1 - \cos\phi) - \frac{\hbar}{2e}I(t)\phi, \qquad (5)$$

where $I(t)$ is the external current passing through the junction.

The general theory of Josephson effect [3] is valid when

$$E_Q \ll E_J, \quad E_Q = Q^2/2C. \qquad (6)$$

In this case, the electric charge Q and Josephson phase ϕ can be considered as classical variables. In the opposite limit, such an approach fails and Q and ϕ must be viewed as noncommuting operators [6] such that

$$Q = -2ei(\partial/\partial\phi) \qquad (7)$$

in the ϕ representation.

It was shown [7] that the properties of a small junction depend on the impedance of the environment (external leads). Only in the high-resistance environment, $R_e \gg R_Q$, can Coulomb blockade not be suppressed by charge fluctuations in the leads. The semiclassical theory of Josephson junctions with a small capacitance and a low quasi-particle conductivity under low-current conditions has been worked out in [8]. In this theory, $2e$ periodicity is attributed to Bloch zones of the junction. The lower zone height depends on the ratio κ of the Josephson energy $E_J = \hbar I_c/2e$ to the electrostatic energy $E_Q = e^2/2C$; that is, $\kappa = E_J/E_Q$. We will assume that $\kappa \ll 1$, since this condition is necessary for charge effects to be observed in experiments. In the single-zone approximation, the equation for quasi-charge appears as [8]

$$\frac{dq}{dt} = I(t) - \frac{1}{R}\frac{dE(q)}{dq}. \qquad (8)$$

In a high-resistance environment, equation (8) for the normalized charge $y = q/e$ is recast as

$$\dot{y} + \nu(y) = \nu_e. \qquad (9)$$

Here, the dot means differentiation with respect to time in terms of R_NC, where R_N is the Josephson junction resistance; $E(q)$ is the dispersion law for the lower zone; and $V(q) = dE/dq$.

We also introduce the dimensionless parameters $\nu(q) = V(q)/V_0$, which is the voltage across the junction, and ν_e, which is the external voltage in units $V_0 = e/C$. According to [8], $\nu(q)$ is given by

$$\nu(q) = \frac{(y - y^3)}{((y^2 - 1) + (\kappa/2)^2)^{0.5}}. \qquad (10)$$

DELAY TIME FOR COOPER PAIR TUNNELING IN THE CASE OF VOLTAGE RAMP

Small Josephson junctions have a nonlinear normalized differential capacitance $c_{dif}^{-1} = d\nu(y)/dy$. It is essential that this capacitance may take negative values. Because of this, the properties of small Josephson junctions greatly differ from those of other nonlinear reactive elements, including standard Josephson junctions. As follows from Eq. (8), the behavior of a small Josephson junction is akin to the behavior of a particle in the potential field of the form $U(q) = E(q) - \nu_e q$, where $E(q)$ is the periodic dispersion law for the lower zone.

As for a usual Josephson junction [3, 9, 10], this nonlinear capacitance causes an additional delay of Cooper pair tunneling. However, the delay time involves a factor that depends on the shape of the voltage pulse ν_e. Using (8) and approximating $\nu(y)$ near $y = 1$ as $\nu(y) \approx 2(1 - y^2)/\kappa$, we arrive at the solution to Eq. (9)

$$y = \left(\nu_e - \tanh\frac{2}{\kappa}\tau + \frac{\kappa\nu_e}{2}\tanh\frac{2}{\kappa}\tau\right)\Big/\left(1 - \nu_e\tanh\frac{2}{\kappa}\tau\right). \qquad (11)$$

For delay time, we have

$$\tau_d \approx \kappa. \qquad (12)$$

Here, it is assumed that ν_e becomes larger than the Coulomb blockade edge e/C stepwise. This formula is in good agreement with a more complicated formula obtained by exactly solving Eq. (10) with the nonlinear dependence $\nu(y)$.

Let the rate of external voltage ramp $\alpha = d\nu_e/d\tau$ be such that the tunneling of a Cooper pair takes place near $y = 1$ with negligibly small fluctuations. In terms of the new variables

$$z = \frac{4}{\kappa}\tau - \frac{8}{\alpha\kappa^2},$$

$$\alpha_z = \alpha\kappa^2/16$$

Eq. (9) takes the form

$$\frac{dy}{dz} - \frac{2}{\kappa}y^2 = \alpha_z z \qquad (13)$$

with the asymptotic solutions

$$y = C_1\alpha_z^{2/3}(z - z_0) \quad \text{at} \quad y < 1, \qquad (14a)$$

$$y = \frac{2^{1/3}}{(C_3\alpha_z^{-1/3} - z)^{1/3}} \quad \text{at} \quad y \longrightarrow \infty, \qquad (14b)$$

where $z_0 = C_2\alpha_z^{-1/3}$ is the time instant y becomes equal to unity, C_1–C_3 are constants close to unity, and τ_d is the time it takes for the quasi-charge to become infinitely large. This time is taken to be the delay time in the case of voltage ramp.

Solution (14a) corresponds to the charging of the nonlinear capacitor under voltage ramp (the blockade condition). In this case, the particle moves along the inner slope of the barrier $U(q) = E(q) - \nu_e q$. The motion after the Coulomb blockade edge has been achieved (i.e., the motion along the outer slope of the barrier) is described by solution (14b). Joining the asymptotic solutions at the boundary of their applicability domains, we obtain the time delay under voltage ramp:

$$\tau_d = \frac{2}{\alpha\kappa} + \frac{C_3}{2^{19/3}}\left(\frac{\kappa}{\alpha}\right)^{1/3}. \quad (15)$$

THERMAL FLUCTUATIONS OF THE COULOMB BLOCKADE EDGE

When the voltage rises slowly, the energy barrier $U(q) = E(q) - \nu_e q$ gradually lowers and fluctuations may initiate Cooper pair tunneling. For the probability of tunneling within a time interval t, we can write (see, e.g., [3])

$$w(t) = 1 - \exp\left(-\int_0^t \Gamma(q(t))dt\right), \quad (16)$$

where the tunneling rate Γ can be expressed as a Gaussian peak centered at the point $q = e$ ($y = 1$):

$$\Gamma = \frac{I_c}{e}\frac{\pi\kappa}{2^{7/2}(\pi\gamma)^{1/2}}\exp\left(-\frac{(y-1)^2}{\gamma}\right). \quad (17)$$

Here, I_c is the critical current of a Josephson junction and the parameter $\gamma = 2kCT/e^2$ is small ($\gamma \ll 1$).

If thermal fluctuations are insignificant, we can substitute asymptotic solution (14a) into Eq. (16). Such an approach is justified, since the inertial motion of a particle in the potential filed $U(q) = E(q) - \nu_e q$ corresponds to solution (14a) and the duration of this motion is large compared with that of the motion corresponding to solution (14b). Integration yields

$$w = 1 - \exp\left(-\frac{\pi}{2^{29/6}C_1}\frac{I_e RC}{e}\left(\frac{\kappa}{\alpha}\right)\right)^{2/3} \times\left(1 - \mathrm{erf}\frac{(C_1\alpha_z^{2/3}(z - z_0))}{\gamma^{1/2}}\right), \quad (18)$$

where erf(…) is the error function.

From (18), the charge fluctuation is given by

$$\delta y = \frac{(\gamma\pi)^{1/2}}{2}\frac{1}{\left(1 - \dfrac{2^{29/6}e\left(\dfrac{\alpha}{\kappa}\right)^{2/3}}{\pi I_c RC}\right)}. \quad (19)$$

The fluctuation of the Coulomb blockade edge is found from the relationship

$$\delta V = \frac{e}{C}\delta y. \quad (19')$$

DISCUSSION

As follows from the last formula, fluctuations of the Coulomb blockade edge depend not only the factor γ, which is responsible for thermal fluctuations, but also on the rate of voltage rise α. When the latter parameter increases, so do the fluctuations of the blockade edge. At the same time, as the rate of rise of the current through a usual Josephson junction increases, the fluctuation of the critical current is suppressed [3, 9]. The difference in the behavior of usual and small Josephson junctions is associated with different nonlinearities in the junctions. As is known [3], the equivalent circuit of a usual Josephson junction consists of four parallel-connected currents: Cooper pair current I_S, quasi-particle current I_N, displacement current I_D, and fluctuation current I_F. The Cooper pair current $I_S = I_c\sin\phi$ behaves as a nonlinear reactive "power-consuming" element $U_S(\phi) = -E_J\cos\phi$. If current variations are small, the differential value of the nonlinear inductance is given by

$$L_S^{-1} = \frac{2\pi I_c}{\Phi_0}\frac{di}{d\phi} = \frac{2\pi I_c}{\Phi_0}\cos\phi. \quad (20a)$$

In small junctions, unlike usual Josephson junctions, the current, as was noted above, has three components: nonlinear displacement current I_D, quasi-particle current I_N, and fluctuation current I_F. The nonlinear displacement current I_D behaves as a capacitive element with an energy $U(q) = E(q)$. Its inverse differential value is

$$C_{\mathrm{dif}}^{-1} = C^{-1}d\nu(y)/dy. \quad (20b)$$

The change of the variable ϕ to q changes the inductive nature of a junction to capacitive.

Thus, we studied the effect of Coulomb blockade on the dynamic properties of small Josephson junctions in terms of the resistive model. Formulas for the delay time of Cooper pair tunneling in terms of nonlinear capacitance were derived. In addition, we obtained a formula for the thermal fluctuation of the Coulomb blockade edge when the current through the junction rises linearly.

REFERENCES

1. K. K. Likharev, Mikroélektronika **16**, 195 (1987).
2. P. Silvestrini, in *Quantum Mesoscopic Phenomena and Mesoscopic Devices in Microelectronics*, Ed. by I. O. Kulik and R. Ellialtioglu, NATO Science Series (Kluwer, Dordrecht, 2000), pp. 321–325.

3. K. K. Likharev, *Introduction to the Josephson Junctions* (Nauka, Moscow, 1985).
4. D. V. Averin, K. K. Likharev, and A. B. Zorin, Zh. Éksp. Teor. Fiz. **88**, 692 (1985) [Sov. Phys. JETP **61**, 407 (1985)].
5. D. V. Averin and K. K. Likharev, Zh. Éksp. Teor. Fiz. **90**, 733 (1986) [Sov. Phys. JETP **63**, 427 (1986)].
6. P. W. Anderson, in *Lectures on Many-Body Problems*, Ed. by E. R. Caianiello (1962), p. 113.
7. L. S. Kuzmin, Yu. Pashkin, D. S. Golubov, and A. D. Zaĭkin, Phys. Rev. B **54**, 10074 (1996).
8. K. K. Likharev and A. B. Zorin, J. Low Temp. Phys. **62**, 345 (1986).
9. I. N. Askerzade, Zh. Tekh. Fiz. **68** (9), 129 (1998) [Tech. Phys. **43**, 1123 (1998)].
10. I. N. Askerzade, Zh. Tekh. Fiz. **71** (12), 88 (2001) [Tech. Phys. **46**, 1575 (2001)].

Translated by V. Isaakyan